ATTUALITÀ

LORENZO PULITO

Algoritmi predittivi e valutazione della pericolosità sociale: livelli di rischio alla luce dell'*AI Act* e prospettive interne di impiego*

Il contributo si concentra sugli *artificial intelligence risk assessment tools* per la valutazione della pericolosità sociale e ne indaga i pericoli e le soluzioni per mitigarli, tenuto conto del regolamento europeo sull'intelligenza artificiale. Interrogandosi sulla natura del contributo cognitivo fornito da tali algoritmi previsionali, si sofferma sulle prospettive interne per un loro *fair use*, anche alla luce della recente L. 23 settembre 2025, n. 132, recante «Disposizioni e deleghe al Governo in materia di intelligenza artificiale».

Predictive algorithms and the assessment of social dangerousness: risk levels in light of the AI Act and internal prospects for use.

The paper focuses on artificial intelligence risk assessment tools for the assessment of social dangerousness and investigates both the risks they pose and the potential solutions to mitigate them, taking into account the European regulation on artificial intelligence. Questioning the nature of the cognitive contribution offered by these predictive algorithms, the paper focuses on the internal prospects for their fair use, also in light of the recent Law No. 132 of 23 September 2025, entitled "Provisions and delegations to the Government on artificial intelligence".

1. *Considerazioni introduttive*. I *risk assessment tools* per la previsione e valutazione della pericolosità sociale[1] trovano ormai un'ampia diffusione pratica in diversi ordinamenti[2] A fronte

[1]* Questo lavoro è stato parzialmente sostenuto dal progetto FAIR - Future AI Research (PE00000013), nell'ambito del programma MUR del PNRR finanziato dal NextGenerationEU.

[2] Dal punto di vista degli obiettivi e degli impatti sul processo penale è possibile distinguere tra strumenti predittivi in funzione decisoria (fondati su previsioni statistiche di cui il giudice si dovrebbe servire per emettere una decisione e che servono ad implementarne il patrimonio conoscitivo in funzione del provvedimento da emettere, il quale resta affidato all'umano) e strumenti predittivi decisori (i quali servono ad ipotizzare le future decisioni dei giudici in casi simili). Gli strumenti considerati nel presente lavoro rientrano nella prima categoria. Sulla distinzione, KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Cass. pen.*, 2024, 5, 1647.

[2] Secondo BASILE, *Esiste una nozione ontologicamente unitaria di pericolosità sociale? Spunti di riflessione, con particolare riguardo alle misure di sicurezza e alle misure di prevenzione*, in *Riv. it. dir. proc. pen.*, 2018, 2, 645 ss., la pericolosità sociale funge da "formula magica", suscettibile di essere invocata in plurime e svariate sedi, tanto da porsi in dubbio che costituisca una nozione ontologicamente unitaria. Nell'ambito del presente lavoro, in via approssimativa e semplificativa, si considererà il giudizio sulla pericolosità sociale alla stregua di una valutazione circa la probabilità di commissione di ulteriori reati.

dei vantaggi che tali sistemi intelligenti promettono, il loro impiego può comportare significative compressioni dei diritti fondamentali. È fondamentale inquadrare quali siano i principali rischi che questi *tools* comportano e, in base al loro livello, classificare correttamente tali sistemi algoritmici in base a quanto previsto dall'*AI Act*[3]. La rapidità con cui si vanno diffondendo spinge, poi, a interrogarsi circa gli spazi e le condizioni di un loro eventuale prossimo impiego nell'ambito del nostro sistema processuale penale, anche alla luce della recente L. 23 settembre 2025, n. 132[4].

2. *Gli algoritmi previsionali: un inquadramento.* In un *paper* scientifico già di qualche anno fa si affermava come «[m]ore than 200 AI Risk Assessment Tools (AI RATs)», sistemi che incorporano algoritmi di apprendimento automatico, che generano modelli di rischio basati su enormi volumi di dati, «are used around the world at different settings related to the criminal justice system»[5].

I *risk assessment tools* hanno conosciuto una rilevante diffusione oltreoceano, in particolare negli Stati Uniti d'America, dove vengono adoperati per aiutare il giudice ad assumere decisioni in tutte le fasi del processo: in quella cautelare (*pre-trial decisions*), in quella decisoria (*sentencing*) e in quella esecutiva (*parole decisions*)[6]; in tali contesti questi *software* servono soprattutto per sciogliere prognosi di pericolosità sociale e rischio di recidivanza[7].

Gli strumenti di valutazione del rischio sono giunti alla quarta generazione[8].

Nella prima generazione ("1G"), nota anche come "valutazione clinica non strutturata", l'analisi del rischio veniva effettuata da professionisti clinici.

La seconda generazione ("2G") ha adottato il "metodo attuariale", meccanico e algoritmico[9].

Quelli di terza generazione ("3G"), oltre a utilizzare fattori di rischio statici, incorporano anche fattori di rischio dinamici, teoricamente legati alla recidiva. Infine, la quarta generazione ("4G"),

---

Sulla disciplina processuale che regola l'accertamento della pericolosità sociale e la conseguente scelta della misura di sicurezza, utili riferimenti in CABIALE, *L'accertamento giudiziale della pericolosità sociale fra presente e futuro*, in *Arch. pen. web*, 2022, 2, 1 ss.

[3] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), in *G.U.U.E.*, 12 luglio 2024.

[4] L. 23 settembre 2025, n. 132, recante «Disposizioni e deleghe al Governo in materia di intelligenza artificiale», in *G.U.*, 25 settembre 2025.

[5] CHELIOUDAKIS, *AI Risk Assessment Tools: A judge's best friend?*, in *Proceedings of the Twelfth International Workshop on Juris-informatics (JURISIN 2018)*, Yokohama, 2018, 99.

[6] D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont., Riv. trim.*, 2019, 2, 356.

[7] QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, in *BioLaw Journal*, 2019, 1, 142.

[8] CHELIOUDAKIS, *Risk Assessment Tools in Criminal Justice: Is There a Need for Such Tools in Europe and Would Their Use Comply with European Data Protection Law?*, in *Australian National University Journal of Law & Technology (ANUJOLT)*, 2020, 2, 77 ss.

[9] GROVE-MEEHL, *Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy*, in *Psychology, Public Policy, and Law,* 2016, 2, 293 ss.

pur basandosi su quella precedente, se ne distingue per la capacità di fornire supporto agli operatori giudiziari nell'individuazione delle strategie di intervento per la gestione dei casi, documentando le variazioni delle specifiche esigenze criminogene che potrebbero verificarsi dall'ingresso di un autore di reato nel sistema di giustizia penale fino alla sua uscita. L'obiettivo principale dei sistemi di questa generazione è assicurare soluzioni aderenti ai principi di efficace trattamento e facilitare la supervisione clinica volta a migliorare la protezione pubblica dalla recidiva[10], a fronte della maturata consapevolezza che questa dipende dalla combinazione tra variabili di rischio a livello individuale e quelle a livello di comunità[11].

Il *tool* previsionale più conosciuto è senza dubbio il *Correctional Offender Management Profiling for Alternative Sanctions* (Compas), un software di apprendimento automatico utilizzato dai tribunali statunitensi per misurare la probabilità di recidiva degli imputati[12], la cui notorietà si lega soprattutto alla circostanza di essere stato l'oggetto del *leading case*, culminato con l'altrettanto celebre "sentenza Loomis" emessa dalla Corte suprema del Wisconsin[13], che ha aperto il dibattito sulle condizioni di accettabilità degli output predittivi con il *right to a fair trial*, il *right to be sentence on accurate information*, nonché il *right to an individualized sentence.*

La valutazione del rischio è effettuata attraverso algoritmi statistici, che lo quantificano attribuendo un punteggio. Quest'ultimo scaturisce sulla base di molteplici *data points*, che includono fattori storico-statici (come i precedenti penali, l'età del primo arresto, la partecipazione ad associazioni criminose) e fattori dinamico-criminogenici (come la stabilità residenziale, lo stato occupazionale, i legami con la comunità, l'abuso di sostanze, l'inclusione e le relazioni sociali, la composizione familiare), oltre che sulla base delle risposte, fornite dal soggetto che deve essere valutato, a centotrentasette domande a scelta multipla[14]. Compas ha due modelli di rischio primari: recidiva generale e recidiva violenta. La scala generale del rischio di recidiva viene utilizzata per prevedere nuovi reati; la scala del rischio di recidiva violenta si concentra sulla probabilità di crimini violenti, quali omicidio, violenza sessuale, rapina e aggressione aggravata. I risultati della scala Compas vengono trasformati in punteggi decili, suddivisi in gruppi

[10] ANDREWS-BONTA-WORMITH, *The Recent Past and Near Future of Risk and/or Need Assessment*, in *Crime & Delinquency*, 2006, 1, 7.

[11] BYRNE-PATTAVINA, *Next generation assessment technology: The potential and pitfalls of integrating individual and community risk assessment*, in *Probation Journal*, 2017, 3, 242 ss.

[12] LIU-CHEN-SHEN-CHOO, *FairCompass: Operationalising Fairness in Machine Learning*, in *https://arxiv.org/abs/2312.16726v1*, 27 dicembre 2023.

[13] State v. Loomis, 881 N.W.2d 749 (Wis. 2016). Con riguardo al caso giudiziario v. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale 'predittiva'*, in *Cass. pen.*, 2019, 4, 1750 ss. Per la puntuale ricostruzione della vicenda cfr. altresì CARRER, *Se l'*amicus curiae *è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in *Giur. pen. (web)*, 2019, 4; FIORIO, *Predizione algoritmica e giurisdizione di sorveglianza*, in *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, a cura di Baccari-Felicioni, Milano, 2023, 253 ss.; MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, ivi, 237.

[14] Le domande sono relative a cinque macro-aree (*criminal involvement, relationship/lifestyle, personality/attitudes, family and social exclusion*): v. KEHL-GUO-KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, in *www.dash.harward.edu*, 14 agosto 2017, 11. Sul punto cfr. anche MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, cit., 236.

di uguali dimensioni; in particolare, i punteggi in decili da 1 a 4 sono etichettati come rischio "basso", da 5 a 7 "medio" e da 8 a 10 "alto"[15].

Molti altri, sebbene meno noti, sono gli algoritmi di previsione in questo ambito. Si possono ricordare: *Level of Service Inventory Revised* (Lsi-R), *Public Safety Assessment* (Psa)[16] e *Structured Assessment of Violence Risk in Youth* (Savry)[17].

Appartengono alla quarta generazione l'*Offender Screening Tool* (Ost) e l'*Ohio Risk Assessment System* (Oras)[18]: più precisamente, sono *risk and need assessment tools*, strumenti in grado di fornire una base informativa più completa per la definizione di strategie di gestione del rischio posto dal singolo individuo, attraverso la valutazione dei fattori di rischio, dei bisogni criminogenici e della responsività del soggetto.

Tali strumenti, basati sulla considerazione che la personalità non resta segnata in maniera irrimediabile dal reato commesso in passato, ma continua ad essere aperta alle prospettive di cambiamento, traducono il modello teorico *Risk-Need-Responsivity*, che abbraccia tutte le dimensioni di una previsione del futuro comportamento del reo in relazione alle sue caratteristiche individuali e alla sua capacità di reagire al trattamento[19].

Sebbene siano state sollevate molte questioni giuridiche a proposito della inclusione nei punteggi di rischio di fattori come l'associazione tra pari e le relazioni familiari di un imputato[20], si ritiene il loro «contributo decisivo [...] sul piano informativo», in quanto occasione propulsiva

---

[15] LAGIOIA-ROVATTI-SARTOR, *Algorithmic fairness through group parities? The case of COMPAS-SAPMOC*, in *Ai & Society*, 2023, 2, 463.

[16] Si tratta dello strumento di valutazione del rischio pre-processo più popolare negli Stati Uniti, usato in fase di *pre-trial* per assistere il giudice nella decisione sul rischio di rimettere in libertà l'accusato arrestato prima della definizione del processo. Lo strumento è illustrato da BRITTAIN-GEORGES-MARTIN, *Examining the Predictive Validity of the Public Safety Assessment*, in *Criminal Justice and Behavior*, 2021, 10, 1432.

[17] Consiste in un inventario strutturato di giudizi professionali che mira a stimare il rischio di comportamenti violenti negli adolescenti di età compresa tra dodici e diciotto anni: a riguardo, v. BORUM-LODEWIJKS-BARTEL-FORTH, *The Structured Assessment of Violence Risk in Youth (SAVRY),* in *Handbook of violence risk assessment*², a cura di Douglas-Otto, New York, 2021, 438 ss. In argomento, cfr. altresì KLEEVEN-DE VRIES ROBBÉ-POPMA, *The Validity of Violence Risk Assessment in Young Adults: A Comparative Study of Juvenile and Adult Risk Assessment Tools*, in *Emerging Adulthood*, 2023, 6, 1412 s.

[18] Sul cui funzionamento cfr. LATESSA-LOVINS-LUX, *The Ohio Risk Assessment System*, in *Handbook of Recidivism Risk/Needs Assessment Tools*, a cura di Singh-Kroner-Wormith-Desmarais-Hamilton, Hoboken, 2017, 147 ss. Tra i *risk and need assessment tools* rientra anche *Prisoner Assessment Tool Targeting Estimated Risk and Need* (Pattern), sul quale v. LABRECQUE-HESTER-GWINN, *Revalidation of the First Step Act Risk Assessment: A Test of Predictive Strength, Dynamic Validity, and Racial/Ethnic Neutrality*, in *Crime & Delinquency*, 2025, 2, 299 ss.; STIMSON, *The First Step Act's Risk and Needs Assessment Program: A Work in Progress*, in *www.heritage.org*, 8 giugno 2020.

[19] Tale quadro è stato originariamente proposto da ANDREWS-BONTA, *Risk-need-responsivity model for offender assessment and rehabilitation*, in *https://www.publicsafety.gc.ca/cnt/rsrcs/rsk-nd-rspnsvty/rsk-nd-rspnsvty-eng.pdf*. Cfr. anche ANDREWS, *The risk-need-responsivity (RNR) model of correctional assessment and treatment*, in *Using social science to reduce violent offending*, a cura di Dvoskin-Skeem-Novaco-Duglas, Oxford, 2012, 127 ss.; ANDREWS-BONTA, *The psychology of criminal conduct*⁵, New York, 2010.

[20] KARP, *What Even is a Criminal Attitude? - and Other Problems with Attitude and Associational Factors in Criminal Risk Assessment*, in *Stanford Law Review*, 2023, 6, 1431 ss.

per un dialogo sul trattamento sanzionatorio, soprattutto in fase di commisurazione, che sia «*tailored made* sulle caratteristiche del reo» e proporzionato[21]. Di conseguenza, tali strumenti di intelligenza artificiale, capaci di mettere in luce le caratteristiche della persona a tutto tondo, si rivelano utili «per la definizione di progetti di sanzioni strutturati sulle concrete esigenze criminogeniche poste dal reo»[22], potendo fornire al giudice «uno spettro di elementi circa la personalità dell'autore di reato e le sue condizioni di vita»[23].
Giova ricordare, infine, che esistono diversi *tools* algoritmici per la valutazione del rischio di ripetizione ed escalation della violenza di genere (quali Viogèn, un algoritmo adoperato in Spagna e che trova base giuridica nella Ley Orgánica 28 dicembre 2004, n. 1, ed Epv-R, strumento utilizzato nei tribunali baschi che, come il primo, è ispirato al metodo *Spousal Assault Risk Assessment*): questi strumenti, per la vulnerabilità dei soggetti interessati alle valutazioni e la delicatezza dei delitti che delineano il loro campo d'uso privilegiato, pongono sfide specifiche[24], non scrutinabili in questa sede.

3. *Livelli di rischio.* Tali "arnesi" algoritmici promettono diversi vantaggi: superamento dell'uso di euristiche fuorvianti da parte del giudice, risparmio di tempi e risorse, neutralità, obiettività, alleggerimento del peso della responsabilità di decidere gravante sui giudicanti, soprattutto in un ambito in cui il giudizio è probabilistico e l'incertezza è ontologica.
Ma l'esperienza d'oltreoceano ha consentito anche di metterne a fuoco le gravi problematicità. Prima tra tutte, quella relativa al carattere discriminatorio dei risultati algoritmici: si pensi a quanto emerso, nell'ambito dell'inchiesta realizzata dall'organizzazione *ProPublica*[25], relativamente a Compas, *bias machine* foriera di pregiudizi, derivanti «dalla correlazione tra fattori di rischio *associati* ad una certa provenienza etnica», che scaturisce dalla «rappresentazione stereotipata di un gruppo sociale, frutto di un condizionamento umano difficile da estirpare, [che] inficia l'intero processo di *risk assessment*, producendo un circolo vizioso che si autoalimenta»[26].
Preoccupazioni ancora più profonde sono emerse in tempi recenti: secondo un'interessante ricerca, Compas non sarebbe discriminatorio soltanto nei confronti di determinati gruppi di imputati, bensì nei riguardi di *tutti* gli imputati, dal momento che aumenterebbe le probabilità

[21] MALDONATO, Risk and need assessment tools *e riforma del sistema sanzionatorio: strategie collaborative e nuove prospettive*, in *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, a cura di Di Paolo-Pressacco, Napoli, 2022, 162.
[22] *Ibid.*, 165.
[23] *Ibid.*, 167, sia pur con riferimento alla pena prescrittiva di cui all'articolato del Gruppo di lavoro dell'Associazione italiana dei Professori di Diritto penale, incaricato di avanzare proposte in tema di riforma del sistema sanzionatorio.
[24] PULITO, *Il contributo dell'intelligenza artificiale simbiotica nella protezione delle vittime vulnerabili e nel contrasto della violenza di genere*, in *BioLaw Journal*, 2024, Special Issue 1, 221 ss.
[25] Si v. lo studio condotto da ANGWIN-LARSON-MATTU-KIRCHNER, *How We Analyzed the COMPAS Recidivism Algorithm*, in *www.propublica.org*, 23 maggio 2016. Secondo tale inchiesta, l'algoritmo discrimina gli imputati neri, per i quali la propensione alla recidiva risultava pari al doppio di quella calcolata per i bianchi.
[26] PADUA, *Intelligenza artificiale e giudizio penale: scenari, limiti e prospettive*, in *Proc. pen. giust.*, 2021, 6, 1500 s. L'A. si interroga anche sulle soluzioni (sia *ex ante* che *ex post*) volte a correggere tali distorsioni.

di carcerazione e privilegerebbe fortemente le vittime[27]. È fondamentale notare che tale orientamento decisionale, per quanto eventualmente coincidente con le tendenze politiche espresse della maggior parte degli Stati federali dell'U.S.A., non è ascrivibile ad alcun provvedimento normativo emanato dai loro organi legislativi né è stata discussa pubblicamente, ma si insinua nascostamente attraverso i gangli progettuali del sistema artificiale.

Si profila il nodo dell'opacità di funzionamento dell'apparato algoritmico, dovuta essenzialmente all'indisponibilità dei codici sorgente del *risk assessment* digitale, coperti da *trade secret*, pur accedendo ai quali il funzionamento del sistema potrebbe rimanere poco trasparente[28].

Altre "patologie" di questi *tools* sono rappresentate dall'effetto *anchoring*, dalla dubbia validità predittiva per mancanza di una adeguata teoria scientifica di supporto o, ancora, dalla difettosa traduzione della stessa in modello computazionale[29].

Il quadro si fa ancora più sconcertante di fronte alla tesi, pur'essa sostenuta, che questi strumenti di valutazione del rischio possano produrre comportamenti criminali, piuttosto che meramente prevederli[30].

A fronte di un così elevato livello di rischi è fondamentale comprendere se e come questi *tools* siano classificabili in base al recente regolamento europeo sull'intelligenza artificiale.

---

[27] ENGEL-LINHARDT-SCHUBERT, *Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism*, in *Artificial Intelligence and Law*, 2025, 33, 385 ss.

[28] Come mettono in evidenza BACCARI-PECCHIOLI, *I.A. e giudizio sul fatto: gli strumenti di e-evidence per la cognizione*, in *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, a cura di Baccari-Felicioni, cit., 132 s., è «possibile riscontrare un basilare deficit di trasparenza rispetto al funzionamento dei sistemi di intelligenza artificiale che, per una serie ragioni, rappresenta perciò una vera e propria incognita. Non solo per gli operatori giuridici»; risulta «di cruciale peso la circostanza 'strutturale' per la quale la programmazione dei sistemi di I.A. può basarsi sull'approccio dell'apprendimento automatico, in virtù del quale l'algoritmo stesso risulta in grado di automodificarsi sulla scorta dell'ampliamento progressivo del proprio dataset di riferimento. Di conseguenza, non è assurdo che gli stessi programmatori non siano in grado di spiegare l'iter operativo 'sotterraneo' tramite il quale il software ha elaborato una determinata risposta all'input ricevuto». Su tali aspetti, tra i molti, CANZIO, *Intelligenza artificiale e processo penale*, in *Prova scientifica e processo penale*, a cura di Canzio-Luparia Donati, Milano, 2022, 907; CASONATO, *Giustizia e intelligenza artificiale: considerazioni introduttive*, in *BioLaw Journal*, 2021, 2, 359; CONTISSA-LASAGNI-SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Dir. internet*, 2019, 1, 620; DONATI, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 2020, 1, 427; MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Dir. pen. cont., Riv. trim.*, 2019, 2, 404 ss.; PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. web*, 2022, 1, 4; QUATTROCOLO, *Processo penale e rivoluzione digitale: da ossimoro a endiadi?*, in *mediaLaws*, 2020, 3, 127; EAD., *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale 'predittiva'*, cit., 1752, nonché 1758; EAD., *Equità del processo penale e automated evidence alla luce della convenzione europea dei diritti dell'uomo*, in *Revista Ítalo-Española de Derecho Procesal*, 2019, 1, 117.

[29] Su questi problemi e, in particolare, sulle questioni che la costruzione di uno strumento di rischio attuariale solleva sull'accuratezza, l'uguaglianza e lo scopo della punizione, cfr. EAGLIN, *Constructing Recidivism Risk*, in *Emory L. J.*, 2017, 1, 59 ss.

[30] KHOSROWI-VAN BASSHUYSEN, *Making a Murderer: How Risk Assessment Tools May Produce Rather Than Predict Criminal Behavior*, in *American Philosophical Quarterly*, 2025, 4, 309 ss.

4. *I* tools *di* risk assessment *vietati*. Il testo normativo europeo disciplina i "sistemi di intelligenza artificiale", inclusi quelli che possono intersecare il delicato settore della giustizia penale[31], secondo un approccio di carattere orizzontale e basato sul rischio, con l'obiettivo di introdurre regole armonizzate dirette ai diversi soggetti pubblici e privati che, indipendentemente dalla loro ubicazione, li sviluppano, distribuiscono e utilizzano nel territorio dell'Unione europea.
I propositi, declinati nei primi due dei centottanta *considerando*, appaiono decisamente ambiziosi: migliorare il funzionamento del mercato interno, promuovere la diffusione di un'intelligenza artificiale antropocentrica e affidabile, garantendo allo stesso tempo un livello elevato di protezione dei diritti fondamentali, nonché promuovere l'innovazione, sino a rendere l'Unione stessa un *leader* nell'adozione di un'intelligenza artificiale affidabile.
L'*AI Act* è parte di una più complessa trama normativa che regola l'intelligenza artificiale e concorre, unitamente ad altre fonti, a disegnare un «sistema di limiti all'impiego indiscriminato» della stessa[32].
Il regolamento europeo non definisce l'intelligenza artificiale[33], ma – all'art. 3 – il «sistema di AI», definito come «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».
Il c.d. *risk based approach*[34], metodo di normazione non nuovo nel panorama unionale, commisura il contenuto delle regole all'intensità e alla portata dei rischi che possono essere generati

---

[31] Per un inquadramento, TERESI, *L'AI Act nell'ottica del processual-penalista: uno sguardo preliminare*, in *PenaleDP*, 20 giugno 2024, 1 ss.

[32] In questi termini QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, in *Diritto di Difesa*, 16 gennaio 2025, 12, la quale cita la Carta dei diritti fondamentali, la normativa europea secondaria (come il Gdpr, la direttiva 2016/680 e tutto il "pacchetto" delle direttive di Stoccolma e, in generale, di quelle aventi base giuridica nell'art. 82 Tfue), la Convenzione europea dei diritti dell'uomo e le tradizioni costituzionali comuni. Meritano di essere altresì richiamati il *Digital Service Act*, il *Digital Market Act*, e il *Data Governance Act*. Si ricorda anche che è stata approvata la Direttiva (UE) 2024/2853 del Parlamento europeo e del Consiglio, del 23 ottobre 2024, sulla responsabilità per danno da prodotti difettosi, che abroga la direttiva 85/374/CEE del Consiglio. È stata invece ritirata la Proposta di Direttiva del Parlamento europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, COM (2022) 496 *final*, 28 settembre 2022.

[33] Secondo BALSAMO, *L'impatto dell'intelligenza artificiale nel settore della giustizia*, in *www.sistemapenale.it*, 24 maggio 2024, 1, siffatta definizione segna il venir meno di quella tradizionale (e fuorviante) concezione che ha qualificato l'intelligenza artificiale come simulazione dell'intelligenza umana. Sottolinea che l'enfasi sull'intelligenza "simile a quella umana" potrebbe essere ingannevole KONIAKOU, *From the "rush to ethics" to the "race for governance"*, in *Information Systems Frontiers*, 2023, 25, 74.

[34] Su tale approccio, si v. CANZIO, *AI ACT e processo penale: sfide e opportunità*, in *www.sistemapenale.it*, 14 ottobre 2024, 4; MANCIOPPI, *La regolamentazione dell'intelligenza artificiale come opzione per la salvaguardia dei valori fondamentali dell'UE*, in *Federalismi*, 2024, 7, 123 ss.; PETRUSO-SMORTO, *Il Regolamento europeo sull'intelligenza artificiale: una prima lettura*, in *Nuov. Giur. civ. comm.*, 2024, 4, 989 ss.; TORRE, *Il Regolamento europeo sull'intelligenza artificiale: i profili processuali*, in *Proc. pen. giust.*, 2024, 6, 1539 ss. In precedenza, a proposito della proposta di regolamento del 21 aprile 2021 (COM (2021)206), v. BARONE, *Intelligenza artificiale e processo penale: la linea dura del Parlamento europeo. Considerazioni a margine della risoluzione del Parlamento europeo del 6 ottobre 2021*, in *Cass. pen.*, 2022, 3, 1180 ss.; CASONATO-MARCHETTI, *Prime osservazioni sulla proposta di regolamento*

dai sistemi d'intelligenza artificiale, indipendentemente dalla tecnologia più o meno avanzata che caratterizza ciascun algoritmo[35].

La scala gerarchica dei rischi è solitamente paragonata ad una piramide di quattro livelli, al cui vertice si collocano i sistemi a rischio "inaccettabile", di cui il regolamento vieta «l'immissione sul mercato, la messa in servizio o l'uso»: seguono i sistemi a rischio "alto" (che possono essere adoperati soltanto in conformità a rigorosi presupposti), "limitato" (di cui è lecito servirsi nel rispetto degli obblighi di trasparenza) e, infine, "minimo" (non soggetti a regole particolarmente stringenti).

La prima pratica di intelligenza artificiale vietata dal regolamento (art. 5, lett. a) riguarda i sistemi che utilizzano tecniche subliminali, mediante stimoli audio, grafici e video, o «tecniche volutamente manipolative o ingannevoli», capaci di indurre le persone a adottare comportamenti indesiderati o a prendere decisioni in modo da sovvertirne e pregiudicarne l'autonomia, il processo decisionale e la libera scelta. Tuttavia, perché incorrano nel divieto, è necessario che le pratiche manipolative o di sfruttamento, alle quali questi sistemi facciano ricorso, «provochi[no] o possa[no] ragionevolmente provocare [...] un danno significativo», la cui individuazione è lasciata all'interprete, non fornendosi alcun parametro di riferimento a riguardo.

Sono altresì inaccettabili (art. 5, lett. b), se il rischio lesivo è altrettanto significativo[36], le tecniche che sfruttano le vulnerabilità di una persona (età, disabilità o altra situazione di disagio sociale o economico).

Il divieto di sistemi di intelligenza artificiale che permettono ad attori pubblici o privati di attribuire un *social scoring* alle persone fisiche (art. 5, lett. c) si applica, invece, se il punteggio sociale determina, in contesti sociali non collegati a quelli di originaria generazione e raccolta dei dati, un trattamento pregiudizievole o sfavorevole di persone fisiche o di interi gruppi,

---

dell'Unione Europea in materia di intelligenza artificiale, in *BioLaw Journal*, 2021, 3, 415 ss.; PIZZETTI, *La proposta di Regolamento sull'IA della Commissione europea presentata il 21.4.2021 (COM (2021) 206 final) tra Mercato Unico e competizione digitale globale*, in *Dir. internet*, 2021, 4, 591 ss.; SCHEPISI, *Le "dimensioni" della regolazione dell'intelligenza artificiale nella proposta di regolamento della Commissione*, in *Quad. AISDUE* (Atti convegni), 2022, 16, 330 ss.; SIMONCINI, *La proposta di regolazione europea dell'intelligenza artificiale. Prime riflessioni*, in *Protezione dei dati personali e nuove tecnologie*, a cura di Adinolfi-Simoncini, Napoli, 2022, 1 ss.

[35] La categorizzazione fatta propria dall'*AI Act* è stata criticata in quanto delinea un approccio "statico" al fenomeno dell'intelligenza artificiale. Cfr. CANATO, *Verso il superamento del "legal risk" europeo: intelligenza artificiale e approccio proporzionale al rischio*, in *Leg. pen.*, 2024, 3, 94 ss., secondo la quale il limite dell'*AI Act* è di dare rilevanza al «mero "legal risk"» e di proporre «un pericolo di danno sulla base di standard regolatori individuati a priori senza alcun richiamo al fatto», ravvisandosi «la necessità evidente di attuare un approccio maggiormente flessibile, proporzionale e "semi-quantitativo", funzionale a consentire un aggiornamento e una valutazione "case by case" delle stesse categorie di rischio, così come proposte». L'approccio di valutazione del rischio semi-quantitativo è stato elaborato da NOVELLI-CASOLARI-ROTOLO-TADDEO-FLORIDI, *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, in *Digital Society*, 2024, 3, 1 ss.

[36] La individuazione vaga e indeterminata delle condizioni in presenza delle quali scatta il divieto è stata stigmatizzata in dottrina, per l'ampio spazio di discrezionalità che ne discende sia in capo a fornitori e utilizzatori, sia in capo alle autorità degli Stati chiamate ad applicare la disciplina: MARCHETTI, *La regolazione europea del mercato dell'intelligenza artificiale*, in *Rivista della regolazione dei mercati*, 2024, 1, 3 ss.

oppure, laddove i contesti siano collegati, un trattamento pregiudizievole ingiustificato o sproporzionato rispetto alla gravità del loro comportamento sociale.
Sono collocati tra i sistemi vietati quelli che creano o ampliano le banche dati di riconoscimento facciale mediante *scraping* non mirato di immagini facciali (art. 5, lett. e); quelli che identificano o inferiscono emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici (art. 5, lett. f), in situazioni – si noti – relative al luogo di lavoro e all'istruzione (eccezion fatta per quei sistemi immessi sul mercato esclusivamente per motivi medici o di sicurezza, come i sistemi destinati all'uso terapeutico); i sistemi di categorizzazione biometrica basati sui dati biometrici di persone fisiche per trarre deduzioni o inferenze in merito alle opinioni politiche, all'appartenenza sindacale, alle convinzioni religiose o filosofiche, alla razza, alla vita sessuale o all'orientamento sessuale di una persona (art. 5, lett. g).
È vietato (art. 5, lett. h) l'utilizzo di sistemi di identificazione biometrica remota – cioè di quei sistemi di intelligenza artificiale destinati all'identificazione, tipicamente a distanza, di persone fisiche, senza il loro coinvolgimento attivo, mediante il confronto dei dati biometrici con quelli contenuti in una banca dati di riferimento – operanti con modalità «in tempo reale»[37], in spazi accessibili al pubblico, a fini di attività di contrasto: tuttavia, l'impressione è di trovarsi al cospetto di una "tigre di carta" in quanto il divieto, posto a protezione della libertà di espressione e di riunione dei singoli individui, onde evitarne l'esposizione alla sorveglianza di massa, è "inflaccidito" da ampissime deroghe.
È questo un *topos* ricorrente nel tessuto del regolamento, dove a governare sono le "eccezioni" piuttosto che le regole, che non risparmia neppure i sistemi di intelligenza artificiale "predittiva". Sono proibiti quelli che consentono di «valutare o prevedere il rischio che una persona fisica commetta un reato» (art. 5, lett. d)[38], ciò in quanto tali sistemi sono capaci di infiltrare la presunzione di innocenza e di compromettere il diritto di ciascun individuo di essere giudicato in base al proprio «comportamento effettivo»[39].
Tuttavia, seguendo la stessa traiettoria sopra evidenziata, si è coniato un divieto a "maglie larghe", che vige soltanto se la previsione viene ricavata «unicamente sulla base della profilazione» del soggetto o «della valutazione dei tratti e delle caratteristiche della [sua] personalità», mentre non opera laddove i sistemi "predittivi" siano adoperati a supporto di una «valutazione umana», basata su «fatti oggettivi e verificabili direttamente connessi a un'attività criminosa». Quando non vietati, i *tools* che formulano la previsione sull'attitudine a delinquere di un soggetto sono classificabili come sistemi di intelligenza artificiale ad "alto rischio", dei quali si occupa l'art. 6 dell'*AI Act*. Tuttavia, tale inquadramento non risulta né agevole né condiviso.

---

[37] È necessario chiarire (come fa il diciassettesimo *considerando*), che nel caso dei sistemi *real time*, il rilevamento dei dati biometrici, il confronto e l'identificazione avvengono tutti istantaneamente o quasi e, in ogni caso, senza ritardi significativi, perché tali strumenti sfruttano il "flusso" di materiale «dal vivo» (ad esempio filmati) generato da una telecamera o da un altro dispositivo con funzionalità analoghe. Diversamente, nei sistemi *post remote* i dati biometrici sono già stati rilevati: confronto e identificazione, pertanto, avvengono con un ritardo significativo, comparando tali dati con materiale, come immagini o filmati generati da telecamere a circuito chiuso o da dispositivi privati, che è stato generato prima che il sistema fosse usato in relazione alle persone fisiche interessate.
[38] A commento della disposizione VIGORITO, sub *Art. 5 (comma 1 c-e)*, in *Intelligenza Artificiale. Commentario*, a cura di Mantelero-Resta-Riccio, Milano, 2025, 151 ss.
[39] Così il *considerando* n. 42.

5. L'intelligenza artificiale ad "alto rischio" e i rischi dell'"alto rischio". Prima di affrontare la controversa interpretazione delle pertinenti disposizioni, è opportuno offrire una panoramica più ampia circa la disciplina riguardante i sistemi ad "alto rischio".
La citata disposizione di cui all'art. 6 suddivide i sistemi di intelligenza artificiale ad "alto rischio" in due categorie. La prima, descritta nel paragrafo 1, si rivolge ai sistemi ricompresi nell'ambito applicativo di altre norme europee di armonizzazione, che ne subordinano la circolazione ad una preventiva valutazione di conformità da parte di terzi (cfr. Allegato I al regolamento (UE) 2024/1689). La seconda, di cui al paragrafo 2, individua i sistemi ad "alto rischio" «indipendenti», i quali, sebbene privi delle condizioni di cui all'Allegato I, rientrano nella categoria in forza del «rischio significativo di danno per la salute umana, la sicurezza o i diritti fondamentali delle persone fisiche (cfr. Allegato III)»[40]. Tra i settori individuati in quest'ultimo allegato del regolamento (i quali includono: la biometria; le infrastrutture critiche; l'istruzione e la formazione professionale; l'occupazione, la gestione dei lavoratori e l'accesso al lavoro autonomo; l'accesso a servizi essenziali e la loro fruizione; la migrazione, l'asilo e la gestione del controllo delle frontiere), due rilevano particolarmente per la giustizia penale: i sistemi – consentiti dal diritto dell'Unione e dello Stato membro[41] – impiegati per attività di *law enforcement*[42] (n. 6 dell'Allegato III)[43], nonché i sistemi dedicati all'«amministrazione della giustizia»[44] (n. 8 dell'Allegato III)[45].
Un sistema intelligente classificato ad "alto rischio" è assoggettato a requisiti di conformità e monitoraggio. In particolare, i fornitori di apparati ad "alto rischio" sono tenuti a garantire che i loro prodotti siano progettati ed utilizzati in modo da minimizzare i rischi di violazione dei diritti fondamentali, adottando misure di sicurezza tecniche e organizzative, conducendo

[40] L'efficace sintesi è di FERRETTI, *La relazione di cura e il consenso informato nell'era della Medical Artificial Intelligence*, in *Corti supreme e salute*, 2025, 1, 104.

[41] QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, cit., 10, sottolinea come tale precisazione escluda qualsiasi effetto di armonizzazione rispetto all'uso di tali strumenti nei singoli ordinamenti, discendente dal loro inserimento nell'Allegato.

[42] Per un quadro complessivo e critico, v. SACHOULIDOU, *Harnessing AI for law enforcement: Solutions and boundaries from the forthcoming AI Act*, in *New Journal of European Criminal Law*, 2024, 2 117 ss. Nel settore delle c.d. "attività di contrasto" ricadono, oltre ai sistemi di identificazione biometrica da remoto "in tempo reale", «i sistemi di IA in grado di: a) determinare il rischio, per una persona fisica, di diventare vittima di reati; b) valutare l'attendibilità delle affermazioni delle persone (una sorta di macchina della verità avanzata); c) valutare l'affidabilità degli elementi probatori raccolti nel corso delle indagini o del procedimento; d) determinare il rischio di commissione del reato o di recidiva in relazione a una persona fisica o per valutare i tratti e le caratteristiche della personalità o il comportamento criminale pregresso di persone fisiche o gruppi; e) profilazione delle persone fisiche nel corso dell'indagine, dell'accertamento e del perseguimento di reati»: per tale sintesi, TORRE, *Il Regolamento europeo sull'intelligenza artificiale: i profili processuali*, cit., 1548.

[43] ZOLEA, sub *Allegato III, punto 6*, in *Intelligenza Artificiale. Commentario*, a cura di Mantelero-Resta-Riccio, cit., 968 ss.

[44] In questo "settore" rientrano i «sistemi di IA destinati a essere usati da un'autorità giudiziaria o per suo conto per 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, o a essere utilizzati in modo analogo nella risoluzione alternativa delle controversie».

[45] A commento, ZOLEA, sub *Allegato III, punto 8a*, in*Intelligenza Artificiale. Commentario*, a cura di Mantelero-Resta-Riccio, cit., 984 ss.

valutazioni di impatto sulla protezione dei dati e la trasparenza nelle operazioni algoritmiche[46]. Più precisamente, essi dovranno garantire che i sistemi ad "alto rischio" siano conformi ai requisiti previsti dal regolamento (artt. 8-16) e, pertanto, dovranno: prevedere un sistema di gestione dei rischi (art. 9)[47]; assicurare la qualità e la governance dei dati (art. 10)[48]; redigere e mettere a disposizione delle autorità e degli organismi notificati la documentazione tecnica, volta a dimostrare che il sistema ad "alto rischio" è conforme ai requisiti (art. 11); consentire e conservare la registrazione automatica degli eventi («*log*») per la durata del ciclo di vita del sistema e garantire un livello di tracciabilità del suo funzionamento adeguato alla finalità prevista (art. 12); progettare e sviluppare sistemi trasparenti (onde consentire che gli output generati dall'artefatto possano essere compresi e correttamente interpretati dal *deployer*, a cui si dovranno fornire «istruzioni per l'uso» concise, complete, corrette, chiare, pertinenti, accessibili

[46] CIGNO-SCIMÈ, *La compliance per l'IA ad alto rischio (qualità conformità, ecc.)*, in *Il regolamento europeo sull'intelligenza artificiale*, a cura di Cassano-Tripodi, Santarcangelo di Romagna, 2024, 574.

[47] L'art. 9 del regolamento stabilisce l'obbligo di istituire un «sistema di gestione dei rischi», quale requisito essenziale dei sistemi di intelligenza artificiale 'ad alto rischio', che va «inteso come un processo iterativo continuo pianificato ed eseguito nel corso dell'intero ciclo di vita [...], che richiede un riesame e un aggiornamento costanti e sistematici», da documentare e mantenere nel tempo. Nelle varie fasi in cui si deve articolare la gestione dei rischi rientrano l'identificazione, l'analisi e la valutazione di quelli «noti e ragionevolmente prevedibili che il sistema [...] può porre per la salute, la sicurezza e i diritti fondamentali», sia quando è usato «conformemente alla sua finalità prevista», sia quando lo sia «in condizioni di uso improprio ragionevolmente prevedibile» (lettere a e b), mentre si devono considerare anche quelli «derivanti dall'analisi dei dati raccolti dal sistema di monitoraggio successivo all'immissione sul mercato» (lett. c). L'ultima fase è quella dell'adozione di «misure di gestione dei rischi opportune e mirate» (lett. d), che, secondo il par. 5, sono quelle per cui «i pertinenti rischi residui associati a ciascun pericolo nonché il rischio residuo complessivo dei sistemi sono considerati accettabili». Il legislatore europeo riconosce quindi, espressamente, un'area di "rischi consentiti" o "accettabili"; e l'adeguatezza delle misure che devono (se non eliminarli) "ridurli", è da commisurare alle «conoscenze tecniche, l'esperienza, l'istruzione e la formazione che ci si può aspettare dal *deployer* e [ad] il contesto presumibile in cui il sistema è destinato ad essere usato» (art. 9, par. 5, comma 3). Altra prescrizione da seguire nella gestione dei rischi, che merita di essere segnalata, è quella per cui tali sistemi devono essere «sottoposti a prova al fine di individuare le misure di gestione dei rischi più appropriate e mirate», garantendo che «funzionino in modo coerente per la finalità prevista e che siano conformi ai requisiti» stabiliti dal regolamento stesso (par. 6).

[48] I dati sono la fonte della conoscenza della macchina, sicché, in base al principio "*garbage in garbage out*", se i dati trattati dal sistema non possiedono un'adeguata qualità e sono inficiati da errori o pregiudizi, l'output algoritmico ne risentirà, dando esiti altrettanto scarsi e inattendibili.

e comprensibili) (art. 13)[49], efficacemente sorvegliabili dall'umano durante il periodo in cui sono in uso (art. 14)[50], nonché accurati, robusti e cybersicuri (art. 15).
La vera sfida risiede nel tradurre i suddetti requisiti normativi avanzati in realtà operative concrete e nell'assicurare il rispetto dei parametri giuridici[51].
Infatti, tra gli obblighi gravanti sui fornitori di intelligenza artificiale *high risk* (enunciati all'art. 16), tenuti a istituire un sistema di gestione della qualità conforme all'art. 17 del regolamento,

---

[49] Sotto questo profilo vengono in rilievo gli studi in materia di *explicable AI*, che, seppur in fase di sviluppo, aspirano nel prossimo futuro a risolvere il problema della opacità dei sistemi di giustizia predittiva. Su tali studi, cfr. BARRETO ARRIETA-DIAZ RODRIGUEZ-DEL SER-BENNETOT-TABIK-BARBADO-GARCIA-GIL LOPEZ-MOLINA-BENJAMINS-CHATILA-HERRERA, *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*, in *Information Fusion,* 2020, 58, 82 ss.; BRANTING-PFEIFER-BROWN, FERRO, ABERDEEN, WEISS-PTAFF-LIAO, *Scalable and explainable legal prediction*, in *Artificial Intelligence Law*, 2021, 2, 213 ss.; GUIDOTTI-MONREALE-RUGGIERI-TURINI-GIANNOTTI-PEDRESCHI, *A Survey of Methods for Explaining Black Box Models*, in *ACM Computing Surveys*, 2018, 5, 1 ss.; LIM, *Judicial Decision-Making and Explainable Artificial Intelligence*, in *Singapore Academy of Law Journal*, 2021, 33, 280 ss.
Un interessante rapporto sulla prototipazione dei requisiti di trasparenza del regolamento europeo sull'intelligenza artificiale è in GILS-HEYMANS-OOMS-DE BRUYNE, *From Policy to Practice: Prototyping The EU AI Act's Transparency Requirements*, 2024, s.l.
La differenza tra i termini spiegabilità e interpretabilità è evidenziata da BOLLÉ-CASEY-JACQUET, *The role of evaluations in reaching decisions using automated systems supporting forensic analysis*, in *Forensic Science International: Digital Investigation*, 2020, 34, 6.

[50] L'art. 14 impone un «obbligo di sorveglianza umana», fin dalla fase di progettazione e sviluppo dei sistemi, da realizzare «anche con strumenti di interfaccia uomo-macchina adeguati», in modo tale che essi possano «essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso» (par. 1). Sebbene la sorveglianza umana non possa riguardare ogni singola fase operativa dei sistemi di intelligenza artificiale, la cui utilità è data proprio dalla loro autonomia, nondimeno deve «prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di intelligenza artificiale ad "alto rischio" è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile, in particolare qualora tali rischi persistano nonostante l'applicazione di altri requisiti di cui alla presente sezione» (par. 2).
In senso critico LAUX, *Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act*, in *AI & Society,* 2024, 39, 2853 ss., rileva come l'*AI Act* non fornisca indicazioni su come ottenere un'efficace supervisione, lasciando poco definiti gli obblighi per gli sviluppatori dei sistemi.
Si possono distinguere almeno tre approcci (*optional approach, benchmark approach e feedback approach*) per combinare la valutazione umana con la valutazione quantitativa del rischio: cfr. VAN DIJCK, *Predicting Recidivism Risk Meets AI Act*, in *European Journal on Criminal Policy and Research*, 2022, 3, 419 s.

[51] FOTI, *AI Act: criticità e sfide attuative, tra ambizioni regolatorie e complessità applicative*, in *Altalex*, 4 luglio 2025. Per una panoramica sulle difficoltà nell'operazionalizzazione del rischio di recidiva e della sua incidenza sulla "incoerenza predittiva", si rimanda all'interessante studio di GREENE-SHMUELI-FELL-LIN-LIU, *Forks over knives: Predictive inconsistency in criminal justice algorithmic risk assessment tools*, in *Journal of the Royal Statistical Society Series A: Statistics in Society*, 2022, 2, 701. Gli A. propongono di adottare deviazioni ragionevoli (*forks*) da un singolo percorso di sviluppo (coltelli) per elaborare strumenti di valutazione del rischio di prossima generazione (basati su algoritmi di apprendimento automatico più automatizzati e flessibili con funzionalità integrate di acquisizione dati e selezione delle variabili) che siano riproducibili, spiegabili e democraticamente responsabili.

documentandolo in maniera sistematica e ordinata attraverso procedure e istruzioni scritte, vi è quello di sottoporre tali sistemi alla pertinente procedura di valutazione della conformità (art. 43), prima di immetterli sul mercato o in servizio: ma è proprio su queste procedure che si addensano dubbi circa l'effettiva capacità della disciplina regolamentare di tutelare i diritti fondamentali.

In termini generali, il fornitore può scegliere di effettuare la valutazione di conformità[52] dei sistemi ad "alto rischio" secondo la procedura basata sul controllo interno oppure tramite organismi certificati. Solo alcune scarse eccezioni vincolano la scelta della procedura da seguire, limitandola al modello del controllo preventivo affidato al terzo (art. 43, par. 1, secondo periodo).

Tra queste eccezioni non figura l'intelligenza artificiale impiegabile nell'ambito processuale penale, sottratta pertanto al test di compatibilità esterno.

Così, chi dovrebbe essere controllato finisce per coincidere con il controllore stesso, in quanto il meccanismo fonda su una semplice autodichiarazione: il sistema dei controlli, «accontentandosi [della] c.d. *self-compliance*, che attesti il rispetto della disciplina prudenziale», risulta inadeguato ad assicurarne la puntuale osservanza[53].

Analoghe considerazioni possono farsi per quanto riguarda la «valutazione d'impatto sui diritti fondamentali» (*ex* art. 27)[54]: anche questa valutazione anticipata si risolve in una autodichiarazione dello stesso fornitore, circostanza che, invece di minimizzare i rischi mediante opportune misure di mitigazione, potrebbe indurre a minimizzare la rappresentazione dei pericoli.

Si potrebbe obiettare che l'*AI Act* prevede comunque un sistema di controlli *ex post* da parte delle autorità preposte e meccanismi sanzionatori[55]; ma, come condivisibilmente rilevato, «l'intervento pubblico tardivo è un ossimoro per una disciplina che ha l'ambizione di essere precauzionale e la pretesa di non compromettere i diritti»[56].

Si segnala, non ultimo per importanza, un ulteriore aspetto critico: i sistemi intelligenti riconducibili ai settori individuati nell'Allegato III sono classificati ad "alto rischio" sulla scorta di una presunzione *iuris tantum*, superabile se l'artefatto intelligente non influenza in modo

---

[52] Per commenti preliminari sulle valutazioni di conformità, v. MÖKANDER-AXENTE-CASOLARI-FLORIDI, *Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation*, in *Minds&Machines*, 2022, 32, 241 ss.

[53] DE MINICO, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, in *Rivista AIC*, 2024, 2, 87. Analogamente, hanno espresso critiche sugli schemi autodichiarativi di *self-compliance*, FALLETTA–MARSANO, *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza a artificiale e GDPR*, in *Riv. it. informatica e dir.*, 2024, 1, 124 s.; NOVELLI, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *Federalismi*, 2024, 1, 111.

[54] Per un inquadramento della valutazione dell'impatto sui diritti fondamentali prevista dall'*AI Act* e per la proposta di criteri metodologici che dovrebbero essere seguiti nella sua elaborazione, v. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, 2024, 54, 1 ss.

[55] Sul complesso sistema di governance previsto dal regolamento sull'intelligenza artificiale, che si articola su un doppio livello (quello sovranazionale e quello nazionale), e sulle sue criticità, si rinvia a VILLANI, *Il sistema di vigilanza sull'applicazione dell'AI Act: ognun per sé?*, in *Quad. AISDUE*, 2024, 2, 1 ss.

[56] DE MINICO, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, cit., 88.

sostanziale la decisione umana, in quanto, ad esempio, destinato ad eseguire un «compito procedurale limitato», la cui individuazione – però - è interamente rimessa all'interprete[57].
Pur tra luci e ombre, l'ampio corredo di prescrizioni impartite dal regolamento concorre a delineare una serie di condizioni di liceità dell'impiego, anche in ambito processuale penale, delle tecnologie basate sull'intelligenza artificiale[58].

6. *Classificazione degli AI RATs per la valutazione della pericolosità sociale.* Le criticità che si addensano intorno alle tutele apprestate dall'*AI Act* per mitigare i pericoli dell'intelligenza artificiale ad "alto rischio", unite a quelle proprie dei *tools* in oggetto, fornirebbero già solidi argomenti per condividere l'interpretazione secondo la quale il regolamento (UE) 2024/1689 ne vieterebbe in assoluto l'impiego.
Sulla scorta della premessa secondo cui questi *tools* si basano proprio «sulla valutazione dei tratti e delle caratteristiche della personalità», si è concluso che la disposizione di cui all'art. 5, lett. d), del citato regolamento enuncerebbe un «divieto generalizzato di immissione sul mercato di software di *risk assessment*», e che tale divieto prevarrebbe sulla contraddittoria disposizione dell'*AI Act* contenuta nell'Allegato III, sub n. 6, lett. d), che, classificando gli stessi sistemi come ad "alto rischio", sembrerebbe ammettere ciò che l'art. 5 vieta[59].
L'Allegato III, richiamato dall'art. 6 del regolamento, classifica ad "alto rischio" «i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto o per loro conto» al fine di «determinare il rischio di commissione del reato o di recidiva [...] non solo sulla base della profilazione delle persone fisiche»[60] o «per valutare i tratti e le caratteristiche della personalità o il comportamento criminale pregresso di persone fisiche o gruppi»[61].

---

[57] In tal senso DONATI, *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, in *RivistaAIC*, 2025, 1, 15, che auspica «che la Commissione, dopo avere consultato il Consiglio europeo per l'intelligenza artificiale, adotti al più presto orientamenti per guidare fornitori, utilizzatori e autorità di sorveglianza fornendo altresì un elenco esaustivo di esempi pratici di casi d'uso di sistemi di IA ad alto rischio e non ad alto rischio». Per una più ampia panoramica sui *deficit* di tutela dei diritti fondamentali nell'*AI Act*, cfr. CIRONE, *L'*AI Act *e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in *Quad. AISDUE*, 2024, 2, 51 ss.

[58] GALLUCCIO MEZIO, *Tecnologie di riconoscimento facciale: una riflessione sul loro impiego con finalità investigative e probatorie*, in *Cass. pen.*, 2025, 2, 667, secondo cui la violazione delle specifiche condizioni di impiego delle tecnologie di intelligenza artificiale, previste dal regolamento, ne determina la loro "emarginazione" dal procedimento penale, quantomeno a far data dal momento di effettiva applicabilità delle diverse sezioni che compongono il complesso *corpus* normativo. L'A. ritiene che la violazione delle nuove regole europee, aventi natura *self executing*, «qualifichi come vietato dalla disciplina europea l'impiego di tali tecnologie in termini non compatibili con la suddetta regolamentazione» e «ne preclude tanto l'ammissione e l'utilizzabilità probatoria, tanto l'impiego con finalità puramente investigative».

[59] QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, cit., 6 s.

[60] Per "profilazione", ai sensi dell'art. 4, n. 4 del *General Data Protection Regulation* (GDPR), al quale rinvia l'art. 3, punto 52) dell'*AI Act*, si intende «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica».

[61] A questo ambito sono riconducibili anche le attività, specialmente pre-investigative, proprie della c.d. polizia predittiva (*predictive policing*). I *software* di polizia predittiva sono raggruppabili in almeno due

Sebbene la formulazione di entrambe le disposizioni citate non appaia perspicua, tuttavia, non si ritiene che il legislatore europeo abbia inteso vietare *tout court* i *risk assessment tools*, ma soltanto quelli che si basano "unicamente" sul *profiling* o sulle caratteristiche della personalità, quando – aggiungasi – non siano adoperati a supporto delle valutazioni umane effettuate sulla scorta di elementi fattuali obiettivi e connessi all'attività criminosa.
A siffatta conclusione induce l'eccezione di cui al secondo periodo dell'art. 5, lett. d); sennonché, anche la formulazione di questa "deroga" ha fatto sorgere difficoltà interpretative, in quanto la si è intesa rivolta «a sistemi che non sono *risk assessment*», ovvero riferita a «metodiche di *digital forensics* che mirano a valutare la verosimiglianza di possibili ricostruzioni causali, anche alternative, dei fatti oggetto del procedimento penale»[62].
Il contesto in cui è allocata la disposizione di esonero porta a ritenere, con il conforto delle recenti *Guideliness* sull'art. 5 diffuse dalla Commissione europea[63], che la stessa sia pur sempre riferibile ai sistemi di *risk assessment* di cui qui si discorre, i quali – ove a supporto della valutazione umana "individualizzante" – parrebbero "scivolare" dalle strette del divieto per refluire nella sfera dell'intelligenza artificiale ad "alto rischio"[64].
Donde la conclusione per cui questi *tools*, quando non vietati, costituiscono "pratiche" di intelligenza artificiale ad "alto rischio" e, dunque, assoggettati alla *compliance* stabilita per gli strumenti così classificati, sia pur con tutte le perplessità precedentemente espresse a proposito dei meccanismi di conformità previsti dal regolamento europeo.

---

categorie: i *place-based system,* che forniscono indici di rischio sulla commissione di futuri reati in certi luoghi, e i *person-based system*, volti a delineare i profili dei soggetti maggiormente inclini a divenire possibili autori di condotte criminali.
In argomento, tra i tanti e senza pretesa di esaustività, ALGERI, *Intelligenza artificiale e polizia predittiva*, in *Dir. pen. proc.*, 2021, 6, 730; ID., *Intelligenza artificiale e processo penale*, in *Cybercrime*, diretto da Cadoppi-Canestrari-Manna-Papa, Torino, 2023, 1767; BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in *Diritto penale e intelligenza artificiale* a cura di Balbi-De Simone-Esposito-Manacorda, Torino, 2022, 6; CAMALDO, *Intelligenza artificiale e investigazione penale predittiva*, in *Riv. it. dir. proc. pen.*, 2024, 1, 235; ESPOSITO, *Nuovi strumenti di lotta all'illegalità, riflessioni nell'era dell'intelligenza artificiale*, in *Pen. dir. proc.*, 2024, 3, 406; FERGUSON, *Policing Predictive Policing*, in *Washington University Law Review*, 2017, 5, 1123 ss.; PADUA, *Intelligenza artificiale e giudizio penale*, cit., 1491 ss.; PIETROCARLO, *La* predictive policing *nel regolamento europeo sull'intelligenza artificiale*, in *Leg. pen.*, 2024, 3, 340; POLIDORO, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale*, in *Arch. pen. web*, 2020, 3, 6 s.; SEVERINO, *Intelligenza artificiale e diritto penale*, *Intelligenza artificiale. Il diritto, i diritti, l'etica*, a cura di Ruffolo, Milano, 2020, 536 ss.

[62] QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, cit., 6.

[63] La Commissione europea ha recentemente pubblicato gli *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA)*, 29 luglio 2025, in *https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act*. Le *Guidelines*, in corso di formale adozione da parte della Commisione europea, sebbene non vincolanti, forniscono una panoramica delle pratiche considerate "inaccettabili", mirando a garantire l'applicazione coerente, efficace e uniforme dell'*AI Act* in tutta l'UE.

[64] Cfr. *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA)*, cit., 73 ss., specialmente parr. 192, 201, 204, 206, 214.

Mette conto evidenziare che l'inaccettabilità o l'"accettabilità condizionata" (quella dell'intelligenza artificiale classificabile ad "alto rischio") viene delineata nell'*AI Act* in maniera statica e verticale (i quattro livelli di rischio sono parametrati, in sostanza, agli ambiti operativi che possono mettere in pericolo i diritti fondamentali e, pertanto, sono aprioristicamente identificati[65], trascurandosi le interazioni dinamiche tra le fonti di pericolo e i profili di vulnerabilità dei soggetti che ne sono esposti)[66]. Per queste ragioni si è proposto di approcciare alle quattro categorie di rischio secondo una logica orizzontale, applicando un modello basato su scenari, in modo che lo stesso sistema possa essere flessibilmente categorizzato come "inaccettabile", ad "alto rischio", a "rischio limitato" o a "rischio minimo", tenuto conto delle interazioni tra i vari fattori di rischio presenti in situazioni reali. L'approccio, definibile come "semi-quantitativo", si articola in due distinte fasi: la costruzione degli scenari di rischio e la valutazione quantitativa basata sulla proporzionalità. La costruzione dello scenario di rischio include i seguenti fattori determinanti: *hazard* (H), *exposure* (E), *vulnerability* (V)[67], *response* (R), *extrinsic risks*. Una volta tratteggiato lo scenario di rischio, il secondo passaggio consiste nel valutare, mediante un test di proporzionalità (basato sulla formula del peso di Alexy)[68], se la categorizzazione prevista dal regolamento risulti appropriata – vuoi perché i valori in gioco sono correttamente bilanciati, vuoi perché essa riflette la reale esposizione ai rischi emergente dallo scenario concreto – oppure occorra rivederla. A questo approccio si farà riferimento per trarre alcune conclusioni minime.

7. *Prospettive di impiego: il nodo sulla qualificazione dell'output algoritmico.* Nel nostro sistema penale il vaglio sulla pericolosità sociale di un individuo, se basato sui metodi tradizionali, è «attività complessa ed articolata sulla base di molteplici fattori, condotta dal giudicante secondo un ragionamento logico e razionale», che si interseca o può intersecarsi con quello sulla personalità dell'individuo[69].
L'art. 220, co. 2 c.p.p. sancisce il divieto di c.d. "perizia criminologica", ossia l'attività di osservazione scientifica dell'imputato volta a metterne in luce aspetti della personalità[70].

---

[65] NOVELLI-CASOLARI-ROTOLO-TADDEO-FLORIDI, *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, cit., 11 ss.

[66] In altri termini, la valutazione del rischio sottesa al regolamento europeo non riflette un sufficiente grado di granularità: così CANATO, *Verso il superamento del "legal risk" europeo: intelligenza artificiale e approccio proporzionale al rischio*, cit., 95.

[67] In termini generali, il pericolo si riferisce alle fonti di potenziali effetti avversi sugli elementi esposti; l'esposizione si riferisce all'inventario degli elementi all'interno dell'intervallo della fonte di pericolo; la vulnerabilità si riferisce all'insieme di attributi o circostanze che rendono gli elementi esposti suscettibili agli effetti avversi quando hanno un impatto sulla fonte di pericolo; la risposta riguarda le misure esistenti che contrastano o mitigano il rischio.

[68] ALEXY, *A theory of constitutional rights*, Oxford, 2002; ID., *On balancing and subsumption. A structural comparison*, in *Ratio Juris*, 2003, 4, 433 ss.

[69] MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, cit., 229 s.

[70] Per le ragioni a fondamento di tale divieto cfr. ADORNO, *sub Art. 220 c.p.p.*, in *Codice di procedura penale commentato*[6], a cura di Giarda-Spangher, Milano, 2023, 2997 s.

Nel giudizio, pertanto, la valutazione della personalità dell'imputato trova sparuti spazi, potendo il giudice svolgere tale indagine per la commisurazione della pena (art. 133 c.p.) o l'applicazione delle misure di sicurezza (art. 203 c.p.).
Guardando alla fase delle indagini, con particolare riferimento alle misure cautelari, il profilo attinente alla personalità dell'imputato assume rilievo rispetto all'esigenza di cui all'art. 274, lett. c),c.p.p., desumibile «da comportamenti o atti concreti» o dai «precedenti penali».
Nella fase dell'esecuzione della pena, invece, il giudizio personologico gode di maggiore spazio: l'inciso «salvo quanto previsto ai fini dell'esecuzione della pena» autorizza in tale ambito l'effettuazione di accertamenti di natura criminologica, «in armonia con quella concezione "bifasica" del processo, che colloca in un secondo momento la concreta determinazione del trattamento penale»[71].
Per ciascuna delle suindicate fasi potrebbero immaginarsi ampi e promettenti impieghi degli *AI RATs*, ad esempio, in termini di contributo al raggiungimento dell'agognato obiettivo della «*certezza* del giudizio discrezionale» anche rispetto alla quantificazione della pena[72], da cui discendono a cascata tutta una serie di effetti, tra cui la realizzazione della finalità rieducativa della pena stessa[73]. Prospettive di utilizzo potrebbero intravedersi in relazione alla valutazione del pericolo di reiterazione del reato, *ex* art. 274, lett. c) c.p.p., esigenza – tra quelle contemplate nel sistema cautelare codicistico – divenuta egemonica nella prassi[74]. Invitano a immaginare un tale scenario d'uso i numerosi fallimenti legislativi volti a minimizzare il ricorso alla cautela e a ridurre la sfera di discrezionalità del giudice[75], onde contrastare l'impiego degli strumenti

[71] In questi termini, BOCCELLARI-LIANI, *sub Art. 220 c.p.p.*, in *Codice di procedura penale commentato*[6], a cura di Giarda-Spangher, cit., 3022.
[72] BRICOLA, *La discrezionalità nel diritto penale*, vol. I, *Nozione e aspetti costituzionali*, Milano, 1965, 83 ss.
[73] *Ibid.*, 100.
[74] GUIDO, *Intelligenza artificiale e procedimento penale: ragionando di valutazione del rischio* de libertate, in *Arch. pen. web*, 2023, 1, 10.
[75] Gli interventi legislativi che hanno tentato di circoscrivere il perimetro operativo della disciplina di cui all'art. 274, lett. c) c.p.p., non sono riusciti a rafforzare le garanzie volte a contenere la discrezionalità del giudice. Tra essi, merita di essere ricordato quello portato dall'art. 2 L. 16 aprile 2015, n. 47, con cui il legislatore ha stabilito che il pericolo di reiterazione deve essere non solo attuale ma anche concreto, e ha vietato che lo stesso sia desunto dalla sola gravità del titolo di reato per cui si procede. Sul tema, v. CIAVOLA, *La valutazione delle esigenze cautelari*, in *La riforma delle misure cautelari personali*, a cura di Giuliani, Torino, 2015, 59 ss.; CORTESI, *Interventi sulle misure custodiali*, in *Misure cautelari ad personam in un triennio di riforme*, a cura di Diddi-Geraci, Torino, 2015, 20 ss.; IASEVOLI, *Concretezza ed attualità delle esigenze cautelari*, in *Il rinnovamento delle misure cautelari*, a cura di Bene, Torino, 2015, 19 ss.; TURCO, *La riforma delle misure cautelari*, in *Proc. pen. giust.*, 2015, 5, 107.

cautelari per il perseguimento di finalità spurie[76], con buona pace della presunzione di innocenza[77]. *In executivis*, poi, il magistrato di sorveglianza potrebbe servirsi dei *tools* non solo ai fini dell'ammissione alle misure alternative o della concessione di benefici penitenziari[78], ma anche per strutturare i contenuti e gli obiettivi del programma di trattamento[79]. Il magistrato di sorveglianza accerta la qualità di «persona socialmente pericolosa» sulla base – ancora – dei criteri di cui all'art. 133 c.p.p. (la personalità del reo, la sua condotta di vita, il contesto sociale e familiare di provenienza, le modalità e le caratteristiche del fatto commesso e le sue conseguenze), con valutazione che verte, in modo particolare, sui dati anamnestici, sulle perizie psichiatriche e su quelle criminologiche, sull'esistenza di precedenti penali e di procedimenti pendenti, sugli elementi desumibili dalla sentenza di condanna, sul comportamento *post delictum* e sull'andamento del percorso trattamentale.

L'impiego dell'intelligenza artificiale per valutare la pericolosità in ciascun segmento processuale, tra quelli evocati, pone delle specifiche criticità, per affrontare le quali risulta propedeutico interrogarsi sull'inquadramento da dare al contributo generato dagli *AI RATs*[80].

---

[76] La compatibilità dell'esigenza cautelare in questione con i valori costituzionali è sempre apparsa problematica, in quanto la prognosi di recidivanza anticipa impropriamente il giudizio sulla sua responsabilità, ancora da accertare. Ampie "tracce" di questa tensione si trovano negli scritti – solo per citarne alcuni – de: MAZZA, *Le persone pericolose (in difesa della presunzione di innocenza)*, in *www.penalecontemporaneo.it*, 20 aprile 2012, 7; PASTA, *Lo scopo del processo e tutela dell'innocente: la presunzione di non colpevolezza*, in *Arch. pen. web*, 2018, 1, 19; PAULESU, *La presunzione di non colpevolezza*, Torino, 2008, 119 ss.; PRESUTTI, *Gli incerti confini delle esigenze cautelari: le cautele come forma di anticipazione della pena*, in *Le fragili garanzie della libertà personale per una effettiva tutela dei principi costituzionali*, Milano, 2014, 43 ss.; VIGONI, *La fisionomia tridimensionale della presunzione d'innocenza: profili di sviluppo della disciplina codicistica*, in *Proc. pen. giust.*, 2023, 1, 233 ss.

[77] Sulla presunzione di non colpevolezza imprescindibile il rinvio agli studi monografici di ILLUMINATI, *La presunzione d'innocenza dell'imputato*, Bologna, 1979; PAULESU, *La presunzione di non colpevolezza dell'imputato*, cit.

[78] L'affidamento in prova al servizio sociale è possibile solo quando «assicuri la prevenzione del pericolo che [il condannato] commetta altri reati» (art. 47 ord. penit.); la detenzione domiciliare è accessibile solo se la misura è «idonea ad evitare il pericolo che il condannato commetta altri reati» (art. 47-*ter*, c. 1-*bis* ord. penit.); non diversamente, la detenzione domiciliare speciale può essere concessa solo «se non sussiste un concreto pericolo di commissione di ulteriori delitti (art. 47-*quinquies* ord. penit.). Inoltre, la semilibertà sottintende la valutazione circa la pericolosità sociale, in quanto il condannato può essere ammesso a tale misura «quando vi sono le condizioni per un graduale reinserimento (...) nella società» (art. 50 ord. penit.). Ancora, la pericolosità sociale rileva anche per la concessione dei permessi premio (art. 30-*ter* ord. penit.) perché, accanto al requisito della regolare condotta, si deve valutare l'assenza di pericolosità sociale del condannato.

[79] Di tale avviso CANESCHI, *Intelligenza artificiale e sistema penitenziario*, in *Riv. it. dir. proc. pen.*, 2024, 1, 263 s., secondo la quale «non è seriamente pensabile che l'ingresso di algoritmi predittivi nell'ordinamento nazionale possa essere escluso, tanto più con riferimento alla fase di esecuzione della pena in cui il divieto di esprimere giudizi personologici sul condannato è decaduto». Sul possibile utilizzo degli strumenti di *risk assessment* per la predisposizione del trattamento penitenziario si rinvia a ZARA, *Tra il probabile e il certo. La valutazione del rischio di violenza e di recidiva criminale*, in *www.penalecontemporaneo.it*, 20 maggio 2016, 4 ss.

[80] Sul punto, v. MACRÌ, *I primi passi dell'Italia verso l'impiego dell'IA nel processo penale e il calcolo del rischio di recidiva*, in *Giur. pen. web*, 2025, 2, 11.

Infatti, se si dovesse ritenere, con parte della dottrina, di qualificare l'output generato dal procedimento algoritmico alla stregua di una perizia psicologica e criminologica, se ne dovrebbe concludere, stante il divieto stabilito nell'art. 220, co. 2 c.p.p. per l'inammissibilità dei software di *risk assessment* nel giudizio.

Seppur vero che alla testimonianza esperta fanno riferimento i Paesi in cui si è sedimentata l'applicazione dei *tools* predittivi di cui si discorre (in particolare, Stati Uniti e Canada)[81], in realtà tali sistemi non sono affatto paragonabili ai periti[82], non essendo capaci di esplorare il foro interiore della persona e non essendo interrogabili, mediante la tecnica della *cross-examination* con contestazioni e richiesta di precisazioni[83].

Sembra, piuttosto, che il rischio della valutazione della pericolosità affidata alla macchina risieda nel fatto che «la classificazione dei comportamenti umani [avvenga] per categorie d'autore, senza valorizzazione della singolarità»; di conseguenza, proprio «per il mancato approfondimento delle particolarità del caso specifico», il risultato algoritmico «si assesta ad un livello di generalità tale da distinguersi nettamente da quello della perizia criminologica»[84].

L'insidia che si cela nell'utilizzo degli strumenti predittivi consiste nel basare il giudizio sulle condotte commesse in passato da altri soggetti, diversi dall'interessato, e nell'avviare «forme di determinismo penale», con conseguente passaggio «dal diritto penale del fatto», *ex* art. 25 co. 2 Cost., «ad un diritto penale del profilo d'autore» – in cui «la pericolosità di un soggetto sia stabilita sulla base di modelli comportamentali facenti parte di un certo ambito o di un

---

[81] La giurisprudenza canadese ha ricondotto l'output generato dal *risk assessment tool* al parere offerto dall'esperto, richiedendo che, per la sua ammissibilità, siano soddisfatti i criteri e le condizioni a riguardo cristallizzati dalle sentenze *Mohan* e *White Burgess.* Un requisito concerne l'adeguata qualifica dell'esperto: con riferimento a questo parametro si addensano dubbi e perplessità, specialmente perché la capacità del tribunale di valutare se la macchina esperta possegga una conoscenza tale da riflettere un "apprendimento sistematico" o una "valutazione sistematica dei dati" potrebbe essere compromessa, tra l'altro, dall'impossibilità di sottoporre l'algoritmo a controinterrogatorio. Per ampi riferimenti v. il *paper* della Law Commission of Ontario, *AI in criminal justice project*|Paper 3, *AI and the Assessment of Risk in Bail, Sentencing and Recidivism*, aprile 2025, 32 ss.

[82] Questi particolari testimoni esperti non svolgerebbero il compito cui è chiamato il perito, quello di prospettare, mediante un contributo reso di regola oralmente, una ricostruzione delle leggi scientifiche, tecniche o artistiche idonee ad offrire una spiegazione di elementi relativi al fatto da accertare: in tal senso, BACCARI-PECCHIOLI, *I.A. e giudizio sul fatto: gli strumenti di e-evidence per la cognizione*, cit., 128.

[83] Cfr. MAZZA-TOGNAZZI, *La valutazione della pericolosità sociale* ante *e* post delictum, cit., 215, ai quali «non pare che la raccolta, l'elaborazione e il confronto di dati, anche conferenti la soggettività dell'individuo, possano considerarsi un'esplorazione del foro interiore della persona». In senso analogo, MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, cit., 241, per la quale è «improprio» ricondurre i *r.a.t.* nell'ambito di operatività del divieto di cui all'art. 220, co. 2 c.p.p. Cfr. altresì ROMANÒ, *Intelligenza artificiale come prova scientifica nel processo penale: una sfida tra machine-generated evidence e equo processo*, in *Prova scientifica e processo penale*, a cura di Canzio-Luparia Donati, cit., 932, per il quale «l'output della macchina non può essere accomunato ad una perizia psicologica per almeno due ordini di ragioni: l'operazione computazionale non corrisponde all'attività dello psicologo fondata sul dialogo con il paziente e rappresenta un risultato indicativo di una statistica generale relativa a casi simili avvenuti nel passato».

[84] Cfr. MAZZA-TOGNAZZI, *La valutazione della pericolosità sociale* ante *e* post delictum, in *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, a cura di Baccari-Felicioni, cit., 215 s.

determinato periodo» – che, in quanto tale, contrasta con il principio di individualizzazione del trattamento sanzionatorio (previsto dall'art. 27, co. 1 e 3 Cost.)[85].
Parametri costituzionali, questi ultimi, ineludibili, anche laddove – poco condivisibilmente – si volesse evocare per tali *tools* la prova atipica[86].
Anche a prescindere dalla condivisibilità della prospettiva che colloca il contributo di questi apparati al di fuori degli inquadramenti teorico-concettuali degli apporti gnoseologici dell'intelligenza artificiale in ambito probatorio[87], neppure il richiamo ai mezzi di prova tipici è in grado di restituire la complessità del fenomeno.
Se è ragionevole riferirsi alla prova scientifica basata su una *novel science*[88], questa non necessita di essere qualificata o veicolata attraverso le forme della perizia (o della consulenza tecnica), in

[85] MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, cit., 241 s. In termini non dissimili, v. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *www.penalecontemporaneo.it*, 29 maggio 2019, 21; MANES, *L'oracolo algoritmico e la giustizia penale al bivio tra tecnologia e tecnocrazia*, in *www.discrimen.it*, 15 maggio 2020, 1 ss.
Secondo ECKHOUSE-LUM-CONTI COOK-CICCOLINI, *Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment*, in *Criminal Justice and Behavior*, 2019, 2, 198, la questione centrale è che le valutazioni del rischio di recidiva basate sull'IA facilitano il processo decisionale sulla libertà di un individuo in base al comportamento altrui. Nelle conclusioni, a p. 205, gli A. affermano che: «even if the risk scores were unbiased (which they are not), the numbers do not speak for themselves. We have to use human insight and human judgment to decide what they mean and when we should use them. In doing so, policymakers and judges need to consider all three layers of bias», «algorithmic fairness, data bias, and the inherent justice of using group-based decision-making», «and develop legal frameworks that promote transparency, accurate measurement, and just decision-making».
[86] Come affermato da TOMBELLI, *La tutela della corrispondenza tra atipicità della prova e tentativi di riforma*, in *www.penaledp.it*, 16 gennaio 2025, «non risulta possibile invocare l'art. 189 c.p.p. per consentire l'ingresso di prove atipiche lesive dei diritti costituzionalmente garantiti da riserva di legge e di giurisdizione». Come precisato da DINACCI, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici*, in *Proc. pen. giust.*, 2022, 6, 1630, «le realtà storiche derivanti o ricavabili dalle macchine dell'intelligenza artificiale non possono essere introdotte nel processo attraverso lo schema dell'art. 189 c.p.p.», ciò «in quanto non sembra controvertibile che la "conoscenza" derivante dalla intelligenza artificiale risulti perfettamente inquadrabile nell'atto peritale o, comunque, nella prova tecnica».
[87] Parrebbero di questo avviso, in quanto la predizione di che trattasi «non soccorre quale strumento per la cognizione del giudice in relazione alla *quaestio facti* da accertare», BACCARI-PECCHIOLI, *I.A. e giudizio sul fatto: gli strumenti di e-evidence per la cognizione*, cit., 118 ss., i quali richiamano la distinzione tra *digital evidence* (meccanismi algoritmici come mezzi di ricerca della prova), *e-evidence* (prova algoritmica in senso stretto, anche definita come *automated evidence* o *machine/AI generated evidence*) e *automated evaluation of evidence* (macchine intelligenti che possono coadiuvare il giudice nell'opera di valutazione probatoria). Riconduce l'output dei software di *risk assessment*, invece, al novero delle prove generate automaticamente, QUATTROCOLO, *Prova e intelligenza artificiale*, in *La prova scientifica*, a cura di Conti-Marandola, Milano, 2023, 480, secondo la quale la valutazione automatizzata di pericolosità può utilmente rientrare nel paradigma della prova peritale, sempre che sia possibile estrarre dal software una chiara, attendibile e accurata teoria scientifica, a sua volta validata, e che il suo funzionamento sia trasparente e spiegabile.
[88] Fornendo al giudice non tanto elementi di «valutazione», quanto elementi «da valutare», il risultato dell'intelligenza artificiale è stato inquadrato «alla stregua di una prova scientifica», sia pur «con le dovute

quanto la macchina è autonomamente in grado di elaborare l'output[89], potendosi semmai ricorrere all'istituto peritale se e quando dovessero sorgere questioni relative all'apparato di intelligenza artificiale[90].

Ancora, l'output non appare accostabile neppure ad un documento, categoria a cui sempre più disinvoltamente si fa riferimento quando si è al cospetto di "surrogati tecnologici"[91]; infatti, il prodotto algoritmico non è la semplice cristallizzazione di dati rappresentativi di fatti, persone o cose, ma è qualcosa di più: è il risultato della loro autonoma elaborazione da parte della macchina, persino capace di autoapprendimento, niente affatto paragonabile all'incorporamento che caratterizza il mero documento.

Una suggestiva tesi, sebbene prospettata in altro contesto predittivo, ha proposto di paragonare il prodotto del *tool* algoritmico «ad una sorta di parere amministrativo», che si vorrebbe di natura semi-vincolante (ovvero obbligatorio, ma "non vincolante")[92]. Un siffatto inquadramento – si sostiene -, che considera la macchina come «una sorta di "consigliere *a latere*" del giudice umano», esclude che la previsione algoritmica possa essere trattata alla stregua di un elemento di prova e consente di superare il problema del contraddittorio "per" e "sulla" prova attraverso il ricorso al *Daubert test* e ai criteri enunciati dalla Corte suprema statunitense[93].

L'accostamento della macchina al consigliere *a latere* appare ardito, perché finisce per attribuirle compiti di vero e proprio affiancamento nella decisione, ipotesi evidentemente in contrasto con diversi parametri costituzionali[94]. Meno problematico ravvisare nel contributo dell'intelligenza artificiale quello della figura di "*amicus curiae*" o quello che altre figure "ibride" forniscono al magistrato, supportandone l'attività, con modalità che incrociano (quando non

---

precauzioni», da MAZZA-TOGNAZZI, *La valutazione della pericolosità sociale* ante *e* post delictum, cit., 221. In senso analogo, GIALUZ, *Intelligenza artificiale e diritti fondamentali in ambito probatorio*, in *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milano, 2021, 61 ss., reputa che la prova fondata sull'intelligenza artificiale rientra nell'ambito della prova scientifica.

[89] In tale senso, con riferimento al *facial recognition*, GALLUCCIO MEZIO, *Tecnologie di riconoscimento facciale: una riflessione sul loro impiego con finalità investigative e probatorie*, cit., 649.

[90] UBERTIS, *Perizia, prova scientifica e intelligenza artificiale nel processo penale*, in *www.sistemapenale.it*, 3 giugno 2024, 18.

[91] DEGANELLO, *La "tessera mancante": la prova "per documenti" nel processo penale italiano*, in *www.associazionelaic.it*, 16 dicembre 2022, sottolinea che «oltre al *passepartout* classico di recupero al compendio probatorio di un *quid* non legislativamente predeterminato (ovvero l'art. 189 c.p.p. dalla significativa Rubrica normativa "*Prove non disciplinate dalla legge*" del cui uso disinvolto e/o arbitrario fanno fede i repertori giurisprudenziali), sembrerebbe legittimarsene un altro *id est* l'art. 234, co. 1 c.p.p., in merito al cui eventuale uso non "smodato" è lecito avanzare dubbi *pro futuro*».

[92] BARONE, *Giustizia Predittiva e Certezza del Diritto*, Pisa, 2024, 145 ss.

[93] *Ibid.*, 151 s.

[94] In termini netti SCALFATI, *IA e processo penale: prospettive d'impiego e livelli di rischio,* in *Proc. pen. giust.*, 2024, 6, 1406, secondo cui l'affiancamento dell'intelligenza artificiale nella decisione del giudice «è in contrasto con i valori espressi dagli artt.: 101, co. 2 (la soggezione del giudice solo alla legge vieta l'influenza di fattori non normativi diretti a limitarne le scelte), 104, co. 1 (l'indipendenza della magistratura si erge contro ogni fonte esterna in grado di incidere sull'esercizio della funzione, fenomeno non escluso dagli automatismi dell'IA che, anzi, potrebbero diventare fonte occulta di controllo), 13, 14, 15 e 111, co. 6, della Costituzione (laddove soprattutto letti insieme al predetto art. 101, co. 2 – impongono al giudice di motivare il proprio ragionamento, soprattutto in tema di libertà)».

contaminano) la funzione giurisdizionale. Si allude ai componenti l'Ufficio per il processo[95], non a caso né a torto individuato come "ponte" con l'intelligenza artificiale e la giustizia predittiva[96]. Tuttavia, queste "allusioni", pur suggestive, appaiono fuorvianti in quanto postulano l'ingresso nel processo di un nuovo soggetto (artificiale), dal quale si è ben lungi: per quanti sforzi la tecnologia possa compiere, soltanto l'essere umano possiede la consapevolezza della propria *humana fragilitas*[97]. Quel che è certo è che l'autonomia posseduta dalle macchine "pensanti" complica l'inquadramento della natura del loro contributo attraverso il ricorso alle categorie tradizionali, rendendosi evidentemente necessari interventi normativi *ad hoc*.

8. *Esigenze e prospettive interne di disciplinamento.* La peculiarità dell'intelligenza artificiale, che possiede un livello di autonomia non rinvenibile in nessun'altra tecnologia, esige norme calibrate sulla natura del fenomeno da governare, chiare, precise e coerenti rispetto al "dominio" al quale si intende fare riferimento (nella fattispecie, la giustizia penale).
Qualche spunto, muovendo dalla prospettiva probatoria, proviene ancora una volta da oltreoceano: negli Stati Uniti d'America è stata presentata al Congresso la proposta di legge *Justice in Forensic Algorithms Act of 2024*, che mira a rendere inopponibile alla difesa il segreto commerciale, sancendo la possibilità di accedere al codice sorgente dell'algoritmo, superando così i principi giurisprudenziali affermati nel caso *Chubbs*[98]. Inoltre, la proposta aspira ad introdurre standard (*Computational Forensic Algorithm Testing Standards*) e programmi (*Computational Forensic Algorithm Testing Program*) per testare gli algoritmi da utilizzare in ambito forense. Ancora, si consente che all'imputato vengano divulgati diversi elementi chiave: i risultati dell'analisi algoritmica del software, una copia di tale analisi idonea a dare dimostrazione del codice sorgente e dei risultati, i *files* e i dati pertinenti utilizzati dal software. Infine, si prevede l'*inadmissibility* nel caso in cui l'output algoritmico non sia stato sottoposto al programma di test o i suoi esiti non vengano messi a disposizione della difesa.
Spostandosi nel nostro ordinamento, è stata recentemente emanata la L. 23 settembre 2025, n. 132, recante «Disposizioni e deleghe al Governo in materia di intelligenza artificiale». Nel riservare sempre «al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti»[99], la legge non sembra escludere con chiarezza consultazioni del software e impieghi dell'output algoritmico da parte del giudice[100].
Inoltre, pur ribadendo la medesima visione antropocentrica dell'*AI Act* e operando importanti richiami ai requisiti di trasparenza, conoscibilità e spiegabilità e alla tutela del diritto di difesa[101],

---

[95] Si allude all'Ufficio per il processo, sul quale cfr. FUMU, *L'ufficio per il processo*, in *La riforma Cartabia*, a cura di Spangher, Pisa, 2022, 861 ss.

[96] CRICRÌ-DE MARIA-FRANCESCHINI, *L'Ufficio per il processo: tra vocazione efficientista e salvaguardia della funzione giurisdizionale*, in *Gli snodi problematici della riforma Cartabia con uno sguardo al futuro*, a cura di Colaiacovo-Del Vecchio-Nocerino, Bari, 2023, 272 ss.

[97] Su questi aspetti, a proposito del dibattito sulla responsabilità penale diretta dell'intelligenza artificiale, FLORIO, *Il dibattito sulla responsabilità penale diretta delle IA: "molto rumore per nulla"?*, in *Sist. pen.*, 2024, 2, 13.

[98] *People v. Superior Court (Chubbs)* (No B258569, 2015 WL 139069 (Cal Ct App Jan 9, 2015)).

[99] Cfr. art. 15, co. 1 L. n. 132/2025.

[100] BARONE, *La regolamentazione dell'Intelligenza Artificiale: è "corsa agli armamenti"*, cit., 999.

[101] Si v., in particolare, gli artt. 1, 3, 24, co. 5, lett. e) L. n. 132/2025.

non li presidia con rigidi divieti probatori, né delinea criteri di giudizio e obblighi motivazionali per fronteggiare adeguatamente il fenomeno algoritmico.
Norme prescrittive e cogenti andrebbero almeno previste in sede di attuazione delle deleghe legislative dirette al recepimento dell'*AI Act* e delineate «per adeguare e specificare la disciplina dei casi di realizzazione e impiego illeciti di sistemi di intelligenza artificiale»[102].

9. *Controlli significativi.* In attesa di adeguati interventi legislativi, occorre spostare il fuoco su un altro interrogativo: come si dovrebbe trattare l'output algoritmico. È in questa prospettiva che si condivide quella dottrina secondo cui l'output generato dal sistema algoritmico di *risk assessment* va trattato alla stregua di un mero indizio[103], da corroborarsi sempre con altri elementi di prova[104].
Ciò implica che la decisione giuridica sia resa senza servirsi "unicamente" dei risultati algoritmici e che il loro impiego, per essere giuridicamente accettabile, sia assoggettabile e

[102] V. art. 24L. n. 132/2025.

[103] Come precisa BERNASCONI, *Sistema probatorio e disposizioni generali*, in *Manuale di diritto processuale penale*[5], a cura di Scalfati-Bernasconi-De Caro-Menna-Pansini-Pulvirenti-Triggiani-Valentini-Vigoni, Torino, 2025, 275, il codice si riferisce agli «indizi» in due modi: quali prove "minori" rispetto alla prova "vera e propria" o, in maniera corretta, come metodo di ragionamento basato su una inferenza logica. Secondo la giurisprudenza, l'ontologica differenza tra prova e indizio è costituita dal fatto che, mentre la prima, in quanto si ricollega direttamente al fatto storico oggetto di accertamento, è idonea ad attribuire carattere di certezza allo stesso, l'indizio, isolatamente considerato, fornisce solo una traccia indicativa di un percorso logico argomentativo, suscettibile di avere diversi possibili scenari, e, come tale, non può mai essere qualificato in termini di certezza con riferimento al fatto da provare. La differenza tra indizio e prova non risiede nella tipologia del mezzo da cui deriva l'inferenza logica che costituisce il loro carattere comune, ma nei contenuti che essi esprimono e rappresentano: così Cass., Sez. V, 19 gennaio 2023, n. 2237, in *Proc. pen. giust.*, 2023, 5, 1102 ss., con nota di DEL COCO, *La valutazione dell'indizio nel processo penale*.

[104] GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, cit., 17, sostiene come «accanto all'obbligo di un intervento umano andrebbe ritenuta sussistente quella che, nel lessico processualpenalistico, chiameremmo regola di valutazione, in forza della quale l'output prodotto dall'IA va considerato come un mero indizio, che va sempre corroborato con altri elementi di prova». Ancora, E. GUIDO, *Intelligenza artificiale e procedimento penale: ragionando di valutazione del rischio* de libertate, cit., 11, propende per l'opzione interpretativa secondo cui il dato generato dall'intelligenza artificiale «debba essere trattato alla stregua di un mero indizio», ovvero come una «sorta di indicazione che considerata nell'insieme degli elementi apprezzabili dal giudice – oltre alle specifiche modalità e circostanze del fatto, "comportamenti o atti concreti" del soggetto imputato o indagato oppure i "suoi precedenti penali" – consenta di trarre conclusioni più sicure sul futuro comportamento dell'accusato». È opinione di DINACCI, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici*, cit., 1635, che sia «preferibile trattare il dato conoscitivo derivante dall'intelligenza artificiale per quello che è: un indizio» e, quindi, «un momento di conoscenza equivoco che non può essere utilizzato se non in presenza delle caratteristiche imposte dall'art. 192, co. 2 c.p.p.». Secondo 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. web*, 2021, 1, 32, «l'eventuale output prodotto dall'IA può essere considerato solo come un mero indizio, che va sempre corroborato con altri elementi di prova».

assoggettato a un «controllo umano significativo»[105], i cui requisiti possono essere delineati seguendo la "dottrina *Daubert*".
Se, come si è detto, vi è incertezza sull'inquadramento giuridico dell'output algoritmico, una certa unità di vedute si registra sulla necessità che sia compatibile con i diritti fondamentali del processo e suscettibile di controllo[106], valorizzandosi i parametri emersi nella giurisprudenza nordamericana e finalizzati all'esclusione dal processo della pseudo-scienza[107].
Il «controllo umano significativo», potrà riguardare, tra l'altro, sia la validità generale della teoria che ispira il software, sia le modalità con cui quella teoria è stata codificata nel programma[108], ed implica la possibilità per il giudice del caso concreto di discostarsi dal risultato suggerito e di valorizzare in motivazione le sfumature che rendono il proprio caso differente[109].
Del resto, lo stesso *AI Act* non manca di sottolineare più volte come occorra garantire che «le previsioni, le raccomandazioni o le decisioni del sistema di IA possano essere efficacemente ribaltate e ignorate»[110].
Non meno importante è, altresì, il «controllo pubblico significativo»[111], che richiede l'imposizione di limiti allo sviluppo degli algoritmi. L'*AI Act*, improntato al principio di neutralità tecnologica, non offre direttamente indicazioni su come creare, sviluppare e implementare sistemi intelligenti per garantire tecnicamente l'equità procedurale[112]: al contempo, limitando la discrezionalità delle aziende private nella creazione di algoritmi, concorre indirettamente al raggiungimento di quel «meaningful public control».

---

[105] UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sist. pen.*, 2020, 4, 83 s.

[106] Per questi aspetti, MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, cit., 242 ss., che precisa come «uno spazio a tali strumenti predittivi può essere trovato anche nel nostro sistema penale a patto che sia verificabile la validità del modello matematico utilizzato, la genuinità dei dati impiegati e, in sintesi, venga assicurata la trasparenza dell'algoritmo».

[107] UBERTIS, *Perizia, prova scientifica e intelligenza artificiale nel processo penale*, cit., 10. Cfr. altresì DINACCI, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici*, cit., 1636 s.

[108] QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Teoria e Critica della Regolazione Sociale*, 2021, 1, 280.

[109] COPPOLA, *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo* Ex-Aequo, in *Arch. pen. web*, 2023, 2, 16, richiama la nota sentenza "Franzese" (Cass., Sez. un., 10 luglio 2002, n. 30328, in *C.E.D. Cass.*, n. 222138) sull'utilizzo delle leggi scientifiche nell'accertamento del nesso di causalità.

[110] Cfr. il centoquarantunesimo *considerando*, nonché l'art. 14, par. 4, lett. d), l'art. 60, par. 4, lett. k) e l'art. 61, par. 1, lett. d) del Regolamento (UE) 2024/1689.

[111] TAYLOR, *Justice by Algorithm: The Limits of AI in Criminal Sentencing*, in *Criminal Justice Ethics*, 2023, 3, 193 ss. L'A. distingue tra controllo pubblico significativo, controllo giudiziario significativo e controllo umano significativo. Per una critica alle teorie espresse dall'A., v. RYBERG, *Sentencing, Artificial Intelligence, and Condemnation: A Reply to Taylor*, in *Criminal Justice Ethics*, 2024, 2, 131 ss.

[112] L'equità procedurale, oltre che quella distributiva, nello sviluppo dei modelli di apprendimento automatico è all'attenzione dei ricercatori: si v. a riguardo lo studio di WANG-HUANG-TANG-YAO, *Procedural Fairness and Its Relationship with Distributive Fairness in Machine Learning*, in *https://arxiv.org/abs/2501.06753*, 12 gennaio 2025.

I ricercatori sono impegnati a tradurre gli alti principi espressi nell'*AI Act* in regole, strategie e medotologie di progettazione[113], secondo un approccio incentrato sull'uomo[114]. Il paradigma della *human-centred AI*, declinato nelle forme dell'"intelligenza artificiale simbiotica"[115], ambisce a rafforzare la relazione uomo-macchina e a potenziare gli esseri umani, anziché sostituirli. È fondamentale notare che la costruzione di sistemi simbiotici richiede un approccio multidisciplinare ed implica, come fattore imprescindibile per l'interazione tra esseri umani e sistemi intelligenti, l'allineamento di questi ultimi ai principi giuridici[116]. Si tratta di prospettive importanti, che dovrebbero spingere i giuristi a collaborare proficuamente con i tecnici, anziché arroccarsi su posizioni conservatrici.

10. *Adelante, presto, con juicio.* Il contesto in cui viviamo porta alla mente la proverbiale frase pronunciata a mezza voce, nel capitolo XIII dei Promessi Sposi, da Antonio Ferrer, gran cancelliere di Milano, rivolgendosi al cocchiere Pedro, mentre in mezzo a una folla minacciosa sta andando in carrozza a prelevare il vicario di provvisione, ufficialmente per consegnarlo alla giustizia, concretamente per salvarlo dal linciaggio: "*Adelante, presto, con juicio*". *Adelante*: gli agenti artificiali sono ormai parte integrante del nostro presente; *presto*: la rapidità con cui evolvono e dilagano nell'uso quotidiano non permette indugio; *con juicio*: scelte *sin prudencia* possono concretizzare rischi e pericoli, producendo enormi danni.
Quali le prospettive, in conclusione, per gli *AI RATs* nel nostro sistema?
Non ostandovi l'*AI Act* e la L. n. 132/2025, se ne ritiene astrattamente ammissibile l'uso «per ottenere una previsione ipotetica, ma non una predizione vincolante, sostitutiva della condotta giudiziale»[117].
Alla luce del difetto di "granularità" dell'*AI Act* nella valutazione concreta del rischio è auspicabile che, almeno il legislatore delegato, non prescinda dalla considerazione dello specifico

---

113 CALVANO-CURCI-DESOLDA-ESPOSITO-LANZILOTTI-PICCINNO, *Building Symbiotic AI: Reviewing the AI Act for a Human-Centred, Principle-Based Framework*, in *https://arxiv.org/abs/2501.08046*, 20 maggio 2025.

114 I sistemi di Human-Centred AI (HCAI) sono progettati, sviluppati e valutati coinvolgendo gli utenti nel processo, con l'obiettivo di aumentare le prestazioni e la soddisfazione degli esseri umani in compiti specifici. Per una definizione e per affrondimenti sulla HCAI, cfr. DESOLDA-ESPOSITO-LANZILOTTI-PICCINNO-COSTABILE, *From human-centered to symbiotic artificial intelligence: a focus on medical applications*, in *Multimedia Tools and Applications*, 2025, 27, 32109 ss.

115 *Ibid.*, 32138, l'intelligenza artificiale simbiotica viene definita nei seguenti termini: «Symbiotic AI (SAI) systems are HCAI systems powered by a continuing and deeper collaboration between humans and AI, i.e., a symbiosis of human intelligence and artificial intelligence. Humans and AI mutually augment their capabilities, balancing each other's strengths and weaknesses without hampering neither the autonomy of humans nor the performances of AI, leaving humans in control of the system's decisions. In an SAI system, AI benefits from a continuous stream of new user-provided data to refine itself, while humans benefit from AI's improved performances and knowledge».

116 CALVANO-CURCI-DESOLDA-ESPOSITO-LANZILOTTI-PICCINNO, *Building Symbiotic AI: Reviewing the AI Act for a Human-Centred, Principle-Based Framework*, cit., 18. Gli A. propongono un framework per la creazione di sistemi di "intelligenza artificiale simbiotica" conformi all'*AI Act*, composto da quattro principi (*Trasparenza*, *Equità*, *Livello di Automazione* e *Protezione*) e tre proprietà (*Affidabilità*, *Robustezza* e *Sostenibilità)*, che mirano a promuovere una relazione simbiotica, ciascuno in misura diversa, guidando al contempo i professionisti nella creazione di sistemi conformi alle normative.

117 UBERTIS, *Processo penale telematico, intelligenza artificiale e costituzione*, in *Cass. pen.*, 2024, 2, 444.

contesto o "scenario" operativo, delineando con chiarezza e precisione, in relazione al dominio della giustizia penale, confini, condizioni e limiti d'uso, in generale, dell'intelligenza artificiale e, in particolare, degli strumenti predittivi della recidiva.

Riprendendo il ragionamento svolto a proposito dell'approccio semi-quantitativo, si può pervenire a qualche provvisoria considerazione conclusiva.

Così, in uno scenario che vede l'impiego dei *risk assessment tools* nella fase delle indagini[118], con riferimento all'adozione delle misure cautelari, come ausilio alla valutazione del pericolo di reiterazione del reato, *ex* art. 274, lett. c) c.p.p., occorre considerare il maggior impatto che può avere l'«effet moutonnier»[119], che distoglie il decisore dal prestare la dovuta attenzione ai dettagli del caso concreto.

L'intervento del giudice, cui spetterebbe coniugare la valutazione dell'output algoritmico con quella dei "comportamenti" o degli altri "atti concreti", che il disposto normativo assume come elementi indefettibili del vaglio prognostico sulla personalità dell'indagato, potrebbe non bastare a garantire *tout court* l'autenticità del suo giudizio umano, considerata la difficoltà di discostarsi dalla previsione del software, specialmente nel caso in cui il responso macchinico propenda per la sussistenza dell'esigenza cautelare. Già in uno scenario "tradizionale" la previsione umana è tendenzialmente sfavorevole per il prevenuto, in quanto risulta più facile per il giudice scegliere l'opzione che lo pone facilmente al riparo dal rischio di smentita, riconoscendo la sussistenza del pericolo di reiterazione[120]. Né l'effetto distorsivo, esacerbato dal suggerimento della macchina, sarebbe agevolmente compensabile agendo sul piano della motivazione, quand'anche rafforzata. Occorrerebbero soluzioni "compensative" (R), ulteriori, come quella d'impiegare il *tool* solo per confutare la conclusione del giudice che abbia ravvisato un rischio di reiterazione elevato; in altri termini, si potrebbe fare in modo che l'autorità giudiziaria, per le determinazioni di competenza, valuti l'output algoritmico quando questo attenua, anziché aggravare, la posizione della persona soggetta a valutazione[121].

---

[118] Secondo UBERTIS, *Perizia, prova scientifica e intelligenza artificiale nel processo penale*, cit., 18, le condizioni del controllo umano significativo non sarebbero ragionevolmente imponibili durante le indagini, richiedendosi in ogni caso il rispetto dei diritti fondamentali, «mentre sarebbe comunque inutilizzabile nella fase del giudizio quanto trasgredisse i canoni indicati quali necessari attributi del controllo umano significativo».

[119] GARAPON-LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthrpologique*, Paris, 2018, 239-

[120] MAZZA, *Distopia del processo artificiale*, in *Arch. pen. web*, 2025, 1, 17, dubita del fatto che il giudice, già refrattario attualmente a respingere la proposta di decisione proveniente dai *law clerks*, sarà capace in futuro di opporsi al responso di una macchina addestrata e di assumersi «la responsabilità di disattendere i giudizi prognostici di pericolosità elaborati, in materia cautelare o di misure di prevenzione, da algoritmi predittivi che tengono conto di una serie di variabili nemmeno calcolabili dall'uomo».

[121] Si v. il "modello ristretto" proposto da RYBERG, *Artificial intelligence and criminal justice: How to use algorithmic sentencing support in real life (and ethically non-ideal) penal systems?*, in *AI and Ethics*, 2025, 3, 3255 ss. In particolare, il "Restricted Application Model" prevede quattro fasi che illustrano come un giudice dovrebbe avvalersi dell'assistenza algoritmica: 1) il giudice determina quella che ritiene la pena appropriata per l'imputato in un dato caso; 2) questa pena iniziale determinata dal giudice viene quindi inserita nell'algoritmo di determinazione della pena insieme a tutte le informazioni pertinenti specifiche del caso; 3) sulla base delle informazioni sul reato in questione, l'algoritmo fornisce una raccomandazione di pena basata sulle condanne di precedenti casi simili (tuttavia, cosa fondamentale, l'algoritmo fornisce la propria raccomandazione solamente se la pena iniziale suggerita dal giudice è più severa di quella

Non è escluso che, in difetto di misure mitigatrici, si debba concludere – valutazione che dovrebbe svolgere il legislatore delegato[122] – che il rischio sia in effetti inaccettabile, nonostante i *risk assessment tools* in oggetto possano classificarsi in base all'*AI Act* come sistemi ad "alto rischio", e che di conseguenza sia preferibile «non contaminare il segmento cautelare» con strumenti di valutazione algoritmica del rischio[123].

Anche nell'ambito del giudizio l'impiego di un supporto algoritmico per la valutazione della pericolosità sociale dovrebbe tenere conto dello scenario concreto. Pericoli (R) e impatti (E) connessi al loro utilizzo, come l'effetto ancoraggio, solo in parte contrastabile attraverso gli oneri di motivazione[124], risultano indubbiamente meno rilevanti in una condizione quale quella rappresentata dal modello di *sentencing* introdotto dalla c.d. "riforma Cartabia", soluzione processuale che vede una netta cesura tra la fase di decisione sulla responsabilità e quella di commisurazione della pena.

La riforma ha previsto, infatti, un apposito spazio processuale, disciplinato dall'art. 545-*bis* c.p.p., che si presta ad essere "sede ospitante" per esperimenti di convivenza e interazione armonica tra "intelligenze", nella quale l'intelligenza artificiale potrebbe aumentare e valorizzare le capacità cognitive umane anziché sostituirle, secondo una logica di scambio e apprendimento reciproco, con conseguenti benefici per l'intero ecosistema processuale.

Un «apporto rilevante dell'intelligenza artificiale» potrebbe collocarsi proprio nell'ambito delle valutazioni sulla sostituibilità della pena, «con specifico riferimento ai parametri maggiormente di natura prognostica»[125], come il giudizio circa la pericolosità qualificata ed il concreto pericolo di violazione delle condizioni imposte, richiesto dalla normativa[126].

---

stabilita dall'algoritmo; in caso contrario, l'algoritmo sarà progettato per confermare la raccomandazione formulata dal giudice; 4) sulla base delle informazioni sul reato, della sua proposta di pena iniziale e della raccomandazione dell'algoritmo, il giudice determina definitivamente la pena da infliggere al trasgressore. In termini non dissimili, ROMANÒ, *La prova algoritmica nel procedimento penale alla luce dell'Ai Act*, in *Prova scientifica e processo penale*[3], a cura di Canzio-Luparia, Miano 2025, 1043; MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra* evidence based practices *e tutela dei diritti fondamentali*, cit., 35.

[122] I criteri direttivi che rilevano in ambito penale sono, quanto alla delega conferita al Governo per il recepimento dell'*AI Act*, quello di cui all'art. 24, co. 2, lett. h), che impone la «previsione di un'apposita disciplina per l'utilizzo di sistemi di intelligenza artificiale per l'attività di polizia»; quanto alla delega conferita «per adeguare e specificare la disciplina dei casi di realizzazione e impiego illeciti di sistemi di intelligenza artificiale», quello di cui all'art. 24, co. 5, lett. e), che richiede la «regolazione dell'utilizzo dei sistemi di intelligenza artificiale nelle indagini preliminari, nel rispetto delle garanzie inerenti al diritto di difesa e ai dati personali dei terzi, nonché dei princìpi di proporzionalità, non discriminazione e trasparenza».

[123] BELVINI, *Intelligenza artificiale e circuito investigativo*, Bari, 2025, 245.

[124] Di questo avviso FIORELLI, *Calcolabilità del giudizio giuridico e processo penale*, Torino, 2025, 209 ss. Anche BELVINI, *Intelligenza artificiale e circuito investigativo*, cit., 137 ss., ritiene che la motivazione, "rafforzata" e irrobustita dall'esposizione delle specifiche e autonome ragioni che inducono il giudice a servirsi dei risultati provenienti dall'intelligenza artificiale o, viceversa, a illustrare i motivi per i quali si aderisce a una conclusione smentita dal computo automatizzato, rappresenti un valido arnese per non smarrire la dimensione antropica del decidere.

[125] Così, condivisibilmente, CAMERA, *A proposito dell'indagine conoscitiva sull'impatto dell'intelligenza artificiale nel settore della giustizia*, in *www.sistemapenale.it*, 15 aprile 2024, 10.

[126] Cass., Sez. II, 14 febbraio 2024, n. 8794, in *C.E.D. Cass.*, n. 286006.

Secondo l'approccio adottato, il *risk assessment* è un processo di valutazione del rischio e non un metodo in grado di fornire una predizione esatta; è un mezzo e non un fine in sé stesso[127].
Nella strategia collaborativa uomo-macchina, ispirata al paradigma simbiotico, l'algoritmo predittivo può fornire un contributo conoscitivo ampio, esteso non solo ai rischi ma anche ai bisogni individuali, arricchendo le modalità di risposta al reato nel rispetto dei principi fondamentali di proporzionalità e dignità della persona[128], realizzando le istanze di individualizzazione della pena, dando concretezza alla sua funzione rieducativa e migliorando, in definitiva, l'esperienza umana[129].
Questo vale a maggior ragione ipotizzando l'impiego *in executivis* dei *tools*, laddove sembrano sorgere minori difficoltà. La prospettiva di ricorrere all'impiego di strumenti algoritmici per la valutazione circa la probabilità di commissione di ulteriori reati, nel che si traduce il giudizio sulla pericolosità sociale, è stata presa in considerazione dall'amministrazione penitenziaria in una recente circolare[130] ed è oggetto di attenzione in seno al Consiglio europeo per la cooperazione penologica[131]. Sebbene quest'ultimo abbia rilevato che l'apporto dell'intelligenza artificiale non costituisca, per lo stato attuale delle conoscenze tecnologiche, un rimedio per risolvere il problema del sovraffollamento[132], allo stesso tempo ne ha auspicato l'utilizzo nei percorsi riabilitativi.

---

[127] MALDONATO, Risk and need assessment tools *e riforma del sistema sanzionatorio: strategie collaborative e nuove prospettive*, cit., 146 s. L'A. precisa che l'output non è una predizione finale, ma richiede una nuova valutazione. Ove ci si affidasse esclusivamente agli strumenti di intelligenza artificiale, si ridurrebbero le persone a mera entità numerica, ciò in contrasto con il diritto fondamentale e inviolabile del rispetto della dignità umana, argine a qualsiasi reificazione dell'uomo. Secondo QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, cit., 258, «il *risk assessment* è un metodo e non un fine, uno strumento di valutazione del rischio di esordio o persistenza di una condotta antisociale e non la risposta ad essa». Per APPLEGARTH-LEWIS-RIEF, *Imperfect Tools: A Research Note on Developing, Applying, and Increasing Understanding of Criminal Justice Risk Assessments*, in *Criminal Justice Policy Review*, 2023, 4, 331, «part of what needs to be understood, participants contend, is that risk assessments are not solutions but tools for agencies to use to help them make decisions regarding individuals, policies, or procedures».

[128] Per consentire questo appare fondamentale che gli strumenti e le tecniche di modellazione predittiva si adattino alla natura complessa delle persone, del diritto e della società, mentre sarebbe errato seguire il processo inverso, ovvero pretendere che siano le persone, il diritto e la società ad adattarsi alla natura rigida della modellazione predittiva: in questi termini, GREENE-SHMUELI-FELL-LIN-LIU, *Forks over knives: Predictive inconsistency in criminal justice algorithmic risk assessment tools*, cit., 715.

[129] MALDONATO, Risk and need assessment tools *e riforma del sistema sanzionatorio: strategie collaborative e nuove prospettive*, cit.,169.

[130] Circolare D.a.p. del 5 ottobre 2023, n. 0062758.U "L'esecuzione penale esterna quale sistema di probation. Linee di indirizzo e indicazioni operative", in *www.sistemapenale.it*, 16 ottobre 2023, 7.

[131] Si tratta di un organismo istituito in seno al Consiglio d'Europa e dedicato alla definizione di standard e principi nell'ambito dell'esecuzione delle sanzioni penali, il quale ha elaborato una *Ethical, Strategic and Operational Guidance on the Use of Artificial Intelligence in Prison and Probation Services and the Private Companies acting on their Behalf*, 10 settembre 2021, consultabile alla pagina *https://rm.coe.int/pc-cp-2021-9-e-rev-the-use-of-artificial-intelligence-in-prison-and-pr/1680a3f8cb*.

[132] I casi d'uso sono concentrati, per lo più, sugli aspetti "gestionali" all'interno dei penitenziari: si v. ROWLAND-SHAH-CHANDRA, *AI in Corrections: The Basics and a Way to Experiment*, in *Federal Probation*, 2023, 3, 4 ss.