



IA PREDITTIVA E PROCESSO PENALE: POTENZIALITÀ, RISCHI E LIMITI LEGALI

LUISS GUIDO CARLI

DIPARTIMENTO DI GIURISPRUDENZA

CATTEDRA DI MACCHINE INTELLIGENTI E DIRITTO

CANDIDATO: FRANCESCO GUERRIERI (170673)

RELATORE: GIOVANNI BUONOMO

CORRELATORE: GIANLUIGI CIACCI

A.A. 2025-2026

Indice.....	2
Premessa.....	5
Capitolo primo: Intelligenza artificiale e processo penale.....	9
1.1) Intelligenza artificiale, sistemi esperti, <i>machine learning</i> e metodi di apprendimento.....	10
1.2) IA predittiva: nozione, campi applicativi e criticità	12
1.3) IA predittiva e processo penale nel mondo	15
1.3.1) Lo stato dell'arte.....	16
1.3.2) Il dibattito negli Stati Uniti: il caso <i>Loomis</i> e <i>COMPAS</i>	19
1.3.3) Il dibattito in Europa.....	24
1.4) Il quadro normativo	27
1.4.1) AI Act, direttiva LED e l. 132 2025	28
1.4.2) Il diritto penale italiano	32
Capitolo secondo: Potenziali campi applicativi dell'IA predittiva nel processo penale	35
2.1) Introduzione.....	35
2.2) Applicazione dell'IA predittiva ai fini delle esigenze cautelari	36
2.2.1) Valutazione dell'esigenza cautelare dell'inquinamento probatorio.....	36
2.2.2) Valutazione dell'esigenza cautelare del pericolo di fuga	38
2.2.3) Valutazione dell'esigenza cautelare del pericolo di reiterazione del reato.....	40
2.3) L'allontanamento dalla casa familiare e il divieto di avvicinamento alla persona offesa.....	43
2.4) Applicazione dell'IA predittiva ai fini delle misure precautelari	47
2.4.1) Introduzione.....	47
2.4.2) Valutazione dei presupposti dell'arresto obbligatorio e dell'arresto facoltativo	47
2.4.3) Valutazione dei presupposti del fermo di indiziato di delitto	50
2.4.4) Valutazione dei presupposti dell'allontanamento d'urgenza dalla casa familiare.....	51
2.5) Applicazione dell'IA predittiva ai fini delle misure di sicurezza personali	53
2.5.1) Introduzione.....	53
2.5.2) L'IA predittiva come strumento per superare la crisi della pericolosità sociale.....	53
2.5.3) Pericolosità sociale e capacità a delinquere: affinità e differenze	54
2.5.4) Valutazione della pericolosità sociale del minore	56
2.6) Applicazione dell'IA predittiva ai fini delle misure di prevenzione	59
2.6.1) Il ruolo dell'IA nella componente prognostica del giudizio di pericolosità.....	59
2.6.2) Il problema dell'indeterminatezza della pericolosità generica	60

2.7) Applicazione dell'IA predittiva ai fini della determinazione della pena	62
2.7.1) Intelligenza artificiale predittiva e discrezionalità giudiziale: potenzialità e rischi ...	62
2.7.2) Criteri di commisurazione della pena in cui l'IA predittiva può offrire supporto	64
2.7.3) Gli indici della capacità a delinquere	66
2.7.4) Valutazione dei presupposti della sospensione condizionale della pena	69
2.8) L'IA al servizio del PM: selezione delle notizie di reato ed esercizio dell'azione penale	71
2.8.1) Introduzione	71
2.8.2) Applicazione dell'IA per selezionare le notizie di reato da trattare con precedenza.	72
2.8.3) Applicazione dell'IA nei presupposti per l'esercizio dell'azione penale	74
2.9) Applicazione dell'IA nell'accertamento del nesso di causalità	77
Capitolo terzo: Principali rischi	81
3.1) Il pericolo di discriminazione	81
3.1.1) Le critiche rivolte a <i>COMPAS</i>	81
3.1.2) I criteri di equità algoritmica	82
3.1.3) Il principio di uguaglianza e il divieto di discriminazione	86
3.1.4) Le garanzie presenti nel diritto vigente	88
3.2) L'opacità algoritmica	90
3.2.1) L'opacità dovuta al problema della <i>black box</i>	90
3.2.2) L'opacità dovuta alla proprietà industriale	93
3.2.3) L'impatto dell'opacità sulle garanzie processuali	96
3.3) I pregiudizi cognitivi	101
3.3.1) Introduzione	101
3.3.2) La <i>MC Namara Fallacy</i>	102
3.3.3) L' <i>automation bias</i>	102
3.3.4) Le garanzie presenti nel diritto vigente	104
3.4) I limiti dell'AI Act nel tutelare i diritti fondamentali	107
3.5) Dal diritto penale del fatto al diritto penale d'autore?	110
3.5.1) Introduzione	110
3.5.2) IA e pericolosità sociale: misure di sicurezza e misure di prevenzione	111
3.5.3) Diritto penale d'autore nella determinazione della pena?	114
3.6) Possibili violazioni del principio di offensività	115
3.7) Possibili violazioni della presunzione di innocenza	116

3.8) Compatibilità con l'individualizzazione della pena e del trattamento cautelare.....	118
3.9) Compatibilità con l'obbligatorietà dell'azione penale	120
Capitolo quarto: Limiti legali.....	124
4.1) Il divieto di perizia criminologica (art.220, comma 2, c.p.p.)	124
4.1.1) Introduzione: il divieto, le deroghe e il giudizio sulla personalità	124
4.1.2) Analisi comparatistica e storica di un divieto peculiare	125
4.1.3) Le <i>rationes</i> alla base del divieto e l'assimilabilità all'IA predittiva	127
4.2) I limiti imposti dalla direttiva 2016/680/UE.....	130
4.2.1) Il divieto di processo decisionale automatizzato: introduzione	130
4.2.2) Il divieto di processo decisionale automatizzato: i singoli requisiti.....	131
4.2.3) Le garanzie spettanti all'interessato qualora il divieto sia derogato	135
4.2.4) Decisioni e profilazioni discriminatorie	136
4.3) regolamento 2024/1689/UE: tra pratiche vietate e sistemi ad alto rischio.....	138
4.3.1) Il divieto di cui all'art. 5, par.1, lett. d)	138
4.3.2) I sistemi ad alto rischio rilevanti nel processo penale.....	140
4.3.3) Le difficoltà ermeneutiche nel coordinare l'art. 5 con l'allegato III	142
4.4) L'art. 15 della legge 132/2025.....	143
4.4.1) La riserva al magistrato nelle attività incidenti nel merito	143
4.4.2) Spazi residui di ricorso all'IA nelle attività non incidenti nel merito	146
Conclusioni	148
Bibliografia.....	151

PREMESSA

Negli ultimi anni, stiamo assistendo a uno sviluppo dell'intelligenza artificiale (IA) senza precedenti. Nel dibattito pubblico, sono sempre più frequenti discussioni in merito alle potenzialità dell'IA generativa, intesa come la produzione di contenuti di varia natura, quali ad esempio, video, audio o testo, a partire da un richiamo testuale¹. Si pensi al successo su larga scala di assistenti virtuali quali *ChatGPT*, *Copilot* e *Gemini*.

Allo stesso tempo, sia le imprese private che le pubbliche amministrazioni fanno sempre più spesso ricorso all'IA predittiva, intesa come la stima del verificarsi di un evento futuro, al fine di orientare una decisione da adottare nel presente². Un esempio è *SyRI* (*Systeem Risico Indicatie*), un sistema di profilazione predittiva che è stato utilizzato nei Paesi Bassi al fine di segnalare alla pubblica amministrazione che un determinato individuo presenta un rischio di frode particolarmente elevato³. Tale tendenza a ricorrere all'IA predittiva per supportare le decisioni adottate dalle autorità pubbliche riguarda anche il processo penale. Ad esempio, negli Stati Uniti, il sistema predittivo del rischio di recidiva *COMPAS* (*Correctional Offender Management Profiling for Alternative Sanctions*) viene impiegato nella determinazione della pena⁴.

Nel contesto statunitense, un ampio dibattito è sorto in merito al caso *Loomis*, vicenda giudiziaria in cui l'imputato era stato condannato ad una pena determinata, oltre che dalla gravità del reato e dai precedenti giudiziari del reo, in base al sistema predittivo *COMPAS*, il quale aveva stimato un rischio di recidiva particolarmente elevato⁵. Sebbene la Corte Suprema del Wisconsin abbia considerato legittimo il ricorso a tale strumento, il caso ha sollevato numerose perplessità.

In primo luogo, la difesa non ha potuto avere accesso al funzionamento del sistema poiché è coperto dal segreto industriale. Allo stesso tempo, il diritto ad una pena individualizzata potrebbe essere stato violato, poiché l'imputato potrebbe essere stato giudicato in base all'appartenenza ad un gruppo sociale, anziché in base alla sua situazione individuale, in quanto il sistema potrebbe aver stimato un rischio di recidiva particolarmente elevato sulla base del gruppo sociale di appartenenza dell'imputato⁶.

In Unione europea, e in particolare in Italia, si riscontra un approccio decisamente più cauto in merito all'utilizzo dell'IA nel processo penale. Nel regolamento 2024/1968/UE (noto come AI Act), l'impiego di intelligenza artificiale nel processo penale viene, a seconda dei casi, classificato come pratica vietata (art. 5, par. 1, lett. d) o come sistema ad alto rischio (allegato III n. 6 e n. 8 lett. a)⁷. L'impiego di IA in ambito penalistico potrebbe rientrare poi nel divieto di processo decisionale automatizzato, previsto dall'art. 11, par. 1,

¹ G. BUONOMO, *Macchine intelligenti, processo induttivo e apprendimento autonomo*, in G. BUONOMO -G. CIACCI -A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, p. 5.

² *Ibid.*, p. 5.

³ A. NARAYANAN -S. KAPOOR, *AI Snake Oil*, Princeton University Press, Princeton, 2024, pp. 48-49.

⁴ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 14.

⁵ *Ibid.*, par. 85.

⁶ Sulla possibile natura discriminatoria di *COMPAS*, si veda J. ANGWIN -J. LARSON -S. MATTU, *et al.*, *Machine Bias: Risk Assessments in Criminal Sentencing*, *ProPublica*, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, consultato il 3 agosto 2025.

⁷ Per approfondire, si veda il par. 4.3.

della direttiva 2016/680/UE (nota come direttiva LED)⁸. In Italia, è stata avanzata la tesi secondo la quale il ricorso all'IA per valutare la personalità dell'imputato sia precluso dal divieto di perizia criminologica previsto dall'art. 220, comma 2 del codice di procedura penale (c.p.p.)⁹.

Da ultimo, è recentemente entrata in vigore la legge 23 settembre 2025, n. 132, "*recante disposizioni e delega al governo in materia di intelligenza artificiale*". All'art. 15, comma 1, si è stabilito che "*nei casi di impiego dei sistemi di IA nell'attività giudiziaria è riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti*"¹⁰.

Nel presente lavoro, verranno analizzate le potenzialità, i rischi e i limiti legali dell'eventuale introduzione dell'intelligenza artificiale predittiva al fine di supportare la polizia giudiziaria, i PM e i giudici nella formulazione di giudizi prognostici nel processo penale italiano.

Non verranno invece specificamente analizzate le ipotesi di ricorso all'IA nel processo penale che non rientrano nella nozione di IA predittiva. Ad esempio, si considererà il possibile impiego dell'IA per calcolare il rischio di recidiva, in quanto si tratta di IA predittiva. Non sarà invece oggetto di specifica analisi nella presente disamina l'utilizzo dell'IA per valutare l'affidabilità di elementi probatori, perché tali sistemi IA non sono funzionali a stimare la probabilità di verificazione di eventi futuri.

Verrà analizzata la possibilità che l'IA supporti i giudizi prognostici formulati da persone fisiche quali gli agenti di polizia giudiziaria, i PM e i giudici. Non verrà considerata la possibilità che l'IA sostituisca radicalmente l'essere umano, in quanto tale ipotesi, soprattutto nel processo penale, appare ancora lontana, nonché in evidente contrasto con numerosi principi di rilevanza costituzionale e sovranazionale. Ad esempio, tale eventualità pare in contrasto con il diritto di accesso al giudice in caso di limitazione della libertà personale, previsto dagli artt. 5, par. 2 e 3, della Carta dei diritti fondamentali dell'Unione europea, e 5, par. 3 e 4, della Convenzione europea dei diritti dell'uomo. Diversa è invece l'ipotesi dell'impiego dell'IA a supporto dell'essere umano: qui la questione risulta maggiormente complessa e merita un'analisi maggiormente approfondita nella presente disamina. Non è un caso che, negli Stati Uniti, l'IA predittiva è stata impiegata nel processo penale per supportare il giudice (ad esempio, con riferimento a *COMPAS*), mentre non risulta che sia stata utilizzata altresì per sostituirlo radicalmente (quanto meno nel processo penale all'interno di ordinamenti in cui vige lo stato di diritto).

Per quanto concerne la struttura del presente lavoro, nel primo capitolo, verranno fornite delle nozioni preliminari per comprendere il rapporto tra intelligenza artificiale predittiva e processo penale, con particolare riferimento al concetto di intelligenza artificiale, alle attuali applicazioni dell'IA predittiva, al dibattito sul ricorso all'IA nel processo penale sia in Europa che negli Stati Uniti e al contesto normativo di riferimento.

⁸ Per approfondire, si veda il par. 4.2.

⁹ S. QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI ACT*, in *Diritto di Difesa*, 2023, p. 5. Cfr. par. 4.1.

¹⁰ Cfr. par. 4.4.

Riguardo alle potenzialità, analizzate specificamente nel secondo capitolo, verranno illustrati gli ambiti dell'ordinamento penale italiano in cui l'IA predittiva potrebbe avere un ruolo. Si cercherà di mettere in relazione gli algoritmi attualmente impiegati all'interno di ordinamenti stranieri con i possibili ambiti normativi italiani in cui tali sistemi potrebbero essere adottati. Ad esempio, *COMPAS* è attualmente in uso negli Stati Uniti per supportare la decisione giudiziale circa il rilascio o la detenzione dell'indagato in attesa del processo¹¹. Questo sistema predittivo potrebbe essere impiegato nel sistema italiano al fine di valutare la sussistenza delle esigenze cautelari a fondamento di una misura cautelare personale, ai sensi dell'art. 274 c.p.p. Verranno esaminati, in particolare, i seguenti campi applicativi in cui l'IA predittiva potrebbe avere un ruolo: esigenze cautelari (artt. 273 c.p.p. e ss.), allontanamento dalla casa familiare (art. 282-*bis* c.p.p.), divieto di avvicinamento ai luoghi frequentati dalla persona offesa (art. 282-*ter* c.p.p.), misure precautelari (artt. 381 e ss. c.p.p.), misure di sicurezza (artt. 201 e ss. c.p.), facendo riferimento altresì alla pericolosità sociale del minore (art. 37, comma 2, d.p.r. 22 settembre 1988 n. 448), misure di prevenzione (d.lgs. 6 settembre 2011, n.159), commisurazione della pena, con particolare riferimento al giudizio sulla capacità a delinquere, (art. 133, comma 2, c.p.), sospensione condizionale (art. 164 c.p.), selezione delle notizie di reato da trattare con precedenza (art. 1, comma 6, d.lgs. 20 febbraio 2006, n. 106) ed esercizio dell'azione penale (art. 50, comma 1, e art. 408 c.p.p.). Questi campi applicativi, per quanto eterogenei, sono accumulati dalla circostanza che, a seconda dei casi, la PG, il PM o il giudice sono chiamati a formulare giudizi prognostici che potrebbero essere supportati dall'IA predittiva.

Verrà anche esaminata la possibilità di impiegare l'IA per l'accertamento del nesso di causalità. In questo caso, le informazioni di natura statistica che l'IA è in grado di fornire potrebbero essere impiegate non tanto per supportare un giudizio prognostico quanto per coadiuvare l'accertamento di un fatto verificatosi nel passato. Nel capitolo dedicato alle potenzialità dell'IA predittiva, verranno analizzate le principali argomentazioni a sostegno di tale tecnologia. Ad esempio, mediante il ricorso all'IA, la valutazione sulla personalità dell'imputato potrebbe risultare maggiormente accurata. Si dubita infatti che il giudice, allo stato dell'arte, abbia a disposizione gli strumenti necessari per effettuare tale accertamento, in quanto, ai sensi dell'art. 220, comma 2, c.p.p., è vietata la perizia al fine di stabilire la personalità dell'imputato.

Nel terzo capitolo, verranno invece discussi i principali rischi. In primo luogo, tali algoritmi potrebbero risultare discriminatori nei confronti delle minoranze, come evidenziato da *ProPublica* con riferimento a *COMPAS*¹². Il funzionamento di tali algoritmi potrebbe poi non essere sufficientemente trasparente per ragioni di proprietà industriale, oppure a causa del problema della *black box*, intesa come la difficoltà tecnica nel ricostruire i processi logici in base ai quali il sistema ha prodotto un determinato risultato (detto *output*)¹³.

¹¹ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025, p.27.

¹² J. ANGWIN -J. LARSON -S. MATTU *et al.*, *op. cit.*

¹³ G. BUONOMO, *op. cit.* pp. 15-18.

Inoltre, l'autorità giudiziaria potrebbe essere soggetta a pregiudizi cognitivi, come l'*automation bias*, la tendenza umana ad accettare acriticamente i suggerimenti offerti dalle macchine¹⁴.

In questo quadro si inserisce l'AI Act, il regolamento europeo sull'intelligenza artificiale. Per quanto possa essere uno strumento utile per mitigare tali rischi, in dottrina sono stati avanzati dubbi quanto alla sua idoneità a garantire una tutela effettiva dei diritti fondamentali¹⁵.

Vi sono altresì pericoli di natura strettamente penalistica. Il ricorso all'IA potrebbe favorire la deriva verso un diritto penale d'autore in quanto l'imputato verrebbe giudicato primariamente in relazione alla sua personalità e, soltanto in via secondaria, sulla base al fatto contestato. Infatti, sistemi di IA si basano anche su informazioni che afferiscono meramente alla personalità del reo anziché sui fatti di reato oggetto dell'imputazione, come ad esempio i precedenti penali¹⁶. Integrando un diritto penale d'autore, anche il principio di offensività rischierebbe di essere compromesso, perché, in un giudizio in cui la pericolosità sociale assume un ruolo preminente e il fatto di reato un ruolo secondario, potrebbero essere puniti dei fatti privi di reale offensività. Inoltre, il giudizio sulla personalità dell'imputato offerto dall'IA predittiva potrebbe influenzare la valutazione giudiziale sulla colpevolezza dell'imputato, minando la presunzione di innocenza¹⁷. Potrebbe essere poi compromesso il principio di individualizzazione della pena secondo cui l'imputato deve essere giudicato sulla base delle sue condizioni individuali. L'imputato potrebbe essere giudicato in relazione all'appartenenza ad un gruppo sociale a cui il sistema di IA associa un determinato rischio di reiterazione del reato¹⁸.

Infine, verrà analizzata l'ipotesi del ricorso all'IA per supportare la scelta del Pubblico Ministero tra esercizio dell'azione penale ed archiviazione (art. 50, comma 1, e 408 c.p.p.) e ne verrà valutata la compatibilità con l'obbligatorietà dell'azione penale, ai sensi dell'art. 112 della Costituzione (Cost.).

Nel quarto capitolo, si valuterà se, e in che misura, la normativa vigente precluda l'introduzione dell'IA predittiva nel processo penale italiano. Sarà specificamente analizzata la compatibilità del ricorso all'IA come strumento di accertamento della personalità dell'imputato con il divieto di perizia criminologica ai sensi dell'art. 220, comma 2, c. p. p. In secondo luogo, si analizzerà se, e in che misura, il divieto di processo decisionale automatizzato, ai sensi della direttiva LED, limiti l'impiego dell'IA nel processo penale. Ne seguirà un'analisi circa quali sistemi di IA sono vietati e quali ad alto rischio, ai sensi dell'AI Act e verrà spiegato l'ambito di applicazione del divieto di cui all'art. 15, comma 1, della legge 23 settembre 2025, n. 132, contenente la riserva di determinate valutazioni al magistrato anziché all'IA.

Per quanto riguarda gli strumenti utilizzati, il presente lavoro si basa, innanzitutto, sull'analisi della normativa rilevante, ad esempio, al fine di individuare i campi applicativi dell'IA nel processo penale, nonché allo scopo

¹⁴ K. L. MOSIER -L. J. SKITKA, *Automation Use and Automation Bias*, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, n. 3, 1999, p. 344.

¹⁵ M. ALMADA -N. PETIT, *The EU AI Act: between the rock of product safety and the hard place of fundamental rights*, in *Common market law review*, vol. 62, 2025, pp. 1114-1118. V. par. 3.4.

¹⁶ L. NOTARO, *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021, pp. 65-66.

¹⁷ *Ibid.*, pp. 68-69.

¹⁸ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, p. 21.

di ricostruire i limiti legali all'introduzione dell'IA nel processo penale. La presente disamina tiene altresì conto della giurisprudenza rilevante, come il caso *Loomis*. Un altro strumento utilizzato è l'analisi della dottrina, resa necessaria sia per comprendere pienamente la normativa rilevante che per analizzare il dibattito in merito alle potenzialità e ai rischi del ricorso all'IA nel processo penale. Per comprendere in che modo l'IA potrebbe essere introdotta nel processo penale, tale ricerca tiene conto anche dell'analisi comparativa di ordinamenti stranieri in cui l'IA predittiva è già stata introdotta (come quello statunitense)¹⁹. Da ultimo, il presente lavoro si basa su di un approccio multidisciplinare volto a considerare anche dati extragiuridici. Si pensi al problema della *black box* il quale, per essere compreso, richiede conoscenze dello stato dell'arte in ambito informatico.

¹⁹ Ad esempio, l'analisi comparativa emerge nei par. 2.1, 2.2, 2.3 in cui si confrontano le misure cautelari italiane con le condizioni per la detenzione preprocessuale (*pretrial detention*) statunitense.

CAPITOLO PRIMO

INTELLIGENZA ARTIFICIALE E PROCESSO PENALE

1.1) Intelligenza artificiale, sistemi esperti, machine learning e metodi di apprendimento

Nel presente capitolo introduttivo occorre, innanzitutto, spiegare cosa si intenda con l'espressione "intelligenza artificiale (IA)", in merito alla quale, sussiste un ampio dibattito multidisciplinare che in questa sede può essere illustrato soltanto nei suoi tratti essenziali.

Si possono distinguere quattro approcci all'intelligenza artificiale, secondo quanto sostenuto da Peter Norvig e Stuart J. Russel, figure fondamentali nel campo dell'intelligenza artificiale, note principalmente per il manuale *Artificial Intelligence: A Modern Approach*. Secondo il primo approccio, detto *acting humanly*, si considera intelligente una macchina che agisce nel modo in cui si comporterebbe un essere umano²⁰. Tale approccio è alla base del test teorizzato da Alan Turing, pioniere dell'informatica, nonché precursore dell'intelligenza artificiale. In tale esperimento, vi è un interrogatore umano che pone delle domande a un altro soggetto e a una macchina. Il test viene superato se l'umano non è in grado di stabilire se le risposte provengono dalla macchina o da una persona²¹.

Secondo un'altra concezione dell'intelligenza artificiale, detta *thinking humanly*, una macchina viene considerata intelligente se pensa nello stesso modo di un essere umano²². Tuttavia, secondo John R. Searle, uno dei maggiori esponenti della filosofia del linguaggio e della mente del secolo scorso, una macchina non può pensare in maniera analoga ad una persona in quanto si limita ad associare simboli sulla base delle istruzioni ricevute, senza essere consapevole del significato delle domande che le vengono poste. A supporto di tale tesi, Searle ha ideato l'esperimento della stanza cinese in cui una persona che parla soltanto inglese risponde meccanicamente a delle domande scritte in cinese, solamente grazie ad un libretto di istruzioni in inglese che spiega come associare quei simboli cinesi a delle risposte. Sebbene dall'esterno possa sembrare che il soggetto parli cinese, in realtà si limita ad associare dei simboli senza comprenderne il significato, in modo analogo a quanto avviene in un computer. Non sarebbe quindi sufficiente che una macchina comunichi in modo indistinguibile da un altro essere umano per essere considerata intelligente²³.

Un'ulteriore concezione dell'intelligenza artificiale, sintetizzabile con l'espressione *thinking rationally*, considera una macchina intelligente se applica in modo corretto i principi fondamentali della logica classica, come, ad esempio, il principio di identità e di non contraddizione. È il caso della tradizione logicista, in base alla quale sono stati sviluppati programmi in grado di risolvere ogni problema descritto in notazione logica.

²⁰ S. RUSSELL -P. NORVIG, *Artificial Intelligence: A Modern Approach*, 3ª ed., Prentice Hall, Upper Saddle River (NJ), 2010, pp. 2-3.

²¹ A.M. TURING, *Computing Machinery and Intelligence*, in *Mind*, vol. LIX, n. 236, 1950, pp. 433-460.

²² S. RUSSELL -P. NORVIG, *op cit.*, p. 3.

²³ J.R. SEARLE, *Minds, Brains, and Programs*, in *Behavioral and Brain Sciences*, vol. III, n. 3, 1980, pp. 417-457.

La principale difficoltà consiste nel formulare la conoscenza nei termini formali della notazione logica, soprattutto quando la conoscenza non è certa al 100%²⁴.

Norvig e Russell considerano preferibile l'approccio dell'agente razionale, sintetizzabile con l'espressione *acting rationally*. Un agente razionale agisce in modo tale da ottenere il miglior risultato, oppure, in una situazione di incertezza, il miglior risultato atteso. Questo approccio implica il pensare in maniera razionale in quanto per raggiungere il miglior risultato è spesso necessario usare la logica. Agire razionalmente comprende però anche azioni che non riguardano il pensiero razionale. Ad esempio, scostarsi da una stufa bollente è un riflesso che risulta più efficace rispetto al pensare razionalmente alle alternative possibili. L'*acting rationally approach* è dunque più generale del *thinking rationally approach* poiché non comprende soltanto il pensiero razionale. Tale approccio risulta poi più solido scientificamente degli approcci basati sul comportamento umano in quanto la razionalità è più facilmente definibile in termini matematici²⁵.

Dopo aver esposto i vari approcci all'intelligenza artificiale, occorre illustrare le due principali famiglie in cui l'IA si divide: i sistemi esperti e il *machine learning* (Apprendimento automatico). In questa tesi, ci si concentrerà sul *machine learning* e sulle problematiche che tale tecnologia comporta, con la conseguenza che i sistemi esperti verranno esposti al solo fine di illustrarne le differenze.

Questi sistemi si basano sul ragionamento deduttivo, in base al quale, partendo da premesse di carattere generale, si formula un giudizio relativo ad un caso particolare. Si pensi al noto sillogismo aristotelico in forza del quale dalla premessa maggiore "tutti gli uomini sono mortali" e dalla premessa minore "Socrate è un uomo" si perviene alla conclusione "Socrate è mortale". Nel metodo deduttivo, se le premesse sono vere sarà necessariamente vera anche la conclusione. I sistemi esperti sono in grado di applicare il ragionamento deduttivo per giungere a delle conclusioni relative ad un caso particolare sulla base delle informazioni di carattere generale fornite dal programmatore²⁶. Per sviluppare tali sistemi, occorre tradurre in un linguaggio eseguibile dalla macchina la conoscenza di un esperto in un determinato ambito del sapere, come la medicina o il diritto. Ad esempio, il software *Turbotax* viene utilizzato per la preparazione della dichiarazione dei redditi. Gli sviluppatori del software, consultandosi con degli esperti in materia tributaria, hanno tradotto il significato e la logica delle norme in ambito tributario in regole formali eseguibili dalla macchina²⁷.

Il principale limite dei sistemi esperti è che non sono in grado di fornire alla macchina le stesse capacità di un essere umano qualora non sia possibile codificare tutte le ipotesi che la macchina dovrà gestire²⁸. Il *machine learning* permette di superare tale limite in quanto apprende migliorando le proprie prestazioni sulla base dell'osservazione dell'ambiente circostante²⁹. L'apprendimento automatico si basa infatti sul metodo induttivo,

²⁴ S. RUSSELL -P. NORVIG, *op cit.*, p. 4.

²⁵ *Ibid.*, pp. 4-5.

²⁶ G. BUONOMO, *Macchine intelligenti, processo induttivo e apprendimento autonomo*, in G. BUONOMO -G. CIACCI -A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, pp. 25-27; R. ROVATTI, *Il processo di apprendimento algoritmico e le applicazioni nel settore legale*, in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp.34-36.

²⁷ H. SURDEN, *Artificial Intelligence and Law*, Cambridge University Press, Cambridge, 2019, pp. 1316-1317.

²⁸ R. ROVATTI, *op. cit.*, pp. 32-33.

²⁹ S. RUSSELL -P. NORVIG, *op cit.*, p. 693.

ossia sull'osservazione di una serie di casi particolari al fine di formulare conclusioni di carattere generale. A differenza del metodo deduttivo, si giunge ad una conclusione che, anche quando le premesse sono vere, non è necessariamente vera, bensì meramente probabile. Potrebbero esistere dati ancora non osservati in grado di confutare la conclusione precedentemente formulata sulla base delle limitate informazioni disponibili. Ad esempio, con l'osservazione del primo cigno nero in Australia è stata confutata la precedente convinzione secondo cui i cigni sono necessariamente bianchi³⁰.

Il *machine learning* presenta applicazioni pratiche di varia natura. Ad esempio, un sistema di apprendimento automatico può imparare a riconoscere un oggetto sulla base delle sue caratteristiche ricorrenti, anche dette *pattern* (come la forma e il colore). Calando queste considerazioni nel campo di indagine oggetto della presente trattazione, si consideri il caso di un modello predittivo che impara a valutare il rischio di reiterazione del reato sulla base delle caratteristiche personali di coloro che in passato sono risultati recidivi. È il caso del modello predittivo statunitense *COMPAS* che sarà oggetto di un'analisi maggiormente approfondita nei paragrafi successivi³¹.

I sistemi di *machine learning* si classificano in base al metodo di apprendimento. I metodi più noti sono l'apprendimento supervisionato, l'apprendimento non supervisionato e per rinforzo. Nell'apprendimento supervisionato, al sistema vengono forniti dei dati in *input* etichettati, ossia associati ad un *output*. Sulla base delle caratteristiche ricorrenti (*pattern*) che permettono di associare *input* e *output*, viene formulata una funzione in grado di produrre degli *output* a partire da dati in *input* non etichettati³². Un esempio di apprendimento supervisionato è la classificazione che consiste nell'associare i dati in *input* ad un numero finito di *output* possibili. Si pensi ad un algoritmo che deve stabilire se il tempo locale sarà soleggiato, nuvoloso o piovoso: in questo caso, gli output possibili sono di un numero finito, pari a tre³³. Nell'apprendimento non supervisionato, il sistema individua in autonomia delle caratteristiche ricorrenti all'interno dei dati in *input*, in assenza di riscontri o di istruzioni fornite dal programmatore. Una delle tecniche più note in questo ambito è il *clustering* che consiste nel raggruppare i dati in *input* in gruppi di elementi con caratteristiche simili, denominati *cluster*. Nell'apprendimento per rinforzo, il sistema impara sulla base di ricompense (*reward*) o di penalità (*punishment*) che il sistema riceve a seconda che l'azione compiuta contribuisca o meno a raggiungere l'obiettivo assegnato³⁴.

1.2) IA predittiva: nozione, ambiti di applicazione e criticità

Per IA predittiva si intende una stima del verificarsi di un evento futuro al fine di orientare una decisione da adottare nel presente. In contrapposizione, con intelligenza artificiale generativa si intende la produzione di

³⁰ M. BUMFORD, *Friday Essay: a Rare Bird — how Europeans Got the Black Swan so Wrong*, in *The Conversation*, 2021, <https://theconversation.com/friday-essay-a-rare-bird-how-europeans-got-the-black-swan-so-wrong-161654>, consultato il 3 agosto 2025.

³¹ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 15.

³² S. RUSSELL -P. NORVIG, *op cit.*, p. 695.

³³ *Ibid.*, p. 696.

³⁴ *Ibid.*, p. 695.

contenuti di varia natura, quali video, audio o testo, a partire da un richiamo testuale (denominato *prompt*)³⁵. Negli ultimi anni, l'IA generativa ha avuto uno sviluppo particolarmente significativo, con il successo di assistenti virtuali quali ChatGPT, Copilot e Gemini.

Arvind Narayanan e Sayash Kapoor, nell'opera *AI Snake Oil*, hanno analizzato criticamente le capacità e i limiti dell'intelligenza artificiale. In merito all'IA generativa, hanno sostenuto che vi sono dei progressi reali, pur permanendo delle criticità. L'IA predittiva non viene invece ritenuta dagli studiosi ugualmente affidabile, in ragione dei suoi limiti strutturali quali l'insufficienza dei dati disponibili e la difficoltà di catturare tutte le variabili rilevanti, soprattutto quando si tratta di prevedere le vicende umane³⁶.

Sebbene le criticità non manchino, l'IA predittiva viene impiegata per orientare le decisioni delle imprese private e delle pubbliche amministrazioni. In questo paragrafo, ne verranno illustrate le principali applicazioni, senza pretesa di esaustività e senza accennare alle possibili problematiche che saranno invece affrontate nel paragrafo successivo.

Quanto all'IA predittiva nel settore privato, si può citare il *credit scoring*, che aiuta la valutazione dell'affidabilità finanziaria di chi richiede un prestito o un altro prodotto finanziario. Esempi di piattaforme di *credit scoring* sorte di recente sono SCHUFA (*Schutzgemeinschaft Für Allgemeine Kreditsicherung*) in Germania o FICO (*Fair Isaac Corporation*) negli Stati Uniti³⁷.

Quanto al settore pubblico, l'IA predittiva è stata impiegata per prevenire frodi nel settore della sicurezza sociale. Rientra in questa ipotesi SyRI (*Systeem Risico Indicatie*), un modello predittivo che è stato utilizzato nei Paesi Bassi per segnalare che un determinato individuo presenta un elevato rischio di frode, lasciando alla pubblica amministrazione la scelta se avviare o meno un'indagine nei suoi confronti. Tale sistema è stato oggetto di forti critiche per aver portato le autorità olandesi ad accusare ingiustamente di frode numerose persone³⁸.

Un altro ambito di applicazione rilevante dell'intelligenza artificiale predittiva è la cosiddetta giustizia predittiva, intesa come previsione probabilistica sulle decisioni giudiziarie future. Di particolare interesse è lo studio a cura del Professor Nikolaos Aletras, uno dei pionieri dell'IA predittiva, avente ad oggetto la possibilità di impiegare il *machine learning* per prevedere l'esito delle sentenze. In particolare, mediante strumenti di elaborazione del linguaggio naturale (*natural language processing, NLP*), è stato sviluppato un modello capace di analizzare il testo delle precedenti sentenze della Corte EDU e di prevedere se, a seguito di un ricorso, la Corte accerterà o meno la violazione di determinati diritti previsti nella CEDU, con un livello di accuratezza pari a circa il 79%.³⁹

³⁵ G. BUONOMO, *op. cit.*, p. 5.

³⁶ S. RUSSELL -P. NORVIG, *op. cit.*, p. 98. Il prof. Narayanan insegna scienze informatiche alla Princeton University, ove è anche direttore del Center for Information Technology Policy. Sayash Kapoor è ricercatore presso il dipartimento di scienze informatiche della stessa università.

³⁷ G. SPINDLER *Algorithms, Credit Scoring, and the New Proposals of the EU for an AI Act and on a Consumer Credit Directive*, in *Law and Financial Markets Review*, 2023, pp. 239-240.

³⁸ A. NARAYANAN – S. KAPOOR, *op. cit.*, pp. 48-49.

³⁹ N. ALETRAS -D. TSARAPANIS -D. PREOTIUC, *et al.*, *Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective*, in *PeerJ Computer Science* 2, 2016, <https://doi.org/10.7717/peerj-cs.93>,

Analogamente, un modello predittivo a cura del Professor Daniel Martin Katz, uno degli studiosi più influenti sull'uso dell'IA applicata al diritto, ha ottenuto, sulle decisioni della Corte Suprema degli Stati Uniti, un'accuratezza del 70.2% riguardo l'esito del caso e del 71.9% circa la decisione dei singoli giudici⁴⁰. In questo caso, la previsione era basata su dati raccolti nel *database* della Corte Suprema⁴¹.

L'IA è stata poi impiegata ai fini della polizia predittiva, intesa come previsione statistica circa i luoghi in cui verranno commessi dei reati, oppure riguardo i potenziali autori o le potenziali vittime, al fine di orientare le forze dell'ordine nella prevenzione del crimine⁴². Un noto esempio è *Predpol*, un sistema di *machine learning* impiegato dalla polizia statunitense al fine di identificare i luoghi in cui risulta più probabile che si verificheranno reati sulla base di variabili quali la data e l'ora, il luogo e il tipo di crimine⁴³.

La polizia predittiva può influenzare il processo penale in quanto, indirizzando le forze dell'ordine verso determinati luoghi o persone, ha un impatto indiretto sul materiale probatorio che viene raccolto e messo a disposizione dell'autorità giudiziaria. Tuttavia, la polizia predittiva non è direttamente riconducibile al processo penale in quanto è funzionale all'attività di prevenzione dei reati e non nell'ambito delle indagini. La polizia predittiva non sarà quindi ulteriormente approfondita in questa tesi che ha invece come oggetto principale il processo penale.

La presente disamina ha piuttosto come obiettivo primario quello di approfondire l'impiego dell'IA predittiva nel processo penale, con particolare riferimento alla gestione di rischi quali la reiterazione del reato o il pericolo di fuga. Ad esempio, *COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)* stima il rischio generale di recidiva generale, di recidiva violenta e di recidiva antecedente al processo⁴⁴. Un altro esempio è *EPV-R*, uno strumento di valutazione del rischio impiegato dai giudici nei Paesi Baschi (Spagna) al fine di stimare la probabilità che un individuo commetta atti di grave violenza nei confronti del proprio *partner*⁴⁵.

Nonostante l'impiego dell'IA predittiva sia particolarmente diffuso sia nel settore privato che nel pubblico, non mancano criticità. Innanzitutto, anche se il modello predittivo presenta un'accuratezza elevata, la decisione presa sulla base di tale strumento non è necessariamente corretta. L'accuratezza potrebbe essere

consultato il 22 settembre 2025, pp. 1-3. Nikolaos Aletras è un professore di Elaborazione del Linguaggio Naturale (*Natural Language Processing, NLP*) presso il Dipartimento di Informatica dell'Università di Sheffield, con affiliazione anche al gruppo di *Machine Learning, ML*.

⁴⁰ D. M. KATZ -M.J. BOMMARITO -J. BLACKMAN *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, in *PLOS ONE* 12, 2017, pp. 1-3. Daniel Martin Katz è un professore di diritto presso Illinois Tech – Chicago-Kent College of Law dove insegna i seguenti corsi: *Artificial Intelligence & Law, Practice & Professionalism (Professional Responsibility), E-Discovery, Blockchain Cryptocurrency & Law, Introduction to Legal Technology & Innovation, Legal Project Management, Legal Process Improvement e Civil Procedure*.

⁴¹ *Ibid.*, p. 4.

⁴² E. PIETROCARLO. *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali in Sistema penale*, 2023, pp.1-2.

⁴³ I. MUGARI -E. OBIOHA, *Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing*, in *Social Sciences*, 2021, <https://doi.org/10.3390/socsci10060234>, consultato il 25 settembre 2025.

⁴⁴ Corte Suprema del Wisconsin, *op. cit.*, par. 14.

⁴⁵ A. VALDIVIA -C. HYDE VAAMONDE -J. GARCÍA MARCOS, *Judging the Algorithm: a Case Study on the Risk Assessment Tool for Gender-Based Violence Implemented in the Basque Country*, in *AI & Society*, vol. 40 n. 4, 2025, p.1.

misurata sulla base di dati raccolti in un contesto diverso rispetto a dove la decisione verrà presa⁴⁶. Ad esempio, un modello predittivo aveva erroneamente associato ai pazienti affetti da asma bronchiale una minore probabilità di contrarre la polmonite. La ragione è che tale sistema era stato addestrato sulla base di dati raccolti in ospedale dove coloro che soffrono di asma vengono immediatamente indirizzati nei reparti di terapia intensiva. In tale contesto, chi soffre di asma contrae la polmonite più raramente di chi non ne soffre in quanto questi ultimi non vengono immediatamente collocati nei reparti di terapia intensiva⁴⁷. Il caso dimostra che l'IA predittiva potrebbe presentare un'accuratezza elevata soltanto nello specifico contesto in cui i dati sono stati raccolti. Al di fuori di tale ambito, potrebbe suggerire delle decisioni erranee, come sottovalutare il rischio che un soggetto che soffre di asma contragga la polmonite⁴⁸.

Un altro rischio è l'eccessiva automazione, intesa come il rischio di delegare decisioni critiche a sistemi IA, in assenza della sorveglianza umana necessaria per rimediare agli errori algoritmici.

Per mitigare tale rischio, l'art. 11, par. 1 della direttiva 2016/680/UE prevede che “una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato sia vietata”. L'avverbio “unicamente” suggerisce che è necessario un grado minimo di intervento umano affinché il processo decisionale non sia considerato come completamente automatizzato⁴⁹.

Peraltro, la sorveglianza umana potrebbe non essere un rimedio efficace in quanto, nell'interazione uomo-macchina, le persone sono spesso vittime di un pregiudizio denominato *automation bias*. Con tale espressione, si fa riferimento al ricorso all'automazione come scorciatoia cognitiva, anziché ricercare e rielaborare attentamente le informazioni ricevute.⁵⁰ La sorveglianza umana potrebbe allora essere vana poiché la previsione fornita dall'IA potrebbe essere accettata acriticamente. Un esempio significativo di tale problematica è rappresentato dal già citato *SyRI (Systeem Risico Indicatie)*, in forza del quale le autorità olandesi hanno accusato ingiustamente numerose persone di frode fiscale⁵¹. In questo caso, la sorveglianza umana è stata inefficace poiché le autorità pubbliche hanno accettato acriticamente l'*output* prodotto dall'IA. Riprendendo l'analisi condotta da Narayanan e Kapoor, questi hanno evidenziato che vi sono alcuni ambiti in cui non bisognerebbe ricorrere all'IA per prevedere il futuro. Le ragioni sono molteplici. In alcuni casi, per giungere ad una previsione accurata occorrerebbe raccogliere una quantità di dati che supera le capacità attualmente disponibili. In un esperimento denominato *Fragile Families Challenge*, si è tentato di prevedere i risultati personali di bambini svantaggiati, giungendo a risultati inaccurati. Il *dataset* considerato, pur essendo nell'ordine delle migliaia di dati, era insufficiente per prevedere gli esiti di vita dei bambini che sono

⁴⁶ A. NARAYANAN -S. KAPOOR, *AI Snake Oil*, Princeton University press, Princeton, 2024, pp. 43-35.

⁴⁷ C. ROSS -B. HERMAN *Denied by AI: How Medicare Advantage Plans Use Algorithms to Cut Off Care for Seniors*, in *STAT*, 2023, <https://www.statnews.com/2023/03/13/medicare-advantage-plans-denial-artificial-intelligence/>, consultato il 4 agosto 2025.

⁴⁸ A. NARAYANAN -S. KAPOOR *op. cit.*, pp. 43-45.

⁴⁹ V. anche par. 4.2.2.

⁵⁰ K. L. MOSIER, L. J. SKITKA. *Automation Use and Automation Bias*, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, n. 3, 1999, p. 344. Per approfondimenti in merito all'*automation bias*, si rinvia al par. 3.3.3.

⁵¹ A. NARAYANAN -S. KAPOOR *op. cit.*, pp. 48-49.

determinati da un numero di variabili estremamente elevato⁵². Narayanan e Kapoor ipotizzano che, per giungere a dei risultati accurati, sarebbe stato necessario raccogliere una quantità di dati superiore al numero di esseri umani sulla Terra. Un'ulteriore criticità dell'IA predittiva è che le variabili in gioco potrebbero essere troppo specifiche e legate al caso concreto per poter essere osservate. Nel caso di *Fragile Families Challenge*, un bambino ha ottenuto dei risultati superiori rispetto a quelli previsti dal sistema a causa dell'inaspettato aiuto di un generoso vicino di casa. Si tratta di una variabile che difficilmente poteva rientrare tra i dati raccolti nell'esperimento. Per queste ragioni, Narayanan e Kapoor dubitano dell'accuratezza dell'IA nel prevedere le condizioni di vita delle persone⁵³.

Un'altra criticità dell'IA predittiva riguarda il pericolo di discriminazione che dipende dalla rappresentatività dei dati con cui il sistema è stato addestrato⁵⁴. Ad esempio, l'organizzazione *no-profit ProPublica* ha segnalato che *COMPAS* presenta dei *bias* di natura razziale nei confronti di imputati afroamericani. Ne conseguirebbe che, a parità di altre condizioni, alcuni imputati avrebbero una maggiore probabilità di essere erroneamente considerati ad alto rischio recidiva⁵⁵.

Infine, l'intelligenza artificiale predittiva può essere affetta dal problema della cosiddetta *black box* (scatola nera). Tale questione riguarda soprattutto il *deep learning*, una tecnica di apprendimento automatico che si basa su di una molteplicità di strati di rielaborazione. Questa complessità rende estremamente difficile ricostruire i processi logici che hanno condotto alla produzione dell'*output* poiché questi passaggi possono avvenire negli strati più profondi di rielaborazione il cui funzionamento non è noto. In caso di malfunzionamento, non è possibile sapere quali sono le ragioni che hanno determinato l'errore algoritmico⁵⁶. In sintesi, l'uso dell'IA predittiva da parte delle imprese e delle pubbliche amministrazioni solleva diverse questioni: l'importanza del contesto in cui i dati sono stati raccolti, l'eccessiva automazione, le difficoltà nel prevedere le condizioni di vita, la discriminazione e il problema della *black box*. Molte di queste criticità verranno trattate più diffusamente nel terzo capitolo, dedicato ai possibili rischi che l'intelligenza artificiale predittiva porrebbe qualora venisse implementata nel processo penale italiano.

1.3) Intelligenza artificiale predittiva e processo penale nel mondo

1.3.1) Lo stato dell'arte

Dopo aver illustrato i rischi dell'intelligenza artificiale predittiva in generale, occorre fare specifico riferimento al suo possibile impiego nel processo penale.

Negli Stati Uniti, è oggi particolarmente frequente il ricorso ai *risk assessment tools*, ossia strumenti di valutazione del rischio che mirano a stimare la probabilità che un individuo commetta nuovamente reati o

⁵² *Ibid.*, pp. 70-78.

⁵³ *Ibid.*, pp. 97-98.

⁵⁴ *Ibid.*, pp. 53-57. Per approfondire in merito al pericolo di discriminazione, si rinvia ai paragrafi 3.1.1-3.1.4.

⁵⁵ J. ANGWIN -J. LARSON -S. MATTU *et al.*, *Machine Bias: Risk Assessments in Criminal Sentencing*, *ProPublica*, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, consultato il 14 luglio 2025.

⁵⁶ G. BUONOMO, *op. cit.*, pp. 15-18. Sul problema della *black box*, si vedano i paragrafi 3.2.1-3.2.3.

comunque violi obblighi imposti dal sistema giudiziario. Questi strumenti producono dei *risk score*, ossia un punteggio di rischio indicante il livello di probabilità di verifica del rischio.

Per comprendere lo stato dell'arte è necessario ripercorrere le quattro generazioni di strumenti di valutazione del rischio.

La prima generazione era basata su giudizi clinici non strutturati, soggetti ad errori e pregiudizi, la cui valutazione clinica era rimessa sostanzialmente alla stima soggettiva dei singoli giudici. La seconda si basava su scale a punteggio cumulativo maggiormente standardizzate ma oggetto di critiche perché fondate sul consenso tra gli studiosi, anziché su metodi statistici solidi. La terza generazione si è basata sui bisogni dei detenuti anziché soltanto sul livello di rischio, mentre la quarta è funzionale ad un'integrazione maggiormente trasversale nel sistema giudiziario, non soltanto al fine di prevenire il rischio ma anche per selezionare modalità ed obiettivi di intervento e valutare i progressi della riabilitazione⁵⁷.

Per quanto riguarda i campi applicativi di tali strumenti, questi vengono utilizzati negli Stati Uniti per supportare l'autorità giudiziaria nella scelta tra il rilascio e la detenzione di chi è indagato in attesa di giudizio⁵⁸.

Un noto esempio è il *Public Safety Assessment (PSA)*, uno strumento di valutazione del rischio utilizzato principalmente in tre Stati: Arizona, Kentucky e New Jersey⁵⁹. PSA stima la probabilità di mancata comparizione in giudizio, di nuovo arresto per reato violento e di nuovo arresto per reato non violento. Si basa su di un numero particolarmente ristretto di fattori in quanto rilevano soltanto l'età e i precedenti penali⁶⁰.

I *risk assessment tools* sono utilizzati negli Stati Uniti anche per supportare i giudici nella determinazione della pena. In almeno cinque Stati, il ricorso a *risk assessment tools* nella determinazione della pena è obbligatorio: Arizona, Oklahoma, Ohio, Pennsylvania e Kentucky⁶¹. In Arizona, si richiede che nella relazione pre-sentenza (*presentencing report*) si faccia riferimento ai bisogni e a rischi criminogeni documentati da un *risk assessment tool*⁶². Vi sono invece degli Stati in cui questi strumenti di valutazione del rischio sono meramente permessi, senza essere obbligatori, come nell'Idaho, nella Louisiana e nell'Indiana⁶³.

Un noto esempio è il *Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)*, utilizzato nel caso *Loomis* dalla Circuit Court della Contea di Clark, al fine di supportare i giudici nella determinazione della pena. *COMPAS* si basa su diverse scale di rischio, ossia strumenti per misurare e

⁵⁷ T. BLOMBERG -W. BALES -K. MANN *Validation of the Compas Risk Assessment Classification Instrument*, Center for Criminology and Public Policy Research College of Criminology and Criminal Justice Florida State University, Tallahassee, Florida, 2010, pp. 2-4.

⁵⁸ Bureau of Justice Assistance, *What is Risk Assessment?*, in Bureau of Justice Assistance <https://bja.ojp.gov/program/psrac/basics/what-is-risk-assessment>, consultato il 4 agosto 2025. Per approfondimenti circa l'uso dell'IA nella scelta circa la detenzione preprocessuale, si rinvia ai parr. 2.2.1, 2.2.2, 2.2.3.

⁵⁹ D. KEHL -P. GUO -S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2017, pp. 22-23. pp.10-14.

⁶⁰ *Ibid.*, p. 10. Il *risk assessment tool PSA* è stato trattato più diffusamente nei parr. 2.2.2 e 2.2.3.

⁶¹ *Ibid.*, p. 16.

⁶² Codice dell'Amministrazione giudiziaria dello stato dell'Arizona, par. 6, 6-201.01(J)(3).

⁶³ D. KEHL -P. GUO -S. KESSLER *op. cit.* p. 16.

classificare la probabilità che un evento negativo si verifichi. Specificamente, si misura il rischio generale di recidiva e il rischio di recidiva violenta⁶⁴.

Altri *risk assessment tools* vengono invece adottati al fine di stimare il rischio di recidiva e i bisogni individuali dei detenuti, al fine di offrire a ciascuno di essi un programma adeguato. Questo approccio viene denominato *Risk Need Responsivity (RNR)* ed è basato su tre principi. Il *risk principle* implica che il trattamento deve essere adeguato al livello di rischio che ciascun soggetto presenta. Il *need principle* richiede che il trattamento si basi sui bisogni criminogeni dell'individuo (o fattori dinamici), ossia sulle cause che determinano il comportamento criminale. Il *responsivity principle* consiste nel massimizzare la capacità del reo di imparare sulla base delle sue specifiche modalità di apprendimento, della sua motivazione e dei suoi punti di forza⁶⁵. Ad esempio, il *Missouri Department of Corrections* utilizza l'*Ohio Risk Assessment System (ORAS)* per identificare i fattori che spingono le persone verso un comportamento criminale o comunque negativo⁶⁶.

Altri sistemi di valutazione del rischio vengono impiegati per valutare se disporre o meno misure alternative alla detenzione, come nel caso *HART (Harm Assessment Risk Tool)*, utilizzato sino a pochi anni fa nel Regno Unito. Tale sistema classificava i rei in tre categorie: rischio elevato, rischio moderato e rischio basso. Un reo veniva classificato ad alto rischio se si prevedeva che sarebbero stati commessi reati gravi quali l'omicidio volontario, la violenza sessuale o la rapina. Con rischio moderato si intendeva invece che, sebbene sussistesse il rischio di reiterazione del reato, questo era limitato a reati non gravi. Venivano classificati a rischio basso coloro che si prevedeva non commetterebbero reati nei due anni successivi. *HART* è stato usato per supportare gli operatori di custodia nel valutare se sottoporre o meno il reo al programma *Checkpoint*. Tale programma offriva misure riabilitative alternative alla detenzione nei confronti di soggetti classificati come a rischio moderato⁶⁷. La Durham Constabulary, la polizia territoriale nella contea di Durham, ha dichiarato di aver interrotto l'utilizzo di tale *risk assessment tool* nel 2021 in ragione della mancanza delle risorse necessarie per garantire la conformità agli standard etici e legali appropriati⁶⁸.

Vi sono poi *risk assessment tools* impiegati per assistere i giudici nell'adozione di misure cautelari in caso di violenza di genere. Un esempio è *EPV-R (Escala de Predicción del Riesgo de Violencia Grave contra la Pareja*

⁶⁴ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025, pp. 27-28. Per approfondimenti, si rinvia al par. 1.3.2 in cui si analizza come *COMPAS* è stato utilizzato nel caso *Loomis*.

⁶⁵ J. BONTA -D. A. ANDREWS, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, Public Safety Canada, Carleton University, 2007, p. 1.

⁶⁶ Missouri Department of Corrections, *Ohio Risk Assessment System (ORAS)*, in *Missouri Department of Corrections*, 2019, <https://doc.mo.gov/justice-reinvestment-initiative/oras>, consultato il 15 luglio 2025.

⁶⁷ M. OSWALD -J. GRACE -S. URWIN, et. al., *Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality*, in *Information & Communications Technology Law*, Taylor & Francis, 2018, p. 227.

⁶⁸ Fair Trials, *FOI reveals over 12,000 people profiled by flawed Durham police predictive AI tool*, in *Fair Trials*, 2022, <https://www.fairtrials.org/articles/news/foi-reveals-over-12000-people-profiled-by-flawed-durham-police-predictive-ai-tool/>, consultato il 12 gennaio 2026; Oxford Institute of Technology and Justice, *Tech and Justice Atlas*, Oxford Institute of Technology and Justice – Blavatnik School of Government, 2025, <https://www.techandjustice.bsg.ox.ac.uk/research/united-kingdom>, consultato il 12 gennaio 2026.

– *Revisada*), utilizzato nel sistema giudiziario dei Paesi Baschi (in Spagna). *EPV-R* produce un *risk score* circa il pericolo che un individuo commetta violenza grave nei confronti del proprio partner⁶⁹.

In sintesi, l'IA predittiva è oggi già utilizzata all'interno del processo penale di ordinamenti stranieri nella scelta tra il rilascio e la detenzione in attesa del processo, nella determinazione della pena, nell'individuazione di un programma di riabilitazione adeguato al singolo detenuto e nell'adozione di misure cautelari in caso di violenza di genere.

1.3.2) Il dibattito negli Stati Uniti: il caso *Loomis* e *COMPAS*

Come è stato illustrato nel paragrafo precedente, molti dei *risk assessment tools* attualmente in uso nel mondo sono stati adottati negli Stati Uniti, con la conseguenza che dibattito statunitense in materia risulta di particolare interesse. Occorre fare particolare riferimento al noto caso *Loomis*, in cui sono state affrontate diverse questioni problematiche che vengono discusse ancora oggi anche in Europa. Il caso riguardava un imputato condannato dalla Circuit Court della Contea di Clark (Wisconsin, USA) per aver messo dolosamente in pericolo la pubblica sicurezza, aver tentato di eludere la polizia stradale, aver guidato un veicolo a motore senza il consenso del proprietario, ed essere stato in possesso di un'arma da fuoco e di un fucile. Tali reati erano stati commessi nell'ambito di una sparatoria alla guida che, tuttavia, non era stata specificamente contestata all'imputato⁷⁰.

Nell'impugnare la sentenza della Circuit Court dinanzi alla Corte Suprema del Wisconsin, la difesa contestò la possibilità di utilizzare lo strumento di valutazione del rischio *COMPAS* ai fini della determinazione della pena.

COMPAS produce tre *risk score*, rappresentati graficamente su dei grafici a barre, relativamente al rischio generale di recidiva, al rischio di recidiva violenta e al rischio di recidiva pre-processuale, inteso come la probabilità che una persona commetta un nuovo reato prima dell'inizio del processo⁷¹. Nel caso *Loomis*, la valutazione del rischio generata da *COMPAS* era contenuta nel *presentence investigation report (PSI report)*, una relazione compiuta da un funzionario incaricato che viene utilizzata dal giudice per determinare la pena. Nello *PSI report*, si evidenziava che i *risk score* non sono stati sviluppati al fine di quantificare la pena o per decidere se una persona deve essere incarcerata o meno ma per identificare degli imputati che dovrebbero essere sottoposti a particolari misure riabilitative, in quanto considerati ad alto rischio⁷².

Dinanzi alla Corte Suprema del Wisconsin, *Loomis* lamentò che l'utilizzo di *COMPAS* nella determinazione della pena violava il diritto dell'imputato al *due process*, il giusto processo ai sensi del Quattordicesimo Emendamento della Costituzione statunitense⁷³. In primo luogo, sarebbe stato violato il diritto a essere

⁶⁹ A. VALDIVIA -C. HYDE VAAMONDE -J. GARCÍA MARCOS, *Judging the Algorithm: a Case Study on the Risk Assessment Tool for Gender-Based Violence Implemented in the Basque Country*, in *AI & Society*, vol. 40 n. 4, 2025. *EPV-R* è stato trattato più diffusamente nel par. 2.3.

⁷⁰ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 14.

⁷¹ *Ibid.*, par. 14.

⁷² *Ibid.*, parr. 15-17.

⁷³ *Ibid.*, par. 34.

giudicato sulla base di informazioni accurate, con particolare riferimento al diritto a rivedere e verificare l'informazione presente nello PSI report. Tale diritto era stato riconosciuto dalla stessa Corte Suprema del Wisconsin nel caso *State v. Scaff*⁷⁴. La difesa di Loomis evidenziava che il funzionamento di *COMPAS* era coperto dal segreto industriale, impedendo all'imputato di contestare la validità scientifica di *COMPAS*.

La Corte negò l'applicabilità di *State v. Scaff* al caso di specie, in quanto in tale controversia l'imputato non aveva nessuna possibilità di integrare e confutare quanto contenuto nello PSI report. Nel caso di specie, la difesa aveva invece avuto accesso ai dati di addestramento di *COMPAS*. Tali informazioni erano state fornite dallo stesso Loomis mediante un questionario o comunque erano state tratte da registri il cui accesso era pubblico⁷⁵. Nonostante il diritto a rivedere e verificare l'informazione non fosse applicabile al caso di specie, la Corte richiese che nello PSI report venisse indicato che la società proprietaria di *COMPAS* aveva invocato il segreto industriale in relazione al funzionamento del modello predittivo. Nelle intenzioni della Corte, tale informazione avrebbe dovuto garantire che *COMPAS* venisse impiegato con adeguata cautela⁷⁶.

Tale conclusione non ha soddisfatto una parte della dottrina statunitense. Si è osservato che la possibilità di rivedere e contestare le singole informazioni su cui il sistema di IA si basa non implica essere in grado di mettere in discussione le modalità attraverso cui il *risk score* è stato prodotto⁷⁷. È stato anche evidenziato che l'impossibilità per i giudici di avere accesso al funzionamento del sistema potrebbe impedire loro di comprendere il *risk score*⁷⁸. Il pericolo è che la commisurazione della pena venga influenzata in maniera eccessiva dalle imprese private che sviluppano tali algoritmi⁷⁹.

Si è allora proposto di richiedere che l'affidabilità del *risk assessment tool* venga valutata *ex ante* da un'autorità indipendente, prima che il modello predittivo venga messo a disposizione dell'autorità giudiziaria. Si potrebbe così verificare la validità di tale modello senza compromettere il segreto industriale⁸⁰.

Il problema della proprietà industriale è quanto meno attenuato nel caso di *PSA (Public Safety Assessment)*, in quanto i fattori utilizzati sono noti. Tali parametri sono infatti pubblicamente consultabili sul sito dell'*Arnold Foundation*, una fondazione filantropica privata statunitense che non ha scopo di lucro, a differenza della *Northpointe*, la società proprietaria di *COMPAS* (oggi *Equivant*)⁸¹.

Tornando al caso *Loomis*, un'ulteriore criticità sollevata dalla difesa riguardava la scarsa accuratezza di *COMPAS*. La Corte non ha negato che ci siano dei dubbi quanto alla sua accuratezza, avendo evidenziato che

⁷⁴ Corte Suprema del Wisconsin, *State v. Scaff*, n. 152 N.W. 2d 48, 1989, parr. 57-58.

⁷⁵ Corte Suprema del Wisconsin, *State v. Loomis*, *op. cit.*, parr. 46-56.

⁷⁶ *Ibid.*, par. 66.

⁷⁷ D. KEHL -P. GUO -S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2017, pp. 22-23.

⁷⁸ N. STRAUSS, *Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment*, in *Columbia Law Review*, vol. 120, 2020, pp. 1682-1685.

⁷⁹ *Ibid.*, pp. 1709-1710.

⁸⁰ U. RUFFOLO, *La macchina sapiens come "avvocato generale" ed il primato del giudice umano*, in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, p. 221. Il tema dell'opacità algoritmica dovuta alla proprietà intellettuale è stato affrontato più diffusamente nel par. 3.2.2.

⁸¹ Advancing Pretrial Policy & Research, *Public Safety Assessment – Risk factors*, AdvancingPretrial.org, <https://advancingpretrial.org/psa/factors/>, consultato il 14 luglio 2025.

COMPAS produce il *risk score* comparando l'imputato con dati raccolti a livello federale, anziché con specifico riferimento allo Stato del Wisconsin. La Corte ha anche rimarcato che l'accuratezza deve essere costantemente monitorata al fine di tenere conto dei cambiamenti nella popolazione. Infine, è stato sottolineato il rischio che il modello predittivo classifichi coloro che appartengono a delle minoranze come ad alto rischio di recidiva, con un *bias* discriminatorio. La Corte non ha quindi escluso le perplessità quanto alla natura discriminatoria di *COMPAS*, pur escludendo che l'uso di strumenti di valutazione del rischio ai fini della determinazione della pena sia necessariamente incompatibile con il *due process* statunitense. Al fine di garantire il rispetto del *due process*, la Corte ha richiesto che nello *PSI* vengano contenute determinate informazioni al fine di garantire che i giudici utilizzino il *risk assessment tool* con adeguata cautela. I giudici devono essere informati che, secondo alcuni studi, *COMPAS* classifica coloro che appartengono a delle minoranze come ad alto rischio di recidiva, in maniera non proporzionale⁸².

Si noti che dei dubbi sull'accuratezza di *COMPAS* sono stati avanzati anche da Hany Farid, professore di informatica presso l'Università di Berkeley. Si è confrontata l'accuratezza di *COMPAS* nel prevedere il rischio di recidiva con l'accuratezza delle previsioni effettuate da persone selezionate *online* che si presume non abbiano particolare esperienza nella giustizia penale. Tali previsioni presentano la stessa accuratezza e gli stessi *bias* di *COMPAS*. Inoltre, sebbene le previsioni effettuate si basino su 137 fattori, tale sistema presenta la stessa accuratezza di un semplice classificatore che tiene conto soltanto di due caratteristiche. Gli autori sostengono che questi risultati sollevano dei dubbi non soltanto sull'affidabilità di *COMPAS*, ma più in generale sulla possibilità di prevedere la recidiva mediante algoritmi⁸³.

Riprendendo l'analisi del caso *Loomis*, la difesa aveva anche sostenuto che il *due process* fosse stato violato relativamente al diritto ad una pena individualizzata⁸⁴. La difesa aveva argomentato che Loomis sarebbe stato giudicato sulla base dell'appartenenza ad un determinato gruppo di persone, anziché sulla base dei fatti contestati e sull'unicità della personalità dell'imputato⁸⁵. Tale rischio si concretizza quando i *risk assessment tools* individuano delle correlazioni statistiche tra l'appartenenza ad un determinato gruppo sociale, ed una maggiore o minore probabilità di recidiva. Nella sentenza si fa riferimento alla Guida Pratica a *COMPAS* in cui si afferma che i *risk assessment tools* sono funzionali a prevedere il comportamento di gruppi di persone e non di singoli individui⁸⁶. Il *Wisconsin Department of Corrections (DOC)* (l'ente responsabile per la gestione dei detenuti in Wisconsin) ha suggerito di utilizzare il proprio giudizio professionale nel valutare il *risk score* e ha previsto che la previsione algoritmica non dovrebbe essere accolta, in media, nel 10% dei casi, in modo da evitare automatismi⁸⁷.

⁸² Corte Suprema del Wisconsin, *State v. Loomis op. cit.*, parr. 58-66.

⁸³ H. FARID -J. DRESSEL, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, in *Science Advances* vol. 4, n.1, American Association for the Advancement of Science, 2018, p. 1.

⁸⁴ Corte Suprema del Wisconsin, *State v. Loomis op. cit.*, par. 34. Con riferimento, invece, alla possibile violazione del principio di individualizzazione della pena nel caso di introduzione dell'IA predittiva nel processo penale italiano si rinvia al par. 3.7.

⁸⁵ *Ibid.*, par. 67.

⁸⁶ *Ibid.*, par. 69.

⁸⁷ *Ibid.*, par. 70.

La Corte Suprema del Wisconsin, già nel caso *Gallion*, aveva riconosciuto che il diritto a una pena individualizzata costituisce uno dei principi fondamentali della giustizia penale del Wisconsin⁸⁸. La Corte aveva anche sostenuto che i giudici necessitano di informazioni maggiormente complete ai fini della determinazione della pena⁸⁹. Nel caso *Loomis*, la Corte ha ritenuto che il bisogno di informazioni più complete possa essere soddisfatto mediante il ricorso a *risk assessment tools*. Come sostenuto anche dalla Corte Suprema dell'Indiana, i *risk assessment tools* permettono ai giudici di applicare in maniera maggiormente efficace i criteri di determinazione della pena previsti dalla legge⁹⁰.

In definitiva, la Corte Suprema del Wisconsin distingue tra due ipotesi. Da un lato, viene riconosciuto che l'impiego di un *risk assessment tool* come fattore determinante ai fini della determinazione della pena potrebbe essere in contrasto con il diritto a una pena individualizzata⁹¹. Allo stesso tempo, la Corte considera legittimo il ricorso ad un *risk assessment tool* qualora questo non costituisca il fattore determinante ai fini della determinazione della pena, ma si limiti a rappresentare un'informazione aggiuntiva da valutare congiuntamente con altri fattori, come nel caso *Loomis*⁹².

Anche tale soluzione presenta delle criticità evidenziate dalla dottrina. La Corte ribadisce che i giudici dovrebbero valutare il *risk score* con cautela, tenendo conto anche di altri fattori rilevanti ai fini della determinazione della pena⁹³. Tuttavia, parte della dottrina ritiene che la Corte si affidi in maniera eccessiva alla cautela dei singoli giudici, in quanto gli esseri umani tendono a fidarsi delle soluzioni offerte dalle macchine in misura decisamente maggiore di quanto dovrebbero⁹⁴. Tale pregiudizio, noto come *automation bias*, consiste nella tendenza a fidarsi maggiormente delle capacità di una macchina rispetto alle proprie abilità, anche in presenza di evidenze circa l'inaccuratezza del sistema⁹⁵.

Riprendendo l'esame del caso *Loomis*, un ulteriore motivo di impugnazione difensiva riguardava il diritto dell'imputato a non essere giudicato sulla base del genere. In *State v. Harris*, la Corte Suprema del Wisconsin aveva sancito il diritto dell'imputato a non essere giudicato sulla base di fattori non appropriati, quali l'etnia o il genere⁹⁶. Nel caso *Loomis*, le parti concordavano sulla sussistenza di evidenze statistiche quanto alla maggiore probabilità di recidiva degli imputati uomini rispetto alle imputate donne, ma la difesa contestava la legittimità di ogni generalizzazione basata sul genere, anche qualora dovesse essere statisticamente accurata⁹⁷.

⁸⁸Corte Suprema del Wisconsin, *State v. Gallion*, n.270 N.W. 2d 535, 2004, par. 48.

⁸⁹*Ibid.*, par. 34.

⁹⁰Corte Suprema dell'Indiana, *Malenchik v. State*, n. 928 N.E.2d 564, 2010, p. 574.

⁹¹Corte Suprema del Wisconsin, *State v. Loomis*, *op. cit.* par. 68.

⁹²*Ibid.*, par. 72.

⁹³*Ibid.*, par. 74.

⁹⁴W. H. GRAVETT, *Judicial Decision-Making in the Age of Artificial Intelligence* in H. SOUSE ANTUNES *Multidisciplinary Perspectives on Artificial Intelligence and the Law, Law, Governance and Technology Series*, Springer International Publishing, Cham, 2023, p. 291.

⁹⁵*Ibid.*, p. 290.

⁹⁶Corte Suprema del Wisconsin, *State v. Harris*, n. 326 N.W 2d 685, 2010, par. 33.

⁹⁷Corte Suprema del Wisconsin, *State v. Loomis op. cit.* parr. 78-79.

La Corte Suprema del Wisconsin ha ripreso lo stesso caso *State v. Harris* in cui si era ribadito che è onere dell'imputato dimostrare che i giudici hanno effettivamente determinato la pena sulla base del genere⁹⁸. Nel caso di specie, la Corte ha sostenuto che l'imputato non aveva assolto tale onere. La Circuit Court (il giudice *a quo*) aveva motivato sulla base di fattori diversi rispetto al genere, quali la gravità del reato, la storia criminale e l'alto rischio di recidiva suggerito da COMPAS⁹⁹.

Sono stati sollevati dubbi circa la compatibilità di *risk assessment tools* che tengono conto del genere con la clausola di pari protezione sancita dal Quattordicesimo Emendamento della Costituzione statunitense¹⁰⁰. Tali perplessità si fondano sul caso *Craig v. Boren*, in cui la Corte Suprema degli Stati Uniti ha considerato illegittima una legge che prevedeva, in base al genere, due diversi limiti di età ai fini del consumo di determinate tipologie di birre. La Corte Suprema ha riconosciuto che gli uomini presentano una maggiore probabilità statistica di guidare in stato di ebbrezza rispetto alle donne, ma la legge è stata ritenuta illegittima in quanto generalizzava sulla base del genere¹⁰¹. Sulla base di *Craig v. Boren*, si è quindi sostenuto che il ricorso a *risk assessment tools* che tengono conto del genere potrebbe essere illegittimo, anche qualora basarsi sul genere dovesse migliorare l'accuratezza del modello predittivo. Vi sarebbe quindi una tensione tra l'accuratezza, in base alla quale gli algoritmi di *machine learning* vengono valutati, e la giustizia su cui l'ordinamento giuridico si fonda¹⁰².

Nel dibattito statunitense, si è fatto riferimento, tra i possibili fattori di discriminazione, non solo al genere ma anche allo status socioeconomico e l'etnia. La Corte Suprema degli Stati Uniti, nel caso *Bearden v. Georgia*, ha sancito che la povertà dell'imputato non può essere considerata un fattore nel valutare il rischio di recidiva, con la conseguenza che ad un soggetto povero viene comminata una pena maggiore¹⁰³. Qualora un sistema di IA si basi sullo status socioeconomico, si potrebbe applicare dunque la sentenza *Bearden v. Georgia*¹⁰⁴.

È pacifico che i giudici non possano tenere conto dell'etnia nella determinazione della pena e nessun *risk assessment tool* attualmente in uso si basa esplicitamente su di essa. Può però accadere che l'IA discrimini indirettamente in ragione dell'etnia, sebbene il sistema non sia stato specificamente addestrato a considerare tale fattore¹⁰⁵. Tuttavia, la Corte Suprema degli Stati Uniti, nel caso *Washington v. Davis*, ha richiesto che, affinché una legge formalmente neutra sia illegittima, deve essere provato l'intento discriminatorio di natura razziale.¹⁰⁶ Risulta dunque difficile provare l'illegittimità costituzionale del ricorso ad un modello predittivo

⁹⁸ Corte Suprema del Wisconsin, *State v. Harris*, *op. cit.*, par. 64.

⁹⁹ Corte Suprema del Wisconsin, *State v. Loomis*, *op. cit.* par. 85-86.

¹⁰⁰ D. KEHL -P. GUO -S. KESSLER *op. cit.*, pp. 25- 26. Per quanto riguarda, invece, l'analoga questione sulla compatibilità dell'impiego di algoritmi discriminatori con il principio di uguaglianza (art. 3 Cost.), il divieto di discriminazione (art. 14 CEDU) e il diritto di non-discriminazione (art. 21 CDFUE), si rinvia al par. 3.2.3.

¹⁰¹ Corte Suprema degli Stati Uniti, *Craig v. Boren*, n. 429 U.S. 190, 1976, p. 210.

¹⁰² D. KEHL -P. GUO -S. KESSLER *op. cit.*, pp. 25- 26.

¹⁰³ Corte Suprema degli Stati Uniti, *Bearden v. Georgia*, n. 461 U. S. 660, 1983, pp. 661-662.

¹⁰⁴ D. KEHL -P. GUO -S. KESSLER *op. cit.* p. 26.

¹⁰⁵ *Ibid.*, p. 24.

¹⁰⁶ Corte Suprema degli Stati Uniti, *Washington v. Davis*, n. 426 U.S. 229, 1976, pp. 239-245.

che non è stato esplicitamente addestrato a tenere conto dell'etnia, ma che potrebbe discriminare indirettamente in base ad essa¹⁰⁷.

Lo stesso *COMPAS* è stato oggetto di critiche particolarmente severe quanto al rischio di discriminazione di natura razziale da parte dell'organizzazione *no profit ProPublica*, in quanto classificherebbe erroneamente gli imputati afroamericani come futuri criminali, con una frequenza pari al doppio degli imputati bianchi. Si è anche osservato che gli imputati bianchi sarebbero stati erroneamente classificati come a basso rischio, con una frequenza superiore agli imputati afroamericani¹⁰⁸.

La *Northpointe*, la società all'epoca proprietaria di *COMPAS*, ha risposto alle critiche valutando l'equità del modello predittivo sulla base degli stessi dati considerati da *ProPublica*. Secondo *Northpointe*, *ProPublica* non avrebbe tenuto conto della diversa frequenza di recidiva tra gli imputati afroamericani e gli imputati bianchi. Tenendo conto di tale differenza, i dati confuterebbero la sussistenza di un *bias* razziale¹⁰⁹.

In conclusione, dall'analisi del caso *Loomis* e dello strumento di valutazione del rischio *COMPAS* emergono molte delle problematiche che hanno caratterizzato il dibattito in merito all'uso dell'IA predittiva nel processo penale come la proprietà industriale, il diritto a una pena individualizzata e il pericolo che l'IA produca *output* discriminatori. Queste criticità verranno discusse in maniera maggiormente analitica nel terzo capitolo.

1.3.3) Il dibattito in Europa

In Europa, il ricorso all'IA predittiva nella giustizia penale è complessivamente meno frequente rispetto agli Stati Uniti, con la conseguenza che dibattito in materia ha avuto natura prevalentemente teorica. Questa differenza è stata sottolineata anche dalla Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ), secondo cui l'utilizzo di algoritmi di IA nei sistemi giudiziari europei rimane prevalentemente un'iniziativa commerciale, affidata al settore privato.¹¹⁰ Le principali applicazioni dell'intelligenza artificiale nei sistemi giudiziari europei consistono in motori di ricerca giurisprudenziali avanzati, risoluzione di controversie online, assistenza nella redazione di atti e analisi predittiva, ossia l'uso di modelli di IA per prevedere esiti giudiziari. Alcuni sistemi di analisi predittiva, come *Prédicte* e *Case Law Analytics*, vengono usati in Francia nell'ambito di processi diversi da quello penale¹¹¹. Lo strumento di valutazione del rischio *HART*, il quale riguarda invece l'ambito penalistico, non è più in uso nel Regno Unito, come si è visto¹¹².

¹⁰⁷ D. KEHL -P. GUO -S. KESSLER *op. cit.* pp. 24-25.

¹⁰⁸ J. ANGWIN -J. LARSON -S. MATTU, *et al.*, *Machine Bias: Risk Assessments in Criminal Sentencing*, *ProPublica*, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, consultato il 3 agosto 2025.

¹⁰⁹ W. DIETERICH -C. MENDOZA -T. BRENNAN, *COMPAS Risk Scales: Demonstrating Accuracy, Equity and Predictive Parity*, Northpointe Inc., 2016, <https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html>, consultato il 4 agosto 2025, pp.1-3. Per approfondire in merito alle criticità sollevate da *ProPublica* e alla risposta della *Northpointe*, si rinvia al par. 3.1.1.

¹¹⁰ Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ), *Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, Consiglio d'Europa, 2018, p. 15.

¹¹¹ *Ibid.*, pp. 15-16.

¹¹² Fair Trials, *FOI reveals over 12,000 people profiled by flawed Durham police predictive AI tool*, in *Fair Trials*, 2022, <https://www.techandjustice.bsg.ox.ac.uk/research/united-kingdom>, consultato il 12 gennaio 2026; Oxford Institute of Technology and Justice, *Tech and Justice Atlas*, Oxford Institute of Technology and Justice – Blavatnik School of Government, 2025, <https://www.techandjustice.bsg.ox.ac.uk/research/united-kingdom>, consultato il 12 gennaio 2026, v. par. 1.3.2.

La stessa CEPEJ ha adottato una Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari, contenente cinque principi non vincolanti che dovrebbero guidare l'azione degli Stati membri del Consiglio d'Europa.

In primo luogo, si richiede che il ricorso all'intelligenza artificiale sia compatibile con i diritti fondamentali garantiti dalla Convenzione europea dei diritti dell'uomo (CEDU), con particolare riferimento alle garanzie del diritto ad un processo equo, sancito dall'art. 6 della Convenzione stessa¹¹³.

In secondo luogo, la Carta etica sancisce il principio di non discriminazione: gli Stati membri dovrebbero prevenire lo sviluppo o l'intensificazione della discriminazione tra persone o gruppi di persone¹¹⁴. Si pensi al dibattito negli Stati Uniti sulla natura discriminatoria di *COMPAS*, a danno della popolazione afroamericana¹¹⁵. La Carta etica richiede che gli algoritmi di IA non conducano a risultati deterministici e individua la vigilanza umana come soluzione. Si evidenzia poi che il pericolo di discriminazione è particolarmente significativo quando vengono trattati dati sensibili quali l'origine razziale, la fede religiosa o l'appartenenza sindacale¹¹⁶.

Inoltre, viene adottato il principio della qualità e della sicurezza dei dati. I dati attraverso cui i sistemi di IA sono addestrati devono essere selezionati all'interno di fonti affidabili e non raccolti generalmente da internet. La sicurezza fa riferimento all'intangibilità e all'integrità del sistema rispetto ad attacchi informatici¹¹⁷.

Il quarto principio sancito dalla Carta riguarda la trasparenza, l'imparzialità e l'equità. L'imparzialità implica che il sistema di IA deve essere privo di pregiudizi, mentre l'equità fa riferimento alla necessità di far prevalere gli interessi della giustizia rispetto ad interessi privati. Infine, la trasparenza è intesa come la possibilità di avere accesso al funzionamento del sistema di IA, con la conseguente necessità di un bilanciamento non facile tra proprietà intellettuale e trasparenza.

Vengono illustrate due possibili soluzioni. In primo luogo, il codice potrebbe essere *open source* e dunque non soggetto al segreto industriale. In alternativa, delle autorità indipendenti potrebbero essere incaricate di certificare la validità dei metodi adottati dal sistema di IA senza compromettere il segreto industriale¹¹⁸. La proprietà intellettuale non costituisce il solo ostacolo alla trasparenza algoritmica, in quanto il sistema di IA potrebbe essere affetto dal problema della *black box*, intesa come la difficoltà tecnica di ricostruire i processi logici in base ai quali il modello ha prodotto un determinato *output*¹¹⁹. Affinché l'IA sia realmente trasparente è necessario che anche i non esperti siano in grado di comprenderne il funzionamento. La Carta etica richiede quindi che il funzionamento del sistema sia spiegato in modo chiaro e familiare¹²⁰.

¹¹³ Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) *op. cit.*, p. 7.

¹¹⁴ Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) *op. cit.*, p. 8.

¹¹⁵ J. ANGWIN -J. LARSON -S. MATTU *et al.*, *Machine Bias: Risk Assessments in Criminal Sentencing*, ProPublica, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, consultato il 3 agosto 2025.

¹¹⁶ Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) *op. cit.* p. 8.

¹¹⁷ *Ibid.*, p. 10.

¹¹⁸ *Ibid.*, p. 11. In materia, si veda il par. 3.2.2.

¹¹⁹ G. BUONOMO, *Macchine Intelligenti, processo induttivo e apprendimento autonomo*, in G. BUONOMO, G. CIACCI, A. COSTANZO, *Macchine intelligenti e diritto*, Giappichelli, Torino, 2025, pp. 15-18. Cfr. par. 3.2.1.

¹²⁰ Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ), *op. cit.*, p. 11.

Da ultimo, viene stabilito il principio del controllo dell'utilizzatore (*under user control*). Si richiede di evitare un approccio prescrittivo, in cui gli operatori della giustizia sono vincolati alla soluzione offerta dall'IA. Le persone devono invece essere in grado di valutare criticamente l'*output* prodotto dall'IA. Lo stesso principio implica che l'utilizzatore abbia il controllo delle proprie scelte e che debba essere informato in modo chiaro e comprensibile delle diverse opzioni disponibili, con particolare riferimento all'obbligo di informare del diritto di accesso a un tribunale, ai sensi dell'art. 6 della CEDU¹²¹.

La CEPEJ evidenzia le componenti del diritto ad un processo equo (art. 6 della CEDU) che potrebbero essere compromesse dall'intelligenza artificiale¹²². In primo luogo, occorre rispettare il diritto di accesso ad un tribunale, riconosciuto dalla Corte EDU nella sentenza *Golder c. Regno Unito*¹²³. L'art. 6 della CEDU non permette l'introduzione di un giudice robot che sostituisca il giudice umano perché in tal caso le parti sarebbero private del diritto di accesso ad un tribunale¹²⁴.

In secondo luogo, l'IA potrebbe compromettere il principio del contraddittorio, in quanto le parti, a causa della proprietà industriale e dell'opacità algoritmica, potrebbero non avere accesso al funzionamento del sistema di IA. Le parti sarebbero allora private della possibilità di fare delle osservazioni volte a porre dei dubbi sulla validità del sistema di IA¹²⁵. Il principio del contraddittorio è stato infatti definito dalla Corte EDU come il diritto di conoscere e commentare tutte le prove e le osservazioni compiute al fine di influenzare la decisione del tribunale¹²⁶.

L'utilizzo dell'IA potrebbe comportare altresì degli squilibri tra le parti, con un possibile impatto negativo sul principio della parità delle armi, anch'esso riconosciuto dalla Corte EDU come parte del diritto ad un processo equo¹²⁷. L'IA potrebbe favorire le istituzioni e le grandi società dotate di competenze informatiche e di risorse finanziarie, e sfavorire una parte della popolazione avente poca dimestichezza nell'uso del computer¹²⁸.

La CEPEJ ipotizza poi delle limitazioni circa l'imparzialità e l'indipendenza giudiziaria. L'IA potrebbe mettere pressione nei confronti dei giudici affinché non si discostino dalla soluzione offerta dal modello predittivo. Tale soluzione spesso si basa sull'orientamento prevalente in giurisprudenza. Secondo la CEPEJ, non si può escludere che, in alcuni ordinamenti dove l'indipendenza giudiziaria è a rischio, l'esecutivo controlli i giudici affinché non si discostino da tale tendenza maggioritaria¹²⁹.

¹²¹ *Ibid.*, p. 12.

¹²² *Ibid.*, pp. 33-34.

¹²³ Corte EDU, *Golder c. Regno Unito*, ric. n. 4451/70, 1975, par. 36.

¹²⁴ J. ULENAERS, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?* in *Asian Journal of Law and Economics*, Vol.11, n.2, p. 33.

¹²⁵ V. STEPONENAITE -P. VALCKE, *Judicial Analytics on Trial: An Assessment of Legal Analytics in Judicial Systems in the Light of the Right to a Fair Trial*, in *Maastricht Journal of European and Comparative Law*, Vol. 27, n. 6, SAGE Publications, 2020, pp. 766-767.

¹²⁶ Corte EDU, *Reinhardt e Slimane-Kaïd c. Francia*, ric. n. 23043/93 e 22921/93, 1998, par. 103. Si veda il par. 2.3.3 sulle garanzie processuali che potrebbero essere violate a causa dell'opacità algoritmica, in cui si tratta anche delle possibili violazioni del principio del contraddittorio.

¹²⁷ Corte EDU, *Dombo Beheer B.V. c. Paesi Bassi*, ric. n. 51761/22, 52189/22, 52564/22, 52761/22, 54140/22, 54165/22, 54177/22, 57850/22, 57978/22, 93/23, 2394/23, 2395/23, 2397/23, 4470/23, 4472/23, 4476/23, 6337/23, 6340/23, 37336/23, 42349/23, 1993, par. 33.

¹²⁸ Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) *op. cit.* pp. 33-34.

¹²⁹ *Ibid.*, p. 34.

Infine, la CEPEJ analizza delle questioni che riguardano specificamente la giustizia penale. L'intelligenza artificiale rischia di favorire l'insorgenza di dottrine deterministiche, ossia quelle teorie secondo cui il comportamento criminale sarebbe determinato da fattori fissi. Potrebbero allora essere compromesse le teorie di individualizzazione della pena secondo cui la persona deve essere giudicata con riferimento al fatto compiuto e alla sua specifica situazione personale¹³⁰. Il pericolo di determinismo penale potrebbe insorgere in quanto una persona potrebbe essere giudicata sulla base dell'appartenenza ad un determinato gruppo sociale. L'imputato non verrebbe giudicato primariamente in relazione al fatto di reato oggetto dell'imputazione ma in ragione della sua personalità¹³¹.

Dopo aver analizzato come il rapporto tra IA e giustizia penale è stato discusso dal Consiglio d'Europa, occorre sinteticamente illustrare il dibattito all'interno dell'Unione europea, con particolare riferimento al regolamento 2024/1968/UE, noto come AI Act.

Sin dal 2019, la Commissione europea si è posta l'obiettivo di creare fiducia nell'intelligenza artificiale antropocentrica. Con tale espressione, si intende che l'IA non deve essere fine a sé stessa, ma uno strumento a servizio delle persone avente il fine ultimo di migliorare il benessere degli esseri umani¹³². Il Gruppo di esperti ad alto livello sull'intelligenza artificiale, istituito dalla stessa Commissione europea, ha individuato i requisiti che i sistemi di IA devono presentare affinché siano affidabili: intervento e sorveglianza umana, robustezza tecnica e sicurezza, riservatezza e governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale, e *accountability* (principio di responsabilità)¹³³. Alcuni di questi requisiti sono stati tradotti in obblighi normativi previsti dall'AI Act per i sistemi ad alto rischio, come nel caso della trasparenza o della sorveglianza umana (artt. 13-27).

Nel 2020, lo Studio del Comitato Libe, una commissione permanente del Parlamento Europeo, ha analizzato l'impatto dell'IA sui diritti fondamentali. Lo studio considera il rischio che i sistemi IA favoriscano decisioni di natura discriminatoria, con conseguente lesione del diritto a non essere discriminati ai sensi dell'art. 21 della Carta dei diritti fondamentali dell'Unione europea (CDFUE)¹³⁴. Si considera inoltre il rischio di ledere il diritto ad un ricorso effettivo ai sensi dell'art. 47 CDFUE, qualora non sia possibile contestare una decisione adottata sulla base dell'intelligenza artificiale (si pensi al caso in cui l'IA non sia trasparente a causa della proprietà intellettuale o al problema della *black box*)¹³⁵.

Lo studio ha fornito poi delle raccomandazioni che hanno guidato la disciplina da parte dell'Unione europea in materia di intelligenza artificiale, influenzando il contenuto dell'AI Act, richiedendo una verifica del quadro

¹³⁰ *Ibid.*, p. 34.

¹³¹ L. NOTARO, *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021, pp. 65-66.

¹³² Commissione europea, *Comunicazione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni – Creare fiducia nell'intelligenza artificiale antropocentrica*, COM (2019) 168 final, Bruxelles, 2019, p. 2.

¹³³ Gruppo di esperti ad alto livello sull'intelligenza artificiale, *Orientamenti etici per un'IA affidabile*, Commissione europea, 2019, pp. 15-23.

¹³⁴ Parlamento europeo, Comitato LIBE, *Studio sull'impatto dell'intelligenza artificiale sui diritti fondamentali nella U.E.*, 2020, pp. 40-42; Il Comitato LIBE è una commissione permanente del Parlamento Europeo che si occupa di tutte le politiche dell'Unione europea che riguardano i diritti fondamentali.

¹³⁵ *Ibid.*, p. 43.

normativo all'epoca esistente¹³⁶ ed evidenziando la necessità di adattare la regolazione dell'intelligenza artificiale alle specifiche esigenze della giustizia penale, con particolare riferimento al problema della trasparenza. Lo studio ha proposto inoltre di adottare regole speciali per alcune pratiche, come, ad esempio, il riconoscimento facciale e la necessità di rafforzare la fiducia nei sistemi di IA finanziati dall'Unione europea, mediante l'adozione di sistemi soggetti al controllo pubblico¹³⁷.

Anche alla luce di tale analisi, nel 2021, la Commissione europea ha inviato al Parlamento e al Consiglio dell'UE la proposta di regolamento in materia di IA. Nel 2024, al termine di un articolato confronto, il Parlamento e il Consiglio hanno raggiunto un accordo, con la conseguente approvazione del testo definitivo. Una sua caratteristica fondamentale è il *risk based approach*, in quanto vengono disposte delle regole proporzionali al livello di rischio che una determinata pratica o un certo sistema presenta¹³⁸. In base alla tipologia del rischio, il regolamento distingue tra pratiche vietate (art. 5), sistemi ad alto rischio (allegato III), sistemi che presentano un livello di rischio limitato (soggetti soltanto agli obblighi di trasparenza, ai sensi dell'art. 50) e sistemi che esulano dall'ambito di applicazione del regolamento, definito dall'art. 2 (come nel caso della mera ricerca scientifica antecedente all'immissione del prodotto nel mercato).

Inoltre, l'AI Act adotta il cosiddetto *new approach*, mutuato da altre normative europee¹³⁹. Con tale espressione si intende che la Commissione europea richiede ad Organismi di Normazione Europei (ESOs) di sviluppare delle norme armonizzate indicanti specifici requisiti tecnici la cui aderenza comporta la presunzione di conformità al regolamento (art. 40). L'obiettivo è promuovere la certezza del diritto a beneficio degli sviluppatori di sistemi IA che, basandosi su norme armonizzate, riducono il rischio di non risultare conformi all'AI Act.

1.4) Il quadro normativo

1.4.1) AI Act, l. 132/2025 e direttiva LED

Nei precedenti paragrafi, si è illustrato il dibattito in materia di IA e processo penale. Nei seguenti paragrafi, verrà evidenziato il quadro normativo applicabile qualora l'IA predittiva dovesse essere introdotta nel processo penale italiano.

Innanzitutto, assume rilievo il già citato regolamento 2024/1968/UE (AI Act) il cui ambito di applicazione è limitato ai sistemi di IA che soddisfino la definizione di cui all'art. 3, n.1 del regolamento: *“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”*.

¹³⁶ *Ibid.*, pp. 67-68.

¹³⁷ *Ibid.*, pp. 68-69.

¹³⁸ T. EVAS, *The EU Artificial Intelligence Act: Advancing Innovation for Trustworthy AI*, in *Journal of AI Law and Regulation*, vol. 1, n. 1, 2024, pp. 99-100.

¹³⁹ S.DE VRIES -O. KANEVSKAIA -R.DE JAGER, *Internal Market 3.0: The Old “New Approach” for Harmonising AI Regulation*, in *European Papers*, vol. 8, n. 2, 2023, pp. 594-595.

In primo luogo, i sistemi di IA devono presentare “*livelli di autonomia*”, cioè un significativo livello di indipendenza rispetto all’uomo che li rende in grado di operare senza il coinvolgimento di persone fisiche. Tale requisito esclude software semplici, basati su regole definite esclusivamente da esseri umani (considerando 12).

Un secondo requisito, è la c.d. “*adattabilità*”. Ciò implica che i sistemi di IA possono modificare il proprio comportamento durante l’utilizzo (considerando 12). L’art. 3, n.1 richiede poi che tali sistemi siano in grado di fare delle inferenze. Devono, cioè, essere in grado di produrre *output* quali previsioni, raccomandazioni, decisioni, a partire dai dati forniti in *input*. Tali *output* possono influenzare l’ambiente circostante, inteso come il contesto in cui i sistemi di IA operano (considerando 12).

Un’autorità giudiziaria che usi un sistema di IA potrebbe qualificarsi come *deployer*, ai sensi dell’art. 3, par. 4 che lo definisce come: “*una persona fisica o giuridica, un’autorità pubblica, un’agenzia o un altro organismo che utilizzi un sistema di IA sotto la propria autorità*”. Ai sensi dell’art.2, par.1 lett. b), il regolamento si applica ai *deployer* di sistemi IA che hanno il loro luogo di stabilimento o sono situati all’interno dell’Unione. Pertanto, qualora l’autorità giudiziaria italiana impiegasse l’IA nel processo penale, si rientrerebbe nell’ambito di applicazione del regolamento.

L’art. 5 del regolamento fornisce un elenco di pratiche vietate. Ai fini del processo penale, rileva l’art. 5, par. 1 lett. d), in cui viene proibita “*l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazione del rischio relative a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità*”. Tale divieto viene escluso quando ricorrono “*sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa*”¹⁴⁰.

Il regolamento prevede diversi sistemi di IA che vengono classificati come ad alto rischio. In sostanza, è richiesto che il sistema rientri nelle ipotesi di cui all’allegato III del regolamento e che presenti “*un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche*” (art. 6).

Tra i sistemi ad alto rischio, di particolare rilievo ai fini della presente disamina è l’ipotesi di cui all’allegato III, n. 6, lett. d): i sistemi IA destinati ad essere usati “*per determinare il rischio di commissione del reato o di recidiva in relazione a una persona fisica non solo sulla base della profilazione delle persone fisiche*”. Oppure, ai sensi dell’allegato III, n. 8, n. 3, lett. a), sono sistemi ad alto rischio “*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*”¹⁴¹.

La qualifica di un sistema come ad alto rischio determina l’applicabilità degli obblighi di cui agli artt. 8-27 del regolamento.

¹⁴⁰ Cfr. par. 4.3.1.

¹⁴¹ Cfr. par. 4.3.2.

L'art. 9 prescrive di adottare un *“sistema di gestione dei rischi”*. In questo modo, è possibile individuare e minimizzare i rischi che l'IA può presentare a danno dei diritti fondamentali (considerando 65).

L'art. 10 richiede invece che i dati di addestramento, convalida e prova devono rispettare determinati criteri di qualità. Tale requisito è funzionale ad evitare che, a causa della scarsa qualità dei dati, l'IA sia fonte di discriminazione vietata dall'Unione (considerando 67).

L'art. 11 richiede di redigere una documentazione tecnica, al fine di dimostrare che il sistema ad alto rischio sia conforme al regolamento. Ai sensi dell'art. 12, i sistemi ad alto rischio devono consentire automaticamente la registrazione degli eventi. In questo modo, si garantisce la tracciabilità del sistema, al fine di comprendere le ragioni di eventuali malfunzionamenti (considerando 71).

L'art. 13 prevede che i sistemi di IA devono essere sufficientemente trasparenti da garantire che il *deployer* sia in grado di interpretare l'output del sistema ed utilizzarlo correttamente. Tale norma è stata pensata per rispondere alle preoccupazioni relative all'opacità dell'IA (considerando 72).

L'art. 14 richiede che i sistemi ad alto rischio siano progettati in modo tale da essere efficacemente supervisionati da un essere umano durante il loro uso. La sorveglianza umana serve a garantire che i sistemi siano utilizzati come previsto e ad affrontare il loro impatto sui diritti fondamentali durante tutto il ciclo di vita e non solo in fase di progettazione (considerando 73).

Infine, l'art. 15 richiede che tali sistemi conseguano un adeguato livello di robustezza, accuratezza e cybersicurezza. L'accuratezza è la capacità del sistema di produrre risultati corretti. La robustezza tecnica è la resilienza del sistema rispetto ad eventi indesiderati quali errori o guasti. Invece, la cybersicurezza è la resilienza rispetto a tentativi di terzi malintenzionati di compromettere la sicurezza del sistema (considerando 75-76).

Un'altra norma specifica per i sistemi ad alto rischio concerne *“il diritto alla spiegazione dei singoli processi decisionali”*, ai sensi dell'art. 86. Tale diritto sussiste qualora una decisione *“produca effetti giuridici o in modo analogo incida significativamente su tale persona in un modo che essa ritenga avere un impatto negativo sulla sua salute, sulla sua sicurezza o sui suoi diritti fondamentali”*. Il diritto consiste nella possibilità di ottenere *“dal deployer spiegazioni chiare e significative sul ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione adottata”*.

In Italia, è recentemente entrata in vigore la legge 23 settembre 2025, n. 132, avente ad oggetto *“disposizioni e deleghe al governo in materia di intelligenza artificiale”*. Tale legge mira a disciplinare l'IA nell'ordinamento italiano, introducendo norme di complemento rispetto a quanto già previsto dall'AI Act.

La struttura della legge è la seguente: gli artt. 1-6 sono dedicati ai principi e alle finalità, gli artt. 7-18 concernono disposizioni di settore, come l'art. 15 che disciplina l'uso dell'IA nell'attività giudiziaria¹⁴², gli artt. 19-24 riguardano la strategia nazionale, le autorità nazionali, nonché le azioni di promozione, l'art. 25

¹⁴² Cfr. par. 4.4.1.

disciplina il diritto d'autore, l'art. 26 introduce delle modifiche al codice penale mediante la previsione di nuovi reati e aggravanti e gli artt. 27 e 28 presentano rispettivamente disposizioni finanziarie e finali.

Il legislatore italiano, all'art. 20 di tale legge, ha scelto di individuare quali autorità nazionali per l'intelligenza artificiale l'Agenzia per l'Italia Digitale (AGID) e l'Agenzia per la cybersicurezza nazionale (ACN). Più specificamente, l'AGID è stata qualificata come autorità di notifica, secondo la definizione di cui all'art. 3, n. 19, dell'AI Act: *“l'autorità responsabile dell'istituzione e dell'esecuzione delle procedure necessarie per la valutazione, la designazione e la notifica degli organismi di valutazione della conformità e per il loro monitoraggio”*.

Invece, l'ACN è designata come autorità di vigilanza sul mercato, cioè *“l'autorità nazionale che svolge le attività e adotta le misure a norma del regolamento”*. Tale autorità *“è responsabile per la vigilanza, ivi incluse le attività ispettive e sanzionatorie, dei sistemi di intelligenza artificiale, secondo quanto previsto dalla normativa nazionale e dell'Unione europea”* (art. 20, comma 1, lett.b), della legge 23 settembre 2025, n. 132). Un'altra normativa rilevante ai fini dell'uso dell'IA nel processo penale, è la direttiva 2016/680/UE, nota come *Law Enforcement Directive* o direttiva LED, attuata dal legislatore italiano con il d.lgs. 18 maggio 2018, n. 51. Tale direttiva si applica al trattamento dei dati personali da parte delle autorità competenti per le finalità di *“prevenzione, indagine, accertamento, e perseguimento dei reati o esecuzione di sanzioni penali, incluse la salvaguardia e la prevenzione di minacce alla sicurezza pubblica”* (artt. 1, par. 1 e. 2, par. 1). Quando il trattamento di dati personali riguarda tali attività, non è invece applicabile il regolamento 2016/679/UE, noto come GDPR (lo prevede espressamente l'art. 2, par. 2, lett. d) dello stesso GDPR). Il GDPR verrà dunque trattato soltanto al fine di fare un confronto con la direttiva LED, in modo da mettere in risalto le specificità di quest'ultima¹⁴³.

L'art. 11 della direttiva LED impone agli Stati membri di vietare *“una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato”*. Tuttavia, viene fatta salva la possibilità per il legislatore nazionale o euro-unitario di prevedere delle eccezioni, a condizione che vi siano delle *“garanzie adeguate per i diritti e le libertà fondamentali dell'interessato”*. La sola garanzia che viene espressamente richiesta dalla direttiva è il *“diritto di ottenere l'intervento umano”*.

Tale articolo è stato attuato dall'art. 8, comma 1 del d.lgs. 18 maggio, n. 51, il quale, almeno in parte, ne replica il contenuto: *“sono vietate le decisioni basate unicamente su un trattamento automatizzato, compresa la profilazione, che producono effetti negativi nei confronti dell'interessato, salvo che siano autorizzate dal diritto dell'Unione europea o da specifiche disposizioni di legge”*¹⁴⁴.

¹⁴³ In particolare, nel par. 4.2.2, nell'analizzare il *“processo decisionale automatizzato relativo alle persone fisiche”* di cui all'art. 11 della direttiva LED, si farà spesso il confronto con l'analogo *“Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione”*, ai sensi dell'art. 22 del GDPR.

¹⁴⁴ Per un'analisi maggiormente approfondita delle differenze tra la disposizione in esame e l'art. 11 della direttiva LED, si veda il par. 4.2.2.

In conclusione, l'AI Act, la direttiva LED e la legge 23 settembre 2025, n. 132 potrebbero limitare o addirittura vietare l'uso dell'IA predittiva nella giustizia penale. La portata di tali limiti dipende, inevitabilmente, da come le relative norme verranno interpretate. Pertanto, nel quarto capitolo si analizzerà se e in che misura tali normative precludano l'adozione dell'IA predittiva nel processo penale italiano.

1.4.2) Il diritto penale italiano

Dopo aver illustrato il quadro euro-unitario, occorre illustrare le norme rilevanti all'interno dell'ordinamento penale italiano. Vi sono numerose disposizioni che, sebbene siano state ideate prima dell'avvento dell'intelligenza artificiale, potrebbero presentare nuove applicazioni alla luce di tale tecnologia.

Un esempio emblematico è l'art. 220, comma 2 del codice di procedura penale (c.p.p.) che vieta la perizia criminologica *“per stabilire l'abitudine o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche”*. Risulta quindi ammessa la perizia soltanto per provare la sussistenza di patologie psichiche. Ai sensi dell'art. 220, comma 2, vi sono però due ipotesi in cui la perizia criminologica è in ogni caso ammissibile: in sede di esecuzione della pena, a norma dell'art. 678 c.p.p., e in tema di misure di sicurezza, ex art. 679 c.p.p.

Una parte della dottrina sostiene che tale divieto dovrebbe comprendere l'uso di intelligenza artificiale al fine di valutare la personalità dell'imputato, in quanto anche tali tecnologie si basano su teorie criminologiche¹⁴⁵. Secondo un altro orientamento, non è affatto scontato che tali strumenti si qualificano come perizie criminologiche. Ne consegue che, anziché escluderli a priori, occorre piuttosto assicurarsi che tali sistemi di IA non precludano le garanzie processuali fondamentali¹⁴⁶. La questione presenta indubbiamente degli spunti di riflessione e verrà approfondita più diffusamente nel quarto capitolo, dedicato ai limiti legali dell'uso dell'IA predittiva nel processo penale.¹⁴⁷

Nella parte restante di questo paragrafo, si introdurranno i potenziali campi applicativi dell'IA predittiva nel processo penale, partendo dal presupposto che l'art. 220, comma 2, c.p.p. non pone problemi in tal senso.

In primo luogo, l'IA predittiva potrebbe essere impiegata per valutare la sussistenza delle esigenze cautelari ex art. 274 c.p.p., le quali costituiscono condizione necessaria ma non sufficiente per l'applicazione di una misura cautelare personale. Le esigenze cautelari comprendono il pericolo di inquinamento della prova, il pericolo di fuga ed il pericolo che vengano commessi determinati reati¹⁴⁸.

L'IA predittiva potrebbe essere utilizzata anche per fondare l'esigenza cautelare dell'allontanamento dalla casa familiare (art. 282-bis c.p.p.) e del divieto di avvicinamento ai luoghi frequentati dalla persona offesa (art.

¹⁴⁵ S. QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI ACT*, in *Diritto di Difesa*, 2023, p. 5.

¹⁴⁶ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, pp. 19-23.

¹⁴⁷ Si veda il par. 4.1.

¹⁴⁸ Si veda il par. 2.2.

282-ter c.p.p.). Potrebbero essere utili algoritmi predittivi del pericolo che vengano commessi reati violenti nei confronti del proprio partner, come *EPV-R*¹⁴⁹.

Si può anche ipotizzare l'uso dell'IA per supportare l'autorità giudiziaria e la polizia giudiziaria nel valutare la sussistenza dei presupposti di alcune misure precautelari. Un esempio è il fermo di indiziato di delitto, ai sensi dell'art. 384 c.p.p. Tale misura viene disposta dal pubblico ministero quando sussistono specifici elementi che fanno ritenere fondato il pericolo di fuga, a condizione che si proceda per determinati gravi delitti. In questo caso, l'IA predittiva potrebbe assistere il pubblico ministero nella valutazione del pericolo di fuga. Altri esempi sono l'arresto facoltativo *ex art.* 381 c.p.p. e l'allontanamento d'urgenza dalla casa familiare, ai sensi dell'art. 384-bis c.p.p.¹⁵⁰.

Un altro potenziale campo applicativo dell'IA predittiva sono le misure di sicurezza.

Il presupposto oggettivo, definito dall'art. 203, comma 1, c.p., è la pericolosità sociale, che sussiste “*quando è probabile che commetta nuovi fatti preveduti dalla legge come reati*”. Tale qualità deve essere desunta sulla base della gravità del reato e della capacità a delinquere, in quanto l'art. 203 c.p. rinvia ai criteri fattuali di commisurazione della pena di cui all'art. 133 c.p. L'IA predittiva potrebbe aiutare il giudice nel valutare la sussistenza della pericolosità sociale, con il possibile effetto positivo di ridurre la discrezionalità giudiziale¹⁵¹. L'IA potrebbe supportare i giudici anche ai fini delle misure di prevenzione *ante delictum*.

Tali misure sono oggi disciplinate dal d.lgs. 6 settembre 2011 n. 159, noto come codice antimafia. All'art. 1 si prevedono i soggetti destinatari, ossia: *a) coloro che debbano ritenersi, sulla base di elementi di fatto, abitualmente dediti a traffici delittuosi; b) coloro che per la condotta ed il tenore di vita debba ritenersi, sulla base di elementi di fatto, che vivono abitualmente, anche in parte, con i proventi di attività delittuose; c) coloro che per il loro comportamento debba ritenersi, sulla base di elementi di fatto, (comprese le reiterate violazioni del foglio di via obbligatorio di cui all'articolo 2, nonché dei divieti di frequentazione di determinati luoghi previsti dalla vigente normativa), che sono dediti alla commissione di reati che offendono o mettono in pericolo l'integrità fisica o morale dei minorenni, la sanità, la sicurezza o la tranquillità pubblica*. Anche l'accertamento sulla sussistenza dei presupposti di cui all'art. 1 del codice antimafia potrebbe essere supportato dall'IA¹⁵².

Tale tecnologia potrebbe coadiuvare il giudice ai fini della commisurazione della pena. L'art. 133 del codice penale (c.p.) disciplina i criteri fattuali di commisurazione della pena: gravità del reato e la capacità a delinquere. Ai sensi dell'art. 133, comma 2, c.p., la capacità a delinquere deve essere desunta “*dai motivi a delinquere e dal carattere del reo*”, “*dai precedenti penali e giudiziari e, in genere, dalla condotta e dalla vita del reo antecedenti al reato*”, “*dalla condotta contemporanea o susseguente al reato*”, “*dalle condizioni di*

¹⁴⁹ A. VALDIVIA -C. HYDE VAAMONDE -J. GARCÍA MARCOS, *Judging the Algorithm: a Case Study on the Risk Assessment Tool for Gender-Based Violence Implemented in the Basque Country*, in *AI & Society*, vol. 40 n. 4, 2025; si veda il par. 2.3.

¹⁵⁰ Si veda il par. 2.4.

¹⁵¹ Si veda il par. 2.5.

¹⁵² Si veda il par. 2.6.

vita, individuale, familiare e sociale del reo”. L’intelligenza artificiale dovrebbe idealmente essere basata su tali elementi, per poter fornire un valido supporto al giudice¹⁵³.

La probabilità della recidiva calcolata dall’IA potrebbe assumere rilievo anche ai fini della concessione della sospensione condizionale della pena. Il giudice è tenuto a valutare se *“il colpevole si asterrà dal commettere ulteriori reati”* (art. 164, comma 1, c.p.)¹⁵⁴.

L’IA potrebbe supportare il pubblico ministero (PM) nella scelta tra la richiesta di archiviazione e l’esercizio dell’azione penale. Ai sensi dell’art. 50, comma 1, c.p.p., il PM esercita l’azione penale quando non sussistono i presupposti per la richiesta di archiviazione. Tali presupposti vengono definiti dall’art. 408 c.p.p., il quale stabilisce che il PM presenta al giudice la richiesta di archiviazione quando *“gli elementi acquisiti nel corso delle indagini preliminari non consentono di formulare una ragionevole previsione di condanna”*. Dal combinato disposto di tali disposizioni, si può desumere che, quando gli elementi acquisiti nel corso delle indagini preliminari consentono di formulare una ragionevole previsione di condanna, il PM è tenuto ad esercitare l’azione penale¹⁵⁵. Il PM potrebbe tenere conto della previsione algoritmica quanto al probabile esito del processo per valutare se è possibile formulare una ragionevole previsione di condanna¹⁵⁶.

Infine, l’IA potrebbe sostenere l’accertamento giudiziale del nesso di causalità, ai sensi dell’art. 41 c.p., elaborando informazioni statistiche inerenti a fatti analoghi a quelli oggetto del caso di specie, ovviamente, in funzione meramente ausiliaria e senza sostituire il giudizio controfattuale demandato al giudice¹⁵⁷.

In conclusione, l’uso dell’IA predittiva nel processo penale italiano potrebbe rientrare, almeno in alcuni casi, nel divieto di cui all’art. 220, comma 2, c.p.p. Tale questione verrà analizzata analiticamente nel capitolo 4, dedicato ai limiti legali nell’adozione dell’IA predittiva nel processo penale italiano. Assumendo che tale limite non sussista, nel secondo capitolo si procederà ad analizzare in maniera maggiormente analitica i potenziali campi applicativi dell’IA predittiva nel processo penale che sono stati introdotti in questo paragrafo.

¹⁵³ Si veda il par. 2.7.

¹⁵⁴ Si veda il par. 2.7.4.

¹⁵⁵ P. TONINI -C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, pp. 670-671.

¹⁵⁶ Si veda il par.2.8.2.

¹⁵⁷ Si veda il par. 2.9.

CAPITOLO SECONDO

POTENZIALI CAMPI APPLICATIVI DELL'IA PREDITTIVA

2.1) Introduzione

Vi sono diverse ragioni per cui oggi il ricorso all'intelligenza artificiale potrebbe risultare particolarmente promettente. In diversi ambiti, si invoca il ricorso all'IA per contrastare alcuni abusi nell'esercizio della discrezionalità da parte dell'autorità giudiziaria. In sede di commisurazione della pena, il ricorso all'IA renderebbe la valutazione giudiziale maggiormente standardizzata, riducendo l'influenza di fattori soggettivi¹⁵⁸. Nell'esercizio dell'azione penale, una parte della dottrina ha segnalato alcuni abusi da parte dei pubblici ministeri. Qualora tale valutazione fosse supportata dall'IA, si conseguirebbe l'effetto di restringere tale discrezionalità, in quanto il giudizio diverrebbe meno incerto e, pertanto, anche meno soggetto a decisioni arbitrarie¹⁵⁹.

In secondo luogo, le valutazioni giudiziali potrebbero divenire maggiormente uniformi, in quanto giudici diversi dovrebbero comunque far riferimento al medesimo termine di paragone¹⁶⁰. Si pensi ad una situazione in cui, nell'ambito della commisurazione della pena, vi siano alcuni giudici che, statisticamente e a parità di condizioni, tendano ad irrogare una pena particolarmente severa, e altri che siano soliti disporre una pena particolarmente mite. Nell'ipotesi in cui tali giudici siano supportati dall'IA, tali differenze potrebbero attenuarsi. I giudici statisticamente più severi potrebbero essere spinti dal sistema di IA a rivalutare le loro posizioni in senso favorevole all'imputato. Viceversa, i giudici più clementi potrebbero essere spinti dall'IA a superare la loro tendenza a adottare pene particolarmente miti, in ragione dell'elevata capacità a delinquere rilevata dal modello predittivo.

Il ricorso all'IA sembra essere particolarmente promettente in quei settori in cui assume rilievo la pericolosità sociale (artt. 202 e 203 c.p.), o altri istituti facenti, comunque, riferimento al futuro compimento di reati, come la capacità a delinquere (art. 133, comma 2, c.p.) o il pericolo di reiterazione del reato (art. 274, lett. c), c.p.p.). Sussiste un forte scetticismo, sia da parte della dottrina che da parte del legislatore, nei confronti della capacità delle scienze criminologiche a prevedere il futuro compimento di reati. Tale sfiducia si rinviene nella scelta del legislatore di vietare la perizia criminologica al fine di accertare la personalità dell'imputato, ai sensi dell'art. 220, comma 2, c.p.p.¹⁶¹. L'IA potrebbe supplire ai limiti delle scienze criminologiche, fornendo un supporto al giudice il quale, altrimenti, rischia di effettuare giudizi sulla probabilità che vengano commessi nuovi reati senza una base statistica solida¹⁶².

¹⁵⁸ F. COPPOLA, *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo*, in *Archivio Penale*, 2023, p. 10.

¹⁵⁹ V. GRAMUGLIA, *Obbligatorietà dell'azione penale e ruolo dell'IA nelle scelte del pubblico ministero*, in *Archivio Penale*, Giuffrè Francis Lefebvre, Milano, 2024, p. 9.

¹⁶⁰ F. COPPOLA *op. cit.*, p. 15.

¹⁶¹ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, pp. 19-20.

¹⁶² M. CAIANIELLO *potenzialità e rischi derivanti dall'interazione tra I. A. e giustizia penale preventiva* in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 273.

L'IA predittiva potrebbe supportare l'autorità giudiziaria (e talvolta anche la polizia giudiziaria) nel formulare dei giudizi prognostici, ossia una valutazione su dei fatti futuri alla luce degli elementi presenti. Si tratta, ad esempio, del giudizio sulla capacità a delinquere *ex art. 133, comma 2, c.p.*, oppure del giudizio sul pericolo di reiterazione del reato ai fini delle misure cautelari, ai sensi dell'*art. 274, lett. c), c.p.p.* Tale tipologia di giudizi si prestano ad essere supportati dall'IA predittiva (oggetto specifico della presente disamina), intesa come una stima del verificarsi di un evento futuro al fine di orientare una decisione presa nel presente¹⁶³.

Non ogni possibile uso dell'intelligenza artificiale nel processo penale concerne specificamente l'IA predittiva. Ad esempio, l'intelligenza artificiale può essere impiegata nella valutazione delle prove, come nel caso del *software VERIPOL*, usato dalla polizia spagnola per valutare la veridicità di denunce di rapine, con un'accuratezza del 91%, superiore del 15% rispetto all'accuratezza della polizia giudiziaria sul medesimo *dataset*¹⁶⁴. L'IA può anche essere utile per ricostruire la scena del crimine, come nel caso di *HEMOSPAT*, un *software* utilizzato per l'analisi delle macchie di sangue (*Bloodstain Pattern Analysis, BPA*), al fine di determinare la postura di sospettati e vittime, sequenziare gli eventi e ricostruire la scena nel complesso¹⁶⁵.

Si tratta di ipotesi di ricorso all'intelligenza artificiale nel processo penale di indubbio interesse ma che non rientrano nella nozione di IA predittiva oggetto del presente lavoro. In sede di valutazione delle prove, non si tratta di stimare il verificarsi di un evento futuro bensì di valutare la sussistenza o meno di fatti avvenuti nel passato sulla base di elementi presenti. Per questa ragione, il ricorso all'IA al fine di supportare la valutazione delle prove non verrà specificamente analizzato in questo capitolo.

2.2) Applicazione dell'IA predittiva ai fini delle esigenze cautelari

2.2.1) Valutazione dell'esigenza cautelare dell'inquinamento probatorio

In primo luogo, occorre analizzare le potenzialità dell'IA predittiva come strumento di supporto nella valutazione giudiziale circa la sussistenza delle misure cautelari di cui all'*art. 274 c.p.p.*: il pericolo di inquinamento della prova, il pericolo di fuga e il pericolo di reiterazione del reato. La sussistenza di almeno una delle suddette esigenze cautelari costituisce condizione necessaria, seppur non sufficiente, per disporre una delle misure cautelari personali, disciplinate dal titolo I del libro IV del codice di procedura penale (ad esempio, la custodia cautelare in carcere *ex art. 285 c.p.p.*). Le altre condizioni sono la sussistenza di gravi indizi di colpevolezza e l'insussistenza di cause di giustificazione, di punibilità o di estinzione del reato (*art. 273 c.p.p.*). Le misure cautelari coercitive possono essere applicate soltanto quando si procede per delitti per i

¹⁶³ G. BUONOMO, *Macchine intelligenti, processo induttivo e apprendimento autonomo*, in G. BUONOMO. G. CIACCI. A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, p. 5.

¹⁶⁴ L. QUIJANO SANCHEZ, F. LIBERATORE, J. CAMACHO COLLADOS, *Applying Automatic Text-based Detection of Deceptive Language to Police Reports: Extracting Behavioral Patterns from a Multi-step Classification Model to Understand how We Lie to the Police*, in *Knowledge-Based Systems*, vol. 149, 2018, pp. 155–168.

¹⁶⁵ K. MALONEY, J. KILEEN, A. MALONEY, *The Use of HemoSpat to Include Bloodstains Located on Nonorthogonal Surfaces in Area-of-Origin Calculations*, in *Journal of Forensic Identification*, Vol. 59, 2009, https://www.researchgate.net/publication/285982869_The_Use_of_HemoSpat_To_Include_Bloodstains_Located_on_Nonorthogonal_Surfaces_in_Area-of-Origin_Calculations, consultato il 20 ottobre 2025, pp. 513-524.

quali la legge stabilisce la pena dell'ergastolo o della reclusione superiore nel massimo a tre anni (art. 280, comma 1, c.p.p.).

La prima delle esigenze cautelari previste dall'art. 274 c.p.p. è il pericolo di inquinamento probatorio, inteso come pericolo per l'acquisizione (rispetto ad un possibile occultamento) o la genuinità della prova (rispetto ad una possibile alterazione). In particolare, l'art. 274, lett. a), c.p.p. richiede la sussistenza di “*specifiche ed inderogabili esigenze attinenti alle indagini relative ai fatti per i quali si procede, in relazione a situazioni di concreto ed attuale pericolo per l'acquisizione o la genuinità della prova, fondate su circostanze di fatto espressamente indicate nel provvedimento a pena di nullità rilevabile anche di ufficio*”.

Qualora l'IA predittiva dovesse essere impiegata per prevedere il rischio di inquinamento probatorio, il requisito delle specifiche circostanze di fatto su cui si deve fondare tale esigenza cautelare rischia di essere compromesso, soprattutto nel caso in cui il sistema non dovesse essere trasparente. In tale circostanza, non sarebbe possibile sapere quali sono le circostanze di fatto su cui la previsione algoritmica di inquinamento probatorio si fonda. Inoltre, la mancanza di motivazione su tali circostanze di fatto, determinerebbe la nullità del provvedimento, ex art. 274, lett. a), c.p.p. Tale considerazione non preclude necessariamente il ricorso all'IA in tale ambito, ma richiede che la sola previsione algoritmica sul rischio di inquinamento probatorio non venga considerata di per sé sufficiente a fondare tale esigenza cautelare, dovendosi altresì individuare le specifiche circostanze di fatto alla base della misura.

A differenza del pericolo di fuga e del pericolo di reiterazione del reato, l'inquinamento probatorio non sembra essere attualmente oggetto di valutazione mediante *risk assessment tools*, neppure negli Stati Uniti. Il ricorso all'IA predittiva per valutare l'inquinamento probatorio presenta dunque ancora natura meramente speculativa, non essendo stato sperimentato in altri ordinamenti.

Nel sistema giudiziario statunitense, l'istituto più simile all'inquinamento probatorio è l'*obstruction*, che consiste sostanzialmente nel rischio serio che il soggetto possa ostacolare o tentare di ostacolare la giustizia, oppure minacciare, ferire o intimidire, o tentare di minacciare, ferire o intimidire, un potenziale testimone o giurato. In tal caso, può essere disposta la detenzione preventiva (*pretrial detention*), a norma del *Bail Reform Act* § 3142(f)(2)(b), anziché concedere il rilascio con condizioni (*conditions of release*), sempre che il rischio non possa essere mitigato da misure alternative.¹⁶⁶ Nonostante esista un istituto analogo all'inquinamento probatorio anche negli Stati Uniti (sia pur con le dovute differenze tra i rispettivi ordinamenti), i *risk assessment tools* attualmente in uso negli Stati Uniti nell'ambito della *pretrial detention* sono impiegati principalmente per valutare il rischio di non comparire in processo e il rischio di arresto durante le indagini¹⁶⁷. Qualora l'accertamento del pericolo di inquinamento probatorio fosse supportato dall'intelligenza artificiale, non sarebbe dunque possibile fare affidamento sull'esperienza maturata negli Stati Uniti.

¹⁶⁶ Legge federale statunitense *Bail Reform Act*, 1984, 18 U.S.C. § 3142(f)(2)(b); J. WOOD, *The Bail Reform Act of 1984*, 4th Edition, Federal Judicial Center, Washington (DC), 2022, pp. 19-35.

¹⁶⁷ M. T. STEVENSON -S. G. MAISON, *Pretrial Detention and Bail*, in E. LUNA, *Reforming Criminal Justice: A Report by the Academy for Justice*, Arizona State University, Phoenix, Arizona, 2017, p. 34.

2.2.2) Valutazione dell'esigenza cautelare del pericolo di fuga

Il pericolo di fuga, previsto dall'art. 274, lett. b, c.p.p., ricorre *“quando l'imputato si è dato alla fuga o sussiste concreto e attuale pericolo che egli si dia alla fuga, sempre che il giudice ritenga che possa essere irrogata una pena superiore a due anni di reclusione”*.

L'accertamento sulla sussistenza di un concreto e attuale pericolo che l'imputato si dia alla fuga, costituisce un giudizio prognostico che può essere supportato dall'IA predittiva. Il riferimento all'attualità del pericolo è stato aggiunto dalla legge 16 aprile 2015, n. 47, in quanto la giurisprudenza antecedente aveva ritenuto sussistente tale esigenza cautelare in ipotesi in cui il pericolo di fuga non era imminente¹⁶⁸. Qualora tale pericolo dovesse essere valutato dall'IA predittiva, sarebbe fondamentale che il sistema sia addestrato con riferimento a circostanze recenti e non si esaurisca in una profilazione della storia criminale del soggetto basata su fatti particolarmente risalenti nel tempo che non permetterebbe di accertare l'attualità del pericolo. In alternativa, il sistema di IA potrebbe supportare la valutazione sulla potenziale attitudine del soggetto a darsi alla fuga, mentre la valutazione sull'attualità di tale pericolo potrebbe rimanere una prerogativa esclusiva del giudice, fondata su circostanze recenti. Ad esempio, l'IA potrebbe stimare (in astratto) un elevato pericolo di fuga sulla base della storia criminale dell'imputato, mentre il giudice potrebbe accertare l'attualità di tale pericolo in ragione della circostanza che l'imputato ha acquistato un biglietto aereo per un paese estero.

In giurisprudenza, il requisito della concretezza del pericolo è stato interpretato nel senso che occorre indicare specificamente i fatti da cui tale esigenza cautelare viene desunta¹⁶⁹. Il giudice non quindi potrebbe limitarsi ad argomentare sulla base del *risk score* prodotto dall'IA, ma dovrebbe anche indicare i fatti da cui desumere la sussistenza di tale pericolo.

Per comprendere quali fattori potrebbero essere rilevanti nell'addestrare un modello funzionale a prevedere il pericolo di fuga, risulta opportuno riepilogare quali sono, secondo la giurisprudenza, gli elementi rilevanti sulla cui base si innesta tale esigenza cautelare. Primo elemento rilevante in tal senso è la latitanza che costituisce un ostacolo all'esecuzione di provvedimenti incidenti sulla libertà personale ed è indice di una capacità organizzativa idonea a eludere l'attività della polizia giudiziaria¹⁷⁰. Il preesistente stato di detenzione permette di escludere tale pericolo soltanto nel caso in cui è possibile stabilire in maniera certa che non verranno adottate misure alternative alla detenzione¹⁷¹. Non è invece possibile fondare il pericolo di fuga solamente sulla circostanza che un imputato straniero è stato trovato con documenti falsi¹⁷².

Inoltre, l'art. 274 c.p.p. prevede espressamente che *“le situazioni di concreto e attuale pericolo non possono essere desunte esclusivamente dalla gravità del titolo di reato per cui si procede”*. Le Sezioni Unite hanno ulteriormente chiarito che non è neppure possibile basarsi soltanto sulla particolare gravità della pena inflitta

¹⁶⁸ G. CONSO -V. GREVI -M. BARGIS *Compendio di procedura penale*, 11° edizione, CEDAM, Milano, 2023, pp. 368-370.

¹⁶⁹ A. GAITO – A. BARGI *et al.*, *Codice di procedura penale commentato*, UTET Giuridica, Torino, 2023, disponibile su OneLegale, consultato il 25 settembre 2025, *sub* art. 274.

¹⁷⁰ Cass. pen., Sez. II, 23 marzo 1998, n. 2012; A. GAITO, *op cit.*

¹⁷¹ Cass. pen., Sez. IV, 17 novembre 2010, n. 45408; A. GAITO, *op cit.*

¹⁷² Cass. pen., Sez. III, 28 giugno 1995, n. 3420; A. GAITO, *op cit.*

in primo o in secondo grado¹⁷³. Nel caso dell'IA predittiva, ciò non implica che la gravità del reato non possa essere uno dei fattori attraverso cui addestrare il modello predittivo ma che questo elemento non può avere un peso tale da determinare, da solo, un livello di rischio particolarmente significativo. Altrimenti, si rischierebbe di eludere la norma di cui sopra perché il giudice fonderebbe una misura cautelare su di un sistema di IA che, nella sostanza, si basa esclusivamente sulla gravità del reato.

Negli Stati Uniti, l'intelligenza artificiale è già stata impiegata per valutare la sussistenza di un rischio analogo (seppur non coincidente) al pericolo di fuga: il *risk of non-appearance* o rischio di mancata comparizione in tribunale. Ai sensi del § 3142 del *Bail Reform Act*, il giudice è tenuto a ordinare la detenzione prima del processo se ritiene che, altrimenti, non sia possibile garantire la comparizione della persona in tribunale¹⁷⁴. Tuttavia, il *risk of non-appearance* statunitense non coincide perfettamente con il pericolo di fuga italiano, in ragione delle differenze strutturali dei due ordinamenti processuali, rispettivamente di *common law* e *civil law*. Lo stesso *Bail Reform Act* del 1984, al § 3146, considera reato la non comparsa consapevole dinanzi ad un tribunale come richiesto dalle condizioni di rilascio¹⁷⁵. Nell'ordinamento statunitense, la mancata presenza dell'imputato è dunque un fatto illecito. Coerentemente, la detenzione processuale per *risk of non-appearance* non mira soltanto a prevenire il pericolo di fuga al fine di sottrarsi alla detenzione in caso di condanna, ma è altresì funzionale ad evitare che l'imputato semplicemente non si presenti in udienza.

Nell'ordinamento italiano, l'assenza dell'imputato in giudizio non costituisce reato. Se si verifica tale evenienza, in sostanza, il giudice è tenuto a verificare la presenza del suo difensore (art. 420, comma 1, c.p.p.) e che l'assenza dell'imputato sia consapevole (art. 420 *bis*, comma 2, c.p.p.) Infatti, l'esigenza cautelare del pericolo di fuga presenta un ambito di applicazione ristretto rispetto al *risk of non-appearance*, in quanto non include il pericolo che l'imputato non si presenti in giudizio, essendo tale comportamento pienamente lecito. Questa differenza tra i due ordinamenti implica che i *risk assessment tools* impiegati negli Stati Uniti per valutare il *risk of non-appearance* non possono essere direttamente utilizzati nell'ordinamento italiano. Si rischierebbe di attribuire rilevanza al mero rischio di mancata comparizione in tribunale, che, se non riconducibile al pericolo di fuga, esula dall'ambito di applicazione dell'art. 274 c.p.p. L'esperienza statunitense rimane comunque un punto di riferimento che non dovrebbe essere ignorato qualora si volesse introdurre algoritmi analoghi anche in Italia.

Ancora, sul versante statunitense, spunti di riflessione sono offerti dal già citato strumento di valutazione del rischio *COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)*. Si può fare specifico riferimento al *Pretrial Release Risk*, una scala di rischio funzionale a stimare, con unico *risk score*,

¹⁷³ Cass. pen., Sez. Un., 11 luglio 2001, n. 34537; A. GAITO, *op cit.*

¹⁷⁴ legge federale statunitense *Bail Reform Act*, 1984, 18 U.S.C. § 3142(e)(1); J. WOOD *The Bail Reform Act of 1984 4th Edition*, Federal Judicial Center, Washington (DC), 2022, pp. 18-35.

¹⁷⁵ legge federale statunitense *Bail Reform Act*, 1984, 18 U.S.C § 3146 (a) (1); J. WOOD *op. cit.*, pp. 68-69.

il generico rischio di rilascio antecedente al processo, sulla base del rischio di mancata comparizione in tribunale, oltre che del rischio di arresto per un reato grave (*felony*) prima del processo¹⁷⁶.

Questo strumento è stato elaborato sulla base di fattori predittivi del fallimento preprocessuale, alla luce della ricerca in materia. Uno studio a cura di Marie Van Nostrand, una criminologa e statistica statunitense, ha individuato i seguenti fattori predittivi di tale rischio: imputazione corrente, imputazioni pendenti, arresti precedenti, precedenti fallimenti preprocessuali, stabilità residenziale, stato occupazionale, legami con la comunità e abuso di sostanze¹⁷⁷. Questi elementi sono poi stati ripresi dalla *Northpointe* nello sviluppare *COMPAS*. È interessante notare come non necessariamente questi fattori si pongono quali esclusivamente manifestazioni del rischio di mancata comparizione in tribunale. Si tratta piuttosto di elementi indicativi del rischio di fallimento preprocessuale in genere, che quindi comprende anche il rischio di arresto per un reato grave.

Un altro *risk assessment tool* impiegato per prevedere il rischio di mancata comparizione in tribunale è *PSA* (*Public Safety Assessment*). Una delle tipologie di *risk score* che questo strumento è in grado di produrre è denominata *failure to appear (FTA)* (mancata comparizione in tribunale) e consiste in un punteggio da 0 a 6 che cresce all'aumentare del rischio¹⁷⁸. *PSA* si basa sulla sussistenza o meno di un'altra imputazione pendente al momento dell'arresto, su eventuali precedenti condanne nonché su eventuali antecedenti mancate comparizioni in tribunale. In quest'ultima ipotesi, rileva se le mancate comparizioni si sono verificate negli ultimi due anni, nonché il loro numero. Tutti questi fattori riguardano specificamente il rischio di mancata comparizione in tribunale e non anche il rischio di arresto antecedente al processo, in quanto quest'ultimo pericolo è oggetto di un punteggio distinto, denominato *New Criminal Arrest (NCA)*¹⁷⁹.

In conclusione, l'IA potrebbe supportare la valutazione sul pericolo di fuga ma occorre assicurarsi che tale accertamento avvenga sulla base di circostanze recenti, che non si basi esclusivamente sulla gravità del reato e che il giudice indichi specificamente i fatti in base ai quali la circostanza viene desunta. L'esperienza statunitense mostra come non si tratti di un'ipotesi meramente speculativa, a differenza di quanto precedentemente osservato riguardo all'inquinamento probatorio. Occorre però segnalare che un'eventuale introduzione dell'IA in questo ambito dovrebbe tenere conto della differenza tra il pericolo di fuga *ex art. 274 c.p.p.* e il *risk of non appearance* statunitense.

2.2.3) Valutazione dell'esigenza cautelare del pericolo di reiterazione del reato

¹⁷⁶ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025, p. 27.

¹⁷⁷ M. VAN NOSTRAND, *Assessing Risk among Pretrial Defendants in Virginia: The Virginia Pretrial Risk Assessment Instrument*, Virginia Department of Criminal Services, Richmond, Virginia, 2007, <https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/assessing-risk-among-pretrial-defendants-virginia-virginia-pretrial-risk-assessment-instrument.pdf>, consultato il 25 settembre 2025.

¹⁷⁸ Advancing Pretrial Policy & Research *Public Safety Assessment – Risk factors*, in AdvancingPretrial.org, <https://advancingpretrial.org/psa/factors/>, consultato il 14 luglio 2025.

¹⁷⁹ *Ibid.*

La terza esigenza cautelare prevista dall'art. 274 c.p.p. è il pericolo di reiterazione del reato.

Con la già citata legge 16 aprile 2015 n. 47, il legislatore ha cercato di specificarne le condizioni in maniera maggiormente analitica, con l'obiettivo di evitare richiami eccessivamente generici alla pericolosità sociale dell'imputato. Il legislatore intende in questo modo prevenire censure da parte della Corte costituzionale sotto il profilo della determinatezza¹⁸⁰.

Tale circostanza sussiste “*quando, per specifiche modalità e circostanze del fatto e per la personalità della persona sottoposta alle indagini o dell'imputato, desunta da comportamenti o atti concreti o da suoi precedenti penali, sussiste il concreto e attuale pericolo che questi commetta*” determinati reati. Nello specifico, si limita l'esigenza cautelare a categorie di reati, come gravi delitti di criminalità organizzata, nonché sulla base della pena prevista. In particolare, “*se il pericolo riguarda la commissione di delitti della stessa specie di quello per cui si procede, le misure di custodia cautelare sono disposte soltanto se trattasi di delitti per i quali è prevista la pena della reclusione non inferiore nel massimo a quattro anni*”.

Tale disposizione è stata criticata in dottrina in quanto non indicherebbe in maniera sufficientemente stringente i futuri reati che si ritiene l'imputato possa commettere. La norma si limita piuttosto ad individuare categorie generali di reato, lasciando, quindi, ampia discrezionalità giudiziale nel delineare il rischio specificamente attribuito all'imputato. Autorevole dottrina ha quindi sostenuto che il possibile ricorso all'IA predittiva potrebbe costituire l'occasione per riformare l'art. 274 c.p.p., al fine di dare a tale accertamento maggiore prevedibilità¹⁸¹.

Inoltre, la Cassazione ha precisato che il giudizio sul pericolo di reiterazione del reato deve congiuntamente basarsi sia sulle specifiche modalità del fatto che sulla personalità della persona sottoposta alle indagini o dell'imputato, censurando sentenze di merito in cui il giudice *a quo* aveva basato il proprio giudizio esclusivamente sulle modalità del fatto o sulla personalità¹⁸². Anche un modello predittivo del pericolo di reiterazione del reato dovrebbe idealmente essere basato non soltanto su elementi inerenti alla personalità, ma anche su fattori riconducibili alle specifiche modalità del fatto. Altrimenti, dovrebbe essere il giudice a motivare dettagliatamente l'idoneità delle specifiche modalità del fatto a far emergere il pericolo di reiterazione del reato. In definitiva, l'IA non può essere utilizzata quale mezzo per eludere la disciplina legislativa, la quale, basando tale giudizio sulle specifiche modalità del fatto, mira ad evitare che vengano disposte misure cautelari sulla base di un generico pericolo di reiterazione del reato desunto esclusivamente dalla personalità.

Il requisito della concretezza del pericolo implica che la possibilità che il reato venga reiterato deve essere desunta da elementi non meramente congetturali¹⁸³. Un sistema di IA potrebbe limitarsi a prevedere un

¹⁸⁰ G. CONSO -V. GREVI -M. BARGIS *Compendio di procedura penale*, 11° edizione, CEDAM, Milano, 2023, pp. 368-369.

¹⁸¹ M. CAIANIELLO *potenzialità e rischi derivanti dall'interazione tra I. A. e giustizia penale preventiva* in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 277-279.

¹⁸² Cass. pen., Sez. IV, 1° aprile 2004, n. 37566; A. GAITO – A. BARGI *et al.*, *Codice di procedura penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub art.* 274.

¹⁸³ Cass. pen., Sez. feriale, 8 settembre 1994, n. 3683; A. GAITO *op. cit.*, *sub art.* 274.

significativo pericolo di reiterazione del reato, senza che sia possibile conoscere gli elementi in base ai quali tale pericolo è stato desunto, a causa del problema della *black box*. In tale ipotesi, dovrebbe essere il giudice a verificare che vi siano effettivamente degli elementi concreti, per evitare di fondare una misura cautelare su di un pericolo meramente astratto, eludendo il requisito della concretezza.

Occorre poi analizzare i fattori in base ai quali un sistema di IA potrebbe legittimamente desumere il pericolo di reiterazione del reato. Al riguardo, la giurisprudenza è unanime nel riconoscere che si può fare riferimento ai criteri di cui all'art. 133 c.p.¹⁸⁴. È stata infatti attribuita rilevanza sia a dati sintomatici di natura oggettiva¹⁸⁵ (inerenti alla gravità del reato) che di natura soggettiva (relativi alla capacità a delinquere)¹⁸⁶. Pertanto, un sistema di IA potrebbe desumere il pericolo di reiterazione del reato sulla base degli indici della capacità a delinquere, ai sensi dell'art. 133, comma 2, c.p. Si tratta dei motivi a delinquere, dei precedenti penali e giudiziari e, in genere, della condotta e della vita del reo antecedente al reato, della condotta contemporanea o susseguente al reato e delle condizioni di vita individuale, familiare e sociale del reo.

Negli Stati Uniti, l'IA è stata utilizzata per valutare la sussistenza del pericolo per la comunità (*danger to the community*) che è l'istituto più simile al pericolo di reiterazione del reato di cui all'art. 274 c.p.p., lett. c). Il giudice è tenuto ad ordinare la detenzione prima del processo se nessuna condizione di rilascio, o nessuna combinazione di condizioni di rilascio sarebbero idonee a garantire la sicurezza di qualunque altro membro della comunità (§ 3142 (e) (1) del *Bail Reform Act*)¹⁸⁷.

Come già accennato, il *Pretrial Risk Scale* di *COMPAS* è stata sviluppato in modo tale da prevedere il rischio di fallimento del rilascio preprocessuale sulla base del pericolo di mancata comparizione in tribunale e il rischio di arresto per reato grave (*felony*). Ai fini della valutazione in merito al pericolo per la comunità, assume chiaramente rilievo la valutazione del pericolo di arresto per reato grave, anziché il rischio di mancata comparizione in tribunale¹⁸⁸. Il *Pretrial Risk Scale* genera un unico punteggio predittivo riferito genericamente al rischio di fallimento del rilascio preprocessuale.

Il *PSA*, invece, genera tre *risk score* distinti: il *Failure to Appear (FTA)*, il *New Criminal Arrest (NCA)* e il *New Violent Criminal Arrest (NVCA)*. Mentre il *Failure to Appear* rileva ai fini del rischio di mancata comparizione in tribunale, il *New Criminal Arrest* e il *New Violent Criminal Arrest* rilevano, evidentemente, ai fini della valutazione in merito al pericolo per la comunità. Il *NCA* si basa sui seguenti fattori: età al momento dell'arresto, imputazione pendente al momento dell'arresto, condanne per reato poco grave (*misdemeanor*), condanne per reato grave (*felony*), condanne per reati violenti, precedente mancata comparizione in tribunale,

¹⁸⁴ A. GAITO *op. cit. sub art.* 274.

¹⁸⁵ I dati sintomatici di natura oggettiva si riferiscono alla gravità del reato, ai sensi dell'art. 133, comma 1, c.p.

¹⁸⁶ Cass. pen., Sez. III, 29 maggio 1996, n. 2439; A. GAITO *op. cit. sub art.* 274.

¹⁸⁷ legge federale statunitense *Bail Reform Act*, 1984, 18 U.S.C. § 3142(e)(1), 1984; J. WOOD *The bail reform Act of 1984 4th Edition*, Federal Judicial Center, Washington (DC), 2022, pp. 18-35.

¹⁸⁸ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025, p. 27.

precedente condanna alla detenzione. Per quanto riguarda l'età al momento dell'arresto, vengono aggiunti due punti se, quando la persona è stata arrestata, aveva meno di 23 anni¹⁸⁹.

Il *NVCA*, per valutare il pericolo di arresto per reato violento, si basa sui seguenti fattori: imputazione pendente al momento dell'arresto, precedente condanna, precedente condanna per reato violento. In quest'ultimo caso, rileva anche se le condanne sono superiori o inferiori a tre. Rileva anche la natura violenta del reato contestato, ed è previsto un punteggio maggiormente severo se il reato è stato commesso quando il soggetto aveva meno di 20 anni¹⁹⁰. Anche in questo caso, la giovane età rileva dunque come elemento indice di maggiore pericolosità sociale.

In conclusione, l'intelligenza artificiale potrebbe supportare la valutazione giudiziale sul pericolo di reiterazione del reato, purché tale pratica non si traduca nell'elusione del requisito della concretezza del pericolo e tale giudizio tenga conto anche delle specifiche modalità e circostanze del fatto, e non soltanto della personalità dell'imputato. Come per il pericolo di fuga, alla luce dell'esperienza statunitense, il ricorso all'IA in questo ambito non risulta meramente congetturale.

2.3) L'allontanamento dalla casa familiare e il divieto di avvicinamento alla persona offesa

Oltre ad essere utile nell'ambito delle esigenze cautelari, l'IA predittiva potrebbe essere impiegata per fondare due specifiche misure cautelari: l'allontanamento dalla casa familiare (art. 282-*bis* c.p.p.) e il divieto di avvicinamento ai luoghi frequentati dalla persona offesa (art. 282-*ter* c.p.p.). Tali misure vengono spesso predisposte con particolare riferimento a reati di violenza nelle relazioni familiari o nei confronti del proprio partner, sebbene non vi sia una norma che ne limiti l'applicazione a tali ipotesi.¹⁹¹ Nel valutare se adottare tali misure, il giudice potrebbe avvalersi di strumenti di valutazione del rischio funzionali a calcolare la probabilità che un individuo commetta violenza grave nei confronti del proprio partner.

Il ricorso all'IA in questo ambito si collocherebbe nel quadro delle norme penali introdotte al fine di contrastare in maniera efficace la violenza di genere, come il d.lgs. 14 agosto 2013 n. 93, recante "*Disposizioni urgenti in materia di sicurezza e per il contrasto della violenza di genere, nonché in tema di protezione civile e di commissariamento delle province*". Sul piano internazionale, assume poi rilievo la "*Convenzione del Consiglio d'Europa sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica*", adottata a Istanbul l'11 maggio 2011 (CETS n. 2010), il primo trattato internazionale vincolante contro la violenza sulle donne ratificato con la legge 27 giugno 2013 n. 77.

Qualora le previsioni dei sistemi IA si rivelassero sufficientemente accurate, tale tecnologia potrebbe rivelarsi decisiva nel prevenire la violenza di genere adottando le misure dell'allontanamento dalla casa familiare e del divieto di avvicinamento ai luoghi frequentati dalla persona offesa.

¹⁸⁹ Advancing Pretrial Policy & Research, *Public Safety Assessment – Risk factors*, in AdvancingPretrial.org, <https://advancingpretrial.org/psa/factors/>, consultato il 14 luglio 2025.

¹⁹⁰ *Ibid.*

¹⁹¹ P. TONINI -C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, p.475.

Per comprendere il ruolo dell'IA in questo ambito, occorre analizzare il quadro normativo in cui questi strumenti si inserirebbero. Per quanto riguarda l'allontanamento dalla casa familiare, si prevede che *“il giudice prescrive all'imputato di lasciare immediatamente la casa familiare, ovvero di non farvi rientro, e di non accedervi senza l'autorizzazione del giudice che procede. L'eventuale autorizzazione può prescrivere determinate modalità di visita”*. (art. 282-bis, comma 1, c.p.p.). Ad esempio, l'autorizzazione può limitare l'accesso a determinati orari, nonché prescriverne la durata¹⁹².

Si prevede poi che *“il giudice, qualora sussistano esigenze di tutela dell'incolumità della persona offesa o dei suoi prossimi congiunti, può inoltre prescrivere all'imputato di non avvicinarsi a luoghi determinati abitualmente frequentati dalla persona offesa”*, come, ad esempio, il luogo di lavoro, oppure il domicilio della famiglia di origine (art. 282-bis, comma 2, c.p.p.). In ogni caso, il legislatore non ha fornito un elenco tassativo di luoghi¹⁹³. Viene comunque fatto salvo, per espressa previsione legislativa, il caso in cui la frequentazione dei luoghi da parte dell'imputato sia necessaria per motivi di lavoro.

Inoltre, su richiesta del Pubblico Ministero, il giudice *“può ingiungere il pagamento periodico di un assegno alle persone che, per effetto della misura cautelare disposta, rimangano prive di mezzi adeguati”* (art. 282-bis, comma 3, c.p.p.).

L'art. 282-bis, comma 6, c.p.p., prevede poi delle norme specifiche applicabili per determinati delitti di violenza alla persona in ambito familiare¹⁹⁴. Vi rientrano la violazione di obblighi di assistenza familiare (art. 570 c.p.), i maltrattamenti contro familiari o conviventi (art. 572 c.p.) e la violenza sessuale (art. 609-bis c.p.). Questo regime si applicherebbe anche nel caso in cui un giudice usi un modello predittivo per prevedere il rischio di violenza di genere, come nell'ipotesi considerata in questo paragrafo.

In questi casi, è consentito al giudice di applicare la misura del braccialetto elettronico, definito dall'art. 275-bis c.p.p. come *“una procedura di controllo mediante mezzi elettronici o altri strumenti tecnici”*. In concreto, si tratta di un rilevatore di presenza che invia periodicamente un segnale ad un computer della sala operativa degli organi di polizia. Se il soggetto non si trova dove dovrebbe essere, scatta l'allarme¹⁹⁵. Qualora l'imputato non presti il consenso al braccialetto elettronico, il giudice dispone l'applicazione della misura cautelare in carcere (art. 275, comma 1, c.p.p.).

Per quanto concerne il divieto di avvicinamento, ai sensi dell'art. 282-ter, primo comma, c.p.p., *“il giudice prescrive all'imputato di non frequentare determinati luoghi abitualmente frequentati dalla persona offesa, ovvero di mantenere una determinata distanza, comunque non inferiore a 500 metri, da tali luoghi o dalla persona offesa”*. Anche in questo caso, può essere disposto il braccialetto elettronico. Qualora dovessero esserci ulteriori esigenze di tutela, il giudice può altresì prescrivere il divieto di avvicinamento ai luoghi frequentati da persone legate alla persona offesa da relazione affettiva, nonché ai luoghi frequentati dai suoi

¹⁹² A. GAITO – A. BARGI *et al.*, *Codice di procedura penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub* art. 282-bis.

¹⁹³ *Ibid.*

¹⁹⁴ P. TONINI -C. CONTI, *op. cit.* pp. 475-476.

¹⁹⁵ *Ibid.*, p. 478.

prossimi congiunti o conviventi (art. 282-ter, comma 2, c.p.p.). Il giudice può anche vietare all'imputato di comunicare, con qualsiasi mezzo, con la persona offesa, con chi intrattiene con essa relazioni affettive, con i suoi prossimi congiunti e con i conviventi (art. 282-ter, comma 3, c.p.p.).

Sia nel caso dell'art. 282-bis c.p.p., che nell'ipotesi di cui all'art. 282-ter c.p.p., il giudice è tenuto a riempire ulteriormente di contenuto le misure cautelari con integrazioni prescrittive, al fine di raggiungere l'obiettivo cautelare. Tali misure dovranno quindi essere modellate in relazione alle specifiche dinamiche dell'illecito¹⁹⁶. Ad esempio, le Sezioni Unite hanno chiarito che il giudice è tenuto ad indicare specificamente i determinati luoghi frequentati dalla persona offesa a cui l'imputato non si può avvicinare¹⁹⁷.

Nel caso in cui l'IA dovesse prevedere che verranno commessi reati violenti in ambito familiare, non sarebbe sufficiente motivare sulla base di tale previsione algoritmica. Piuttosto, il giudice dovrà comprendere le specifiche dinamiche dell'illecito sulla base delle quali disporre le integrazioni prescrittive necessarie. Ad esempio, un'elevata probabilità di commettere atti di violenza nei confronti del proprio partner, non fornisce indicazioni circa i luoghi in cui tale violenza potrebbe verificarsi. Spetterebbe comunque al solo giudice l'accertamento in merito ai luoghi abitualmente frequentati dalla persona offesa limitatamente ai quali viene disposto il divieto di avvicinamento.

Un'altra integrazione prescrittiva che il giudice può disporre riguarda l'art. 282-bis, comma 1, c.p.p., a norma del quale il giudice può autorizzare l'imputato ad accedere alla casa familiare, prescrivendo determinate modalità di visita. L'individuazione delle precise modalità di visita non potrebbe essere effettuata dal modello predittivo, che si limiterebbe ad individuare un determinato livello di rischio, bensì dovrebbe essere oggetto di un'autonoma valutazione giudiziale sulla base della specifica situazione di specie.

Per comprendere meglio le potenzialità dell'IA in questo ambito, è necessario fare riferimento all'esperienza della Spagna, con particolare riferimento alla Comunità autonoma dei Paesi Baschi. In Spagna, il processo penale è regolato dalla *Ley de Enjuiciamiento Criminal (LECrim)*. L'art. 544-ter, primo comma, disciplina l'ordine di protezione per le vittime di violenza domestica. Tale misura può essere disposta soltanto se vi sono gravi indizi sulla commissione di determinati reati, come, ad esempio, delitti contro la libertà sessuale o contro l'incolumità della persona. Si richiede poi che da tale reato emerga una situazione oggettiva di rischio per la vittima¹⁹⁸.

L'ordine di protezione, nella maggior parte dei casi, comporta l'adozione di misure analoghe a quelle di cui agli artt. 282-bis e 282-ter c.p.p. Può essere infatti vietato all'indagato o accusato di comunicare con la vittima

¹⁹⁶ A. GAITO, *op. cit.*, sub art. 282-ter.

¹⁹⁷ Cass. pen., Sez. Un., 29 aprile 2021, n. 39005.

¹⁹⁸ *Ley de Enjuiciamiento Criminal (LECrim)*, approvata con *Real Decreto* di 14 settembre 1882, Spagna, art. 544-ter, comma 1, di cui si riporta di seguito l'originale: “*El Juez de Instrucción dictará orden de protección para las víctimas de violencia doméstica en los casos en que, existiendo indicios fundados de la comisión de un delito o falta contra la vida, integridad física o moral, libertad sexual, libertad o seguridad de alguna de las personas mencionadas en el artículo 173.2 del Código Penal, resulte una situación objetiva de riesgo para la víctima que requiera la adopción de alguna de las medidas de protección reguladas en este artículo*”.

o ai suoi parenti o altre persone determinate dal giudice¹⁹⁹. Tale prescrizione è analoga al divieto di comunicare con qualsiasi mezzo con la vittima (oppure, con i suoi congiunti) di cui all'art. 282-ter, comma 3, c.p.p. nell'ordinamento italiano.

Inoltre, in Spagna, il giudice può vietare all'imputato di avvicinarsi alla vittima, ai suoi parenti o altre persone indicate dal giudice²⁰⁰. In questo caso, si può fare un parallelismo con la misura cautelare italiana del divieto di avvicinamento ai luoghi frequentati dalla persona offesa *ex art. 282-ter c.p.p.*

Infine, il giudice può vietare all'imputato di tornare nel luogo in cui il reato è stato commesso, oppure nel luogo in cui la vittima o la sua famiglia risiedono, se divergono²⁰¹. La seconda ipotesi è simile, seppur non coincidente, alla misura cautelare italiana dell'allontanamento dalla casa familiare, *ex art. 282-bis c.p.p.*

In questo contesto normativo, i giudici dei Paesi Baschi usano *EPV-R*, uno strumento di valutazione del rischio predittivo del rischio di recidiva in materia di violenza di genere.

In seguito alla denuncia della vittima, la polizia scrive una relazione in merito al rischio di violenza utilizzando *EPV-R*. Tale relazione viene trasmessa al giudice che decide se prendere provvedimenti e, in caso, quale misura cautelare adottare, tenendo conto della previsione algoritmica.

Questo strumento si basa su venti voci che possono essere classificate in cinque categorie. La prima categoria riguarda i dati personali. Ad esempio, si considera indice di maggiore pericolosità se l'indagato (o imputato) oppure la vittima sono immigrati²⁰². Questo elemento si espone a critiche, in quanto, di fatto, a parità di condizioni, un immigrato viene considerato maggiormente pericoloso. La seconda categoria riguarda l'eventuale stato di relazione di coppia, assumendo rilievo la circostanza che il soggetto si è recentemente separato. La terza categoria concerne il tipo di violenza, in quanto vengono considerate eventuali aggressioni sessuali nella relazione di coppia o l'aver commesso violenza fisica in grado di causare lesioni. Vi sono poi altre voci relative al profilo psicologico del soggetto, con specifico riferimento all'abuso di alcool o di droghe o al comportamento fortemente geloso o controllante nei confronti del partner. Rileva anche la vulnerabilità della vittima accertata sulla base della percezione di essere stata in pericolo di vita nell'ultimo mese²⁰³.

L'utilizzo di *EPV-R* nel sistema giudiziario è stato oggetto di numerose critiche, in quanto i giudici non sarebbero stati sufficientemente formati, con la conseguenza che si sarebbero affidati in maniera eccessiva a tale strumento²⁰⁴.

In conclusione, l'IA potrebbe essere impiegata per rendere il contrasto alla violenza di genere maggiormente efficace valutando l'applicabilità o meno della misura dell'allontanamento dalla casa familiare (art. 282-bis c.p.p.) o del divieto di avvicinamento ai luoghi frequentati dalla persona offesa (art. 282-ter c.p.p.). Il caso di

¹⁹⁹ A. VALDIVIA -C. HYDE VAAMONDE -J. GARCÍA MARCOS, *Judging the Algorithm: a Case Study on the Risk Assessment Tool for Gender-Based Violence Implemented in the Basque Country*, in *AI & Society*, vol. 40 n. 4, 2025, <https://arxiv.org/abs/2203.03723>, consultato il 25 settembre 2025, p. 7; R. IGUALADA, *Domestic Violence and Family Crises: A perspective from Spanish Law in Familia. Il diritto della famiglia e delle successioni in Europa*, 2018, pp. 389-391.

²⁰⁰ A. VALDIVIA -C. HYDE VAAMONDE -J. GARCÍA MARCOS, *op. cit.*, p. 7.

²⁰¹ *Ibid.*

²⁰² *Ibid.*, pp. 16-18.

²⁰³ *Ibid.*

²⁰⁴ *Ibid.*, pp. 11-13.

EPV-R, nonostante le menzionate criticità, fa emergere che non si tratta di una mera speculazione, bensì di una concreta prospettiva di sviluppo.

2.4) Applicazione dell'IA predittiva ai fini delle misure precautelari

2.4.1) Introduzione

Un altro settore in cui l'intelligenza artificiale potrebbe avere un ruolo sono le misure precautelari, disciplinate dagli artt. 379 e ss. c.p.p. Queste misure sono caratterizzate da una stretta connessione funzionale con le indagini, il che spiega la collocazione codicistica all'interno del V libro, dedicato alle indagini preliminari. Il legislatore attribuisce infatti il potere di adottare le misure precautelari agli organi investigativi (p.g. o PM), anziché al giudice. Tali misure si caratterizzano per la loro provvisorietà, essendo previste delle rigide scadenze temporali entro cui devono essere convalidate, e per la loro strumentalità rispetto all'applicazione di misure cautelari coercitive, nel senso che sono funzionali all'adozione di quest'ultime²⁰⁵.

Le misure precautelari previste dal codice di procedura penale sono: l'arresto obbligatorio in flagranza (art. 380 c.p.p.), l'arresto facoltativo in flagranza (art. 381 c.p.p.), il fermo di indiziato di delitto (art. 384 c.p.p.) e l'allontanamento d'urgenza dalla casa familiare (art. 385 c.p.p.).

Nei casi in cui tali misure richiedono alla polizia giudiziaria di effettuare dei giudizi di natura prognostica, l'IA predittiva potrebbe fornire un valido supporto rendendo tali valutazioni meno incerte. La polizia giudiziaria, essendo tenuta a compiere tale valutazione nell'immediato, potrebbe non avere il tempo per effettuare tutte le considerazioni necessarie. L'IA è invece in grado di elaborare tutte le informazioni rilevanti nell'arco di pochissimo tempo. Inoltre, la pressione a cui la p.g. è sottoposta in sede operativa può rendere l'analisi sulla pericolosità poco lucida, con la conseguenza di sopravvalutare il rischio. Il ricorso all'IA predittiva potrebbe rivelarsi particolarmente utile in quei casi in cui, data la mancanza di un pericolo giuridicamente rilevante di reiterazione del reato o di fuga, non è necessario limitare la libertà personale.

Si tratta di un tema particolarmente discusso anche negli Stati Uniti dove è stato osservato che sarebbe opportuno facilitare il rilascio diretto da parte della stazione di polizia (cosiddetto *station house release*) limitando l'arresto ai casi più gravi. Il procedimento necessario ai fini dell'arresto richiede infatti un significativo dispendio di risorse e potenzialmente può ledere il rapporto tra le forze dell'ordine e la comunità. Gli indagati che presentano un rischio particolarmente basso dovrebbero essere rilasciati a seguito della stima del rischio da parte del *risk assessment tool*. Si può citare il caso del Kentucky dove sono stati usati degli strumenti di valutazione del rischio per identificare indagati idonei al cosiddetto *station house release*.²⁰⁶

2.4.2) Valutazione dei presupposti dell'arresto obbligatorio e dell'arresto facoltativo

²⁰⁵ G. CONSO -V. GREVI -M. BARGIS, *Compendio di procedura penale*, 11°ed. CEDAM, Milano, 2023, pp. 518-520.

²⁰⁶ M. STEVENSON -S.G. MAYSON *Pretrial Detention and Bail* in E. LUNA *Reforming criminal justice*, vol. 3, Sandra Day O'Connor College of Law, Arizona State University, 2017, p. 31.

La prima delle misure precautelari previste dal codice di procedura penale è l'arresto obbligatorio. Ai sensi dell'art. 380, primo comma, c.p.p., il presupposto è l'essere colto *“in flagranza di un delitto non colposo, consumato o tentato, per il quale la legge stabilisce la pena dell'ergastolo o della reclusione non inferiore nel minimo a cinque anni e nel massimo a venti anni”*. Ai sensi dell'art. 382, comma 1, c.p.p., con stato di flagranza si intende l'essere colto nell'atto di commettere il reato (cosiddetta flagranza propria), oppure l'essere inseguito dalla polizia giudiziaria subito dopo il reato, o essere sorpreso con cose o tracce dalle quali appaia che egli abbia commesso il reato immediatamente prima (sono entrambe ipotesi di cosiddetta flagranza impropria). L'art. 8, comma 1-ter, legge 12 dicembre 1989, n. 401, ha introdotto la cosiddetta flagranza differita, che ricorre qualora il soggetto risulti l'autore del reato sulla base di documentazione videofotografica dalla quale emerga inequivocabilmente il fatto.

Nel caso dell'arresto obbligatorio, la flagranza costituisce il solo presupposto che la polizia giudiziaria è tenuta ad accertare, non assumendo rilievo la gravità del fatto o la pericolosità sociale dell'arrestato. Pertanto, l'IA predittiva non può fornire particolare ausilio ai fini di tale accertamento, in quanto il giudizio sulla flagranza non presenta natura prognostica, ma consiste in un mero accertamento di una situazione di fatto. L'accertamento dei presupposti dell'arresto obbligatorio rimarrebbe plausibilmente oggetto di una valutazione spettante esclusivamente alla p. g., anche nel caso in cui l'IA predittiva venisse introdotta nel processo penale. Nell'arresto facoltativo, invece, a norma dell'art. 381, comma 1, c.p.p., la polizia giudiziaria non ha l'obbligo, bensì la facoltà di *“arrestare chiunque è colto in flagranza di un delitto non colposo, consumato o tentato, per il quale la legge stabilisce la pena della reclusione superiore nel massimo a tre anni ovvero di un delitto colposo per il quale la legge stabilisce la pena della reclusione non inferiore nel massimo a cinque anni”* (altrimenti, ricorrerebbero i presupposti per l'arresto obbligatorio).

In dottrina è stato notato che l'uso del termine “facoltativo” sarebbe improprio, in quanto l'art. 381 c.p.p. esprimerebbe, in realtà, un potere discrezionale²⁰⁷, giacché *“si procede all'arresto in flagranza soltanto se la misura è giustificata dalla gravità del fatto ovvero dalla pericolosità del soggetto desunta dalla sua personalità o dalle circostanze del fatto”* (art. 381, comma 4, c.p.p.). Il giudizio sulla gravità del reato consiste in un accertamento su fatti avvenuti nel passato; anch'esso non è dunque suscettibile di valutazione mediante l'IA predittiva. Al contrario, il giudizio sulla pericolosità del soggetto, avendo ad oggetto fatti futuri, presenta natura prognostica ed è quindi suscettibile di essere accertato con l'ausilio dell'intelligenza artificiale predittiva.

Occorre poi tenere conto dell'incertezza circa il concetto di pericolosità di cui l'art. 381, comma 4, c.p.p. non fornisce una definizione, a differenza dell'art. 203 c.p. che, in materia di misure di sicurezza, definisce la pericolosità sociale come la probabilità che vengano commessi nuovi reati.

²⁰⁷ L. D'AMBROSIO, *sub* art. 381 in M. CHIAVARIO *Commentario al nuovo Codice di procedura penale*, UTET, Torino, 1990, pp. 380-381.

Tale incertezza si attenua se si basa il giudizio sulla pericolosità in relazione ai criteri di cui all'art. 133 c.p. (indici di commisurazione della pena), come è stato suggerito in dottrina²⁰⁸. Ciò avrebbe un effetto positivo anche nel caso in cui questo giudizio fosse supportato dall'intelligenza artificiale, in quanto si avrebbero dei criteri maggiormente stringenti che gli sviluppatori di tali sistemi sarebbero tenuti a seguire. In ogni caso, permane l'incertezza applicativa dovuta alla mancanza di una definizione legislativa di pericolosità, con la conseguenza che sarebbe difficile individuare quali eventi dovrebbero essere previsti da un sistema di IA funzionale a supportare tale giudizio.

Una possibile soluzione interpretativa potrebbe consistere nel ricondurre la pericolosità giustificante l'arresto facoltativo all'esigenza cautelare di cui all'art. 274, lett. c), c.p.p.: il pericolo di reiterazione del reato. Tale interpretazione muove dal presupposto che, siccome l'arresto in flagranza deve essere convertito in una misura cautelare sulla base del pericolo di fuga, inquinamento probatorio o pericolo di reiterazione del reato (art. 391, comma 5, c.p.p.), occorre attribuire rilievo a tali criteri già in sede di applicazione della misura precautelare²⁰⁹. Accogliendo tale interpretazione, la pericolosità di cui all'art. 391, comma 4, c.p.p., potrebbe essere accertata da uno strumento di valutazione del rischio funzionale a prevedere il pericolo di reiterazione del reato, come nel caso del *Pretrial risk scale* di COMPAS oppure del *New Criminal Arrest* di PSA. Tali strumenti stimano il rischio che una persona venga arrestata per la commissione di un nuovo reato²¹⁰. In definitiva, basare la pericolosità, ai fini dell'arresto facoltativo sul pericolo di reiterazione del reato renderebbe maggiormente agevole lo sviluppo di un sistema IA che supporti tale valutazione.

Un altro rischio da considerare è che tale valutazione potrebbe essere, nei fatti, delegata al sistema di IA senza un sostanziale controllo da parte della p.g. Tale criticità potrebbe essere particolarmente accentuata alla luce della rapidità con la quale le forze dell'ordine sono tenute a compiere tali valutazioni. Per contrastare tale pericolo, è necessario che la polizia giudiziaria sia tenuta a motivare le ragioni alla base del provvedimento disposto. La giurisprudenza prevalente richiede già oggi che la p. g. indichi le ragioni che l'hanno indotta a esercitare il proprio potere, sebbene non sia necessario che il provvedimento venga redatto con una specifica motivazione. Risulta sufficiente che dal contesto descrittivo del verbale emergano le ragioni del provvedimento, in modo tale che queste possano essere valutate dal giudice in sede di convalida, ex art. 319 c.p.p.²¹¹. L'autorità giudiziaria si limita ad effettuare un controllo sulle ragioni addotte dalla polizia e non può sostituirsi ad essa nell'individuare i presupposti di fatto e di diritto legittimanti la misura²¹².

Nel caso in cui l'IA dovesse essere introdotta in questo settore, si potrebbe rafforzare tale onere motivazionale, richiedendo alla p.g. di non limitarsi ad indicare il *risk score* prodotto dall'IA, bensì di addurre specificamente gli elementi di fatto da cui tale pericolosità è stata desunta. Il giudice, in sede di convalida, sarebbe così nelle

²⁰⁸ E. APRILE, *sub art. 381*, in A. GIARDA - G. SPANGHER *Codice di procedura penale commentato*, 4° edizione, Ipsoa, Milano, 2010.

²⁰⁹ L. D'AMBROSIO *op. cit.*, *sub art. 381*.

²¹⁰ Cfr. par. 2.2.3.

²¹¹ Cass. pen., Sez. VI, 6 maggio 2009, n. 3128; A. GAITO – A. BARGI *et al.*, *Codice di procedura penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub art. 384*.

²¹² Cass. pen., Sez. VI, 21 dicembre 1995, n. 4749; A. GAITO – A. BARGI, *op. cit.*

condizioni di sindacare la sussistenza dei presupposti della misura. Non sarebbe invece possibile sindacare tali presupposti nel caso di una motivazione che si limiti ad indicare il livello di rischio stimato dal modello predittivo, in quanto le ragioni sottostanti a tale *output* potrebbero non essere conoscibili dallo stesso giudice, per via del problema della cosiddetta *black box*.

In conclusione, l'IA potrebbe fornire un supporto alla p. g. nell'accertare i presupposti dell'arresto facoltativo, in quanto, a differenza dell'arresto obbligatorio, è presente un giudizio prognostico, essendo richiesto di giustificare la misura in ragione della pericolosità. A tal fine, sarebbe opportuno interpretare la pericolosità di cui all'art. 391, comma 4, c.p.p. facendo riferimento all'esigenza cautelare del pericolo di reiterazione del reato e rafforzare l'onere motivazionale a carico della p. g., richiedendo di indicare le specifiche circostanze di fatto da cui tale pericolo è stato desunto.

2.4.3) Valutazione dei presupposti del fermo di indiziato di delitto

Un'altra misura precautelare rilevante ai fini della presente disamina è il fermo di indiziato di delitto (art. 384 c.p.p.). Il fermo è un provvedimento che, di regola, può essere disposto esclusivamente dal pubblico ministero, quando sono presenti le seguenti condizioni: che sussistano “*specifici elementi che, anche in relazione alla impossibilità di identificare l'indiziato fanno ritenere fondato il pericolo di fuga*”; che la persona sia “*gravemente indiziata*”; che gli indizi riguardino “*un delitto per il quale la legge stabilisce la pena dell'ergastolo o della reclusione non inferiore nel minimo a due anni e superiore nel massimo a sei anni ovvero di un delitto concernente le armi da guerra e gli esplosivi o di un delitto commesso per finalità di terrorismo, anche internazionale, o di eversione dell'ordine democratico*” (art. 384, comma 1, c.p.p.).

A norma del secondo e terzo comma dello stesso articolo, può eccezionalmente procedere anche la polizia giudiziaria nelle seguenti ipotesi: prima che il PM abbia assunto le indagini; qualora si sia successivamente individuato un indiziato fino ad allora ignoto; in caso di fondato pericolo di fuga in cui la situazione di urgenza non permetta di attendere il provvedimento del PM²¹³.

L'intelligenza artificiale predittiva potrebbe supportare la valutazione in merito alla fondatezza del pericolo di fuga, sia nel caso in cui tale potere venga esercitato dal PM che nel caso di fermo disposto dalla polizia giudiziaria. L'IA potrebbe rivelarsi uno strumento particolarmente prezioso nei casi di urgenza previsti dall'art. 384, comma 4, c.p.p., in cui l'impossibilità di attendere il provvedimento del PM rischia di incidere negativamente sulla capacità della polizia giudiziaria di effettuare una valutazione lucida e ponderata circa la sussistenza di tale pericolo.

Quanto alla nozione di pericolo di fuga, in giurisprudenza è stato definito come la ragionevole probabilità che, ove non si intervenisse, l'inquisito farebbe perdere le proprie tracce²¹⁴. Si tratta dunque di un pericolo maggiormente preciso rispetto alla pericolosità idonea a giustificare l'arresto facoltativo *ex art. 381 c.p.p.*, con la conseguenza che sarebbe anche maggiormente agevole sviluppare un sistema IA *ad hoc*.

²¹³ P. TONINI -C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, pp. 549-550.

²¹⁴ Cass. pen., Sez. I, 27 marzo 1991.

Vi sono poi diversi fattori che, pur senza automatismi, possono contribuire a fondare il pericolo di fuga e che dovrebbero essere alla base di un sistema IA che valuti tale pericolo in conformità con il dato legislativo e giurisprudenziale. Ad esempio, *ex art. 384, comma 1 c.p.p.*, il pericolo di fuga può essere fondato anche sull'impossibilità di identificare l'indiziato. La Cassazione ha chiarito che tale impossibilità non fonda necessariamente il pericolo di fuga, ma che occorre valutare in concreto l'idoneità di questo elemento a fondare il rischio di sottrazione alla giustizia, senza automatismi²¹⁵.

L'art. 384, comma 4, c.p.p. considera il possesso di documenti falsi tra gli elementi che possono rendere fondato il pericolo che l'indiziato stia per darsi alla fuga. Anche in questo caso, si deve ritenere che dal possesso di documenti falsi non discende necessariamente il pericolo di fuga, ma che si tratta di un elemento sintomatico, da valutarsi insieme ad altri fattori²¹⁶.

Anche la gravità del reato è un elemento indiziante ma non necessariamente decisivo. Altrimenti, come affermato dalle Sezioni Unite, si finirebbe per sancire una presunzione di pericolo di fuga per i reati più gravi, quando, invece, tale pericolo dovrebbe essere sempre accertato in concreto²¹⁷.

In definitiva, l'impossibilità di identificare l'indiziato, il possesso di documenti falsi e la gravità del reato sono elementi di cui un sistema IA dovrebbe tenere conto, sebbene non siano da soli decisivi, per una previsione del pericolo di fuga che sia conforme alla legge e alla giurisprudenza.

Occorre poi rilevare come tale pericolo sia sostanzialmente corrispondente al pericolo di fuga *ex art. 274, lett. b), c.p.p.* Infatti, anche se in materia di fermo non si specifica che il pericolo di fuga deve essere concreto, tale requisito si desume dalla specificità degli elementi da cui il pericolo deve essere desunto. Si tratta dunque di presupposti alquanto simili, seppur non perfettamente corrispondenti alla luce della natura maggiormente articolata del giudizio in sede cautelare, in cui è possibile, ad esempio, graduare le misure coercitive adeguandole alle esigenze cautelari²¹⁸. Tenendo conto della somiglianza di tali accertamenti, potrebbero essere usati i medesimi sistemi IA in entrambi i casi. Si pensi ai *risk assessment tools* impiegati negli Stati Uniti al fine di valutare il pericolo di mancata comparizione in tribunale, come *COMPAS* o *PSA*²¹⁹.

In conclusione, l'IA potrebbe supportare il PM o la p.g. nel valutare la fondatezza del pericolo di fuga, tenendo conto, tra gli altri, di fattori come il possesso di documenti falsi, la gravità del reato e l'impossibilità di identificare l'indiziato. Come per il pericolo di fuga in materia cautelare, anche in materia di fermo di indiziato, l'esistenza di strumenti di valutazione del rischio come *COMPAS* e *PSA* consentono di ritenere che si tratti di un concreto e possibile utilizzo dell'IA e non di una mera speculazione.

2.4.4) Valutazione dei presupposti dell'allontanamento d'urgenza dalla casa familiare

²¹⁵ Cass. pen., Sez. II, 7 aprile 2012, n. 1531.

²¹⁶ A. GAITO, *op. cit.*, *sub art. 384*.

²¹⁷ Cass. pen., Sez. Un., 11 maggio 1993, n. 9.

²¹⁸ L. D'AMBROSIO *sub art. 381* in M. CHIAVARIO *Commentario al nuovo Codice di procedura penale*, UTET, Torino, 1990, pp. 404-405.

²¹⁹ Cfr. par. 2.2.2.

Il legislatore, con la legge 15 ottobre 2013, n. 119, ha aggiunto un'ulteriore misura precautelare, con l'obiettivo di predisporre un efficace strumento per il contrasto alla violenza domestica: l'allontanamento d'urgenza dalla casa familiare, ai sensi dell'art. 384-*bis* c.p.p. La *ratio* della misura è analoga al già citato allontanamento dalla casa familiare *ex* art. 282 *bis* e 282-*ter* c.p.p., in quanto entrambi gli istituti sono stati introdotti facendo specifico riferimento a reati in materia di violenza nelle relazioni familiari. Anche in questo ambito, l'IA potrebbe rendere la valutazione dei presupposti dell'art. 384 *bis* c.p.p. maggiormente affidabile, in modo da contrastare efficacemente la violenza in ambito familiare²²⁰.

Ai sensi dell'art. 384-*bis*, comma 1, c.p.p., la polizia giudiziaria può disporre tale misura, previa autorizzazione del PM, se il soggetto è colto nell'atto di commettere uno dei reati di violenza alla persona elencati nell'art. 282-*bis*, comma 6, c.p.p., come il reato di maltrattamenti familiari (art. 572 c.p.) o il reato di violenza sessuale (art. 609-*bis* c.p.). Come già osservato, lo stato di flagranza non consiste in un giudizio prognostico suscettibile di ausilio mediante l'IA predittiva²²¹.

Il secondo presupposto, previsto dal medesimo comma, è che “*sussistano fondati motivi per ritenere che le condotte criminose possano essere reiterate ponendo in grave ed attuale pericolo la vita o l'integrità fisica o psichica della persona offesa*”. In questo caso, si tratta di una valutazione chiaramente prognostica, che pertanto, può essere supportata dall'IA predittiva.

La legge 24 novembre 2023, n. 168, recante “*Disposizioni per il contrasto della violenza sulle donne e della violenza domestica*”, ha aggiunto un'ulteriore ipotesi di allontanamento d'urgenza che può essere disposta soltanto dal pubblico ministero. *Ex* art.384-*bis*, comma 2-*bis*, c.p.p., il PM può disporre tale misura anche al di fuori dei casi di flagranza, purché il soggetto risulti gravemente indiziato per determinati reati, quali, ad esempio, i maltrattamenti contro familiari (art. 572 c.p.). Inoltre, non deve essere possibile, per la situazione di urgenza, attendere il provvedimento del giudice. Si richiede poi la sussistenza di fondati motivi per ritenere che le condotte criminose possano essere reiterate, ponendo in grave e attuale pericolo la vita o l'integrità fisica o psichica della persona offesa. Anche nella nuova ipotesi di allontanamento d'urgenza, sussiste quindi il medesimo giudizio prognostico, rispetto al quale l'IA potrebbe fornire un valido supporto.

Risulta possibile distinguere due valutazioni prognostiche che la p.g. o il PM sono tenuti ad effettuare: il pericolo di reiterazione del reato e il conseguente grave e attuale pericolo per la vita o l'integrità fisica della persona offesa. Circa il pericolo di reiterazione del reato, è possibile basarsi su modelli predittivi specificamente sviluppati per prevedere il rischio di violenza di genere, come *EPV-R*, a cui si è fatto precedentemente riferimento. Tale strumento, sebbene sia stato oggetto di critiche, mostra come stimare il rischio di un crescendo di violenza impiegando *risk assessment tools* sia una concreta possibilità di sviluppo²²².

²²⁰ Cfr. Par. 2.4.

²²¹ Cfr. par. 2.3.2.

²²² A. VALDIVIA -C. HYDE VAAMONDE -J. GARCÍA MARCOS, *Judging the Algorithm: a Case Study on the Risk Assessment Tool for Gender-Based Violence Implemented in the Basque Country*, in *AI & Society*, vol. 40 n. 4, 2025, <https://arxiv.org/abs/2203.03723>, consultato il 25 settembre 2025, p.1.

In secondo luogo, occorre valutare l'idoneità della reiterazione del reato a porre in grave e attuale pericolo la vita o l'integrità fisica della persona offesa. Questo ulteriore pericolo dovrà essere accertato dalla p.g. o dal PM, alla luce delle circostanze del caso concreto, senza l'ausilio dell'IA. In molti casi, dall'accertamento del rischio di reiterazione del reato, consegue necessariamente anche il pericolo per la persona offesa. Ad esempio, se si ritiene sussistente il pericolo di reiterazione del reato di lesioni, ne consegue il pericolo per l'integrità della persona offesa. Rimane però opportuno tenere presente che si tratta di due giudizi prognostici distinti, al fine di evitare di rendere il requisito del rischio per la persona offesa una mera formula di stile.

In conclusione, sistemi IA potrebbero essere utilizzati per prevedere il pericolo di reiterazione del reato in materia di violenza di genere, al fine di fondare la misura precautelare dell'allontanamento d'urgenza dalla casa familiare.

2.5) Applicazione dell'IA predittiva ai fini delle misure di sicurezza personali

2.5.1) Introduzione

Un altro ambito del processo penale in cui l'intelligenza artificiale potrebbe avere un ruolo sono le misure di sicurezza, ai sensi degli artt. 199 e ss. c.p. Le misure di sicurezza sono una categoria di sanzioni penali ulteriore rispetto alla pena, impennate sull'idea di pericolosità delle persone e delle cose.²²³ Le misure di sicurezza incidenti sulla libertà personale sono denominate misure di sicurezza personali (artt. 199 e ss. c.p.), in contrapposizione alle misure di sicurezza patrimoniali, incidenti, invece, sul patrimonio (artt. 236 e ss. c.p.). In questo paragrafo, si tratterà delle misure di sicurezza personali, basate sulla pericolosità sociale, oggetto di accertamento mediante l'IA.

I presupposti per disporre una misura di sicurezza, a norma dell'art. 202, commi 1 e 2, c.p., sono l'aver commesso un reato o un quasi reato, nonché la pericolosità sociale. Per "quasi reato", in dottrina, si intendono i casi in cui la legge deroga espressamente alla regola fissata dall'art. 202, comma 1, c.p., secondo il quale la misura di sicurezza può essere disposta soltanto se è stato commesso un fatto previsto dalla legge come reato. Si tratta del reato impossibile (art. 49 c.p.), dell'accordo per commettere un delitto che poi non viene commesso (art. 115, commi 1 e 2, c.p.), dell'istigazione a commettere un reato, se l'istigazione viene accolta ma il reato non viene commesso (art. 115, comma 3, c.p.) e dell'istigazione a commettere un delitto, se l'istigazione non viene accolta (art. 115, comma 4, c.p.).

2.5.2) L'IA predittiva come strumento per superare la crisi della pericolosità sociale

Il secondo presupposto per applicare una misura di sicurezza, a norma dell'art. 203 c.p., è la pericolosità sociale, intesa come la probabilità che la persona commetta nuovi reati. La pericolosità sociale si accerta mediante un giudizio bifasico: in una prima fase, si analizza la personalità del soggetto, mentre, in una seconda

²²³ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022, pp. 883-884.

fase, si formula la prognosi criminale²²⁴. Un sistema di IA potrebbe essere impiegato per calcolare il livello di pericolosità sociale e adottare di conseguenza una misura di sicurezza.

Un'argomentazione a sostegno di tale possibilità riguarda la cosiddetta crisi della pericolosità sociale. In dottrina, sussiste una forte sfiducia in merito alla validità e all'affidabilità delle prognosi criminali, con la conseguenza che diversi autori considerano superata la stessa categoria concettuale della pericolosità sociale²²⁵. Alla base di tale crisi, vi è un forte scetticismo nei confronti delle scienze psicologiche criminologiche. Tale scetticismo si rinviene anche nella scelta dello stesso legislatore di vietare l'accertamento della personalità mediante perizie criminologiche, ai sensi dell'art.220, comma 2, c.p.p.²²⁶ (sebbene in dottrina si sia anche sostenuto che tale divieto sia piuttosto funzionale a tutelare la presunzione di innocenza)²²⁷.

Un'altra ragione della crisi della pericolosità sociale è che, a causa dello stesso art. 220, comma 2, c.p.p., il giudice deve compiere tale accertamento da solo, non essendo ammessa la perizia criminologica. Il giudice può avvalersi di un perito soltanto per stabilire se il soggetto è incapace di intendere o di volere, in quanto è prevista una deroga legislativa qualora si accertino qualità psichiche dipendenti da cause patologiche. Soltanto in questo ultimo caso, il giudice può porre al perito un ulteriore quesito in merito alla pericolosità sociale del soggetto²²⁸.

Dinanzi a tale sfiducia rispetto all'affidabilità delle prognosi criminali, si possono distinguere due approcci sulle potenzialità dell'IA in materia di misure di sicurezza. Secondo un primo approccio, lo stesso art. 220, comma 2, c.p.p. preclude il ricorso all'intelligenza artificiale nel valutare la personalità dell'imputato²²⁹. Al contrario, secondo una diversa impostazione, l'intelligenza artificiale potrebbe fornire un valido in grado di supplire alle difficoltà che il giudice potrebbe incontrare nel compiere tale accertamento senza potersi avvalere di un perito²³⁰.

Questo secondo approccio convince maggiormente alla luce delle potenzialità dell'IA che è in grado di elaborare una quantità di dati decisamente superiore rispetto a quanto potrebbe mai fare un singolo giudice, giungendo a conclusioni maggiormente solide sul piano statistico.

In definitiva, le odierne difficoltà nell'accertare la pericolosità sociale costituiscono un argomento a favore e non contro il ricorso all'IA in questo settore.

2.5.3) Pericolosità sociale e capacità a delinquere: affinità e differenze

²²⁴ *Ibid.*

²²⁵ *Ibid*; L. FORNARI, *Misure di sicurezza e doppio binario: un declino inarrestabile?* in *Rivista italiana di diritto e procedura penale*, Milano, 1993, p. 637.

²²⁶ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, <https://archivioldpc.dirittopenaleuomo.org/d/6702-quando-la-justizia-penale-incontra-l-intelligenza-artificiale-luci-e-ombre-dei-risk-assessment-too>, consultato il 3 agosto 2025, pp. 19-20.

²²⁷ G. AMATO, *Individuo e autorità nella disciplina della libertà personale*, Giuffrè, Milano, 1967, p. 380.

²²⁸ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 891.

²²⁹ S. QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI ACT*, in *Diritto di Difesa*, 2023, p. 5.

²³⁰ M. CAIANIELLO, *potenzialità e rischi derivanti dall'interazione tra I. A. e giustizia penale preventiva* in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, p. 273.

Tornando all'analisi della normativa rilevante in materia di misure di sicurezza, ai sensi dell'art. 203, comma 2, c.p.p., la pericolosità sociale si accerta in base ai criteri di cui all'art. 133 c.p.: la gravità del reato e la capacità a delinquere. Quest'ultimo criterio risulta affine alla pericolosità sociale, in quanto, secondo una parte della dottrina, sussiste un rapporto di genere e specie tra la capacità a delinquere, intesa come la mera possibilità che il soggetto commetta un reato, e la pericolosità sociale, da intendersi come la probabilità della commissione di nuovi reati. Tale ricostruzione muove dal presupposto che la capacità a delinquere sia rivolta al futuro, in quanto attiene all'attitudine a commettere nuovi reati²³¹. Questa tesi trova conferma anche in giurisprudenza, in quanto la Cassazione ha fatto riferimento al rapporto di genere e specie per affermare che l'accertamento sulla pericolosità sociale implica un accertamento implicito sulla capacità a delinquere. Nel caso di specie, ne conseguiva che, essendo stata disposta una misura di sicurezza, e avendo quindi accertato la pericolosità sociale, non era possibile disporre la sospensione condizionale, in quanto la capacità a delinquere era stata già implicitamente accertata²³².

Sulla base della tesi del rapporto di genere e specie, si può sostenere che, in virtù dell'affinità tra la pericolosità sociale e capacità a delinquere, il giudizio sulla pericolosità sociale dovrebbe tenere conto, in misura preminente, degli indici di capacità a delinquere indicati nell'art. 133, comma 2, c.p.: i motivi a delinquere, il carattere del reo, i precedenti penali e giudiziari, la condotta antecedente, contemporanea o susseguente al reato e le condizioni di vita individuale, familiare e sociale del reo.

I fattori che un sistema di IA dovrebbe considerare nell'accertare la pericolosità sociale e la capacità a delinquere dovrebbero essere gli stessi. La differenza risiederebbe piuttosto nella rilevanza che il giudice dovrebbe attribuire all'*output* prodotto dal modello predittivo. Nel giudizio sulla capacità a delinquere, il giudice deve commisurare la pena su di una scala di gravità. Il giudizio sulla pericolosità sociale presenta invece natura binaria: se il livello di rischio è tale da ritenere come probabile che la persona commetterà reati, il giudice deve disporre una misura di sicurezza, mentre, se tale livello di rischio non dovesse essere raggiunto, il giudice negherà l'applicazione.

L'affinità tra la capacità a delinquere e la pericolosità sociale implica anche che, potenzialmente, gli stessi algoritmi possono essere impiegati sia per determinare la pena che per valutare se disporre una misura di sicurezza. Ad esempio, lo strumento di valutazione del rischio *COMPAS* utilizzato nel caso Loomis per determinare la pena²³³ si presta, potenzialmente, sia ad accertare la pericolosità sociale che la capacità a delinquere. Una delle scale di rischio generate da *COMPAS* è il *general recidivism* che calcola il rischio che vengano commessi nuovi reati entro due anni dalla data in cui tale strumento è stato utilizzato²³⁴. Tale

²³¹ F. BASILE, *Esiste una nozione ontologicamente unitaria di pericolosità sociale? Spunti di riflessione con particolare riferimento alle misure di sicurezza e di prevenzione*, in *La pena, ancora fra attualità e tradizione*, Giuffrè, 2018, pp. 975-978; M. RONCO – B. ROMANO, *Codice penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub* art. 203.

²³² Cass. pen., Sez. I, 13 maggio 1991.

²³³ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 14.

²³⁴ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025, p. 27.

previsione del rischio di recidiva è simile, seppur non pienamente coincidente, alla definizione di pericolosità sociale come probabilità che vengano commessi nuovi reati. Le misure di sicurezza si applicano anche a persone che non hanno commesso nessun reato, assumendo rilevanza anche i quasi reati. In dottrina, si è osservato che sarebbe più corretto definire la pericolosità sociale più genericamente come la probabilità di commettere reati, anziché come la probabilità di commetterne nuovi, proprio in ragione del fatto che le misure di sicurezza si possono applicare anche a soggetti che non ne hanno commesso alcuno. Nei confronti di chi non ha commesso reati, utilizzare uno strumento predittivo del rischio di recidiva potrebbe dunque non essere efficace, perché, in questo caso, non si tratta di prevedere il rischio di recidiva, bensì il rischio che venga commesso un crimine per la prima volta.

Occorre anche segnalare che, negli Stati Uniti, la scala di rischio *general recidivism* viene impiegata nella commisurazione della pena e non in materia di misure di sicurezza, con la conseguenza non è possibile basarsi sull'esperienza statunitense per avere un quadro chiaro sulle potenzialità e i rischi in questo specifico settore. In conclusione, la pericolosità sociale e la capacità a delinquere sono concetti strettamente collegati, essendo entrambi orientati alla previsione di comportamenti futuri, ed essendo possibile in entrambi i casi fare affidamento a strumenti predittivi della recidiva. Tuttavia, le misure di sicurezza si applicano anche a soggetti che non hanno commesso reati, con la conseguenza che, in questa particolare ipotesi, il riferimento al rischio di recidiva sarebbe improprio.

2.5.4) Valutazione della pericolosità sociale del minore

L'intelligenza artificiale potrebbe fornire un valido supporto anche in materia di pericolosità sociale del minore, per la quale sussiste una definizione di pericolosità sociale *ad hoc*. Il legislatore, nel d.p.r. 22 settembre 1988, n. 448, recante “*disposizioni sul processo penale a carico di imputati minorenni*”, ha infatti definito in maniera maggiormente puntuale la pericolosità sociale del minore.

Tale accertamento assume rilievo ai fini dell'affidamento coattivo del minore in una comunità educativa. Il codice penale del 1930 ha previsto il riformatorio giudiziario come principale misura che deve essere disposta nei confronti del minore, salvo che la stessa debba essere eseguita in tutto o in parte quando il minore abbia compiuto gli anni 21, essendo in tal caso prevista la libertà vigilata (art. 223, commi 1 e 2, c.p.). Il riformatorio giudiziario è stato oggetto di forti critiche, in quanto presenta un regime del tutto simile a quello carcerario, inducendo il legislatore, con il già citato d. p. r. 448/1988, a stabilire che tale misura è applicata soltanto in relazione a determinati delitti, indicati dall'art. 23, comma 1 dello stesso d. p. r. (art. 36, comma 2). Per effetto di tale norma, il riformatorio giudiziario presenta natura residuale, essendo di regola applicata la libertà vigilata (art. 36, comma 1)²³⁵. Si è anche previsto che il riformatorio giudiziario si esegue nella forma del collocamento coattivo del minore in una comunità educativa (artt. 22 e 36, comma 2). Ne consegue che, in concreto, il riformatorio giudiziario non trova più attuazione.

²³⁵ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022, p. 907.

La pericolosità sociale del minore sussiste “*quando, per le specifiche modalità e circostanze del fatto e per la personalità dell'imputato, sussiste il concreto pericolo che questi commetta delitti con uso di armi o di altri mezzi di violenza personale o diretti contro la sicurezza collettiva o contro l'ordine costituzionale, ovvero gravi delitti di criminalità organizzata*” (art. 37, comma 2, d.p.r. 22 settembre 1988 n. 448).

Tale nozione di pericolosità sociale risulta maggiormente precisa rispetto a quella delineata dall'art. 203 c.p.,²³⁶ per diverse ragioni. In primo luogo, si richiede che la pericolosità sociale sia desunta dalle specifiche modalità e circostanze del fatto e non soltanto dalla personalità dell'imputato. Chiedendo al giudice di valorizzare il fatto compiuto nei suoi connotati oggettivi, si mitiga il rischio di un diritto penale d'autore, in cui è la personalità ad assumere particolare rilievo.

In secondo luogo, il pericolo deve essere concreto. Viene dunque richiesto uno standard di pericolosità sociale maggiormente stringente, in maniera analoga alla concretezza che deve caratterizzare il pericolo di inquinamento probatorio, di fuga e di reiterazione del reato, ai sensi dell'art. 274 c.p.p.

Infine, la pericolosità sociale del minore è limitata al pericolo che vengano commessi reati particolarmente gravi, come i delitti con uso di armi. La maggiore precisione e ristrettezza della norma ha comportato, negli ultimi anni, un numero particolarmente basso di minori sottoposti all'affidamento coattivo in comunità. Ad esempio, nel 2018, tale misura ha riguardato soltanto 18 minori²³⁷.

Dalla maggiore precisione della nozione di pericolosità sociale del minore rispetto alla definizione di pericolosità sociale *ex art. 203 c.p.*, conseguono maggiori tutele a favore dei potenziali destinatari di tali misure. La puntualità della definizione di pericolosità sociale del minore riduce il rischio che tale accertamento risulti arbitrario. Tale garanzia sarebbe particolarmente importante nel caso in cui l'IA fosse impiegata per supportare tale valutazione. La definizione di pericolosità sociale del minore, essendo maggiormente stringente e precisa rispetto all'art. 203 c.p., suggerisce al giudice di adottare particolare cautela nell'applicare le misure di sicurezza, come confermato dalla scarsa frequenza con cui le misure di sicurezza vengono applicate ai minori (si veda sopra). Anche in presenza di un elevato livello di pericolosità sociale stimato dall'intelligenza artificiale, il giudice sarebbe comunque prudente nell'applicare una misura di sicurezza, perché sarebbe tenuto a verificare che, nel caso concreto, il minore rientri nella puntuale e stringente definizione di pericolosità sociale individuata dal d. p.r. 1988 n. 448.

In sede di applicazione di una misura di sicurezza *ex art. 203 c.p.*, l'IA predittiva potrebbe invece riempire di significato nel caso concreto a una nozione di pericolosità sociale non puntualmente definita. La valutazione sulla pericolosità sociale potrebbe essere sostanzialmente delegata all'intelligenza artificiale, senza valutare, ad esempio, le specifiche modalità del fatto da cui la pericolosità sociale è stata desunta o la sussistenza o meno di un pericolo concreto.

Sarebbe quindi auspicabile che il legislatore delinei la pericolosità ai sensi dell'art. 203 c.p. in maniera analoga alla pericolosità sociale del minore, in modo tale da rendere la disposizione di cui all'art. 203 c.p.

²³⁶ *Ibid.*, p. 908.

²³⁷ *Ibid.*

maggiormente precisa. Si potrebbe richiedere, quanto meno, la concretezza del pericolo e la rilevanza delle specifiche modalità e circostanze del fatto. Una siffatta riforma sarebbe particolarmente urgente nel caso in cui l'IA dovesse essere introdotta nell'ambito delle misure di sicurezza, proprio per mitigare il rischio che tale accertamento venga sostanzialmente delegato all'IA.

Inoltre, anche nel caso specifico della pericolosità del minore, non mancano studi volti ad analizzare le potenzialità dell'IA. La letteratura spagnola si è concentrata su *SAVRY*, un *risk assessment tool* che viene impiegato per prevedere la probabilità che un minore di età compresa tra i 12 e i 18 anni commetta reati violenti. *SAVRY* si basa su fattori di rischio e fattori protettivi. Con fattori di rischio, si fa riferimento, ad esempio, a problemi comportamentali, nonché alla storia familiare e al contesto sociale. Tra i fattori protettivi, idonei a ridurre il rischio di recidiva, viene compreso il supporto familiare, le abilità sociali e l'impegno scolastico²³⁸.

In uno studio condotto in Catalogna, è stato effettuato un confronto tra diversi algoritmi di *machine learning* e *SAVRY* il quale, invece, è uno strumento di valutazione del rischio tradizionale che non si basa sull'apprendimento algoritmico. Se da un lato gli algoritmi di *machine learning* considerati sono risultati maggiormente accurati rispetto a *SAVRY*, dall'altro il *machine learning* tende ad essere discriminatorio nei confronti degli uomini e degli stranieri, a differenza di *SAVRY* che è risultato complessivamente più equo²³⁹. Sarebbe idealmente preferibile utilizzare algoritmi specificamente addestrati per accertare la pericolosità sociale del minore, come nell'esperimento sopra citato. Non si potrebbe utilizzare, ad esempio la scala di rischio di *COMPAS* predittiva del rischio di recidiva²⁴⁰, non solo perché non è stata pensata specificamente per i minorenni, ma anche perché non risultano esserci degli studi che accertino che l'accuratezza di *COMPAS* sia stata validata con specifico riferimento ai minori.

Inoltre, un eventuale modello predittivo della pericolosità sociale del minore non dovrebbe basarsi su di un generico pericolo di recidiva, ma specificamente sul pericolo che vengano commessi i reati di cui all'art. 37, comma 2, d.p.r. 448/1988. Anche gli algoritmi di *machine learning* sviluppati nel citato studio condotto in Catalogna non sarebbero dunque adatti allo scopo, in quanto predittivi di un generico rischio di recidiva.

In conclusione, la maggiore precisione e ristrettezza della pericolosità sociale del minore attenua il rischio che i giudici di fatto delegino tale accertamento all'IA, in contrapposizione alla maggiore ampiezza della pericolosità sociale ai sensi dell'art. 203 c.p. Al contrario, la mancanza di una definizione puntuale di pericolosità sociale negli artt. 202 e 203 c.p. rende tale rischio maggiormente significativo nel caso di misure di sicurezza disposte nei confronti di maggiorenni.

²³⁸ S. TOLAN -M. MIRON -E. GOMEZ., *et al.*, *Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in Catalonia*, in *Proceedings of the 17th International Conference on Artificial Intelligence and Law (ICAAIL '19)*, Montreal, Canada, Association for Computing Machinery, 2019, <https://doi.org/10.1145/3322640.3326705>, consultato il 6 ottobre 2025.

²³⁹ *Ibid.*

²⁴⁰ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, [Practitioners-Guide-to-COMPAS-Core | DocumentCloud](#), consultato il 25 settembre 2025, p. 27.

Lo studio condotto in Catalogna suggerisce poi che la possibilità di calcolare la pericolosità del minore con strumenti di IA sia concreta, seppur non esente da rischi. Occorre comunque assicurarsi che il relativo modello predittivo sia specificamente sviluppato per i minori e che la previsione riguardi soltanto i gravi reati di cui all'art. 37, comma 2, d.p.r. 448/1988.

2.6) Applicazione dell'IA predittiva ai fini delle misure di prevenzione

2.6.1) Il ruolo dell'IA nella componente prognostica del giudizio di pericolosità

Rimanendo nell'ambito della giustizia preventiva, occorre esaminare le potenzialità dell'IA nell'accertamento dei presupposti delle misure di prevenzione. A differenza delle misure cautelari o di sicurezza, le misure di prevenzione si caratterizzano per il fatto che non presuppongono la previa commissione di un reato. Per tale ragione, vengono denominate misure *ante delictum* o *praeter delictum*²⁴¹.

Tali misure sono previste nel d.lgs. 6 settembre 2011, n. 159, noto come codice antimafia. I soggetti destinatari sono previsti dall'art. 4. Molte delle ipotesi contenute in tale articolo fanno riferimento a soggetti indiziati della commissione di determinati reati, tra cui, ad esempio, il reato di associazione mafiosa di cui all'art. 416-bis c.p.

L'art. 4, lett. c). (il quale, a sua volta, rinvia all'art. 1) fa invece riferimento a soggetti genericamente pericolosi: *“coloro che debbano ritenersi, sulla base di elementi di fatto, abitualmente dediti a traffici delittuosi; coloro che per la condotta e per il tenore di vita debba ritenersi, sulla base di elementi di fatto che vivono abitualmente, anche in parte, con i proventi di attività delittuose; coloro che per il loro comportamento debba ritenersi, sulla base di elementi di fatto, che sono dediti alla commissione di reati che offendono o mettono in pericolo l'integrità fisica o morale dei minorenni, la sanità, la sicurezza o la tranquillità pubblica”*.

La Cassazione, in una sentenza del 2014, ha chiarito che il giudizio di prevenzione non si esaurisce in un giudizio prognostico in merito alla pericolosità sociale, bensì comprende anche l'apprezzamento di fatti storici indicanti la possibilità di iscrivere il soggetto in una delle categorie criminologiche individuate dalla legge²⁴². La giurisprudenza sembrerebbe così aver superato il tradizionale insegnamento della dottrina, secondo cui le misure di prevenzione, a differenza delle misure di sicurezza, non necessitano della previa commissione di un reato. Tra l'altro, tale orientamento dottrinale era sorto in presenza di un dato legislativo differente da quello attuale, in quanto antecedente al d.lgs.159/2011²⁴³.

Nella citata sentenza, la Cassazione scinde il giudizio di pericolosità in due componenti. In primo luogo, vi è una componente constatativa, in cui si verifica che vi siano degli elementi idonei a rappresentare che siano

²⁴¹ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* pp. 930-931.

²⁴² Cass. pen., Sez. I, 11 febbraio 2014, n. 23641.

²⁴³ G. VASSALLI, *Misure di prevenzione e diritto penale*, in *Studi in onore di Biagio Petrocelli*, vol. III, Giuffrè, 1972; F. 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 *Rivista italiana di diritto e procedura penale*, vol. 61, n. 2, Milano, 2018, pp. 974-975.

avvenute delle condotte contrarie alla convivenza sociale. A tale elemento si affianca la componente prognostica che consiste nel valutare la probabilità del ripetersi di condotte antisociali.

Questa distinzione tra la componente constatativa e la componente prognostica è particolarmente rilevante ai fini della presente analisi, poiché l'intelligenza artificiale predittiva non potrebbe fornire particolare ausilio nella fase constatativa, in quanto si tratta di accertare la sussistenza di fatti avvenuti nel passato, mentre, come precedentemente illustrato, l'IA predittiva mira a prevedere fatti futuri e non a ricostruire accadimenti passati. Allo stesso tempo, l'intelligenza artificiale potrebbe supportare il giudice nella fase prognostica, in quanto, in questo caso, si tratta di valutare la probabilità di eventi futuri.

In definitiva, nell'ipotesi di introduzione dell'IA in materia di misure di prevenzione, vi sarebbe un giudizio constatativo spettante esclusivamente al giudice e un giudizio prognostico in cui l'autorità giudiziaria sarebbe affiancata dall'intelligenza artificiale.

2.6.2) Il problema dell'indeterminatezza della pericolosità generica

Una delle argomentazioni avanzate in dottrina a sostegno del ricorso all'intelligenza artificiale in questo settore è che, alla luce dell'indeterminatezza della pericolosità generica ai sensi dell'art. 1 del codice antimafia, l'intelligenza artificiale potrebbe rappresentare uno strumento per rafforzare la valutazione giudiziale e ridurre la discrezionalità²⁴⁴. In questo paragrafo, verrà dunque analizzato il problema dell'indeterminatezza di tale disposizione, al fine di comprendere le implicazioni di un eventuale accertamento della c.d. pericolosità generica con l'ausilio dell'IA.

In generale, il principio di determinatezza implica che le norme penali devono descrivere fatti suscettibili di essere accertati e provati nel processo²⁴⁵. In merito alle misure di prevenzione, la questione era sorta inizialmente nella vigenza della legge 27 dicembre 1956, n. 1423, recante “*Misure di prevenzione nei confronti delle persone pericolose per la sicurezza e per la pubblica moralità*”, oggi sostituita dal d.lgs. 6 settembre 2011, n. 159. All'art. 1, nel determinare i soggetti destinatari delle misure di prevenzione, si faceva riferimento a manifestazioni che avessero dato luogo o dessero fondato motivo di ritenere che fossero “*proclivi a delinquere*”. La Corte costituzionale aveva censurato l'eccessiva indeterminatezza della disposizione, ritenendo che il principio di legalità e, quindi, anche il principio di determinatezza, dovesse applicarsi anche alle misure di prevenzione. Il fondamento di tale principio non si deve però individuare nell'art. 25, comma 2, Cost., come per le pene, bensì nell'art. 25, comma 3²⁴⁶. L'esercizio del potere giurisdizionale in materia di misure di prevenzione deve dunque trovare un fondamento legale e determinato e una conseguente riduzione dei margini di discrezionalità/arbitrarietà della decisione. Nel caso di specie, l'espressione “*proclivi a*

²⁴⁴ L. MACRI, *I primi passi dell'Italia verso l'impiego dell'IA nel processo penale e il calcolo del rischio di recidiva* in *Giurisprudenza penale web*, 2025, <https://www.giurisprudenzapenale.com/2025/02/06/i-primi-passi-dellitalia-verso-limpiego-dellia-nel-processo-penale-e-il-calcolo-del-rischio-di-recidiva/>, consultato il 12 settembre 2025, p. 41.

²⁴⁵ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, pp. 88-89.

²⁴⁶ Corte cost., 16 dicembre 1980, n. 77.

delinquere” non soddisfaceva tale standard di determinatezza ed è stata, quindi, oggetto della declaratoria di incostituzionalità.

Il legislatore, con la legge 3 agosto 1988, n. 327, è intervenuto introducendo il requisito degli “*elementi di fatto*” in grado di dimostrare la dedizione alla commissione di delitti, ovvero la circostanza di vivere con i relativi proventi. Nelle intenzioni del legislatore, tale requisito avrebbe dovuto rendere il giudizio sulla pericolosità maggiormente determinato, in quanto ancorato ad elementi fattuali. Si tratta di un elemento presente anche nell’attuale formulazione dell’art. 1 del d.lgs. 6 settembre 2011, n. 159. Tuttavia, mentre una parte della dottrina sostiene che l’attuale formulazione soddisfa maggiormente le esigenze di determinatezza,²⁴⁷ un’altra parte evidenzia come non vi sia in realtà un sostanziale progresso rispetto alla formula censurata dalla Corte costituzionale²⁴⁸.

La Corte europea dei diritti dell’uomo, nella causa *De Tommaso c. Italia*, ha condannato l’Italia per la violazione dell’art. 2 del “*Protocollo n. 4 alla Convenzione per la salvaguardia dei diritti dell’uomo e delle libertà fondamentali*”, firmato a Strasburgo il 16 settembre 1963 (CETS n. 46), in cui si sancisce che la libertà di circolazione non può essere sottoposta a restrizioni diverse da quelle previste dalla legge²⁴⁹. Ad avviso della Corte, la legge italiana non contiene indicazioni sufficientemente dettagliate su che tipo di condotta sia da considerare espressiva di pericolosità sociale. L’attuale formulazione dell’art. 1 sembrerebbe non soddisfare le esigenze di determinatezza, intesa come suscettibilità dei fatti di essere accertati e provati nel processo.

Alla luce di tale indeterminatezza, occorre comprendere se il ricorso all’intelligenza artificiale può assumere un ruolo positivo o meno. Da un lato, come è stato osservato in dottrina, l’intelligenza artificiale può rappresentare uno strumento per supportare l’accertamento giudiziale e dunque, ridurre la discrezionalità, assumendo un ruolo di *double check* degli elementi raccolti manualmente dall’autorità²⁵⁰. Tale argomentazione è analoga alle osservazioni già svolte in materia di misure di sicurezza, in quanto, anche per le misure di prevenzione l’IA potrebbe affiancare il giudice nel compiere un accertamento altrimenti caratterizzato da una significativa incertezza²⁵¹.

Nelle misure di sicurezza, il pericolo di indeterminatezza, se sussistente, è comunque quanto meno attenuato rispetto alle misure di prevenzione. Come già osservato, l’art. 203, comma 2, c.p., rinvia ai criteri di determinazione della pena ai sensi dell’art. 133 c.p., con la conseguenza che un sistema di IA potrebbe essere addestrato con riferimento a fattori riconducibili a tali criteri. Ai fini delle misure di prevenzione, sarebbe invece molto difficile individuare, sia pur in termini generali, i criteri attraverso cui addestrare un sistema di IA predittivo di una siffatta pericolosità sociale. Non si può allora escludere che, di fatto, sarebbero gli sviluppatori di tali algoritmi a determinare le condotte espressive di pericolosità sociale.

²⁴⁷ A. GAITO -B. ROMANO *et al.* *Misure di prevenzione*, UTET Giuridica, 2013, p. 99.

²⁴⁸ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 933.

²⁴⁹ Corte europea dei diritti dell’uomo, *De Tommaso c. Italia*, 23 febbraio 2017, ric. n. 43395/09.

²⁵⁰ L. MACRI, *op. cit.*, p. 41.

²⁵¹ Cfr. par. 2.4.2.

In conclusione, l'IA potrebbe supportare l'autorità giudiziale nell'accertare la pericolosità in materia di misure di prevenzione e ridurre la discrezionalità, ma sarebbe opportuno riformulare l'art. 1 del d.lgs. 6 settembre 2011, n. 159, in conformità alle indicazioni della Corte costituzionale e della Corte europea dei diritti dell'uomo.

2.7) L'applicazione dell'IA predittiva in sede di determinazione della pena

2.7.1) Intelligenza artificiale predittiva e discrezionalità giudiziale: potenzialità e rischi

Fino ad ora, sono stati esaminati i potenziali campi di applicazione dell'intelligenza artificiale riconducibili alla giustizia preventiva. In questo paragrafo, verranno invece illustrate le potenzialità dell'IA in sede di determinazione della pena. Come precedentemente accennato, il dibattito statunitense si è concentrato primariamente su tale ambito, soprattutto in merito all'accesa discussione sorta a seguito del caso *Loomis*.

Una possibile argomentazione a sostegno del ricorso all'IA in questo settore è che si potrebbe restringere la discrezionalità giudiziale in sede di commisurazione della pena, rendendo tale giudizio maggiormente standardizzato e quindi, meno arbitrario.

A norma dell'art. 132, comma 1, c.p. *“nei limiti fissati dalla legge, il giudice applica la pena discrezionalmente; esso deve indicare i motivi che giustificano l'uso di tal potere discrezionale”*. La discrezionalità si deve alla scelta del legislatore di individuare delle fattispecie generali e astratte per poi attribuire al singolo giudice il compito di modulare le stesse in relazione alle specifiche circostanze del caso concreto²⁵². Tale indeterminatezza è intenzionale, in quanto il legislatore non ha voluto esaurire, mediante la previsione normativa, tutte le modalità giuridicamente rilevanti del comportamento²⁵³. In dottrina, si è discusso sulla natura libera o vincolata della discrezionalità penale, preferendo la seconda ipotesi. La discrezionalità incontra i cosiddetti limiti esterni, consistenti nella predeterminazione del tipo di pena da comminare, nonché della cornice edittale all'interno della quale la pena deve essere commisurata. La discrezionalità presenta poi dei limiti interni, relativi al rispetto dei criteri di cui all'art. 133 c.p.²⁵⁴.

L'art. 132, comma 1, stabilisce altresì che il giudice *“deve indicare i motivi giustificanti l'uso di tale potere discrezionale”*. La norma svolge sia una funzione di garanzia rispetto a decisioni arbitrarie che una funzione di controllo, in quanto si rende possibile sindacare la decisione giudiziale in sede di impugnazione²⁵⁵.

La dottrina è spesso critica in merito alle modalità in base alle quali l'obbligo di motivazione viene esplicito nella prassi giurisprudenziale. Secondo una parte della giurisprudenza, il giudizio di commisurazione della pena sarebbe infatti il risultato di un'intuizione derivante dalla valutazione complessiva delle circostanze del fatto e della personalità dell'imputato²⁵⁶. In senso critico, una parte della dottrina osserva che l'osservanza di

²⁵² G. MARINUCCI -E. DOLCINI -G.L. GATTA, *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022, p. 795; F. BRICOLA, *La discrezionalità nel diritto penale. Nozione ed aspetti costituzionali*, vol.1, Giuffrè, Milano, 1965, p. 100.

²⁵³ F. CORDERO, *Le situazioni soggettive nel processo penale*, Utet, Torino, 1956, p. 161.

²⁵⁴ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 100.

²⁵⁵ G. LATTANZI -E. LUPO, *Codice penale. Rassegna di giurisprudenza e dottrina*, Giuffrè, Milano, 2010, p.9.

²⁵⁶ Cass. pen., sez. VI, 7 maggio 1982; Cass. pen., Sez. II, 16 dicembre 1982; G. LATTANZI -E. LUPO, *op. cit.*

vincoli esegetici predeterminati, anziché il ricorso all'intuizione giudiziale, risulta maggiormente coerente con il principio di legalità su cui si basa l'intero diritto penale²⁵⁷. Secondo questa impostazione, sarebbe preferibile l'orientamento giurisprudenziale che sostiene che la commisurazione della pena non si può basare sull'intuizione giudiziale, in quanto occorre indicare verificabili criteri razionali. In tal senso, a Cassazione ha affermato che i parametri di cui all'art. 133 c.p. sono elusi qualora nella motivazione si faccia genericamente riferimento alle circostanze di cui all'art. 133 c.p.²⁵⁸. Si tratta di una prassi giurisprudenziale fortemente criticata dalla dottrina, la quale evidenzia che, talvolta, i giudici ricorrono a delle mere formule motivazionali di stile.

La stessa giurisprudenza di legittimità è stata criticata per aver, di fatto, avallato tale prassi nel caso in cui la pena venga applicata nel minimo. Quando la pena viene applicata nel minimo, in misura prossima al minimo, o comunque in misura inferiore al medio edittale, sarebbero sufficienti le tradizionali formule sintetiche quali “appare congruo” o “equo”, oppure “adeguato”²⁵⁹. Secondo questo orientamento, quanto più il giudice intende discostarsi dal minimo edittale, tanto più questi è tenuto a dare ragione del corretto esercizio del proprio potere discrezionale²⁶⁰. Si è altresì affermato che un'interpretazione rigorosa dell'obbligo di motivazione trova spazio solamente nel caso in cui la pena venga applicata in misura compresa fra il medio e il massimo edittale (ipotesi, peraltro, non frequente nella prassi)²⁶¹. Tali principi sono stati criticati in quanto l'obbligo di motivazione ex art. 132, primo comma, c.p., non dovrebbe configurarsi in maniera diversa a seconda del *quantum* della pena, data l'assenza di una previsione legislativa espressa che fondi tale interpretazione²⁶².

Mentre, da un lato, i giudici di merito ricorrono talvolta a delle mere formule di stile, dall'altro, l'orientamento giurisprudenziale prevalente rischia, almeno in alcune ipotesi, di avallare tale prassi. Il giudice potrebbe prima decidere il *quantum* della pena, salvo poi ritenere soddisfatto l'onere motivazionale mediante delle “formulette pigre”, che richiamano l'equità, la correttezza o l'adeguatezza della pena inflitta²⁶³. In definitiva, la determinazione della pena può risultare priva di una motivazione puntuale, con l'ulteriore conseguenza negativa di risultare eccessivamente discrezionale.

In questo contesto, si inseriscono le potenzialità dell'intelligenza artificiale nel rendere tale giudizio meno arbitrario. Tali sistemi potrebbero raccogliere e analizzare moltissime informazioni, con specifico riferimento a quelle relative ai criteri da cui desumere la capacità a delinquere, ai sensi dell'art. 133, comma 2, c.p.²⁶⁴. Verrebbero così forniti dei criteri di valutazione più oggettivi e replicabili rispetto alle singole interpretazioni

²⁵⁷ G. LATTANZI -E. LUPO, *op. cit.*

²⁵⁸ Cass. pen., Sez. I, 14 ottobre 1988; G. LATTANZI -E. LUPO *op. cit.*

²⁵⁹ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 797; Cass. pen., Sez. IV, 5 novembre 2015, n. 46412.

²⁶⁰ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 797.; Cass. pen., Sez. I, 13 marzo 2013, n. 24213.

²⁶¹ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 797; Cass. pen., Sez. V, 27 giugno 2019, n. 35100.

²⁶² G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 797, p. 798.

²⁶³ F. COPPOLA, *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo*, in *Archivio Penale* 2023, n.2, <https://archiviopenale.it/commisurazione-della-pena-e-intelligenza-artificiale-una-ipotesi-di-lavoro-con-lalgoritmo-ex-aequo/articoli/41566>, consultato il 16 settembre 2025, p. 10.

²⁶⁴ R. BLAIOTTA, *Giustizia, errore, intelligenza artificiale*, in *Sistema penale*, 2023, <https://www.sistemapenale.it/it/articolo/blaiotta-giustizia-errore-intelligenza-artificiale>, consultato il 16 settembre 2025, pp. 14-17.

individuali. Ciò consentirebbe di standardizzare parte del processo decisionale, riducendo l'influenza di giudizi soggettivi. Il giudizio sulla commisurazione della pena sarebbe dunque meno arbitrario, e maggiormente conforme alla discrezionalità vincolata di cui all'art. 132 c.p.

Sarebbe anche possibile ridurre almeno in parte la prassi giudiziale consistente nel motivare genericamente facendo riferimento ai criteri di cui all'art. 133 c.p. Il giudice sarebbe tenuto a confrontarsi con l'*output* prodotto dall'IA, in sede di motivazione, indicando le ragioni per le quali il giudizio effettuato dall'intelligenza artificiale risulta convincente o meno nel caso di specie. La motivazione ne risulterebbe maggiormente articolata, inducendo lo stesso giudice ad effettuare una valutazione maggiormente ponderata.

Tuttavia, il noto problema della cosiddetta *black box* potrebbe porsi in contrasto con l'obbligo motivazionale di cui all'art. 132 c.p. Nel caso di un sistema IA che suggerisca di adottare una pena particolarmente elevata, il giudice non sarebbe in grado di individuare quali dei criteri di cui all'art. 133 c.p. sono stati decisivi. L'attuale prassi consistente nel motivare mediante delle mere formule di stile rischia allora di essere sostituita da una nuova tendenza a motivare *per relationem* con un rinvio puro e semplice all'*output* algoritmico. Per questa ragione, anche in caso di impiego dell'IA, il giudice dovrebbe comunque motivare con specifico riferimento ai criteri di cui all'art. 133 c.p. che, secondo il proprio convincimento, risultano decisivi nel determinare il *quantum* della pena. Ciò sarebbe necessario anche al fine di permettere, in sede di impugnazione, di sindacare l'esercizio della discrezionalità giudiziale. Qualora il giudice si limitasse a rinviare ad un sistema IA di cui si ignora il funzionamento per il problema della *black box*, non sarebbe possibile conoscere, e quindi sindacare in sede di impugnazione, come sono stati applicati i criteri *ex art.* 133 c.p.

In conclusione, l'intelligenza artificiale potrebbe restringere la discrezionalità giudiziale, riducendo la rilevanza di giudizi soggettivi a favore di una maggiore standardizzazione. Allo stesso tempo, vi è il rischio che i giudici si limitino a rinviare all'*output* algoritmico, senza individuare i criteri *ex art.* 133 c.p. che sono risultati decisivi nel caso di specie.

2.7.2) Criteri di commisurazione della pena in cui l'IA predittiva potrebbe offrire supporto

Dopo aver illustrato il potenziale ruolo dell'IA nel supportare la discrezionalità giudiziale, occorre analizzare quali dei criteri di commisurazione della pena potrebbero essere applicati con l'ausilio di tale tecnologia.

L'art. 133 c.p. delinea i criteri fattuali di commisurazione della pena: gravità del reato e capacità a delinquere. Secondo parte della dottrina, ai criteri fattuali espressamente previsti dal legislatore, si affiancano i criteri finalistici su cui, invece, il codice penale tace. La commisurazione della pena deve infatti tenere conto delle finalità costituzionalmente rilevanti che la condanna persegue²⁶⁵, come la finalità rieducativa, prevista dall'art. 27, comma 3, Cost. Tale finalità, secondo la giurisprudenza costituzionale, deve essere perseguita, non soltanto in sede di esecuzione, ma anche ai fini della commisurazione della pena²⁶⁶. Non si può escludere che, in futuro, l'IA sia in grado di replicare o imitare anche i giudizi di valore che i giudici compiono nell'applicare i criteri

²⁶⁵ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 798.

²⁶⁶ Corte cost., 2 luglio 1990, n. 313.

finalistici. Si pensi, ad esempio, alla scelta che un giudice deve compiere circa la prevalenza, nel caso concreto, tra la finalità rieducativa o la finalità special-preventiva. A dimostrazione di tale potenzialità, si può citare il modello predittivo *Agatha*, il quale è in grado di formulare dei giudizi di valore per stabilire quale argomento sia preferibile nel caso concreto²⁶⁷. In futuro, l'IA potrebbe dunque avere un ruolo anche nel formulare giudizi di valore rilevanti ai fini dell'applicazione dei criteri finalistici.

Ai fini della presente disamina, occorre comunque concentrarsi sui criteri fattuali di commisurazione della pena, relativamente ai quali l'ipotesi del ricorso all'intelligenza artificiale sembra maggiormente concreta, anche alla luce dell'esperienza statunitense relativa alla previsione del rischio di recidiva al fine di determinare la pena.

L'intelligenza artificiale predittiva non potrebbe fornire particolare ausilio in merito all'accertamento sulla gravità del reato, ai sensi dell'art. 133, comma 1, c.p.p., non essendo un giudizio prognostico volto a valutare il possibile verificarsi di eventi futuri²⁶⁸.

Il giudizio sulla capacità a delinquere sembrerebbe invece avere natura prognostica, essendo invece proiettato verso il futuro, sebbene il punto non sia pacifico in dottrina. Secondo un primo orientamento, la capacità a delinquere, in una visione retrospettivo-retributiva, consiste nella personalità dell'autore rispecchiantesi nell'azione. Il giudizio sulla capacità a delinquere non riguarderebbe dunque il possibile compimento di fatti futuri²⁶⁹ e gli indici della capacità a delinquere si riferirebbero al grado di colpevolezza per il fatto compiuto. In definitiva, la capacità a delinquere guarderebbe al passato e non al futuro²⁷⁰.

Secondo un altro orientamento, l'istituto andrebbe letto in una prospettiva prognostico-preventiva, con riferimento all'attitudine del soggetto a commettere nuovi reati²⁷¹. Si tratta di un dibattito risalente, che, secondo autorevole dottrina, si può superare non tanto argomentando sul piano letterale o storico, bensì alla luce della finalità rieducativa della pena, ai sensi dell'art. 27, comma 3, Cost. Secondo questa impostazione, tra le possibili interpretazioni, prevale quella tesi che tiene conto sia delle esigenze di intimidazione/ammonimento del reo, che dell'opportunità di evitargli inutili rischi di desocializzazione e di aprirgli prospettive di reinserimento nella società²⁷². Il giudizio sulla capacità a delinquere si proietta al futuro, con la conseguenza che gli indici di capacità a delinquere, ai sensi dell'art. 133, comma 2, c.p., devono essere utilizzati ai fini di una prognosi sui futuri comportamenti del reo²⁷³.

²⁶⁷ T. BENCH CAPON -G.SARTOR, *A Model of Legal Reasoning with Cases Incorporating Theories and Values in Artificial Intelligence*, Amsterdam, Paesi Bassi, 2003, p. 99; K. ASHLEY, *Introducing AI & Law and Its Role in Future Legal Practice, Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, Cambridge, 2017, pp. 1-6.

²⁶⁸ Cfr. par. 2.3.2 sul perché, in materia di arresto facoltativo, il giudizio sulla gravità del reato, non presenta natura prognostica, a differenza del giudizio sulla pericolosità.

²⁶⁹ M. RONCO – B. ROMANO, *Codice penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub* art. 133 c. p.

²⁷⁰ B. PETROCELLI, *La funzione della pena* In *Saggi di diritto penale*, CEDAM, Padova, 1952, p. 124; G. BETTIOL *Diritto penale: parte generale*, 11°ed., CEDAM, Padova, 1982, p. 403.

²⁷¹ F. ANTOLISEI *La capacità a delinquere* in *Rivista italiana di diritto penale (RIDP)*, CEDAM, Padova, 1934, p. 68

²⁷² G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 803.

²⁷³ *Ibid.*, p. 803.

Secondo un'altra tesi minoritaria, il giudizio sulla capacità a delinquere presenterebbe natura bidimensionale, in quanto comprenderebbe sia la capacità morale di compiere il reato addebitato che l'attitudine a commettere nuovi illeciti futuri²⁷⁴.

La stessa giurisprudenza di legittimità, in maniera discordante, ha sostenuto, a seconda dei casi, ognuna delle tre tesi. In alcune sentenze, la Cassazione ha riconosciuto espressamente il carattere prognostico della capacità a delinquere²⁷⁵. In altre, ha invece affermato chiaramente che la capacità a delinquere riguarda il passato e non va confusa con la pericolosità sociale²⁷⁶. In altre ancora, la Cassazione ha talvolta attribuito alla capacità a delinquere sia la dimensione prognostico-preventiva che quella etico-retributiva²⁷⁷.

Il ricorso all'IA predittiva per supportare il giudizio sulla capacità a delinquere è incompatibile con la tesi che interpreta tale istituto secondo una visione retrospettivo-retributiva. Il rischio di recidiva individuato da un sistema IA non si riferisce infatti al grado di colpevolezza per il fatto compiuto, con la conseguenza che, in quest'ottica, non offrirebbe particolari indicazioni quanto alla capacità a delinquere.

Al contrario, non vi sarebbe tale incompatibilità con la tesi che interpreta la capacità a delinquere come attitudine del soggetto a commettere nuovi reati, in quanto tale impostazione riconosce chiaramente la natura prognostica di tale giudizio. Sembrerebbe non esserci una netta incompatibilità neppure con l'orientamento che sostiene la natura bidimensionale della capacità a delinquere, in quanto tale tesi ammette la rilevanza di un giudizio prognostico sull'attitudine a commettere nuovi illeciti, sebbene non sarebbe l'unico elemento da tenere conto nell'analizzare la capacità a delinquere del reo.

In conclusione, nell'ambito dei molteplici criteri fattuali e finalistici rilevanti ai fini della commisurazione della pena, le potenzialità dell'IA predittiva riguardano principalmente il giudizio sulla capacità a delinquere, a condizione che sia riconosciuta la natura prognostica di tale valutazione.

2.7.3) Gli indici della capacità a delinquere

In questo paragrafo, verranno illustrati gli indici della capacità a delinquere, ai sensi dell'art. 133, comma 2, c.p. L'intento è quello di ricostruire quali sarebbero i fattori di cui un sistema di IA dovrebbe idealmente tenere conto nel giudicare la capacità a delinquere in maniera conforme alle prescrizioni legislative.

In giurisprudenza è pacifico che non è richiesto motivare con riferimento a tutti i criteri indicati nell'art. 133 c.p., ma è sufficiente fare riferimento soltanto ai criteri che hanno avuto un ruolo determinante nella decisione²⁷⁸.

Un sistema di IA volto a prevedere la capacità a delinquere dovrebbe idealmente essere addestrato tenendo conto di tutti gli indici rilevanti. In concreto, però, solo alcuni di essi potrebbero risultare determinanti nel

²⁷⁴ P. NUVOLONE, *La capacità a delinquere nel sistema del diritto penale*, in *Trent'anni di diritto e procedura penale: studi*, vol. I, CEDAM, Padova, 1969, p. 578.

²⁷⁵ Cass. pen., Sez. I, 11 novembre 1998, n. 5555; Cass. pen., Sez. III, 15 novembre 1992, n. 2280.

²⁷⁶ Cass. pen., Sez. II, 29 ottobre 1981, n. 1580.

²⁷⁷ Cass. pen., Sez. II, 5 giugno 1990, n. 9572.

²⁷⁸ Cass. pen., Sez. I, 25 settembre 2013, n. 3155.

singolo caso, in maniera analoga a quanto avviene per il giudice, poiché anche quest'ultimo, ai sensi dell'art. 133 c.p., considera tutti i criteri previsti, pur limitandosi a motivare solo in relazione a quelli che si rivelano decisivi.

Sviluppare tali sistemi in base agli indici di cui all'art. 133 c.p. garantirebbe che il giudizio sulla capacità a delinquere sia, quanto più possibile, prossimo alle intenzioni del legislatore. Al contrario, non tenere conto di tali indici significherebbe di fatto eludere lo stesso art. 133 c.p.

Rimane però il problema della cosiddetta *black box*, giacché il giudice potrebbe non essere in grado di individuare i criteri che nel caso di specie sono risultati decisivi, con la conseguenza che non sarebbe nelle condizioni di valutare criticamente il *risk score* prodotto dall'IA²⁷⁹.

Ciò premesso, il primo indice previsto dalla norma in parola si riferisce ai “*motivi a delinquere*” e al “*carattere del reo*” (art. 133, comma 2, n. 1, c.p.). I primi devono essere analizzati in base alla loro intensità, all'idoneità a rimanere nel tempo e al valore etico-sociale. Ad esempio, un motivo che di regola produce comportamenti socialmente apprezzabili è indice di una minore capacità a delinquere, rispetto ad un motivo che è solito generare condotte riprovevoli²⁸⁰.

Il giudizio sul carattere del reo consiste in una valutazione complessiva delle componenti innate della personalità dell'agente²⁸¹. Si comprende dunque sia il dato biologico-ereditario che quello del carattere propriamente detto²⁸². La responsabilità penale di un soggetto determinatamente orientato a scelte di vita contrarie alla legalità risulta infatti maggiore di quella di un imputato dal carattere fragile e non ben strutturato, nonché dalle scarse capacità di autocontrollo²⁸³.

Il secondo indice si riferisce ai “*precedenti penali e giudiziari e, in genere, dalla condotta e dalla vita del reo, antecedenti al reato*”. L'espressione “precedenti penali e giudiziari”, dalla formulazione particolarmente ampia, include sia le condanne passate in giudicato che quelle definitive, nonché i procedimenti penali nella fase delle indagini preliminari²⁸⁴. In giurisprudenza, è stata riconosciuta rilevanza alle sentenze di proscioglimento per prescrizione, nonché alle sentenze civili, quali la dichiarazione di fallimento²⁸⁵. Anche lo strumento di valutazione del rischio *PSA* tiene conto, tra gli altri, dei precedenti giurisprudenziali, al fine di calcolare il rischio di un nuovo arresto (*New Criminal Arrest*). Nello specifico, si fa riferimento alle precedenti condanne, graduando la rilevanza della condanna in base alla gravità del reato e alla sua natura violenta o

²⁷⁹ Si vedano i parr. 3.2.1, sul problema della *black box* e 3.3.3, sul pericolo di fare eccessivo affidamento sul suggerimento offerto dalla macchina.

²⁸⁰ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 801.

²⁸¹ *Ibid.*

²⁸² M. ROMANO -G. GRASSO, *Commentario sistematico del Codice penale. Parte generale*, vol. II, Giuffrè, Milano, 2012, p. 355.

²⁸³ M. RONCO – B. ROMANO, *Codice penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub* art. 133.

²⁸⁴ *Ibid.*

²⁸⁵ Corte d'appello di Torino, Sez. II, 08 aprile 2009, n. 1213, in M. RONCO – B. ROMANO, *op. cit.*, *sub* art. 133.

meno²⁸⁶. È plausibile che, anche in Italia, un eventuale sistema di IA predittivo della capacità a delinquere terrebbe conto, in misura significativa, dei precedenti penali e giudiziari.

Alcuni dei fattori impiegati da *COMPAS* per calcolare il rischio di recidiva sono riconducibili alla condotta e alla vita del reo antecedente al reato: la precedente storia criminale, i primi segni di devianza minorile e la storia di inosservanza delle regole²⁸⁷. Il ricorso all'IA predittiva potrebbe dunque rendere maggiormente significativa la rilevanza della condotta e della vita del reo antecedente al reato in sede di commisurazione della pena.

Il terzo indice consiste nella “*condotta contemporanea o susseguente al reato*” (art. 133, comma 2, n.3, c.p.). Con condotta contemporanea, si intende qualunque comportamento tenuto nel periodo che intercorre dall'inizio dell'esecuzione del reato fino alla consumazione dello stesso²⁸⁸. Rientrano, invece, nella condotta susseguente le modificazioni apportate dal reo alle conseguenze del reato, l'atteggiamento serbato nei confronti della vittima, nonché la riparazione, anche parziale, del danno cagionato²⁸⁹.

Risulta poi particolarmente controversa la rilevanza del comportamento processuale dell'imputato. La Cassazione, in maniera ondivaga, in alcuni casi ha ritenuto illegittima la mancata concessione della sospensione condizionale motivata soltanto sulla base dell'atteggiamento non collaborativo in sede processuale²⁹⁰. In altri, ha sancito che dall'esercizio del diritto al silenzio o dalla menzogna, il giudice può trarre il convincimento di una personalità negativa²⁹¹.

Infine, il quarto indice concerne le “*condizioni di vita individuale, familiare e sociale del reo*” (art. 133, comma 2, n.4 c.p.). Si fa riferimento a fattori ambientali di criminogenesi quali, ad esempio, la famiglia disgregata, l'adesione a bande criminali giovanili, l'emarginazione, la disoccupazione o il tipo di educazione ricevuta²⁹².

L'IA potrebbe individuare un elevato rischio di recidiva in ragione del contesto di degrado in cui il soggetto si trova, sebbene risulti controverso se tale circostanza dovrebbe comportare un aumento o una riduzione della pena. In un'ottica generale o special preventiva, la degradazione dell'ambiente familiare o sociale dovrebbe rilevare nel senso di giustificare una pena maggiormente elevata, in quanto tali fattori suggeriscono un rischio maggiore che vengano commessi nuovi reati. Muovendo dalla finalità retributiva della pena, la delinquenza condizionata da un ambiente sociale degradato dovrebbe invece giustificare una sanzione ridotta, in quanto è espressiva di un minor grado di colpevolezza²⁹³.

²⁸⁶ Advancing Pretrial Policy & Research, *Public Safety Assessment – Risk factors*, AdvancingPretrial.org, <https://advancingpretrial.org/psa/factors/>, consultato il 14 luglio 2025.

²⁸⁷ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025, p. 27-28.

²⁸⁸ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 802.

²⁸⁹ M. RONCO – B. ROMANO, *op. cit.*, *sub* art. 133.

²⁹⁰ Cass. pen. Sez. III, 3 giugno 2014, n.38678.

²⁹¹ Cass. pen. Sez. I, 3 marzo 1994, n. 3819.

²⁹² G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 801.

²⁹³ M. RONCO -B. ROMANO, *op. cit.*, *sub* art. 133.

2.7.4) Valutazione dei presupposti della sospensione condizionale della pena

Rimanendo nell'ambito delle pene, un altro settore in cui l'IA predittiva potrebbe assumere un ruolo è quello relativo all'istituto della sospensione condizionale della pena, di cui agli artt. 163 e ss. c.p. È la più tradizionale e risalente misura alternativa alla detenzione, strumento pensato per la cosiddetta lotta alla pena detentiva breve. La *ratio* dell'istituto consiste nel rinunciare, sia pur in maniera condizionata, all'esecuzione della pena, dal momento che l'effetto criminogeno conseguente alla detenzione potrebbe risultare superiore ai benefici²⁹⁴. L'applicabilità della sospensione condizionale è subordinata alla presenza di diversi requisiti. In primo luogo, rileva l'ammontare della pena inflitta, essendo richiesto che la condanna alla reclusione o all'arresto non superi i due anni, oppure, i tre anni se l'imputato è minorenni, o ancora, i due anni e sei mesi se l'imputato ha un'età compresa tra i diciotto e i ventuno (art. 163, commi 1, 2, 3, c.p.).

In secondo luogo, non deve sussistere nessuna delle preclusioni soggettive previste dall'art. 164, commi 2, 4, c.p.: l'aver riportato una precedente condanna per delitto, l'essere stato dichiarato delinquente o contravventore abituale o professionale, o l'aver già beneficiato della sospensione due volte.

Infine, è richiesta la prognosi favorevole sul futuro comportamento del reo. Tale beneficio viene infatti concesso "*soltanto se, avuto riguardo alle circostanze indicate nell'art. 133, il giudice presume che il colpevole si asterrà dal commettere ulteriori reati*" (art. 164, primo comma, c.p.). La sospensione condizionale comporta l'estinzione della pena, se nei termini stabiliti, il reo non commette un delitto o una contravvenzione della stessa indole e rispetta gli obblighi stabiliti dal giudice (art. 167 c.p.).

L'IA predittiva potrebbe supportare il giudice ai fini del terzo presupposto, ossia la prognosi favorevole sul futuro comportamento del reo. Dalla natura prognostica di tale giudizio emergono le potenzialità dell'IA come strumento a supporto della decisione giudiziale, che, altrimenti, rischia di essere particolarmente incerta.

Si tratta di un'ipotesi particolarmente affine al caso *Loomis*. In tale sentenza, il giudice di primo grado ha fatto riferimento al *risk score* prodotto da *COMPAS* per escludere l'applicabilità della cosiddetta *probation* nel caso di specie. La *probation* consiste, in sostanza, nel rilasciare il reo sotto la supervisione delle autorità competenti purché vengano rispettate determinate condizioni²⁹⁵.

La prognosi sul futuro comportamento del reo si formula in base ai criteri di cui all'art. 133 c.p., in maniera analoga alla pericolosità sociale, di cui all'art. 203 c.p. Tra i criteri ex art. 133 c.p., la giurisprudenza di legittimità ha valorizzato soprattutto gli indici della capacità a delinquere²⁹⁶. Ciò non stupisce in quanto, se si concepisce la capacità a delinquere in un'ottica prognostico-preventiva, si spiega come i relativi indici possano fondare anche la prognosi giudiziale sul futuro comportamento del reo di cui all'art. 164, comma 1, c.p. In

²⁹⁴ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, pp. 856-857.

²⁹⁵ R. GRAY, *Probation: An Exploration in Meaning*, National Criminal Justice Reference Service, Washington D.C., 1988, <https://www.ojp.gov/ncjrs/virtual-library/abstracts/probation-exploration-meaning>, consultato il 19 giugno 2025, p. 26.

²⁹⁶ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 861; Cass. pen., Sez. III, 6 luglio 2016, n. 42737.

particolare, la Cassazione ha valorizzato l'atteggiamento di disobbedienza desumibile dalla vicenda processuale²⁹⁷, i precedenti giudiziari, anche non definitivi²⁹⁸, e i precedenti di polizia²⁹⁹.

In dottrina, è stato evidenziato che, nella prassi, il beneficio è stato concesso in maniera quasi automatica. La ragione è che l'art. 133 c.p., nell'individuare gli indici della capacità a delinquere, non fornisce indicazioni univoche. Tale disposizione non è dunque sufficiente a guidare in modo convincente il potere discrezionale del giudice³⁰⁰, con la conseguenza che la prognosi sul futuro comportamento del reo non presenta la capacità selettiva funzionale a garantire la dinamica regola/eccezione che dovrebbe rimanere alla base dell'istituto³⁰¹.

Il ricorso all'IA potrebbe garantire una maggiore capacità selettiva della prognosi sul futuro comportamento del reo. In presenza di un elevato rischio di recidiva stimato dal modello predittivo, il giudice sarebbe tenuto a valutare con particolare cautela se concedere o meno la sospensione condizionale, evitando l'automatismo che talvolta si verifica nella prassi. Si riserverebbe così la sospensione condizionale agli imputati che, effettivamente, non presentano una significativa capacità a delinquere.

Il riferimento all'IA potrebbe altresì rafforzare la motivazione giudiziale affinché non risulti eccessivamente evanescente, dato che i criteri di cui all'art. 133 c.p. non forniscono indicazioni univoche su come accertare la capacità a delinquere nel caso concreto. La Cassazione richiede che il giudice spieghi le proprie scelte nella motivazione della sentenza, sia pur limitandosi a fare riferimento, tra i criteri di cui all'art. 133 c.p., agli indici considerati prevalenti³⁰². Qualora siano presenti nel caso concreto sia elementi favorevoli che ostativi, e il giudice consideri prevalenti i secondi, nella motivazione occorre indicare gli elementi favorevoli e spiegare perché nel caso di specie non sono prevalenti³⁰³.

L'IA predittiva rafforzerebbe la motivazione giudiziale soprattutto nel caso in cui il giudice decidesse di discostarsi dalla valutazione algoritmica. In tale circostanza, il giudice dovrebbe indicare le ragioni per le quali ritiene che la prognosi formulata dalla macchina non sia corretta, rendendo quindi la motivazione maggiormente articolata. Qualora, invece, il giudice dovesse seguire il suggerimento offerto dall'IA, sarebbe necessario che venga comunque fornita una valutazione autonoma e che l'obbligo motivazionale non si risolva in un mero rinvio all'*output* prodotto dall'IA. Bisogna evitare che l'automatismo presente nell'attuale prassi giurisprudenziale, che porta a concedere quasi sempre la sospensione condizionale, venga sostituito da una tendenza altrettanto automatica a seguire l'IA.

In conclusione, l'istituto della sospensione condizionale della pena potrebbe essere supportato dall'intelligenza artificiale, coadiuvando il giudice nella formulazione della prognosi giudiziale sul futuro comportamento del reo, rendendo maggiormente selettiva la concessione del beneficio e rafforzando la motivazione giudiziale.

²⁹⁷ Cass. pen., Sez. III, 3 giugno 2014, n. 38678.

²⁹⁸ Cass. pen. Sez. III, 30 settembre 2015, n.44458.

²⁹⁹ Cass. pen. Sez. II, 5 maggio 2010, n. 18189.

³⁰⁰ P. VENEZIANI, *La punibilità e le conseguenze giuridiche del reato* in C. F. GROSSO -T. PADOVANI -A. PAGLIARO, *Trattato di diritto penale*, Giuffrè, Milano, 2014, p. 347.

³⁰¹ A. MARTINI, *La pena sospesa* in *Itinerari di diritto penale*, Giappichelli, 2001, p. 208.

³⁰² Cass. pen., Sez. V., 7 febbraio 2020, n. 17953.

³⁰³ Cass. pen. Sez. III, 6 luglio 2016, n. 42737.

2.8) L'IA al servizio del PM: selezione delle notizie di reato ed esercizio dell'azione penale

2.8.1) Introduzione

L'IA predittiva presenta delle potenzialità significative non soltanto nel supportare le decisioni giudiziali ma anche nell'assistere l'attività del pubblico ministero (PM). L'art. 112 Cost. sancisce il fondamentale principio secondo cui il PM ha l'obbligo di esercitare l'azione penale. Nei fatti, l'attività del PM presenta comunque una discrezionalità significativa riguardo al se, al come e al quando procedere³⁰⁴. Tale discrezionalità non sussiste soltanto nella circostanza patologica del sottofinanziamento delle attività investigative rispetto al carico di lavoro. Ricorre piuttosto quella discrezionalità propria di ogni attività di governo in cui, data la finitezza delle risorse, occorre fisiologicamente ponderare esigenze contrastanti³⁰⁵.

L'intelligenza artificiale potrebbe supportare il pubblico ministero nell'esercizio di tale discrezionalità, in modo tale che le valutazioni prognostiche spettanti al PM abbiano una base statistica maggiormente solida. Si possono individuare due ambiti di particolare interesse: la selezione di notizie di reato da trattare con precedenza rispetto alle altre (art. 1, comma 6, lett. b), d.lgs. 20 febbraio 2006, n. 106) e l'esercizio dell'azione penale (art. 50, comma 1 e art. 408 c.p.p.).

Entrambi gli ambiti sono stati recentemente disciplinati dal legislatore, rispettivamente con la legge 17 giugno 2022 e con il d.lgs. 10 ottobre 2022, n. 150, nell'ambito della Riforma Cartabia. L'obiettivo è quello di rendere il processo penale maggiormente efficiente. Il legislatore ha introdotto dei criteri di priorità a cui il PM si deve conformare nella trattazione delle notizie di reato e nell'esercizio dell'azione penale (art. 3-*bis* disp. att. c.p.p.). Alla base di tale norma, vi è la tesi secondo cui l'art. 112 Cost. tutela l'interesse alla persecuzione dei reati in modo non assoluto ma relativo, con la conseguenza che l'obbligatorietà dell'azione penale deve essere bilanciata con la necessità di avere un processo efficiente, al fine di garantirne la ragionevole durata (art. 111, comma 2, Cost.)³⁰⁶. La necessità di bilanciare l'obbligatorietà dell'azione penale con l'efficienza processuale, potrebbe giustificare anche il ricorso all'intelligenza artificiale, in quanto l'IA potrebbe razionalizzare l'attività delle procure³⁰⁷.

L'implementazione dell'IA in questo settore sarebbe poi conforme all'impegno che l'Italia ha assunto con il Piano Nazionale di Ripresa e Resilienza (PNRR) di ridurre del 25% i tempi del processo penale nei prossimi cinque anni. Vi è un rapporto inscindibile tra l'efficienza del processo e la sua ragionevole durata, poiché

³⁰⁴ V. GRAMUGLIA, *Obbligatorietà dell'azione penale e ruolo dell'IA nelle scelte del pubblico ministero*, in *Archivio Penale*, Giuffrè Francis Lefebvre, Milano, 2024, <https://archiviopenale.it/obbligatorieta-dellazione-penale-e-ruolo-dellia-nelle-scelte-del-pubblico-ministero/articoli/46820>, consultato il 25 settembre 2025, p. 2.

³⁰⁵ O. DOMINIONI, *Interesse alla persecuzione penale e irrilevanza sociale del fatto nel prisma dell'efficienza giudiziaria*, in *La giustizia penale*, 2010, p. 420.

³⁰⁶ V. GRAMUGLIA, *op. cit.* p. 4; O. DOMINIONI, *Azione obbligatoria penale e efficienza giudiziaria*, in *Diritto penale e procedura penale*, 2020, p. 872.

³⁰⁷ Si veda il par. 3.9.

senza l'una non è possibile raggiungere l'altra³⁰⁸. Occorre altresì ricordare che l'Italia è il primo paese al mondo per numero di condanne per violazione dell'art. 6 della CEDU, in cui si prevede che *“ogni persona ha diritto a che la sua causa sia esaminata equamente, pubblicamente ed entro un termine ragionevole relativamente alla durata dei processi”*. L'Italia è stata complessivamente oggetto di 1202 condanne dovute all'eccessiva durata dei processi³⁰⁹.

Il ricorso all'IA per supportare le scelte discrezionali del pubblico ministero potrebbe, quindi, essere uno strumento per rendere il processo penale maggiormente efficiente. In questo modo, sarebbe possibile ottemperare sia agli obblighi assunti in sede Euro-unitaria nel contesto del PNNR che alle condanne disposte dalla Corte EDU in merito alla durata dei processi. Utilizzare l'IA in questi ambiti potrebbe limitare il dispendio di risorse nell'ambito di procedimenti in cui non è possibile formulare una ragionevole previsione di condanna. Tale risparmio di risorse si tradurrebbe indirettamente in una riduzione dei tempi processuali.

2.8.2) Applicazione dell'IA per selezionare le notizie di reato da trattare con precedenza

In primo luogo, l'IA potrebbe suggerire quali notizie di reato trattare con precedenza. Nel contesto della Riforma Cartabia, il legislatore ha dovuto fare i conti con l'enorme quantità di notizie di reato che pervengono quotidianamente alle procure della Repubblica, con la conseguenza che le procure sono, di fatto, tenute ad individuare quali notizie di reato devono essere portate avanti in via prioritaria rispetto alle altre³¹⁰.

Sono dunque stati modificati i commi 6 e 7 dell'art. 1 del d.lgs. 20 febbraio 2006, n. 106, relativo all'organizzazione dell'ufficio del pubblico ministero. Il procuratore della Repubblica è tenuto a predisporre il progetto organizzativo dell'ufficio, con il quale determina i *“criteri di priorità finalizzati a selezionare le notizie di reato da trattare con precedenza rispetto alle altre”* (comma 6, lett. b). Tali criteri devono essere definiti *“nell'ambito di principi generali indicati dal Parlamento con legge”*. A monte, vengono predeterminati dei criteri generali dai rappresentanti dei cittadini, mentre, a valle, le scelte compiute sono controllabili, giacché devono tenere conto dei criteri sanciti dalle stesse procure³¹¹.

Per assicurare che tali criteri siano conformi alle specifiche esigenze delle singole procure, lo stesso comma 6 richiede di tenere conto del *“numero di affari da trattare, della specifica realtà criminale territoriale e dell'utilizzo efficiente delle risorse tecnologiche, umane e finanziarie disponibili”*. Il ricorso all'IA predittiva per supportare la selezione delle notizie di reato da trattare con precedenza potrebbe essere un esempio di

³⁰⁸ G. L. GATTA, *Riforma della giustizia penale. Contesto, obiettivi e linee di fondo della “legge Cartabia”*, in *Sistema Penale*, <https://www.sistemapenale.it/it/scheda/gatta-legge-2021-134-delega-riforma-giustizia-penale-cartabia>, consultato il 20 settembre 2025 pp. 1-3; G. VARRASO, *La legge Cartabia e il rapporto dei procedimenti speciali al recupero dell'efficienza processuale in Sistema Penale*, <https://www.sistemapenale.it/it/articolo/varraso-legge-cartabia-procedimenti-speciali>, consultato il 20 settembre 2025, pp. 29-32.

³⁰⁹ EUROPEAN COURT OF HUMAN RIGHTS, *Violations by Article & by State*, 2020, https://www.echr.coe.int/documents/d/echr/Stats_violation_1959_2020_ENG, consultato il 20 settembre 2025.

³¹⁰ P. TONINI -C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, pp. 673-674.

³¹¹ M. GIALUZ, *Per un processo penale più efficiente e giusto. Guida alla lettura della riforma Cartabia*. in *Sistema penale*, <https://www.sistemapenale.it/it/scheda/gialuz-per-un-processo-piu-efficiente-e-giusto-guida-alla-lettura-della-riforma-cartabia>, consultato il 26 settembre 2025, pp. 42-43.

utilizzo efficiente delle risorse tecnologiche. Ciò conferma che l'impiego dell'IA in questo settore sarebbe in continuità con l'obiettivo perseguito dalla Riforma Cartabia di rendere maggiormente efficiente il processo penale.

Un altro obiettivo della Riforma è quello di evitare che le scelte di politica repressiva portate avanti dalle procure vengano fatte in maniera non trasparente e autonoma³¹². In quest'ottica, il ricorso all'IA potrebbe ridurre ulteriormente la discrezionalità delle procure. Il PM, nel valutare le notizie di reato da trattare, dovrebbe tenere conto non soltanto dei criteri di priorità definiti nel progetto organizzativo, ma anche dei suggerimenti offerti dalla macchina. Le valutazioni compiute dal pubblico ministero risulterebbero così maggiormente affidabili e meno incerte, in quanto meno influenzate da valutazioni soggettive. L'introduzione dell'IA in questo ambito sarebbe dunque in sintonia con l'obiettivo, già perseguito dal legislatore con la Riforma Cartabia, di ridurre il rischio che le scelte del PM risultino arbitrarie.

Resterebbe comunque salva la possibilità di discostarsi dal suggerimento offerto dal sistema di IA. Si dovrebbe idealmente realizzare un'interazione uomo-macchina virtuosa, che stimoli le procure a valutare criticamente i suggerimenti delle macchine, senza appiattirsi a questi ultimi. Non si può sottovalutare il rischio che le procure, a causa dell'elevato carico di lavoro, si limitino a seguire in maniera acritica il suggerimento offerto dal sistema di IA, senza valutare attentamente le notizie di reato nel merito.

Impiegare l'IA in questo settore sarebbe conforme anche all'obiettivo di efficienza processuale perseguito dalla stessa Riforma Cartabia. Le procure potrebbero individuare, con l'ausilio di tali sistemi, le notizie di reato che presentano maggiori probabilità di portare ad una condanna definitiva, al fine di ottimizzare la gestione delle risorse.

Valutare, anche con l'ausilio dell'IA, quali notizie di reato presentano maggiori probabilità di sfociare in una condanna non consisterebbe in una valutazione anticipata di colpevolezza. L'IA potrebbe fornire un valido ausilio sul piano meramente probabilistico e organizzativo, con riferimento a quei procedimenti che, già sulla base degli elementi inizialmente disponibili (tipologia di reato, quadro indiziario, precedenti giurisprudenziali, standard probatori richiesti), appaiono meritevoli di un più intenso e tempestivo approfondimento investigativo.

Non si tratta neppure di rinunciare a svolgere l'indagine quando il reato presenta una minore probabilità di condanna, in quanto, altrimenti, la mancata indagine si risolverebbe in una negazione dell'obbligatorietà dell'azione penale. L'ipotesi in esame consiste piuttosto nel selezionare determinate notizie di reato da trattare con precedenza rispetto alle altre, in maniera analoga a quanto previsto dai commi 6 e 7 dell'art. 1 del d.lgs. 20 febbraio 2006, n. 106. L'indagine andrebbe quindi aperta in ogni caso, al fine di garantire la compatibilità con il principio di obbligatorietà dell'azione penale. Ad esempio, occorrerebbe comunque iscrivere la notizia di reato nell'apposito registro a norma dell'art. 335 c.p.p.

³¹² *Ibid.*

Il ricorso all'IA in questo ambito si differenzia quindi nettamente rispetto alla scelta tra archiviazione ed esercizio dell'azione penale, in quanto, nell'ipotesi qui considerata, l'IA sarebbe meramente funzionale a garantire un'ottimale gestione delle risorse e non rileverebbe né ai fini della chiusura del procedimento penale, né in merito al rinvio al giudizio. Dinanzi alla mera notizia di reato, in assenza di una vera e propria istruttoria, sarebbe chiaramente prematuro decidere se archiviare o meno. La notizia di reato risulta però già sufficiente ai fini di un giudizio di priorità quanto alle notizie da trattare con precedenza, in ragione del criterio della ragionevole previsione di condanna, ai sensi dell'art. 408 c.p.p.

Si potrebbe obiettare che tale criterio rileva soltanto dopo la conclusione delle indagini e non in sede di selezione delle notizie di reato da trattare. La conseguenza sarebbe che, in tale sede, il PM deve tenere conto soltanto dei criteri di priorità definiti dal progetto organizzativo e non anche della ragionevole previsione di condanna. Non sarebbe dunque legittimo trattare in via prioritaria le notizie di reato che, all'esito della stima resa dall'IA, presentano maggiori probabilità di portare ad una condanna, in quanto tale valutazione rilevarebbe solo in sede di esercizio dell'azione penale.

Questa impostazione pare in realtà eccessivamente formalistica, poiché, come sostenuto in dottrina, nella scelta operata dal legislatore con l'art. 408 c.p.p., è contenuto, sia pure in forma implicita, un criterio che dovrebbe investire, a monte, la stessa notizia di reato³¹³. Il PM potrebbe dunque legittimamente trattare in via prioritaria quelle notizie che, statisticamente, hanno maggiore probabilità di sfociare in una condanna.

In conclusione, l'IA potrebbe essere utilizzata per supportare la selezione delle notizie di reato da trattare con precedenza, con il duplice obiettivo di ridurre la discrezionalità del PM e rendere il processo penale maggiormente efficiente³¹⁴.

2.8.3) Applicazione dell'IA nei presupposti per l'esercizio dell'azione penale

L'intelligenza artificiale potrebbe poi essere impiegata per supportare il pubblico ministero nella scelta tra esercizio dell'azione penale e archiviazione. Ai sensi dell'art. 50, comma 1, il PM "*esercita l'azione penale quando non sussistono i presupposti per l'archiviazione*". La scelta del codice di procedura penale del 1988 è stata quella di non indicare espressamente quale è la quantità degli elementi probatori che sono necessari affinché il PM sia tenuto a formulare l'imputazione. Tali criteri si devono ricavare indirettamente dalla regola di giudizio che il pubblico ministero è tenuto ad osservare in sede di richiesta di archiviazione³¹⁵.

Prima della Riforma Cartabia, l'art. 125 disp. att. c.p.p. disponeva che "*il pubblico ministero presenta al giudice la richiesta di archiviazione quando ritiene infondata la notizia di reato, perché gli elementi acquisiti non sono idonei a sostenere l'accusa in giudizio*". Tale disposizione non aveva chiaramente superato le interpretazioni tradizionali che ricollegavano la nascita dell'obbligo di agire al criterio del *favor actionis*, secondo il quale, in caso di dubbio, occorreva esercitare l'azione penale.

³¹³ V. GRAMUGLIA, *op. cit.* pp. 28-29.

³¹⁴ Per l'analisi delle possibili criticità, con particolare riferimento all'obbligatorietà dell'azione penale, si veda il par. 3.9.

³¹⁵ P. TONINI -C. CONTI, *op. cit.* p. 671.

La Riforma Cartabia supera invece chiaramente tale criterio, abrogando l'art. 125 disp. att. c.p.p. e stabilendo che il PM è tenuto a chiedere l'archiviazione quando *“gli elementi acquisiti nel corso delle indagini preliminari non consentono di formulare una ragionevole previsione di condanna o di applicazione di una misura di sicurezza diversa dalla confisca”*. La *ratio* della norma è duplice: per un verso si vuole ridurre il numero di giudizi che vengono celebrati, per l'altro, l'obiettivo è quello di ridurre il numero di assoluzioni in dibattimento³¹⁶.

La Riforma Cartabia chiarisce poi la natura prognostica del giudizio che il PM è tenuto ad effettuare. Tale aspetto non era altrettanto chiaro nella precedente formulazione che faceva riferimento all'idoneità degli elementi acquisiti nelle indagini preliminari a sostenere l'accusa in giudizio. Trattandosi di un giudizio prognostico, l'IA potrebbe aiutare il pubblico ministero fornendo delle previsioni quanto al più probabile esito del giudizio.

Inoltre, l'introduzione dell'intelligenza artificiale nel processo risulta in continuità con l'obiettivo di efficienza perseguito dalla riforma. L'IA potrebbe fornire al PM una base statistica maggiormente solida in base alla quale effettuare tale prognosi. Il risultato potrebbe essere quello, perseguito dalla stessa Riforma Cartabia, di abbassare il tasso di esiti liberatori in dibattimento, in quanto si eserciterebbe l'azione penale soltanto in presenza di una significativa probabilità statistica di condanna. Anche in questo caso, si avrebbe come effetto positivo indiretto l'evitare un inutile dispendio di risorse processuali e quindi contribuire alla ragionevole durata del processo (art. 11, comma 2, Cost.).

La suddetta Riforma è intervenuta anche nel definire i criteri di priorità nell'esercizio dell'azione penale. Il nuovo art. 3-*bis* disp. att. c.p.p. ha stabilito che, non soltanto in sede di trattazione delle notizie di reato, ma anche in sede di esercizio dell'azione penale, il pubblico ministero si conforma ai criteri di priorità contenuti nel progetto organizzativo dell'ufficio³¹⁷. Il suggerimento offerto dall'intelligenza artificiale costituirebbe allora un ulteriore elemento che, insieme ai criteri di priorità, contribuirebbe a restringere la discrezionalità del PM.

L'uso dell'IA potrebbe supportare il PM nell'orientarsi in un giudizio, come quello *ex art. 408 c.p.p.*, caratterizzato da molteplici profili di incertezza. L'obbligatorietà dell'azione penale, ai sensi dell'art. 112 Cost., non significa che il PM non eserciti discrezionalità in tale sede. Il principio costituzionale si riferisce alla categoria della discrezionalità vincolata, che si caratterizza per la sussistenza di presupposti di natura valutativa al cui accertamento consegue un solo esito decisorio³¹⁸. La sussistenza di un presupposto valutativo, in questo caso la ragionevole previsione di condanna, implica che l'esercizio dell'azione penale non è estraneo a profili di incertezza, non essendo possibile escludere tale aspetto in ragione dell'art. 112 Cost.

In secondo luogo, l'incertezza emerge dalle prassi distorsive della funzione accusatoria nelle quali risulta evidente come l'accusa goda in realtà di significativi margini di discrezionalità circa la strategia investigativa

³¹⁶ M. GIALUZ, *op. cit.* pp. 44-45.

³¹⁷ Cfr. par. 2.7.2.

³¹⁸ F. CORDERO, *Le situazioni soggettive nel processo penale*, Torino, Giappichelli, 1956, pp.158 e ss.

da adottare. Si tratta di prassi patologiche, in cui la discrezionalità vincolata si traduce indebitamente in un giudizio di mera opportunità³¹⁹, come nel caso in cui l'accusa eserciti l'azione penale senza aver svolto previamente le indagini necessarie. La Corte costituzionale ha sancito il principio della completezza delle indagini al fine di contrastare tale prassi³²⁰. L'accusa gode quindi di un margine significativo di discrezionalità nel determinare quando sia necessario procedere ad ulteriori indagini e quando, invece, occorre esercitare l'azione penale, in quanto gli elementi raccolti consentono di formulare una ragionevole previsione di condanna.

Un ulteriore motivo di incertezza si deve alla circostanza che le valutazioni dei PM possono essere viziate dal cosiddetto *confirmation bias* o pregiudizio di conferma: la tendenza a cercare o interpretare le prove di un dato in base a convinzioni preesistenti, anziché esaminare criticamente tali convinzioni valutando anche elementi che potrebbero mostrarne la falsità³²¹. I PM potrebbero essere tentati di cercare esclusivamente elementi che confermino la tesi accusatoria, anziché individuare prove che la confutino. Cercare di confutare la tesi di cui si è convinti risulta infatti maggiormente dispendioso da un punto di vista cognitivo che tentare di confermarla³²².

In definitiva, per le ragioni sopra esposte, la prognosi sulla ragionevole previsione di condanna presenta dei profili di incertezza in presenza della quale l'IA potrebbe offrire un valido supporto. Per quanto l'IA non potrebbe eliminare completamente tale incertezza, potrebbe comunque offrire degli elementi utili per rendere tale giudizio maggiormente affidabile.

Si pensi al caso in cui l'accusa sia in procinto di esercitare l'azione penale pur non avendo svolto le indagini in maniera completa. Un sistema di IA potrebbe rilevare che, allo stato degli atti, il processo presenta una probabilità elevata di concludersi in un'assoluzione. Tale circostanza potrebbe suggerire al PM che gli elementi raccolti non sono sufficienti e che occorre dunque svolgere ulteriori accertamenti.

Oppure, un sistema di IA che preveda l'assoluzione come probabile esito del giudizio, potrebbe allarmare l'accusa circa la possibilità che le indagini siano viziate dal pregiudizio di conferma (*confirmation bias*), portando la stessa a rivedere in maniera critica gli elementi raccolti.

In concreto, i pubblici ministeri potrebbero avvalersi di algoritmi di giustizia predittiva che forniscono delle previsioni su base statistica sul probabile esito di decisioni giudiziarie future³²³. Tra gli esempi vi sono i

³¹⁹ C. VALENTINI *Obbligatorietà dell'azione penale, patologie della prassi e mancanza di controlli*, in *Rivista di diritto processuale*, 2020, pp. 1023 e ss., <https://ricerca.unich.it/handle/11564/729923?mode=complete>, consultato il 22 settembre 2025; V. GRAMUGLIA, *op. cit.*, p. 9.

³²⁰ Corte cost., 15 febbraio 1991, n. 88.

³²¹ G. GULOTTA -P. EGNOLETTI -B. NICCOLAI *et. al.*, *Tendenze generali e personali ai bias cognitivi e la loro ricaduta in campo forense: fondamenti e rimedi*, in *Sistema Penale*, 2021, <https://www.sistemapenale.it/it/articolo/gulotta-et-al-bias-cognitivi-campo-forense>, consultato il 22 settembre 2025, p. 6.

³²² V. GRAMUGLIA, *op. cit.*, pp. 11-14.

³²³ K. D. ASHLEY, *Models for Predicting Legal Outcomes in Artificial Intelligence and Legal Analytics*, Cambridge University Press, Cambridge, Regno Unito, 2017, pp. 107-126.

numerosi modelli predittivi sperimentati per calcolare l'esito del giudizio davanti alla CEDU³²⁴ o l'esito dei ricorsi proposti davanti alla Corte Suprema degli Stati Uniti d'America³²⁵. Questi sistemi di IA sono destinati ad offrire un supporto estremamente significativo ai professionisti, inclusi magistrati e avvocati³²⁶. Gli avvocati potrebbero avvalersi di tali previsioni per adottare di conseguenza la migliore strategia difensiva. I pubblici ministeri potrebbero invece tenere conto del probabile esito del giudizio per valutare se esercitare o meno l'azione penale.

Peraltro, occorre evidenziare che i principali algoritmi di giustizia predittiva attualmente esistenti prevedono l'esito delle sentenze di corti di ultima istanza, come la Corte europea dei diritti dell'uomo o la Corte Suprema degli Stati Uniti, negli esempi precedentemente citati. Vi sono diverse ragioni che spiegano questa circostanza. Le sentenze emesse da tali corti, a differenza di quelle delle corti inferiori, sono più strutturate e, nel caso delle corti di legittimità, tali sentenze risentono in misura inferiore delle circostanze del caso concreto, con l'effetto di rendere le relative previsioni più stabili.

Per introdurre l'intelligenza artificiale in questo ambito occorrerebbe allora verificare, sul piano tecnico, la possibilità di sviluppare anche algoritmi predittivi delle sentenze di primo o secondo grado. Sembrerebbe invece che, allo stato dell'arte, gli algoritmi di giustizia predittiva riguardano principalmente corti di ultima istanza.

In conclusione, l'IA predittiva potrebbe supportare il pubblico ministero nella decisione tra esercizio dell'azione penale e archiviazione, per rendere più affidabile la valutazione della ragionevole previsione di condanna e ridurre l'incertezza derivante da discrezionalità e *bias* cognitivi. Si potrebbe fare ricorso ad algoritmi di giustizia predittiva, sebbene sia ancora da verificare l'accuratezza di tali sistemi quando si tratta di prevedere l'esito di sentenze non emesse da corti supreme.

2.9) L'applicazione dell'IA nell'accertamento del nesso di causalità

In questo capitolo sono stati esaminati i possibili campi applicativi dell'intelligenza artificiale riconducibili a giudizi aventi natura prognostica. In questo paragrafo, verrà invece analizzata una possibile applicazione dell'IA che, a differenza dei casi precedentemente illustrati, riguarda l'accertamento di un fatto verificatosi nel passato e non un giudizio prognostico.

Si tratta del nesso di causalità *ex art. 40, comma 1, c.p.*, in base al quale *“nessuno può essere punito per un fatto preveduto dalla legge come reato se l'evento dannoso o pericoloso da cui dipende l'esistenza del reato non è conseguenza della sua azione od omissione”*. Al comma 2 dello stesso articolo viene invece disciplinata

³²⁴ N. ALETRAS -D. TSARAPANIS -D. PREOTIUC, *et al.*, *Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective*, in *PeerJ Computer Science* 2, 2016, <https://doi.org/10.7717/peerj-cs.93>, consultato il 22 settembre 2025, pp. 1-3.

³²⁵ D. M. KATZ -M.J. BOMMARITO -J. BLACKMAN, *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, in *PLOS ONE* 12, 2017, <https://doi.org/10.1371/journal.pone.0174698>, consultato il 22 settembre 2025.

³²⁶ COMMISSIONE PER L'EFFICIENZA E PER LA GIUSTIZIA DEL CONSIGLIO D'EUROPA (CEPEJ), *Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, Consiglio d'Europa, 2018, p. 42.

la cosiddetta causalità omissiva, stabilendo che “*non impedire un evento, che si ha l’obbligo giuridico di impedire, equivale a cagionarlo*”³²⁷.

L’IA potrebbe fornire delle informazioni statistiche su ciò che avviene nella maggior parte dei casi analoghi e potrebbe essere possibile cogliere quali sono i fattori che determinano l’evento e valutare, sia pur in termini statistici, l’incidenza di ciascuno di essi³²⁸.

Queste informazioni sono utili soprattutto nell’accertamento della causalità omissiva. In tale accertamento occorre verificare sia la causalità reale che la causalità ipotetica. La causalità reale significa che la causa A ha cagionato l’evento B, mentre la causalità ipotetica significa che una condotta doverosa dell’agente avrebbe impedito l’evento³²⁹. Ad esempio, qualora un assistente bagnanti venga considerato responsabile per la morte di un bagnante, non basterà accertare che tale evento sia dovuto all’annegamento (causalità reale). Occorrerà altresì dimostrare che le operazioni salvataggio che l’assistente bagnanti avrebbe dovuto adottare (condotta doverosa) sarebbero state idonee ad impedire l’evento morte (causalità ipotetica).

Nell’accertamento della causalità ipotetica, l’IA potrebbe fornire delle informazioni di natura statistica, utili a stabilire l’idoneità della condotta doverosa ad impedire l’evento. Si pensi al caso in cui si debba verificare se la morte di un paziente sia stata causata dalla mancata somministrazione di un farmaco. L’IA potrebbe fornire delle informazioni utili per comprendere se, statisticamente, dalla mancata somministrazione del farmaco consegue la morte del paziente.

Tuttavia, l’intelligenza artificiale presenta dei limiti nell’accertare il nesso di causalità quando si basa sull’apprendimento statistico, anziché su algoritmi deterministici, come di solito avviene nell’ipotesi del *machine learning*. In questo caso, il sistema di IA formula l’obiettivo in termini di inferenza statistica. Un esempio di apprendimento statistico è il *Bayesian Learning*, il quale si basa sull’osservazione dell’ambiente circostante, al fine di aggiornare continuamente la plausibilità delle ipotesi e migliorare l’accuratezza delle previsioni³³⁰. L’apprendimento statistico è caratterizzato dall’incertezza propria del metodo induttivo, il quale, a differenza del metodo deduttivo, permette di giungere a conclusioni non già certe, bensì meramente probabili³³¹.

La Corte di cassazione, nella nota sentenza *Franzese*³³², ha superato il precedente orientamento secondo cui il giudice può affermare il rapporto di causalità se ha accertato con probabilità (statistica) vicina alla certezza (cioè, vicina al 100%) che quella condotta, azione od omissione è stata causa necessaria del verificarsi

³²⁷ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022, p. 299.

³²⁸ R. BLAIOTTA *Giustizia, errore, intelligenza artificiale*, in *Sistema Penale*, 2020, <https://www.sistemapenale.it/it/articolo/blaiotta-giustizia-errore-intelligenza-artificiale>, p. 13.

³²⁹ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 300.

³³⁰ S. RUSSELL -P. NORVIG, *Artificial Intelligence: A Modern Approach*, 3ª ed., Prentice Hall, Upper Saddle River (NJ), 2010, pp. 802-829.

³³¹ G. BUONOMO, *Macchine intelligenti, processo induttivo e apprendimento autonomo*, in G. BUONOMO -G. CIACCI -A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, p. 28. V. par. 1.1.

³³² Cass. pen., Sez. Un., 11 luglio 2002, n. 30328.

dell'evento³³³. Nel caso *Franzese*, si stabilisce invece che anche coefficienti medio-bassi di probabilità statistica possono essere sufficienti a provare il nesso di causalità, qualora risulti provata la non incidenza nel caso di specie di fattori interagenti in via alternativa. Viceversa, anche in caso di livelli elevati di probabilità statistica, il giudice è comunque tenuto a verificare, nel caso concreto, l'irrelevanza di spiegazioni alternative. Dunque, alla probabilità statistica si affianca, come ulteriore requisito, la probabilità logica, intesa come verifica dell'assenza di altri fattori causali in grado di spiegare l'evento³³⁴.

Nel caso dell'intelligenza artificiale, queste considerazioni implicano che la probabilità statistica, individuata da un modello di *machine learning*, non è di per sé sufficiente a provare il nesso di causalità e quindi, a fondare la responsabilità penale, essendo necessario verificare che non vi siano altre possibili cause degli eventi. La probabilità logica è probabilità baconiana, richiedente cioè un apprezzamento qualitativo. L'IA è invece affine al concetto di probabilità pascaliana, ossia basata su coefficienti numerici. I coefficienti numerici prodotti dall'intelligenza artificiale hanno dunque un ruolo limitato in questo ambito³³⁵. Nella stessa sentenza *Franzese* si evidenzia la differenza metodologica delle scienze naturali, ove la spiegazione statistica presenta spesso carattere quantitativo, e le scienze sociali, come il diritto, in cui, assumendo rilevanza il comportamento umano, risulta inadeguato basarsi su dati meramente quantitativi, anziché enunciare gli stessi in termini qualitativi³³⁶.

I limiti dell'IA nell'accertamento del nesso di causalità emergono anche alla luce dello standard probatorio dell'oltre ogni ragionevole dubbio (art. 533, comma 1, c.p.p.). Tale standard è stato introdotto nel codice di procedura penale soltanto con la legge 20 febbraio 2006, n. 46, sebbene fosse stato già stato accolto dalla giurisprudenza nel caso *Franzese*. Le Sezioni Unite, nell'interpretare l'insufficienza e la contraddittorietà della prova determinante l'esito assolutorio (art. 530, comma 2, c.p.p.), avevano fatto riferimento al “*plausibile e ragionevole dubbio*”³³⁷. La probabilità logica non sussiste qualora vi sia un ragionevole dubbio che, nel caso di specie, l'evento sia stato determinato da un fattore causale alternativo rispetto alla condotta dell'imputato³³⁸. La mera probabilità statistica non soddisfa dunque lo standard probatorio dell'oltre ogni ragionevole dubbio. Ne consegue che l'IA non è in grado, da sola, di accertare il nesso di causalità oltre ogni ragionevole dubbio, in quanto fornisce delle informazioni funzionali ad accertare la mera probabilità statistica. L'intelligenza artificiale non è infatti in grado di escludere, oltre ogni ragionevole dubbio, la sussistenza di fattori causali alternativi.

Anche una parte della dottrina è pervenuta ad una conclusione analoga. Si è osservato che la logica statistica può fornire informazioni sulla frequenza con cui un determinato evento tende a ripetersi nel tempo, ma non è

³³³ Cass. pen. Sez. IV, 28 settembre 2000, n. 1688; Cass. pen. Sez. IV, 28 novembre 2000, n. 14006; in G. MARINUCCI - E. DOLCINI - G.L. GATTA, *op. cit.*, p. 300.

³³⁴ Cass. pen. Sez. Un., 11 luglio 2002, n. 30328, par. 7-8; G. MARINUCCI - E. DOLCINI - G.L. GATTA, *op. cit.*, pp. 264-266.

³³⁵ R. BLAIOTTA *op. cit.*, pp. 11-12.

³³⁶ Cass. pen., Sez. Un., 11 luglio 2002, n. 30328, par. 8.

³³⁷ Cass. pen. Sez. Un., 11 luglio 2002, n. 30328; P. TONINI - C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, pp. 277-281.

³³⁸ P. TONINI - C. CONTI, *op. cit.*, pp. 298-301.

in grado di spiegare il perché. Ad esempio, la circostanza che l'imputato presenta una probabilità statistica di commettere reati particolarmente elevata, non fornisce alcuna indicazione utile quanto alla sua colpevolezza in merito allo specifico fatto di reato oggetto dell'imputazione³³⁹.

In materia di responsabilità penale in capo al produttore di un sistema di intelligenza artificiale, sono stati sollevati dei dubbi in merito alla compatibilità dell'approccio probabilistico alla base del *machine learning* con l'accertamento del nesso di causalità³⁴⁰. Tali dubbi sorgono a causa del già citato problema della *black box* che impedisce di sapere le ragioni per le quali il sistema ha generato un determinato *output*³⁴¹. Si osserva che, in assenza di una spiegazione nota, l'imputazione della responsabilità penale risulta priva di una legge scientifica di copertura. Ciò risulta in contrasto con la stessa sentenza *Franzese* in cui, pur non escludendo il nesso di causalità in caso di probabilità statistica medio-bassa, si richiede comunque una legge scientifica in grado di attribuire l'evento alla condotta dell'imputato³⁴².

In conclusione, l'intelligenza artificiale potrebbe supportare l'accertamento sul nesso di causalità, ma sarebbe necessario che l'IA si limiti a supportare un'autonoma valutazione giudiziale e non costituisca il solo elemento in base al quale venga effettuata tale valutazione. La sola probabilità statistica, su cui si basano i sistemi di *machine learning*, non è da sola sufficiente ad accertare la sussistenza del nesso di causalità oltre ogni ragionevole dubbio, essendo altresì necessario che il giudice, in maniera autonoma rispetto all'IA, escluda la sussistenza di fattori causali alternativi.

³³⁹ M. CAIANIELLO, *potenzialità e rischi derivanti dall'interazione tra I. A. e giustizia penale preventiva* in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 269-270.

³⁴⁰ B. FRAGASSO, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, in *Diritto penale contemporaneo*, 2023, pp. 33-34.

³⁴¹ G. BUONOMO, *op. cit.*, pp. 15-18; v. par.3.2.1.

³⁴² B. FRAGASSO, *op. cit.*, p. 33-34.

CAPITOLO TERZO

PRINCIPALI RISCHI

3.1) Il pericolo di discriminazione

3.1.1) Le critiche rivolte a *COMPAS*

Uno dei problemi emersi nel caso *Loomis* è il pericolo che il modello predittivo possa produrre esiti discriminatori nei confronti di alcune categorie di persone. Ciò si può verificare in quanto il sistema potrebbe attribuire eccessiva rilevanza a fattori che sono al di fuori del controllo dell'imputato come l'educazione familiare o il contesto sociale di appartenenza³⁴³.

Nel caso *Loomis*, la difesa ha contestato la validità di *COMPAS* sulla base delle ricerche svolte dall'organizzazione *no-profit ProPublica* che avrebbero dimostrato che tale strumento di valutazione del rischio sarebbe discriminatorio nei confronti di imputati afroamericani³⁴⁴. Sono stati analizzati i casi di 7000 persone arrestate in Broward County in Florida tra il 2013 e il 2014 ed è stato verificato quanti di essi sono stati effettivamente accusati di reati nei due anni successivi all'arresto. Il risultato ha mostrato che *COMPAS* tende a sovrastimare il rischio di recidiva nei confronti degli imputati afroamericani. Gli imputati di questo tipo che non sono risultati recidivi nei due anni successivi erano stati erroneamente classificati come ad alto rischio di recidiva, con una frequenza statistica pari quasi al doppio rispetto agli imputati caucasici nella stessa situazione (rispettivamente 45% e 23%)³⁴⁵. Viceversa, *COMPAS* attribuisce spesso un rischio di recidiva nei confronti degli imputati di altra origine etnica inferiore rispetto a quello reale. È risultato che gli imputati recidivi nei due anni successivi alla previsione algoritmica erano stati erroneamente classificati come a rischio basso nel 48% dei casi, in contrapposizione al 23% degli imputati afroamericani nella stessa situazione³⁴⁶.

I ricercatori hanno valutato la natura discriminatoria di *COMPAS* neutralizzando l'effetto di alcune variabili, quali i reati precedenti, la probabilità di recidiva futura, l'età e il genere. Anche al netto di tali correzioni, *COMPAS* attribuisce agli imputati afroamericani una probabilità di essere classificati come ad alto rischio di recidiva pari al doppio rispetto agli imputati caucasici³⁴⁷.

Sono stati registrati risultati analoghi anche con riferimento al rischio di recidiva violenta. Gli imputati afroamericani vengono classificati ad alto rischio di commettere reati violenti con una frequenza statistica pari al doppio rispetto agli imputati bianchi. I recidivi violenti di origine caucasica hanno il 63% di probabilità in più di essere erroneamente classificati come a rischio basso³⁴⁸.

La *Northpointe* (la società proprietaria di *COMPAS*, oggi *EQUIVANT*) ha rigettato con forza le critiche di *ProPublica*, argomentando che in tali studi sarebbero stati usati dei metodi statistici inappropriati a valutare

³⁴³ C. KLINGELE, *The Promises and Perils of Evidence-Based Corrections*, in *Notre Dame Law Review*, 2015, p. 577

³⁴⁴ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 62-63.

³⁴⁵ J. LARSON -S. MATTU -L. KIRCHNER *et. al.* *How We Analyzed the Compas Recidivism Algorithm*, *ProPublica*, 2016, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>, consultato il 5 ottobre 2025.

³⁴⁶ *Ibid.*

³⁴⁷ *Ibid.*

³⁴⁸ *Ibid.*

l'equità. *ProPublica* avrebbe basato la propria analisi sulla specificità, ossia la percentuale di recidivi correttamente classificati ad alto rischio e la sensibilità, vale a dire la percentuale di non recidivi classificati correttamente ad alto rischio. Tale analisi risulterebbe incompleta, in quanto non considera il tasso di base di recidiva nella popolazione, cioè il rapporto tra la percentuale di persone che risultano recidive e il resto della popolazione. Il tasso di base non sarebbe omogeneo nella popolazione, essendo maggiore tra gli imputati afroamericani rispetto al resto degli imputati. Se si considera tale differenza, *COMPAS* garantirebbe la parità predittiva, in quanto sussisterebbe la medesima probabilità di recidiva per chi riceve il medesimo livello di rischio, indipendentemente dal gruppo razziale di appartenenza³⁴⁹. La più elevata previsione del rischio di recidiva degli imputati afroamericani rispetto ai caucasici sarebbe dunque stata giustificata dal diverso tasso base di recidiva presente tra i due gruppi³⁵⁰.

In merito a tale questione, si è pronunciata anche la Corte Suprema del Wisconsin, la quale ha riconosciuto che *COMPAS* possa destare delle preoccupazioni circa il pericolo che la previsione algoritmica risulti correlata all'etnia dell'imputato³⁵¹. Tale circostanza non comporterebbe però necessariamente la violazione di diritti dell'imputato. Si richiede soltanto che i giudici, nel tenere conto del *risk score* (punteggio di rischio), in sede di determinazione della pena, siano informati che alcuni studi hanno destato delle preoccupazioni circa il pericolo che questi sistemi attribuiscono un elevato rischio di recidiva a chi appartiene alle minoranze, in maniera non proporzionata³⁵².

3.1.2) I criteri di equità algoritmica

La decisione della Corte Suprema del Wisconsin ha acceso un significativo dibattito sia negli Stati Uniti che in Europa in merito ai criteri da impiegare per valutare se uno strumento di valutazione del rischio risulti discriminatorio o meno. Per comprendere questo dibattito, occorre procedere ad alcune precisazioni terminologiche. Con riferimento ad un modello predittivo del rischio di recidiva, con Veri Positivi (VP), si indicano gli individui con previsione di recidiva che hanno realmente recidivato. Con Falsi Positivi (FP), ci si riferisce alle persone con previsione di recidiva che non hanno realmente recidivato. Con Veri Negativi (VN), si indicano gli individui con previsione di non-recidiva che non hanno recidivato. Con Falsi Negativi (FN), ci si riferisce alle persone con previsione di non-recidiva che hanno recidivato³⁵³.

Un primo criterio per valutare l'equità del modello predittivo concerne la c.d. uguaglianza dell'accuratezza complessiva (*overall accuracy equality*). L'accuratezza complessiva è data dalla seguente formula: $(VP + VN) / (VP + FN + FP + VN)$. Misura dunque il rapporto tra le previsioni corrette e tutte le previsioni. L'uguaglianza

³⁴⁹ W. DIETERICH -C. MENDOZA -T. BRENNAN, *COMPAS Risk Scales: Demonstrating Accuracy, Equity and Predictive Parity*, Northpointe Inc., 2016, <https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html>, consultato il 4 agosto 2025, pp.1-3.

³⁵⁰ F. LA GIOIA -G. SARTOR., *Il sistema Compas: algoritmi, previsioni, iniquità*, in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, p. 234.

³⁵¹ Corte Suprema del Wisconsin, *op. cit.* par. 63.

³⁵² *Ibid*, par. 66.

³⁵³ F. LA GIOIA -G. SARTOR., *op. cit.*, p. 236.

dell'accuratezza complessiva si valuta confrontando l'accuratezza tra i due gruppi. Ad esempio, si può confrontare l'accuratezza delle previsioni circa gli imputati afroamericani rispetto all'accuratezza delle previsioni relative agli imputati caucasici. L'uguaglianza della correttezza complessiva è pienamente raggiunta se l'accuratezza è la stessa con riferimento a ciascun gruppo. Questa definizione muove dal presupposto (da verificare in base all'obiettivo perseguito) che i veri positivi siano desiderabili quanto i veri negativi³⁵⁴.

In numerosi contesti, tra cui il settore della giustizia penale, i veri negativi sono maggiormente auspicabili rispetto ai veri positivi. In materia di misure cautelari, risulta maggiormente grave limitare erroneamente la libertà personale, pur non sussistendo nessuna delle esigenze cautelari di cui all'art. 274 c.p.p., a causa di un falso positivo, rispetto al non disporre erroneamente alcuna misura, pur sussistendone i presupposti, a causa di un falso negativo. La centralità che la libertà personale assume nel nostro ordinamento, ai sensi dell'art. 13 Cost. suggerisce di non valutare soltanto l'accuratezza complessiva, ma di prestare particolare attenzione ai falsi positivi, in quanto potrebbero determinare un'indebita compressione di tale libertà costituzionale. Nel contesto della giustizia penale, l'uguaglianza dell'accuratezza complessiva potrebbe dunque non essere il parametro statistico preferibile per valutare l'equità del modello predittivo.

Un altro criterio per analizzare l'equità algoritmica è la parità statistica. Con parità statistica, si intende che, all'interno di ciascun gruppo, la probabilità che venga effettuata una determinata previsione è la stessa. Ad esempio, si ha parità statistica se un soggetto ha la medesima probabilità di essere classificato come recidivo indipendentemente dal gruppo etnico di appartenenza. In termini matematici, occorre fare il rapporto tra i casi in cui il modello predittivo ha previsto la recidiva e la totalità dei casi, $(VP+FP)/(VP+FN+FP+VN)$, e il rapporto tra le previsioni di non-recidiva rispetto alla totalità delle previsioni, $(VN + FN)/VP + FN +FP +VN)$. Si ha la parità statistica se tali rapporti risultano eguali in entrambi i gruppi. Il limite di questo criterio è che, per ottenere la parità statistica, occorre ridurre l'accuratezza, poiché la parità statistica implica che il numero dei positivi e il numero dei negativi in entrambi i gruppi sia lo stesso, cosa che nella realtà potrebbe non verificarsi. Il problema è che, in presenza di un differente tasso di base tra i due gruppi, cioè di una differente tendenza tra le due popolazioni a risultare recidivi, il numero dei positivi e dei negativi risulta differente. Alternativamente, o si imposta di *default* il medesimo tasso di base, rendendo il modello predittivo meno accurato, oppure si considera che i due gruppi presentano tassi di base differenti, rischiando di compromettere la parità statistica³⁵⁵.

L'equità algoritmica si può altresì misurare in base all'uguaglianza dell'accuratezza condizionata (*conditional procedure accuracy equality*). Si tratta di misurare il rapporto tra le previsioni corrette di recidiva e il totale dei casi in cui le persone sono risultate recidive: $VP/(VP + FN)$, nonché il rapporto tra le previsioni corrette di non-recidiva e la totalità di casi in cui gli individui non hanno recidivato: $VN/(VN + FP)$. L'uguaglianza dell'accuratezza condizionata risulta soddisfatta se tali valori sono uguali in entrambi i gruppi.

³⁵⁴ R.BERK -H. HEIDARI -S. JABBARI *et al.* *Fairness in Criminal Justice Risk Assessments: The State of the Art*, in *Sociological Methods & Research*, 2018, <https://arxiv.org/abs/1703.09207>, consultato il 6 ottobre 2025, p. 13.

³⁵⁵ F. LA GIOIA -G. SARTOR., *op. cit.*, pp. 239-240.

Viene talvolta usata l'espressione eguaglianza di opportunità (*equalized odds*) per evidenziare che, se si soddisfa l'uguaglianza dell'accuratezza condizionata, si garantisce che gli individui che hanno avuto lo stesso comportamento siano trattati allo stesso modo tra i due gruppi³⁵⁶. L'uguaglianza dell'accuratezza condizionata misura, per ciascun gruppo, quanti tra coloro che hanno recidivato sono stati previsti come recidivi, nonché, sempre per ciascun gruppo, quanti tra coloro che non hanno recidivato sono stati previsti come non-recidivi. In questo senso, tale valore si può utilizzare per valutare se, a parità di comportamento effettivo (recidiva o non-recidiva), il modello predittivo considera gli individui nello stesso modo, indipendentemente dal gruppo di appartenenza. Anche l'uguaglianza di opportunità risulta però condizionata dal tasso di base³⁵⁷. Se vi è una maggiore probabilità di essere recidivi in un gruppo rispetto all'altro (e quindi un diverso tasso di base), il modello tenderà a violare l'eguaglianza di opportunità.

Un altro parametro che deve essere considerato è la pari calibrazione, con cui si intende che, tra i due gruppi, deve esserci lo stesso rapporto tra le previsioni di recidiva corrette rispetto alla totalità di previsioni di recidiva. Deve altresì sussistere lo stesso rapporto tra le previsioni corrette di non-recidiva rispetto alla totalità di previsioni di non-recidiva. In termini matematici, $VP/(VP+FP)$ e $VN/(VN+FN)$ devono essere uguali nei rispettivi gruppi. La calibrazione serve a verificare che ci sia eguale proporzione di risultati corretti per ciascun gruppo e tipo di previsione³⁵⁸.

La calibrazione è particolarmente rilevante con riferimento ai *risk assessment tools* utilizzati nell'ambito della giustizia penale³⁵⁹ ed è strettamente correlata alla parità predittiva a cui ha fatto riferimento la *Northpointe* per giustificare l'equità di *COMPAS*. Con tale espressione, si intende che, chi riceve la stessa previsione (recidiva o non-recidiva) deve avere la stessa probabilità di essere recidivo, indipendentemente dal gruppo di appartenenza. La *Northpointe* ha rigettato le accuse secondo cui *COMPAS* sarebbe razzista, argomentando che tale strumento garantirebbe piuttosto la parità predittiva tra i neri e i bianchi³⁶⁰.

L'altra faccia della pari calibrazione è l'uguaglianza del tasso dei falsi, con cui si intende che, tra i due gruppi, deve esserci lo stesso rapporto tra le previsioni errate di recidiva rispetto alla totalità di previsioni di recidiva, nonché lo stesso rapporto tra le previsioni errate di non-recidiva, rispetto alla totalità di previsioni di non-recidiva. In altri termini, $FP/(VP+FP)$ e $FN/(VN+FN)$ devono essere uguali nei rispettivi gruppi. La pari calibrazione è analoga all'uguaglianza del tasso dei falsi, con la sola differenza che, invece di basarsi sulle previsioni corrette, considera le previsioni errate. Si tratta quindi di valori strettamente correlati³⁶¹.

Un ulteriore criterio è la parità di trattamento che richiede che il rapporto tra falsi positivi e falsi negativi (FP/FN) sia lo stesso tra i due gruppi. Tale criterio può essere rilevante come strumento di *policy*, in quanto,

³⁵⁶ *Ibid*; M. HARDT -E. PRICE -N. SREBRO, *Equality of Opportunity in Supervised Learning*, in D.D. LEE, K. SUGIYAMA, U.V. LUXBORG et al., *Equality of Opportunity in Supervised Learning. Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems*, Barcellona, Spagna, 2016, pp. 3315 -3323.

³⁵⁷ F. LA GIOIA -G. SARTOR., *op. cit.*, p. 239.

³⁵⁸ *Ibid*.

³⁵⁹ R. BERK -H. HEIDARI -S. JABBARI, *op. cit.*, p. 14.

³⁶⁰ W. DIETERICH -C. MENDOZA -T. BRENNAN, *op. cit.*, pp. 9-10.

³⁶¹ F. LA GIOIA -G. SARTOR., *op. cit.*, p. 240.

se un determinato errore risulta maggiormente costoso in un gruppo rispetto ad un altro, si può aggiustare il modello predittivo per bilanciare i rapporti³⁶².

Si può parlare di equità totale (*total fairness*) quando vengono soddisfatti contemporaneamente tutti e cinque i menzionati criteri: uguaglianza complessiva dell'accuratezza, parità statistica, uguaglianza dell'accuratezza condizionata (o uguaglianza di opportunità), pari calibrazione (strettamente correlata con il tasso di falsi e quindi considerabile come un unico criterio) e la parità di trattamento. Tuttavia, non risulta concretamente possibile soddisfare contemporaneamente ciascuno dei suddetti criteri³⁶³.

In letteratura si è infatti sostenuto che sussiste un *trade-off* (cioè, un inevitabile compromesso tra obiettivi in conflitto) nella determinazione dei *risk scores*. Si tratta di una questione che non attiene esclusivamente alla giustizia penale, ma sussiste anche in altri settori, come, ad esempio, il calcolo del rischio in ambito medico. Un modello predittivo del rischio di contrarre una malattia potrebbe fornire previsioni diverse in base al genere. Se tale rischio risulta diverso tra i due gruppi, si verificherà alternativamente almeno una delle seguenti distorsioni: miscalibrazione, ossia viene sottovalutato il rischio per almeno un gruppo; disuguaglianza tra sani, vale a dire che, tra le persone sane, un gruppo presenta stime di rischio maggiore rispetto ad un altro; disuguaglianza tra malati, ossia che, tra le persone malate, un gruppo presenta stime di rischio maggiore rispetto ad un altro. Il problema è che, se i diversi gruppi (nell'esempio uomini e donne) presentano un tasso di base differente (cioè, una diversa percentuale di persone che contraggono malattie all'interno del gruppo), il modello non può essere contemporaneamente equilibrato su ciascuna delle dimensioni citate³⁶⁴.

Le scelte in materia di equità algoritmica si concretizzano spesso in dei *trade-off* tra alternative non ideali. Bisogna decidere in che misura si possa rinunciare alla massima accuratezza possibile nella previsione della recidiva, in ragione dell'equità algoritmica³⁶⁵.

La sussistenza di un *trade-off* tra accuratezza e discriminazione è emersa in uno studio relativo alla previsione del rischio di recidiva giovanile in cui si è confrontato *SAVRY*, uno strumento di valutazione del rischio strutturato "tradizionale" con dei sistemi di *machine learning*. Ne è emerso che gli algoritmi di *machine learning* presentano un'accuratezza superiore rispetto a *SAVRY*. Allo stesso tempo, mentre *SAVRY* risulta essere generalmente equo, i sistemi di *machine learning* tendono a discriminare gli uomini e gli stranieri³⁶⁶.

In definitiva, nel valutare l'equità del modello predittivo occorre spesso scegliere tra alternative non ottimali, in quanto l'equità risulta spesso in *trade-off* con l'accuratezza. Si tratta di valutazioni di rilevanza giuridica, in quanto si tratta di bilanciare principi costituzionali in conflitto. A tale scopo, nel seguente paragrafo, verranno ricostruiti i diritti fondamentali in gioco.

³⁶² R. BERK -H. HEIDARI -S. JABBARI, *op. cit.* pp. 14-15.

³⁶³ *Ibid.*, p. 15.

³⁶⁴ J. KLEINBERG -S. MULLAINATHAN -A. M. RAGHAVAN, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, in *arXiv preprint arXiv*, <https://arxiv.org/abs/1609.05807>, consultato il 6 ottobre 2025, p. 17.

³⁶⁵ *Ibid.*, pp. 33-35.

³⁶⁶ S. TOLAN -M. MIRON -E. GOMEZ., *et al.*, *Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in Catalonia*, in *Proceedings of the 17th International Conference on Artificial Intelligence and Law (ICAAIL '19)*, Montreal, Canada, Association for Computing Machinery, 2019, <https://doi.org/10.1145/3322640.3326705>, consultato il 6 ottobre 2025, parr. 1-5.

3.1.3) Il principio di uguaglianza e il divieto di discriminazione

L'impiego di un sistema di IA predittivo del rischio di recidiva, che non soddisfi i criteri di equità algoritmica, sarebbe in tensione con il principio di uguaglianza (art. 3 Cost.), e il divieto di discriminazione desumibile dagli artt. 14 della Convenzione europea dei diritti dell'uomo (CEDU) e 21 della Carta dei diritti fondamentali dell'Unione europea (CDFUE).

L'art. 3, comma 1, Cost., stabilisce che *“tutti i cittadini hanno pari dignità sociale e sono eguali dinanzi alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali”*. Sia in dottrina che in giurisprudenza è pacifico che i fattori di discriminazione indicati nell'art. 3, comma 1 Cost. vengono considerati esemplificativi e non tassativi.³⁶⁷ Vi sarebbe dunque una tensione con il principio di uguaglianza anche nel caso di un sistema di IA che discrimini, ad esempio, sulla base dell'orientamento sessuale, sebbene tale fattore di discriminazione non sia stato espressamente previsto nell'art. 3, comma 1, Cost. Un modello predittivo discriminatorio potrebbe violare l'uguaglianza formale, in quanto, a parità di condizioni, potrebbe essere attribuito un livello di rischio differente, sulla base di fattori quali l'etnia o il sesso. Individui uguali dinanzi alla legge potrebbero quindi essere trattati in maniera diversa, violando il precetto costituzionale.

L'art. 3, comma 2, Cost., stabilisce che *“è compito della Repubblica rimuovere gli ostacoli di ordine economico e sociale, che, limitando di fatto la libertà e l'uguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese”*. Tale principio è noto come uguaglianza sostanziale³⁶⁸. Nel caso di un modello predittivo discriminatorio, l'uguaglianza sostanziale potrebbe essere violata, in quanto il sistema di IA potrebbe riflettere dati storici pregressi, con la conseguenza di riprodurre o aggravare le discriminazioni esistenti. L'impiego dell'IA potrebbe dunque configurarsi come un ostacolo che, di fatto, limita la piena uguaglianza dei cittadini dinanzi alla legge, ai sensi dell'art. 3, comma 2, Cost.

In primo luogo, il modello predittivo potrebbe riprodurre le discriminazioni esistenti, in quanto, se i dati in *input* presentano dei *bias*, i quali a loro volta riflettono i pregiudizi degli esseri umani, anche la macchina tenderà a replicare tali errori, con conseguente compromissione del principio di uguaglianza³⁶⁹.

Non vi è poi soltanto il rischio di riprodurre le discriminazioni esistenti, ma anche di aggravarle per via del cosiddetto effetto farfalla (*butterfly effect*): pregiudizi impliciti nei dati iniziali o piccole deviazioni durante l'addestramento del modello possono portare a risultati fortemente iniqui. Tali *output* spesso colpiscono in

³⁶⁷ R. BIN -G. PITRUZZELLA, *Diritto costituzionale*, Giappichelli, Torino, 2025, pp. 486-488.

³⁶⁸ *Ibid.*

³⁶⁹ P. P. PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Archivio Penale*, 2022, <https://archiviopenale.it/intelligenza-artificiale-e-giustizia-penale-una-lettura-attraverso-i-principi/articoli/34325>, consultato il 6 ottobre 2025, p. 11.

modo sproporzionato i gruppi emarginati, rafforzando le disuguaglianze sociali esistenti³⁷⁰. Risultano così fondate le preoccupazioni avanzate da chi osserva che l'impiego dell'IA nella giustizia penale potrebbe portare all'adozione di misure restrittive della libertà personale nei confronti di soggetti più deboli³⁷¹.

In dottrina, nell'analizzare la compatibilità degli strumenti di valutazione del rischio (*risk assessment tools*) con il principio di uguaglianza, si è altresì sostenuto che l'algoritmo sarebbe per antonomasia anti-egualitario perché considera alcuni fattori di rischio e ne omette altri (ad esempio, età, genere, luogo di residenza o abitudini individuali)³⁷².

In materia, oltre al principio di uguaglianza, rilevano anche l'art. 14 CEDU e l'art. 21 CDFUE i quali stabiliscono, rispettivamente, il divieto di discriminazione e il diritto di non-discriminazione.

L'art. 14 della CEDU prevede che *“il godimento e l'esercizio dei diritti e delle libertà previsti nella presente Convenzione deve essere assicurato senza nessuna discriminazione.”* La stessa disposizione fornisce altresì un elenco, anche in questo caso di carattere esemplificativo, di possibili fattori di discriminazione quali, ad esempio, il sesso, la razza e la religione. L'art. 14 della CEDU non sancisce un diritto assoluto a non essere discriminati, bensì si limita a vietare le discriminazioni nel godimento e nell'esercizio dei soli diritti e libertà previste dalla Convenzione³⁷³.

L'art. 21 della CDFUE *“vieta qualsiasi forma di discriminazione fondata, in particolare, sul sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, la disabilità, l'età o l'orientamento sessuale.”* Anche in questo caso, l'elenco dei fattori di rischio è esemplificativo. Tale disposizione si ispira proprio alla corrispondente previsione nella CEDU³⁷⁴. La differenza rispetto all'art. 14 della Convenzione è che il diritto di cui all'art. 21 della CDFUE non è limitato al godimento e all'esercizio di diritti fondamentali della persona.

L'art. 21 CDFUE è invece soggetto al limite di cui all'art. 51, comma 1, secondo cui tale Carta si applica *“alle istituzioni, organi e organismi dell'Unione nel rispetto del principio di sussidiarietà, come pure agli Stati membri esclusivamente nell'attuazione del diritto dell'Unione.”* Una discriminazione rilevante ai sensi dell'art. 21 CDFUE assume rilievo soltanto a condizione che sia possibile ricondurre il caso di specie alle materie di competenza dell'Unione europea³⁷⁵.

³⁷⁰ E. FERRARA, *The Butterfly Effect in Artificial Intelligence Systems: Implications for AI Bias and Fairness*, in *Machine Learning with Applications*, 2024, <https://www.sciencedirect.com/science/article/pii/S266682702400001X>, consultato il 6 ottobre 2025, p. 1.

³⁷¹ F. SGUBBI, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Bologna, Il Mulino, 2019.

³⁷² V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia.*, in U. RUFFOLO, *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Milano, Giuffrè Francis Lefebvre, 2020, <https://discrimen.it/oracolo-algoritmico-e-la-justizia-penale-al-bivio-tra-tecnologia-e-tecnocrazia/>, consultato il 6 ottobre 2025, pp. 11-13.

³⁷³ W. A. SCHABAS, *The European Convention on Human Rights: A Commentary*, Oxford University Press Incorporated, Oxford, 2015, pp. 555-587.

³⁷⁴ A. SILVEIRA -L. A. COELHO -M. I. COSTA et al., *The Charter of Fundamental Rights of the European Union: a commentary*, JusGov, UMinho Law School, Escola de Direito da Universidade do Minho, Braga, Portogallo, 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4999183, consultato l'8 ottobre 2025, p. 212.

³⁷⁵ *Ibid.*

La Corte di giustizia dell'Unione europea (CGUE) ha distinto tra discriminazione diretta e discriminazione indiretta. La prima sussiste nel caso in cui una persona venga trattata in modo meno favorevole rispetto ad un'altra sulla base di uno dei fattori di discriminazione protetti dalla legge³⁷⁶.

Con discriminazione indiretta si intende, invece, che una previsione normativa, una pratica, o un criterio apparentemente neutro pone determinate categorie di persone in una situazione di particolare svantaggio. Tali previsioni normative, pratiche o criteri vengono però considerati legittimi se sussiste un obiettivo ragionevolmente giustificato e se i mezzi per raggiungere tale obiettivo risultano proporzionati e necessari³⁷⁷. L'Agenzia dell'Unione europea per i diritti fondamentali (o *European Union Agency for Fundamental Rights*, *FRA*) ha segnalato che il ricorso all'intelligenza artificiale costituisce una sfida particolarmente significativa per quanto concerne il rispetto del divieto di discriminazione, a causa dei *bias* di cui i sistemi IA potrebbero essere affetti³⁷⁸.

Nel caso di impiego dell'IA nella giustizia penale, risultano particolarmente significativi soprattutto i rischi di discriminazione indiretta. La discriminazione diretta potrebbe configurarsi soltanto nel caso in cui un sistema fosse programmato per trattare una persona in modo maggiormente svantaggioso rispetto ad un altro. In concreto, anche gli algoritmi che hanno suscitato forti preoccupazioni in merito al pericolo di discriminazione, come *COMPAS*, non vengono esplicitamente programmati al fine di considerare l'etnia, o il genere come fattori rilevanti.

Risulta invece maggiormente significativo il rischio di discriminazione indiretta, in quanto gli algoritmi spesso incorporano dei *bias* presenti nei dati storici con cui sono stati addestrati, con la conseguenza che, in concreto, il sistema di IA tratta in modo meno favorevole alcune categorie di persone rispetto ad altre. In questo caso, il sistema di IA potrebbe configurarsi come un criterio o una pratica apparentemente neutra che pone determinate categorie di persone in una situazione di particolare svantaggio, secondo la definizione di discriminazione indiretta fornita dalla CGUE.

3.1.4) Le garanzie presenti nel diritto vigente

Dopo aver analizzato il rischio che l'impiego dell'IA nella giustizia penale possa condurre ad indebite discriminazioni, occorre illustrare i possibili rimedi esperibili nel diritto vigente, con particolare riferimento al diritto euro-unitario.

³⁷⁶ *Ibid*, p. 216; ad esempio, la Corte di giustizia dell'Unione europea ha riconosciuto la sussistenza di una discriminazione diretta nelle seguenti sentenze: Corte di giustizia dell'Unione europea, *Feryn*, C-54/07, 10 luglio 2008; Corte di giustizia dell'Unione europea, *Association belge des Consommateurs Test Achats ASBL*, C-236/09, 1° marzo 2011; Corte di giustizia dell'Unione europea, *Pensionsversicherungsanstalt c. Christine Kleist*, C-356/09, 18 novembre 2010; Corte di giustizia dell'Unione europea, *NH c. Associazione Avvocatura per i diritti LGBTI – Rete Lenford*, C-507/18, 23 aprile 2020.

³⁷⁷ A. SILVEIRA -L. A. COELHO -M. I. COSTA, *op. cit.*, p. 216; tra i casi in cui la discriminazione indiretta è stata trattata dalla Corte di giustizia si possono citare le seguenti sentenze: Corte di giustizia dell'Unione europea, *Seymour-Smith e Perez*, C-167/97, 9 febbraio 1999; Corte di giustizia dell'Unione europea, *Allonby, Razpredelenie Bulgaria AD c. Komisija za zashtita ot diskriminatsia*, C-256/01, 13 gennaio 2004; Corte di giustizia dell'Unione europea, *CHEZ*, C-83/14, 16 luglio 2015.

³⁷⁸ Agenzia dell'Unione europea per i diritti fondamentali, *Getting the Future Right*, 2021, Lussemburgo, 2020, pp. 68-73.

Come precedentemente illustrato, il regolamento 2024/1968/UE (AI Act) prevede determinati obblighi per i sistemi ad alto rischio. L'art. 10 individua determinati standard che devono essere osservati in materia di dati di addestramento. Al paragrafo 3, si richiede che *“i set di dati di addestramento, convalida e prova siano pertinenti, sufficientemente rappresentativi e, nella misura del possibile, esenti da errori e completi nell'ottica della finalità prevista”* Specificamente, devono possedere *“le proprietà statistiche appropriate anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone relativamente ai quali il sistema di IA ad alto rischio è destinato a essere usato.”*

Il considerando 67 del medesimo regolamento riconduce tali criteri alla finalità di evitare che i sistemi di IA divengano delle fonti di discriminazione vietate dal diritto dell'Unione. Si evidenzia la necessità che i dati presentino probabilità statistiche adeguate al fine di evitare possibili distorsioni, soprattutto nel caso in cui tali sistemi si basino su dei dati storici che potrebbero non essere sufficientemente rappresentativi. Il rischio di distorsioni risulta particolarmente significativo nel caso in cui l'*output* prodotto influenzi l'*input* delle operazioni future (in tal caso ricorre il c.d. *feedback loop* o ciclo di ritorno), con la conseguenza che le distorsioni rischierebbero di innescare un circolo vizioso. Tali distorsioni possono condurre a discriminazioni nei confronti delle persone che appartengono a determinati gruppi vulnerabili, come gruppi razziali o etnici (considerando 67).

Anche l'Agenzia Europea per i Diritti Fondamentali ha segnalato che spesso la scarsa qualità dei dati e i *bias* dei sistemi di IA costituiscono la fonte di discriminazione algoritmica e di trattamento non equo³⁷⁹.

L'obiettivo di rimuovere e correggere i *bias* presenti nei sistemi di IA si colloca nel contesto del diritto europeo antidiscriminatorio, ossia l'insieme delle norme europee funzionali a dare attuazione al citato art. 21 CDFUE. Si pensi, in via esemplificativa alla direttiva 2000/43/CE *“che attua il principio di parità di trattamento tra le persone indipendentemente dalla razza o dall'origine etnica”*. Ancora, si può citare la direttiva 2000/78/CE *“che istituisce un quadro generale per la parità di trattamento in materia di occupazione e condizioni di lavoro.”* In dottrina, si è osservato che, se da un lato pare apprezzabile lo sforzo del legislatore europeo in questo ambito, dall'altro le soluzioni prospettate non presentano natura sistematica e non sono omogenee³⁸⁰.

Occorre poi rilevare che la maggior parte dei fattori di discriminazione protetti dal diritto euro-unitario costituiscono altresì categorie di dati particolari per le quali, in materia di protezione dei dati personali, il trattamento è possibile solo in casi limitati. Ad esempio, la razza o la provenienza etnica costituiscono sia un possibile fattore di discriminazione che una particolare categoria di dati in materia di protezione di dati personali.

Tali dati sono soggetti alla disciplina di cui all'art. 9 del regolamento UE 2016/679 (GDPR), oppure, con specifico riferimento ai dati trattati dalle autorità competenti all'accertamento dei reati, all'art. 10 della direttiva 2016/680/UE (direttiva LED). Quest'ultima disposizione prevede che: *“Il trattamento di dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche o*

³⁷⁹ *Ibid*, p. 69.

³⁸⁰ A. SILVEIRA -L. A. COELHO -M. I. COSTA *op. cit.*, p. 213.

l'appartenenza sindacale, e il trattamento di dati genetici, di dati biometrici intesi a identificare in modo univoco una persona fisica o di dati relativi alla salute o di dati relativi alla vita sessuale della persona fisica o all'orientamento sessuale è autorizzato solo se strettamente necessario, soggetto a garanzie adeguate per i diritti e le libertà dell'interessato e soltanto: a) se autorizzato dal diritto dell'Unione o dello Stato membro; b) per salvaguardare un interesse vitale dell'interessato o di un'altra persona fisica; o c) se il suddetto trattamento riguarda dati resi manifestamente pubblici dall'interessato". Tale limitazione al trattamento dei dati personali è anche funzionale a prevenire ingiuste discriminazioni, come conferma il considerando 38 in cui, con specifico riferimento alla profilazione, si evidenzia il rischio di discriminazione sulla base di particolari categorie di dati personali che per loro natura risultano sensibili sul piano dei diritti e delle libertà fondamentali.

Infine, l'art. 11, par. 2 della direttiva LED prevede che, di regola, il processo decisionale automatizzato relativo alle persone fisiche non si può basare sulle particolari categorie di dati personali di cui all'art. 10 della stessa direttiva. La profilazione che porta alla discriminazione delle persone fisiche è vietata, conformemente al diritto dell'Unione (art. 11, par. 3 direttiva LED)³⁸¹.

In questi casi, sussiste una doppia protezione riconducibile sia al diritto antidiscriminatorio che alla protezione dei dati personali. Queste norme contribuiscono al medesimo obiettivo complessivo di cui all'art. 10 dell'AI Act, consistente nel contrastare le discriminazioni vietate dal diritto dell'Unione³⁸².

3.2) L'opacità algoritmica

3.2.1) L'opacità dovuta al problema della *black box*

Dopo aver analizzato il rischio di discriminazione da parte dei sistemi IA impiegati nel processo penale, occorre considerare il pericolo dell'opacità di tali algoritmi, in quanto il destinatario della decisione adottata con il supporto dell'IA potrebbe non essere posto nelle condizioni di conoscere a pieno la logica alla base di tale sistema. Tale situazione si potrebbe verificare sia per ragioni tecniche, a causa del problema della *black box*, che per via di diritti di proprietà intellettuale. In questo paragrafo, si considera la prima ipotesi.

Con *black box* si fa riferimento alla difficoltà tecnica nel ricostruire i processi logici in base ai quali il sistema di IA ha prodotto un determinato *output*³⁸³. In dottrina, si usa l'espressione "problema dell'opacità" (*opacity problem*), per riferirsi a quella situazione in cui non è possibile capire quale fattore è responsabile per il risultato prodotto, né è possibile conoscere quali sono stati i passaggi tramite i quali tale *output* è stato generato³⁸⁴.

³⁸¹ L'art. 11, par. 2 e 3, della direttiva 2016/680/UE è stato analizzato in maniera maggiormente approfondita nel par. 4.2.3 a cui si rinvia.

³⁸² C. N. PEHLIVAN -N. FORGÒ -P. VALCKE, *The EU Artificial Intelligence (AI) ACT a Commentary*, Kluwer Law International B.V, Olanda, 2025, p. 265.

³⁸³ G. BUONOMO, *Macchine intelligenti, processo induttivo e apprendimento autonomo*, in G. BUONOMO -G. CIACCI -A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, pp. 15-18.

³⁸⁴ B. BROŽEK, *The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making*, in *Artificial Intelligence and Law*, 2024, pp. 429-430.

Sotto questo profilo, i sistemi di IA si distinguono in modelli trasparenti e modelli opachi. Un esempio di modello trasparente è la regressione lineare che si riferisce ad una classe di modelli usati per prevedere un valore continuo. Tali modelli si basano sul presupposto che questo obiettivo è dato dalla combinazione lineare delle variabili predittive³⁸⁵.

I modelli trasparenti spesso non sono tra i più performanti in termini di accuratezza. Gli algoritmi che giungono a risultati maggiormente affidabili sono spesso gli stessi che pongono maggiori difficoltà quando si tratta di comprendere le ragioni per le quali si è giunti ad un determinato *output*. In definitiva, sussiste un *trade-off* (un compromesso tra obiettivi contrastanti), tra trasparenza e accuratezza.

Un esempio sono i *Multilayer Neural Network*, una tipologia di *machine learning* di particolare successo, essendo stata impiegata in settori molto diversi tra loro. Ad esempio, i *Multilayer Neural Network* sono alla base di molti sistemi di raccomandazione che suggeriscono all'utente prodotti o servizi sulla base del comportamento passato dell'utente stesso. Nonostante il risultato prodotto sia apprezzabile, questi modelli risultano difficili da interpretare, in quanto non è chiaro in che modo le variabili interagiscono tra loro, né quali sono state le informazioni che la rete potrebbe aver rilevato dai dati³⁸⁶.

Una soluzione tecnica a questo problema potrebbe essere l'intelligenza artificiale spiegabile (*Explainable AI* o *XAI*). Sono tecniche che mirano a comprendere il funzionamento dell'IA, senza compromettere la *performance* di tali sistemi. Il fine ultimo è quello di promuovere la fiducia degli utenti nell'IA, in quanto gli utenti potrebbero non fidarsi di un sistema di cui ignorano la logica alla base³⁸⁷.

Tali spiegazioni possono essere centrate sul modello (*Model-centric explanation*, *MCE*), oppure centrate sul soggetto (*Subject-centric explanation*, *SCE*). Le spiegazioni centrate sul modello forniscono informazioni generiche sul modello nel suo complesso, senza riferirsi a decisioni specifiche o a dati in *input* particolari. Tramite tali spiegazioni, è possibile ricavare informazioni sui dati utilizzati per addestrare il sistema IA, nonché l'origine di tali dati o le classificazioni che il modello cerca di ottenere. Tuttavia, questa tipologia di spiegazioni, proprio perché incentrate sul modello nel suo complesso, non forniscono informazioni significative ai singoli individui destinatari di decisioni adottate con il supporto dell'IA³⁸⁸.

Per ottenere delle spiegazioni utili ai singoli individui occorre, invece, fare affidamento alle spiegazioni centrate sul soggetto (*subject-centric explanation*). Sono spiegazioni costruite sui dati di un singolo individuo o del singolo caso specifico, anziché sul modello nel suo complesso. Vengono denominate spiegazioni locali, in quanto spiegano soltanto la zona dei dati vicina al caso in esame. Mentre i modelli complessi risultano difficili da interpretare tutti insieme, concentrarsi soltanto su di una parte specifica li rende maggiormente

³⁸⁵ V. BELLE -I. PAPANTONIS, *Principles and Practice of Explainable Machine Learning*, in *Frontiers in Big Data*, 2021, <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2021.688969/full>, consultato l'8 ottobre 2025, pp. 6-8.

³⁸⁶ *Ibid*, pp. 8-9.

³⁸⁷ A. BARREDO ARRIETA, *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*, in *Information Fusion*, 2020, <https://www.sciencedirect.com/science/article/abs/pii/S1566253519308103>, consultato l'8 ottobre 2025, p. 283.

³⁸⁸ L. EDWARDS -M. VEALE, *Slave to the Algorithm? Why 'a Right to an Explanation' is Probably not the Remedy you are Looking for*, in *Duke Law & Technology Review*, vol. 16, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855, pp. 55-56.

comprensibili. Vi è altresì il problema della dimensionalità. Se il sistema usa un numero particolarmente elevato di variabili, il numero di combinazioni cresce esponenzialmente, rendendo tale modello particolarmente difficile da comprendere. La soluzione consiste nell'esaminare da vicino i dati relativi al singolo caso³⁸⁹.

Questa tipologia di spiegazioni fornisce informazioni particolarmente utili ai destinatari delle decisioni adottate sulla base di un suggerimento offerto dall'IA che cerchino di comprendere le ragioni sottostanti. È così possibile sapere quali modifiche ai dati avrebbero cambiato la decisione, oppure quali caratteristiche hanno le persone che hanno ricevuto un trattamento analogo³⁹⁰. Ad esempio, se le persone che hanno ricevuto un trattamento analogo appartengono ad una determinata etnia, si può ipotizzare che il modello predittivo presenti dei pregiudizi di natura razziale.

Nonostante tali tecniche di spiegazione dei sistemi di IA sembrano promettenti, vi sono dei limiti che occorre segnalare. Come precedentemente accennato, sussiste un *trade-off* tra le performance del sistema e la sua interpretabilità. I sistemi maggiormente complessi, che solitamente sono quelli che si basano su di un numero di variabili particolarmente elevato, otterranno risultati maggiormente soddisfacenti, ma saranno particolarmente difficili da interpretare³⁹¹. In presenza di modelli di questo tipo, potrebbe essere particolarmente complesso capire quale è stata la variabile decisiva nel singolo caso.

In secondo luogo, proprio coloro che maggiormente avrebbero bisogno di una spiegazione, potrebbero avere maggiore difficoltà ad ottenerne una tramite l'*Explainable AI*. I sistemi di *machine learning* tendono a commettere errori nei confronti di soggetti che presentano caratteristiche statisticamente poco frequenti. Ad esempio, se un soggetto ha commesso un reato che si verifica poco frequentemente, il modello predittivo potrebbe non avere abbastanza dati per calcolare il rischio di recidiva in maniera accurata con riferimento a tale ipotesi. Paradossalmente, è proprio in questi casi che sarà maggiormente difficile ottenere una spiegazione tramite l'*Explainable AI*. Le spiegazioni centrate sul soggetto selezionano una parte del modello, semplificandolo. Tale semplificazione comporta che non vengono considerate le ipotesi maggiormente complesse, che corrispondono ai casi in cui il sistema tende a sbagliare con maggiore frequenza³⁹². Nell'esempio sopra riportato, se viene stimato il rischio di recidiva nei confronti di un soggetto che ha commesso un reato che si verifica poco frequentemente, non soltanto vi è un maggiore rischio di errore, ma vi è anche la possibilità che non si riesca ad ottenere una spiegazione tramite l'*Explainable AI*.

In sintesi, nonostante le tecniche di *Explainable AI* migliorino la trasparenza, permangono limiti significativi, specialmente per soggetti che presentano delle caratteristiche dalla scarsa frequenza statistica, come nel caso in cui venga stimato il rischio di recidiva di coloro che hanno commesso reati poco frequenti.

³⁸⁹ *Ibid*, pp. 56-57.

³⁹⁰ *Ibid*, p. 58.

³⁹¹ *Ibid*, p. 59.

³⁹² *Ibid*, pp. 60-61.

3.2.2) L'opacità dovuta alla proprietà industriale

Il cosiddetto problema della *black box* non è la sola ragione per cui i destinatari di una decisione adottata con il supporto dell'IA potrebbero non essere in grado di conoscere l'*iter* logico alla base della decisione adottata. Tale mancata conoscenza potrebbe essere dovuta altresì a ragioni di proprietà industriale.

La questione era stata sollevata dalla difesa nel caso *Loomis*, in cui la *Northpointe* aveva invocato la natura proprietaria di *COMPAS* per impedire la divulgazione di informazioni relative a come i fattori vengono ponderati o come vengono determinati i punteggi di rischio³⁹³. La difesa lamentava di non poter contestare la validità scientifica dello strumento. La Suprema Corte del Wisconsin ha invece considerato legittimo l'uso di *COMPAS*, in quanto la difesa aveva avuto accesso ai dati su cui tale sistema si era basato³⁹⁴. Tale conclusione è stata criticata dalla dottrina statunitense, in quanto poter rivedere e contestare le singole informazioni su cui lo strumento di valutazione del rischio si basa, non significa essere in grado di mettere in discussione le modalità attraverso cui il *risk score* è stato prodotto³⁹⁵. Si è poi osservato che non comprendere pienamente il funzionamento del modello predittivo potrebbe essere pregiudizievole anche per gli stessi giudici che vi farebbero affidamento senza la necessaria consapevolezza³⁹⁶.

Anche in Italia, le imprese proprietarie di un sistema di IA predittiva impiegato nel processo penale potrebbero avvalersi della proprietà industriale, al fine di impedire che vengano divulgate informazioni sul suo funzionamento.

Occorre fare specifico riferimento al segreto industriale, ai sensi dell'art. 98 del d.lgs. 10 febbraio 2005, n.30, noto come codice della proprietà industriale. Oggetto del segreto industriale sono le informazioni aziendali o le esperienze tecnico-industriali. Il segreto sussiste qualora “*tali informazioni non siano generalmente note o facilmente accessibili agli operatori del settore per la loro configurazione e combinazione dei loro elementi*”. Ricorre il segreto anche nel caso in cui “*le informazioni abbiano valore economico in quanto segrete o siano sottoposte a misure da ritenersi ragionevolmente adeguate a tenerle segrete*”. Tale definizione di segreto industriale sostanzialmente recepisce la direttiva 2016/943/UE “*sulla protezione del know-how riservato e delle informazioni commerciali riservate*”³⁹⁷.

Nel caso di un sistema di IA, il segreto industriale potrebbe, ad esempio, coprire l'architettura, i dati e i parametri, impedendo a terzi di ricostruire i processi logici sottostanti al sistema di IA, anche qualora ciò fosse tecnicamente possibile.

Occorre bilanciare la tutela della proprietà industriale con l'esigenza del destinatario di una decisione adottata con il supporto dell'IA di conoscerne le ragioni alla base. Come osservato in dottrina, ogni diritto di proprietà

³⁹³ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 100.

³⁹⁴ *Ibid*, parr. 46-56.

³⁹⁵ D. KEHL -P. GUO -S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2017, <https://dash.harvard.edu/entities/publication/73120379-010a-6bd4-e053-0100007fdf3b>, consultato il 26 settembre 2025, pp. 22-23.

³⁹⁶ N. STRAUSS, *Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment*, in *Columbia Law Review*, vol. 120, 2020, pp. 1682-1685.

³⁹⁷ J. PILA -P. TORREMANS *European Intellectual Property Law*, Oxford University Press, Oxford, 2019, pp. 513-515.

intellettuale, essendo inserito in un contesto più ampio, deve essere sempre bilanciato con altri diritti di pari valore³⁹⁸. Tale bilanciamento deve tenere conto della funzione sociale della proprietà intellettuale, riconducibile a un contratto sociale in cui, da un lato, il titolare del diritto può sfruttare economicamente la creazione, mentre dall'altro, la collettività beneficia dell'impatto positivo della creazione stessa nella società³⁹⁹.

Nel caso *Dun & Braadstreet c. Austria*, la Corte di giustizia dell'Unione europea (CGUE) ha dovuto affrontare il tema del bilanciamento tra i diritti di proprietà industriale con l'esigenza del destinatario di una decisione algoritmica di conoscerne le ragioni alla base⁴⁰⁰. Tale sentenza aveva ad oggetto la portata del diritto dell'interessato a “*informazioni significative sulla logica utilizzata*”, in caso di processo decisionale automatizzato, ai sensi dell'art. 15, par. 1, lett. h) del regolamento UE/2016/676 (GDPR). Il giudice *a quo* chiedeva se tale diritto di accesso comprendesse anche le informazioni coperte dal segreto industriale, secondo la definizione fornita dalla direttiva 2016/943/UE. La CGUE ha sancito che, se i dati oggetto del diritto ad informazioni significative sulla logica utilizzata sono coperti dal segreto industriale, il titolare del trattamento è tenuto a comunicare tali informazioni al giudice o alle autorità competenti, a cui spetta ponderare tale diritto e tali interessi nel caso di specie⁴⁰¹.

Per comprendere la portata di tale sentenza, occorre evidenziare che una parte della dottrina aveva sostenuto che il diritto ad informazioni significative sulla logica utilizzata fosse recessivo rispetto alla tutela della proprietà intellettuale⁴⁰². Veniva invocato il considerando 63 del GDPR il quale precisa che le informazioni spettanti all'interessato non possono compromettere i diritti e le libertà altrui, compreso il segreto industriale e la proprietà intellettuale in genere. La CGUE supera tale logica, sancendo che i diritti di proprietà intellettuale in genere non possono aprioristicamente prevalere sul diritto a conoscere la logica impiegata da un sistema di IA.

Il diritto di accesso ad informazioni significative sulla logica utilizzata non è invece ricompreso tra le informazioni che spettano all'interessato, nelle attività di contrasto penale, ai sensi dell'art. 13, par. 1 della direttiva 2016/680/UE (direttiva LED). Tale direttiva si applica, in via alternativa al GDPR, in caso di trattamento di dati personali da parte delle autorità competenti nelle attività di contrasto, ossia a fini di “*prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali*” (art. 2). Ne consegue che, nel caso di impiego dell'IA nel processo penale, gli imputati o indagati non possono esercitare tale diritto, a differenza di quanto avviene quando l'IA è utilizzata da soggetti pubblici o privati al di fuori del processo penale.

³⁹⁸ C. GEIGER, *The Social Function of Intellectual Property Rights, or How Ethics can Influence the Shape and Use of IP Law*, Max Planck Institute for Intellectual Property and Competition Law in G.B. DINWOODIE *Intellectual Property Law: Methods and Perspectives*, Cheltenham, Regno Unito 2013, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2228067, consultato il 9 ottobre 2025, pp. 153-176.

³⁹⁹ *Ibid.*

⁴⁰⁰ Corte di giustizia dell'Unione europea, *Dun & Braadstreet c. Austria*, C-203/22, 27 febbraio 2025.

⁴⁰¹ *Ibid.*, par. 67-76.

⁴⁰² L. EDWARDS -M. VEALE, *op. cit.*, pp. 51-53.

La sentenza *Dun & Braadstreet c. Austria* costituisce comunque un significativo precedente giurisprudenziale anche ai fini della presente disamina, in quanto mette in dubbio l'assunto secondo cui l'esigenza degli interessati di comprendere il funzionamento dell'IA debba necessariamente soccombere dinanzi alla proprietà industriale.

Una questione analoga viene affrontata anche dal regolamento 2024/1689/UE (AI Act). L'art. 86 par. 1, prevede che *“qualsiasi persona interessata oggetto di una decisione adottata dal deployer sulla base dell'output di un sistema di IA ad alto rischio[...] e che produca effetti giuridici o in modo analogo incida significativamente su tale persona in un modo che essa ritenga avere un impatto negativo sulla sua salute, sulla sua sicurezza o sui suoi diritti fondamentali ha il diritto di ottenere dal deployer spiegazioni chiare e significative sul ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione adottata.”*

La formulazione della disposizione è analoga all'art. 22, par. 1 del GDPR nella parte in cui prevede il divieto di processo decisionale automatizzato qualora la decisione *“produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona*. È in presenza di tali decisioni che l'art. 15, par. 1, lett. h del GDPR prevede il diritto *“ad informazioni significative sulla logica utilizzata”* a cui si è fatto precedentemente riferimento (perché l'art. 15 rinvia all'art. 22).

Il diritto di cui all'art. 86, par. 1 dell'AI Act si applica ai sistemi ad alto rischio. Sarebbe dunque rilevante per i sistemi di IA ad alto rischio destinati ad essere utilizzati nel processo penale, come quelli funzionali a *“determinare il rischio di commissione del reato o di recidiva in relazione a una persona fisica”*⁴⁰³.

Tale diritto non si applica qualora siano previste *“eccezioni o limitazioni all'obbligo stabilito in tale paragrafo in virtù del diritto dell'Unione o del diritto nazionale, in linea con il diritto dell'Unione”* (art. 86, par. 2 del suddetto regolamento). Tale disposizione potrebbe far prevalere i diritti di proprietà industriale sul diritto alla spiegazione dei singoli processi decisionali. Il considerando 86 dell'AI Act chiarisce che le informazioni che devono essere fornite alla persona interessata non possono compromettere né diritti di proprietà intellettuale in genere né il segreto industriale in particolare⁴⁰⁴. Non si può comunque escludere che, la CGUE interpreti l'art. 86 dell'AI Act in modo tale da ponderare gli interessi delle parti caso per caso, senza far prevalere automaticamente i diritti di proprietà intellettuale in genere e industriale in particolare, come già avvenuto nella sentenza *Dun & Braadstreet c. Austria* (in quel caso con riferimento al GDPR e non all'AI Act).

In ogni caso, sono state avanzate delle soluzioni funzionali a conciliare le esigenze degli interessati con la proprietà industriale. Si è proposto di sottoporre i sistemi di IA ad un'omologazione autoritativa *ex ante* che ne garantisca l'affidabilità. Tale omologazione potrebbe essere demandata ad un organo statuale, come, il Ministero della Giustizia, un'entità *ad hoc* facente capo al CSM, oppure un'apposita autorità o agenzia. Sarebbe così possibile accertare che i codici di programmazione e le modalità con cui il sistema è stato

⁴⁰³ V. par. 4.3.2.

⁴⁰⁴ C. N. PEHLIVAN -N. FORGÒ -P. VALCKE, *The EU Artificial Intelligence (AI) ACT: a Commentary*, Kluwer Law International B.V, Olanda, 2025, p. 289.

addestrato siano adeguate. Tale procedimento autoritativo sarebbe coperto dal vincolo di segretezza, in modo tale da non ledere i diritti di proprietà industriale. Lo stesso ente incaricato potrebbe essere altresì competente a sorvegliare sulla manutenzione del sistema di IA nel suo ciclo di vita, anche in sede di contestazione dovuta ad un malfunzionamento nel caso concreto⁴⁰⁵.

Questa logica non è estranea all'AI Act il quale, all'art. 78, prevede che “*le autorità di sorveglianza del mercato garantiscono la riservatezza delle informazioni e dei dati ottenuti nello svolgimento dei loro compiti e delle loro attività in modo da tutelare, in particolare: a) i diritti di proprietà intellettuale*”. Le autorità di sorveglianza possono così accertare che i sistemi di IA siano conformi al regolamento (artt. 74-77 AI Act), senza compromettere il segreto industriale.

Un'altra possibile soluzione, non necessariamente alternativa alla precedente, potrebbe consistere nel promuovere lo sviluppo *open source* di modelli predittivi. I sistemi *open source*, in contrapposizione ai sistemi proprietari, permettono di conoscere e, se necessario, contestare i metodi e i dati impiegati, senza violare la proprietà industriale⁴⁰⁶. Il ricorso a tali sistemi sarebbe anche vantaggioso in quanto favorirebbe le collaborazioni interdisciplinari tra studiosi di discipline differenti, dalle scienze computazionali a quelle sociali⁴⁰⁷. Se i dati sono pubblicamente accessibili risulta più facile che esperti di settori differenti possano studiare gli stessi modelli, senza barriere. Ricorrere a sistemi *open source* può comportare una riduzione delle spese, in quanto modelli predittivi commerciali *COMPAS* possono presentare dei costi proibitivi, soprattutto per le giurisdizioni più piccole.⁴⁰⁸

Possono esserci dei rischi in termini di *privacy* e protezione dei dati personali, in quanto i dati contenuti nel casellario giudiziale verrebbero impiegati per addestrare il sistema di IA. Vi è quindi il rischio che uno sviluppo completamente *open source* finisca per rendere pubblici i casellari giudiziali di molti cittadini. Per contrastare questo pericolo è fondamentale che i dati dei cittadini vengano anonimizzati in modo efficace⁴⁰⁹.

In definitiva, l'opacità derivante dalla tutela della proprietà industriale costituisce un serio ostacolo alla comprensibilità delle decisioni adottate con il supporto dell'IA, imponendo la necessità di un difficile bilanciamento tra segreto industriale e diritti fondamentali dell'interessato. Sono di particolare interesse quelle soluzioni che sembrano conciliare tali esigenze altrimenti opposte, come lo sviluppo di modelli *open source* o la sottoposizione dei sistemi di IA ad un'omologazione autoritativa *ex ante*.

3.2.3) L'impatto dell'opacità sulle garanzie processuali

⁴⁰⁵ U. RUFFOLO, *La macchina sapiens come "avvocato generale" ed il primato del giudice umano*, in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, p. 221.

⁴⁰⁶ P. WAGGONER -A. MACMILLEN, *Pursuing Open-Source Development of Predictive Algorithms: The Case of Criminal Sentencing Algorithms*, in *Journal of Computational Social Science*, vol. 5, no. 1, Springer, 2022, <https://doi.org/10.1007/s42001-021-00122-y>, consultato il 25 settembre 2025, pp. 4-5.

⁴⁰⁷ *Ibid*, pp. 3-4.

⁴⁰⁸ *Ibid*, p. 4.

⁴⁰⁹ *Ibid*, pp. 5-6.

Come illustrato nei precedenti paragrafi, i destinatari di decisioni adottate con l'ausilio dell'IA nell'ambito del processo penale potrebbero non essere resi edotti delle ragioni poste alla base dell'*output*, a causa del problema della *black box*, oppure per motivi di proprietà intellettuale.

Tale situazione potrebbe entrare in contrasto con diverse garanzie processuali, di rilevanza sia costituzionale che sovranazionale. Assume rilievo l'art. 24, comma 1, Cost. che sancisce che “*tutti possono agire per la tutela dei propri diritti e interessi legittimi*”. Al comma 2, si prevede che “*la difesa è diritto inviolabile in ogni stato e grado del procedimento*”. Rispetto al passato, il fattore di novità è che oggi le lesioni ai diritti soggettivi e agli interessi legittimi possono derivare dal ricorso all'intelligenza artificiale. In dottrina, si è sostenuto che, in questo nuovo scenario, il diritto alla difesa deve essere interpretato nel senso che ogni volta che un soggetto, pubblico o privato, utilizzi un sistema di IA per adottare una decisione che potrebbe interferire con diritti soggettivi o interessi legittimi, chi subisce tale decisione ha il diritto di comprenderne le ragioni alla base⁴¹⁰.

Questa interpretazione si applicherebbe anche all'impiego dell'IA nel processo penale, in quanto si tratta di soggetti pubblici (il giudice, il PM o la p.g.) che ricorrono all'IA per adottare decisioni che interferiscono con diritti soggettivi (sono tendenzialmente decisioni incidenti sulla libertà personale).

Si potrebbero creare delle zone grigie in cui tecnologie non pienamente comprensibili determinino, di fatto, il contenuto delle sentenze, rendendo poi estremamente difficile sindacarne la legittimità. Si potrebbe così compromettere uno dei capisaldi degli ordinamenti in cui vige il principio di legalità, ovvero di *rule of law*, ossia la possibilità di sindacare le decisioni giudiziarie⁴¹¹.

Sul piano sovranazionale, l'opacità algoritmica si pone in tensione con il diritto ad un ricorso effettivo di cui all'art. 13 della CEDU e 47 della CDFUE. Il primo rileva in caso di violazione dei diritti e libertà sancite dalle Convenzioni. Il secondo non si applica soltanto alle violazioni della suddetta CDFUE, bensì, più in generale, alle violazioni dei diritti e delle libertà sancite dal diritto dell'Unione europea⁴¹².

La tensione tra l'opacità algoritmica e diritto ad un ricorso effettivo è emersa nel caso *Ligue des Droits Humains* in cui la CGUE si è espressa sulla validità della direttiva 2016/681/UE (direttiva PNR) “*sull'uso dei dati del codice di prenotazione (PNR) a fini di prevenzione, accertamento, indagine e azione penale nei confronti dei reati di terrorismo e dei reati gravi*”. La suddetta direttiva, all'art. 6, par. 3, lett.b), sancisce che “*le autorità competenti possano trattare i dati nominativi dei passeggeri per la finalità di prevenire o accertare reati terroristici o altri reati gravi, sulla base di criteri prestabiliti*”⁴¹³. Secondo la Corte, il termine “*prestabiliti*” osta al ricorso a tecnologie di *machine learning*, in quanto tali sistemi possono modificare i criteri di valutazione, senza intervento e controllo umani. Altrimenti, il ricorso al *machine learning*, per via dell'opacità, impedirebbe di comprendere le ragioni sottostanti, con la conseguenza che gli interessati

⁴¹⁰ A. SIMONCINI, *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, in *Rivista di diritto processuale*, 2024, p. 8.

⁴¹¹ *Ibid.*, p. 9.

⁴¹² A. SILVEIRA -L. A. COELHO -M. I. COSTA *et al.*, *The Charter of Fundamental Rights of the European Union: a commentary*, JusGov, UMinho Law School, Escola de Direito da Universidade do Minho, Braga, Portogallo, 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4999183, consultato l'8 ottobre 2025, pp. 439-440.

⁴¹³ *Ibid.*, art. 6, par. 3, lett. b.

perderebbero la possibilità di contestare l'eventuale carattere discriminatorio dei risultati ottenuti. L'opacità del *machine learning* si porrebbe quindi in contrasto con il diritto ad un ricorso effettivo, sancito dall'art. 47 della CDFUE⁴¹⁴.

Queste argomentazioni, sebbene vengano svolte con specifico riferimento alla direttiva PNR, sembrano avere una rilevanza più ampia. In termini più generali, chi subisce una decisione mediante il ricorso all'IA dovrebbe avere diritto a comprendere la logica di tale decisione, al fine di sindacarne la legittimità. Ad esempio, dovrebbe essere in grado di provare l'eventuale natura discriminatoria della decisione. Nel caso *Loomis*, l'opacità dell'IA (in quel caso dovuta anche a ragioni di proprietà intellettuale) aveva reso particolarmente oneroso per l'imputato dimostrare di essere stato discriminato.

Un'ulteriore garanzia processuale suscettibile di violazione mediante il ricorso a sistemi di IA non trasparenti è l'obbligo di motivazione. L'art. 111, comma 6, Cost. sancisce che tutti i provvedimenti giurisdizionali debbano essere motivati. Con riferimento ai provvedimenti giurisdizionali suscettibili di impugnazione, tale principio è funzionale a garantire che la decisione giudiziale possa essere sindacata in sede di impugnazione, in quanto, in assenza di motivazione, non sarebbe possibile contestare le ragioni alla base del provvedimento giurisdizionale. Allo stesso tempo, tale principio non si applica soltanto a provvedimenti suscettibili di impugnazione ma si estende altresì alle sentenze della Corte costituzionale e della Cassazione. Si deve dunque ritenere che l'art. 111, comma 6, Cost. non è meramente funzionale a garantire la sindacabilità dei provvedimenti giurisdizionali in sede di impugnazione, ma serve anche a garantire la trasparenza nell'esercizio della giurisdizione⁴¹⁵.

Il ricorso in sede di motivazione a tecnologie non pienamente comprensibili potrebbe risultare in contrasto con l'art. 111, comma 6 Cost., in quanto l'opacità di tale tecnologia rende, di riflesso, non pienamente comprensibile la motivazione stessa⁴¹⁶. Gli algoritmi di giustizia predittiva possono quindi essere utilizzati dal giudice solo se sono spiegabili nella loro logica⁴¹⁷.

L'opacità dell'IA si pone anche in tensione con il diritto ad una decisione motivata ai sensi dell'art. 6 CEDU. Tale diritto, ancorché non espressamente previsto dall'art. 6, è stato riconosciuto dalla Corte EDU come componente del "*diritto ad un processo equo*", a partire dal caso *H. c. Belgio*⁴¹⁸.

⁴¹⁴ Corte di giustizia dell'Unione europea, *Ligue des Droits Humains*, C-817-19, 21 giugno 2022, parr. 194-195; E. BROUWER, *Challenging Bias and Discrimination in Automated Border Decisions*, *Ligue des droits humains and the Right to Effective Judicial Protection*, in *Verfassungsblog*, 2023, <https://verfassungsblog.de/pnr-border/>, consultato il 10 ottobre 2025; E. BROUWER, *Ligue des droits humains and the validity of the PNR Directive: Balancing individual rights and State powers in times of new technologies*, in *Common Market Review*, 2023, pp. 855-857.

⁴¹⁵ A. SIMONCINI, *op. cit.*, p. 11.

⁴¹⁶ A. SIMONCINI, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, in *Associazione italiana dei costituzionalisti (AIC)*, 2023, <https://www.rivistaaic.it/it/rivista/ultimi-contributi-pubblicati/andrea-simoncini/il-linguaggio-dell-intelligenza-artificiale-e-la-tutela-costituzionale-dei-diritti>, consultato il 10 ottobre 2025, pp. 36-37.

⁴¹⁷ A. SIMONCINI, *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, in *Rivista di diritto processuale*, 2024, p. 8.

⁴¹⁸ Corte europea dei diritti dell'uomo, *H. c. Belgio*, ric. n. 8050/80, 30 novembre 1987, par. 53; S. TRECHSEL, *human rights in criminal proceedings*, vol. 3, Academy of European Law, European University Institute, Oxford University press, Oxford, Regno Unito, 2005, pp. 102-110.

L'opacità algoritmica risulta difficilmente compatibile con la finalità che l'obbligo di motivazione persegue. La Corte EDU ha sancito che i potenziali appellanti devono essere a conoscenza della motivazione della sentenza, in quanto, altrimenti, non sarebbero in grado di presentare delle controargomentazioni in sede di impugnazione⁴¹⁹. Se il giudice nel motivare rinvia ad un sistema di IA non trasparente, le parti non sono in grado di presentare dei controargomenti in sede di impugnazione, poiché ignorano le ragioni in base alle quali il sistema è pervenuto ad una determinata conclusione.

Risulta altresì particolarmente complesso conciliare il ricorso a sistemi di IA non trasparenti con il principio del contraddittorio. Occorre distinguere tra contraddittorio “in senso debole” e contraddittorio “in senso forte”. Il primo si riferisce al principio dell'*audiatur altera pars*, secondo cui è necessario che ciascuna parte possa avere accesso alle argomentazioni e agli atti della controparte e presentare delle contro argomentazioni. Tale principio è oggi espressamente riconosciuto dall'art. 111, comma 2, Cost. e presenta un ambito di applicazione particolarmente ampio, in quanto è valido per ogni processo, e non soltanto ai fini penalistici⁴²⁰.

Con contraddittorio “in senso forte” ai sensi dell'art. 111, comma 3, Cost., ci si riferisce alla necessità che le prove siano formate in dibattimento e non acquisite unilateralmente, salvo le eccezioni previste all'art. 111, comma 5, Cost. Tale principio presenta un ambito di applicazione maggiormente ristretto, in quanto si riferisce soltanto alla formazione della prova e si applica esclusivamente al processo penale⁴²¹.

L'opacità dell'IA potrebbe porsi in tensione con il contraddittorio in senso debole. Si pensi al caso in cui il PM dovesse provare la capacità a delinquere sulla base di un sistema di IA non trasparente, che abbia accertato un rischio di recidiva particolarmente elevato. La difesa, nella sostanza, non avrebbe accesso alle argomentazioni della controparte, perché rimarrebbero ignote le ragioni in base alle quali il modello ha stimato tale rischio di recidiva, con conseguente svalutazione del principio del contraddittorio.

Risulta maggiormente difficile da valutare la compatibilità con il principio del contraddittorio qualora l'*output* prodotto dall'IA non venga allegato da una delle parti, bensì costituisca un mero supporto alla valutazione giudiziale. In dottrina, è stata avanzata una proposta secondo cui il ruolo dell'IA nel processo potrebbe configurarsi come una sorta di avvocato generale, il quale fornirebbe delle soluzioni che il giudice sarebbe tenuto a considerare ma che potrebbe disattendere, motivando⁴²². Si potrebbe sostenere che non vi sarebbero delle peculiari esigenze di contraddittorio da salvaguardare, in quanto il sistema di IA non verrebbe utilizzato da nessuna delle parti processuali ma direttamente dal giudice.

La questione assume particolare rilievo con riferimento alla possibilità di sottoporre i sistemi di IA ad un'omologazione autoritativa *ex ante* che ne accerti l'affidabilità, restando vincolata al segreto procedimentale e quindi, non violando i diritti di proprietà intellettuale. Si è sostenuto che tale soluzione sarebbe sufficiente a

⁴¹⁹ Corte europea dei diritti dell'uomo, *Hadjianastassiou c. Grecia*, ric. n. 12945/87, 16 dicembre 1992, par. 33; S. TRECHSEL, *op. cit.*, pp. 102-110.

⁴²⁰ P. TONINI -C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, pp. 36-38.

⁴²¹ *Ibid*, pp. 38-43.

⁴²² U. RUFFOLO, *La macchina sapiens come “avvocato generale” ed il primato del giudice umano*, in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 207-209.

risolvere i problemi di opacità legati alla proprietà intellettuale, in quanto, nel caso in cui un sistema di IA venga impiegato dal giudice e non dalle parti, non vi sarebbero esigenze di contraddittorio da salvaguardare⁴²³. Tale conclusione sembra quanto meno dubbia se si concepisce il principio del contraddittorio alla luce della giurisprudenza della Corte europea dei diritti dell'uomo. Anche tale principio, sebbene non sia espressamente previsto dall'art. 6 della CEDU, è stato riconosciuto dalla Corte EDU come parte del diritto ad un processo equo⁴²⁴. La Corte aveva originariamente definito il principio del contraddittorio come il diritto a conoscere e a commentare le prove e le osservazioni presentate dalla controparte⁴²⁵. Questa definizione sembrerebbe effettivamente limitare l'ambito di applicazione del contraddittorio alle sole osservazioni e prove presentate dalla controparte.

Tuttavia, la Corte EDU ne ha successivamente riconosciuto una portata più ampia. Nel caso *Lobo Machado c. Portogallo*, il ricorrente aveva lamentato di non aver ottenuto una copia dell'opinione scritta che l'Avvocato generale aveva reso alla Corte. La Corte non ha considerato rilevante se fosse possibile qualificare l'Avvocato generale come una parte. Il principio del contraddittorio si applica infatti anche alle osservazioni presentate da un organo indipendente, distinto dalle parti, come l'avvocato generale, con la conseguenza che sussiste il diritto a conoscere e a commentare tali osservazioni⁴²⁶. Il principio del contraddittorio sembrerebbe dunque applicarsi a tutte le osservazioni legittimamente presentate in processo che siano dirette ad influenzare il giudice, indipendentemente da chi le presenti.

Adottando tale interpretazione, si può concludere che l'impiego di un sistema di IA non trasparente nel processo penale si pone in tensione con il principio del contraddittorio non soltanto nel caso in cui venga utilizzato da una delle parti, ma anche qualora venga impiegato direttamente dal giudice.⁴²⁷ Richiedere che i sistemi di IA siano soggetti ad un'omologazione autoritativa *ex ante*, per quanto possa contribuire a garantirne l'affidabilità, non sembra soddisfare il principio del contraddittorio, perché alle parti sarebbe comunque precluso fornire delle controargomentazioni rispetto alle conclusioni di un sistema di IA che contribuisce ad influenzare la decisione giudiziale.

In conclusione, l'opacità algoritmica, dovuta sia alla proprietà intellettuale che al problema della *black box*, si pone in tensione rispetto a diverse garanzie processuali, quali, il diritto alla difesa (art. 24 Cost.), il diritto ad un ricorso effettivo (art. 13 CEDU e art. 47 CDFUE), e il principio del contraddittorio in senso debole (art. 111, comma 2, Cost. e art. 6 CEDU).

Tale tensione non deve però comportare una rinuncia aprioristica all'impiego dell'IA nel processo penale. L'intelligenza artificiale spiegabile (*Explainable AI*), può, almeno in alcuni casi, aiutare le parti nella

⁴²³ *Ibid*, p. 221.

⁴²⁴ Corte europea dei diritti dell'uomo, *Belziuk c. Polonia*, ric. n. 45/1997 e 829/1035, 25 marzo 1998, par. 37; S. TRECHSEL *op. cit.*, pp. 81-117.

⁴²⁵ Corte europea dei diritti dell'uomo, *Laukkannen e Manninen c. Finlandia*, ric. n. 50230/1999, 28 maggio 2002, par. 34; S. TRECHSEL *op. cit.*, pp. 81-117.

⁴²⁶ Corte europea dei diritti dell'uomo, *Lobo Machado c. Portogallo*, ric. n. 15764/89, 20 febbraio 1996, par. 31.

⁴²⁷ V. STEPONENAITE -P. VALCKE, *Judicial Analytics on Trial: An Assessment of Legal Analytics in Judicial Systems in the Light of the Right to a Fair Trial*, in *Maastricht Journal of European and Comparative Law*, Vol. 27, n. 6, SAGE Publications, 2020, pp. 766-767.

comprensione di tali sistemi. Occorre quindi valutare caso per caso l'idoneità del sistema di IA a soddisfare le garanzie processuali di rilevanza costituzionale e sovranazionale.

Per quanto riguarda l'opacità dovuta alla proprietà intellettuale, è invece necessario adottare un approccio che valuti caso per caso la portata dei diritti delle parti, senza far prevalere aprioristicamente la proprietà intellettuale, come nel caso *Dun & Bradstreet*.

3.3) I pregiudizi cognitivi

3.3.1) Introduzione

Un altro rischio da tenere presente concerne il pericolo che i soggetti del procedimento penale (i giudici, la p.g. o il PM), nell'interazione con i sistemi di IA, siano affetti da pregiudizi (*bias*) cognitivi che ne alterino la capacità di effettuare delle valutazioni corrette.

Occorre fare una premessa sulla differenza tra *bias* cognitivi e *bias* algoritmici. I primi compromettono la capacità dell'essere umano di giungere a valide conclusioni, spesso a causa di trappole conoscitive dovute alla velocità con cui determinate decisioni vengono prese (le cosiddette euristiche)⁴²⁸.

Un noto esempio è la fallacia di Gabler che consiste nella credenza irrazionale secondo cui un evento antecedente condiziona la probabilità di un evento futuro, nonostante i due eventi siano indipendenti⁴²⁹. Ad esempio, il giudizio positivo su di uno studente che in passato è andato bene può portare il docente a sopravvalutare l'esito dell'esame presente.

Un altro esempio è la fallacia del tasso di base che consiste nell'ignorare o sottovalutare la frequenza di un evento nella popolazione (il c.d. tasso di base), concentrandosi su informazioni specifiche prive di rilevanza statistica, come nel caso di una paziente che, dopo che gli venga diagnosticata una determinata malattia, pensi che la patologia sia maggiormente frequente nella popolazione di quanto non lo sia in realtà⁴³⁰.

I *bias* algoritmici riguardano invece le distorsioni sistematiche nel comportamento o nei risultati di un sistema automatizzato, con delle possibili conseguenze discriminatorie nei confronti di alcune categorie.⁴³¹

In dottrina, è stato evidenziato che i pregiudizi cognitivi, di cui i giudici potrebbero essere vittime nell'interazione con l'IA, rischiano di compromettere il rispetto del diritto ad un tribunale imparziale, di cui all'art. 6 della CEDU.⁴³² La Corte EDU ha sancito che l'imparzialità implica che il giudice non sia affetto da *bias* o pregiudizi.⁴³³ Fino ad ora, la Corte EDU ha sancito la violazione del principio di imparzialità con

⁴²⁸ M. DI FLORIO, *Calculate Criminal Law? Criticità nell'uso degli algoritmi di pericolosità sociale*, in *La Legislazione Penale*, 2023, <https://www.la legislazione penale.eu/wp-content/uploads/2023/03/DI-FLORIO-Definitivo-rev..pdf>, consultato il 20 ottobre 2025, p. 14.

⁴²⁹ M. KOVIC -S. KRISTIANSEN, *The Gambler's Fallacy Fallacy (Fallacy)*, in *Journal of risk research*, Routledge, Taylor & Francis Group, 2017, p. 1.

⁴³⁰ M. BAR HILLER, *The base rate fallacy in probability judgements*, *Acta psychologica*, 1980, pp. 211-214.

⁴³¹ V. par. 3.1.1/3.1.2.

⁴³² V. STEPONENAITE -P. VALCKE, *op. cit.*, pp. 770-771.

⁴³³ Corte europea dei diritti dell'uomo, *Piersack c. Belgio*, 1° ottobre 1982, ric. n. 8692/79, parr. 30-31.

riferimento a pregiudizi di natura razziale e non relativamente a *bias* cognitivi⁴³⁴. Non si può però escludere che, in futuro, non soltanto i pregiudizi di stampo razziale, ma anche i *bias* cognitivi, legati all'interazione tra l'uomo e l'intelligenza artificiale, vengano considerati dalla Corte EDU come idonei a violare il diritto ad un tribunale imparziale di cui all'art. 6 CEDU.

3.3.2) La *MC Namara Fallacy*

Un primo pregiudizio cognitivo di cui i soggetti del procedimento penale potrebbero essere vittime è la cosiddetta *MC Namara Fallacy*: la tendenza ad affidarsi in modo eccessivo a misure quantitative, senza considerare elementi che non possono essere misurati. Tale fallacia può essere riassunta con la frase: se non può essere misurato, non è importante⁴³⁵. Il nome viene da Robert S. MC Namara, segretario alla difesa degli Stati Uniti dal 1961 al 1968, il quale venne aspramente criticato per la tattica e la strategia adottata nella guerra in Vietnam, in quanto aveva adottato un metodo incentrato sulla misurazione e sulle metriche, finendo per sottovalutare la rilevanza di elementi che non potevano essere misurati⁴³⁶.

La *MC Namara Fallacy* è stata discussa prevalentemente in ambiti diversi dal diritto, come la difesa militare o la medicina⁴³⁷. Qualora l'IA venisse introdotta nel processo penale, questa fallacia avrebbe altresì delle implicazioni in ambito legale. L'IA predittiva solitamente fornisce delle previsioni in termini quantitativi. Ad esempio, il rischio di recidiva stimato da un *risk assessment tool* come *COMPAS* produce un punteggio (*risk score*) espresso in termini numerici. I giudici, i PM e la p.g. potrebbero fidarsi ciecamente dei dati quantitativi forniti dal sistema di IA, finendo per sottovalutare l'importanza di elementi di natura meramente qualitativa che l'IA potrebbe non essere in grado di valutare.

3.3.3) L'*automation bias*

Oltre alla *McNamara Fallacy*, occorre considerare l'*automation bias*, generalmente definito come la tendenza a fare ricorso all'automazione anziché ricercare ed elaborare attentamente tutte le informazioni necessarie. Una delle conseguenze di tale pregiudizio è che chi assume le decisioni tende a seguire automaticamente il suggerimento offerto dalla macchina, senza verificarne la validità sulla base delle informazioni disponibili⁴³⁸.

⁴³⁴ A titolo esemplificativo, si possono citare i seguenti casi: Corte europea dei diritti dell'uomo, *Sanders c. Regno Unito*, 9 maggio 2000, ric. n. 34129/96, par. 18; Corte europea dei diritti dell'uomo, *Gregory c. Regno Unito*, 25 febbraio 1997, ric. n. 22299/93, par. 96.

⁴³⁵ S. O. MAHONY, *Medicine and the McNamara fallacy*, in *National Library of Medicine*, Edimburgo, Scozia, 2017, <https://pubmed.ncbi.nlm.nih.gov/29465108/>, consultato il 20 ottobre 2025.

⁴³⁶ R. BRENNER, *The MC Namara Fallacy*, <https://chacocanyon.com/pointlookout/230222.shtml>, consultato il 20 ottobre 2025.

⁴³⁷ S. O. MAHONY, op. cit.

⁴³⁸ K. MOSIER L. SKITKA, *Automation Use and Automation Bias* in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1999, pp. 344-348.

I giudici, i PM e la p. g. potrebbero limitarsi a considerare come valide le previsioni fornite dall'IA predittiva, senza valutare criticamente le stesse, con la conseguenza che la decisione verrebbe delegata, nella sostanza, dall'uomo alla macchina⁴³⁹.

Anche questo tema è già emerso, seppur non in questi termini, nel caso *Loomis*. Il *Wisconsin Department of Corrections (DOC)* (l'ente responsabile per la gestione dei detenuti in Wisconsin) aveva suggerito di utilizzare il proprio giudizio professionale nel valutare il *risk score*, stimando che, in media, la previsione algoritmica non dovrebbe essere accolta nel 10% dei casi, in modo tale da evitare automatismi⁴⁴⁰. È quindi emersa la necessità di non dare seguito, in modo automatico, al suggerimento offerto dalla macchina, data la fallibilità dell'intelligenza artificiale. La stessa Corte Suprema del Wisconsin ha cercato di mitigare il rischio che i giudici seguano ciecamente i suggerimenti offerti dall'IA, precisando che il ricorso a strumenti di valutazione del rischio (*risk assessment tools*) è legittimo soltanto qualora non costituiscano il fattore determinante ai fini della determinazione della pena, ma si limitino a rappresentare un'informazione aggiuntiva, da valutare congiuntamente con altri fattori⁴⁴¹. Tale parte della sentenza si presta a critiche, essendo quanto meno dubbio che tale prescrizione sia realmente efficace ad evitare che i giudici siano vittime dell'*automation bias*, in quanto potrebbe tradursi in una garanzia meramente formale.

Anche un procedimento caratterizzato dal coinvolgimento di un essere umano (*human in the loop*) può divenire, nella sostanza, un procedimento quasi automatico, in cui il contributo offerto dall'essere umano è quasi nullo. Questa situazione può verificarsi a causa della scarsa esperienza da parte dell'agente umano coinvolto nel procedimento. Oppure, il contributo umano potrebbe risultare scarsamente rilevante per la mancanza del tempo necessario per controllare il funzionamento del sistema di IA. Potrebbero anche mancare la motivazione o gli incentivi in tal senso⁴⁴².

In uno studio, è stato simulato un processo penale, caratterizzato dall'interazione con l'intelligenza artificiale, in cui i partecipanti hanno ricevuto un suggerimento di IA errato, finalizzato alla valutazione della colpevolezza degli imputati⁴⁴³. Sono stati confrontati i risultati tra chi aveva ricevuto il suggerimento algoritmico prima di emettere il verdetto e chi aveva ricevuto tale supporto soltanto successivamente ed è emerso che, quando il supporto IA è errato, questo influenza negativamente il giudizio. L'accuratezza delle decisioni con supporto IA errato è dunque inferiore all'accuratezza delle decisioni senza supporto IA, poiché i partecipanti che avevano ricevuto un suggerimento algoritmico erroneo hanno commesso maggiori errori dei partecipanti che non hanno ricevuto alcun suggerimento, avendo preso la decisione in autonomia, senza il

⁴³⁹ L. NOTARO, *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021, <https://arpi.unipi.it/handle/11568/1147326>, consultato il 13 luglio 2025, p. 68; F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in G. BALBI -F. DE SIMONE -A. ESPOSITO et al., *Diritto penale e intelligenza artificiale: nuovi scenari*, Torino, Giappichelli Editore, 2022, p. 22.

⁴⁴⁰ Corte Suprema del Wisconsin, *State v. Loomis* op. cit., par. 70.

⁴⁴¹ *Ibid.*, par. 72.

⁴⁴² U. AGUDO -K. G. LIBERAL -M. ARRESE, et al., *The impact of AI errors in a Human-in-the-loop Process*, in *Cognitive Research: Principles and Implications*, 2024, <https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-023-00529-3>, consultato il 22 ottobre 2025, pp. 13-14.

⁴⁴³ *Ibid.*, p. 4.

supporto dell'IA. Questo studio suggerisce che, se il sistema di IA giunge ad una conclusione errata, anche la decisione umana risulta peggiore della decisione che sarebbe stata adottata in assenza del supporto IA.⁴⁴⁴

Coloro che hanno emesso il giudizio prima di conoscere il suggerimento IA hanno mostrato maggiore accuratezza, rispetto a chi ha ricevuto tale *input* prima di emettere il verdetto⁴⁴⁵. Tale risultato sembrerebbe suggerire che sarebbe preferibile che l'essere umano si formi una propria opinione prima di sentire il suggerimento reso dalla macchina, che non il contrario.

La portata dell'*automation bias* può essere altresì influenzata da pregiudizi, quali gli stereotipi di natura razziale.⁴⁴⁶ In una ricerca, è stato mostrato come lo stesso suggerimento offerto dall'IA possa avere un impatto differente a seconda del gruppo razziale di appartenenza del destinatario della decisione. Nel Kentucky, in presenza di un *risk score* basso, i giudici erano soliti deviare dal suggerimento algoritmico con maggiore frequenza qualora l'imputato fosse afroamericano, rispetto al caso in cui l'imputato fosse bianco⁴⁴⁷.

Questa tendenza a dare maggiore peso alle raccomandazioni che confermano i propri pregiudizi viene denominata conformità selettiva (*selective adherence*). In un altro esperimento, l'IA suggeriva ai partecipanti se rinnovare o meno il contratto a degli insegnanti. Il suggerimento di non rinnovare il contratto a dei docenti appartenenti a minoranze etniche veniva accolto con maggiore frequenza rispetto all'ipotesi in cui il medesimo suggerimento riguardava insegnanti caucasici⁴⁴⁸. Queste ricerche mostrano come gli stereotipi, spesso di natura razziale, possono essere amplificati dall'interazione con la macchina a causa dell'*automation bias*, rendendo ancora più concreto il pericolo che le minoranze vengano discriminate.

In sintesi, l'affidamento non critico ai suggerimenti dell'IA potrebbe compromettere l'accuratezza delle decisioni adottate da giudici, PM e p.g. e amplificare pregiudizi esistenti, aumentando anche il rischio di discriminazioni nei confronti delle minoranze.

3.3.4) Le garanzie presenti nel diritto vigente

Quanto precedentemente esposto non implica che i *bias* cognitivi siano necessariamente una ragione per rinunciare a priori al ricorso all'IA nel processo penale. Occorre, invece, adottare garanzie efficaci per evitare che, nel singolo caso i giudizi spettanti ai giudici, la p.g. o i PM siano viziati da tali pregiudizi. Un primo passo fondamentale consiste nell'invitare i giudici, i PM e la p.g. a non aderire acriticamente alla stima resa dall'IA. Tali soggetti del procedimento penale devono essere consapevoli che la macchina può sbagliare e che il suggerimento offerto deve essere sottoposto ad un'attenta valutazione.

⁴⁴⁴ *Ibid*, pp. 4-9.

⁴⁴⁵ *Ibid*, pp. 5-12.

⁴⁴⁶ T. KAZIM -J. TOMLINSON, *Automation Bias and the Principles of Judicial Review* in *Judicial Review*, 2023, p. 9.

⁴⁴⁷ A. ALBRIGHT, *If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions* in *University of Toronto Law Journal*, vol. 73, 2023, pp. 1-6.

⁴⁴⁸ S. ALON-BARKAT -M. BUSUIOC, *Human-AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice*, in *Journal of Public Administration Research and Theory*, vol. 33, 2023, pp. 153-169.

Lo stesso regolamento 2024/1689/UE (AI Act) si muove in questa direzione. L'art. 4 prevede che i fornitori e i *deployer* (gli utilizzatori del sistema IA) sono tenuti ad adottare delle *“misure per garantire nella misura del possibile un livello sufficiente di alfabetizzazione in materia di IA del loro personale nonché di qualsiasi altra persona che si occupa del funzionamento e dell'utilizzo dei sistemi di IA per loro conto, prendendo in considerazione le loro conoscenze tecniche, la loro esperienza, istruzione e formazione, nonché il contesto in cui i sistemi di IA devono essere utilizzati, e tenendo conto delle persone o dei gruppi di persone su cui i sistemi di IA devono essere utilizzati”*. Nel processo penale, l'alfabetizzazione in materia di IA (*AI literacy*) potrebbe consistere nel formare i soggetti del procedimento penale sui rischi che l'IA può comportare in questo settore, con particolare riferimento all'*automation bias*.

Ulteriori garanzie sono poi offerte dall'art. 14 dell'AI Act sull'obbligo di sorveglianza umana per i sistemi ad alto rischio. In particolare, l'art. 14, par. 4 richiede che *“il sistema di IA ad alto rischio è fornito al deployer in modo tale che le persone fisiche alle quali è affidata la sorveglianza umana abbiano la possibilità, ove opportuno e proporzionato, di[...]”* essere consapevoli *“della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio (“distorsione dell'automazione”)”* (lett.b). La versione in lingua italiana usa il termine *“distorsione dell'automazione”* per tradurre l'espressione *“automation bias”*, presente nella versione in inglese. L'*automation bias* compare dunque espressamente tra i rischi, di cui, a norma dell'AI Act, devono essere a conoscenza coloro che sorvegliano su di un sistema IA.

L'art. 14, par. 4, lett d), prevede che la persona fisica a cui è affidata la sorveglianza del sistema di IA deve essere in grado di *“decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio”*. Chi è deputato alla sorveglianza del sistema IA deve quindi essere in grado di esercitare la propria autorità sul sistema, eventualmente discostandosi dal suggerimento offerto⁴⁴⁹.

Si richiede altresì che venga garantita la comprensione da parte di chi è deputato a sorvegliare sul funzionamento di un sistema di IA⁴⁵⁰. L'art. 14, par. 4, lett. a), si riferisce alla *“capacità di comprendere correttamente le capacità e i limiti pertinenti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento, anche al fine di individuare e affrontare anomalie, disfunzioni e prestazioni inattese”*. Deve essere possibile, *“interpretare correttamente l'output del sistema di IA ad alto rischio, tenendo conto ad esempio degli strumenti e dei metodi di interpretazione disponibili”*. Anche questi obblighi, se declinati in ambito giudiziario, afferiscono alla necessità che i giudici, la p.g. e i PM siano formati in maniera adeguata.

⁴⁴⁹ M. FINK, *Human Oversight under art. 14 of the EU AI Act*, Leiden University, Europa Institute, Católica Global School of Law, Central European University (CEU) - Department of Legal Studies, Leiden, Olanda, 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5147196#:~:text=Melanie%20Fink,-Leiden%20University%20-%20Europa&text=Human%20oversight%20in%20the%20AI,building%20trust%20i, consultato il 22 ottobre 2025, p. 11.

⁴⁵⁰ *Ibid*, p. 12.

Sul versante del diritto italiano, questa esigenza è emersa nell'art. 15, comma 4, della legge 23 settembre 2025, n. 132, "*recante disposizioni e delega al Governo in materia di intelligenza artificiale*". Il Ministro della Giustizia "*promuove attività didattiche sul tema dell'intelligenza artificiale e sugli impieghi dei sistemi di intelligenza artificiale nell'attività giudiziaria, finalizzate alla formazione digitale di base e avanzata, all'acquisizione e alla condivisione di competenze digitali, nonché alla sensibilizzazione sui benefici e rischi*". Tali attività didattiche andranno promosse nell'ambito delle linee programmatiche sulla formazione dei magistrati disposte dal Ministro della Giustizia, a norma dell'art. 12, comma 1, lett. a), del d.lgs. 30 gennaio 2006, n. 26 che disciplina la Scuola superiore della magistratura, nonché il tirocinio e la formazione degli uditori giudiziari e la formazione dei magistrati.

Per quanto l'*automation bias* non sia esplicitato dalla disposizione, sarebbe auspicabile che i magistrati vengano formati anche con riferimento a tale *bias* cognitivo, affinché le loro valutazioni non siano viziate in tal senso.

Con specifico riferimento alle misure cautelari, il rischio di *automation bias* potrebbe risultare indirettamente attenuato in virtù dell'art. 292, comma 2, c.p.p. Tale disposizione prevede una motivazione rafforzata nell'ordinanza che dispone la misura cautelare, pena la nullità rilevabile anche d'ufficio. L'ordinanza deve contenere "*l'esposizione e l'autonoma valutazione delle specifiche esigenze cautelari e degli indizi che giustificano in concreto la misura disposta, con l'indicazione degli elementi di fatto da cui sono desunti e dei motivi per i quali essi assumono rilevanza, tenuto conto anche del tempo trascorso dalla commissione del reato*" (lett. c.). Le Sezioni Unite hanno specificato che non sono valide motivazioni *per relationem* appiattite sulla richiesta del pubblico ministero⁴⁵¹.

Qualora l'IA predittiva venisse impiegata per fondare una misura cautelare, l'*output* algoritmico verrebbe plausibilmente indicato nella richiesta del pubblico ministero *ex art.* 291, comma 1, c. p. p., per fondare la misura. Il giudice non potrebbe dunque limitarsi a motivare *per relationem* sulla base della previsione algoritmica, in quanto, altrimenti, la valutazione delle specifiche esigenze cautelari non sarebbe autonoma rispetto alla richiesta del PM, con conseguente violazione dell'art. 292, comma 2, lett. c), c.p.p.

L'autonomia originariamente concepita per evitare appiattimenti sui pubblici ministeri dovrà allora essere assicurata anche con riferimento ai sistemi di intelligenza artificiale di volta in volta adottati in sede cautelare⁴⁵². Imponendo una valutazione autonoma anche rispetto all'*output* prodotto dall'IA, si riduce il rischio di adesione acritica al suggerimento offerto dalla macchina, e quindi, di incorrere nell'*automation bias*. Anche al di fuori dell'ambito delle misure cautelari, risulta necessario che il giudice, il PM o la p.g. individuino in maniera autonoma i fatti a fondamento delle loro decisioni in maniera autonoma rispetto ai sistemi IA. Adottare una motivazione rafforzata, anziché una motivazione *per relationem*, potrebbe quindi risultare uno degli strumenti maggiormente efficaci per contrastare tale pregiudizio cognitivo.

⁴⁵¹ Sez. Un., 21 giugno 2000, n. 17.

⁴⁵² P. P. PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi* in *Archivio Penale*, 2022, <https://archiviopenale.it/intelligenza-artificiale-e-giustizia-penale-una-lettura-attraverso-i-principi/articoli/34325>, consultato il 09/09/2025, pp. 15-16.

In definitiva, non mancano strumenti nel diritto vigente per contrastare tale pregiudizio cognitivo. Le garanzie più efficaci sono la consapevolezza circa il rischio di incorrere nell'*automation bias* e l'obbligo di motivare in maniera autonoma rispetto alla macchina.

3.4) Limiti dell'AI Act nel tutelare i diritti fondamentali

Un altro rischio da considerare è che la tutela dei diritti fondamentali tramite l'AI Act si riveli, in concreto e soprattutto in sede di attuazione, non sufficientemente incisiva.

Con riferimento ai rischi enunciati nei precedenti paragrafi, si è fatto spesso riferimento all'AI act, in quanto tale regolamento potrebbe contribuire a mitigare molti dei rischi fino ad ora analizzati. Ad esempio, il diritto ad una spiegazione dei singoli processi decisionali, di cui all'art. 86, par. 1, potrebbe essere funzionale a superare i problemi di opacità legati ai diritti di proprietà intellettuale, con la conseguenza di tutelare il diritto alla difesa dell'imputato o indagato giudicato sulla base di un sistema IA. Oppure, si è visto che il requisito della rappresentatività dei dati di cui all'art. 10, par. 4, del suddetto regolamento potrebbe mitigare il rischio di discriminazione algoritmica, in osservanza della tutela predisposta dall'art. 21 CDFUE⁴⁵³.

Lo scopo dell'AI Act consiste nel migliorare il funzionamento del mercato interno, garantendo, “*nel contempo, un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea*”, nonché promuovere l'innovazione (considerando 1 e art. 1 dello Stesso regolamento).

Tale regolamento si pone dunque come strumento a tutela di diritti fondamentali, di rilevanza nazionale e sovranazionale, che rischiano di essere minacciati dall'intelligenza artificiale. Sotto questo profilo, l'AI Act recepisce le preoccupazioni sollevate dall'Agenzia per i Diritti Fondamentali dell'Unione europea (FRA) quanto al pericolo che l'IA abbia un impatto negativo sui diritti previsti dalla CDFUE. Tale agenzia ha evidenziato dei rischi in merito alla dignità umana (art. 1), al rispetto della vita privata e della vita familiare (art. 7), alla protezione dei dati di carattere personale (art. 8), alla non-discriminazione (art. 21), al diritto alla sicurezza sociale e all'assistenza sociale (art. 34), al diritto ad una buona amministrazione (art. 41) e al diritto a un ricorso effettivo e a un giudice imparziale (art. 47)⁴⁵⁴.

In primo luogo, l'AI Act mira a tutelare i diritti fondamentali, vietando, all'art. 5 determinate pratiche contrarie sia ai valori dell'Unione che agli stessi diritti sanciti dalla CDFUE (considerando 28). Tale tutela si esplica poi nella previsione di determinati obblighi previsti se il sistema è ad alto rischio. Non a caso, affinché un sistema sia classificabile come ad alto rischio, non è sufficiente che tale sistema sia contenuto nell'elenco tassativo di cui all'allegato I del regolamento. Risulta altresì necessario che l'IA presenti “*un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche*” (art. 6, parr. 1-3). Da quanto fino ad ora esposto, emerge la centralità che i diritti fondamentali assumono nel regolamento.

⁴⁵³ V. parr.3.1.3 e 3.1.4.

⁴⁵⁴ Agenzia dell'Unione Europea per i diritti fondamentali, *Getting the Future Right*, 2021, Lussemburgo, 2020, pp. 60-81.

Nonostante tale tutela, occorre considerare il rischio, che, in concreto, soprattutto in sede di attuazione, l'AI Act si riveli non sufficientemente incisivo.

Una prima ragione per dubitare dell'efficacia della tutela dei diritti fondamentali riguarda il ruolo svolto dalle cosiddette organizzazioni europee di normazione (*European Standardisation Organizations, ESOs*). A norma dell'art. 40, par. 1, i sistemi ad alto rischio che seguono gli standard stabiliti dalle organizzazioni europee di normazione si presumono, salvo prova contraria, conformi alla maggior parte delle prescrizioni previste dal regolamento per i sistemi ad alto rischio⁴⁵⁵. Questo metodo, già ampiamente sperimentato dal legislatore europeo in altre occasioni, è denominato *New Approach* e consiste nell'armonizzare gli standard industriali necessari per essere conformi al diritto dell'Unione europea, garantendo allo stesso tempo la flessibilità necessaria per favorire l'innovazione⁴⁵⁶.

In dottrina è stato però sostenuto che gli organismi europei di normazione, per quanto siano competenti sugli aspetti tecnico-industriali, potrebbero essere privi del necessario *expertise*, in materia di diritti fondamentali, essendo composti solitamente da personale che risulta privo di tale formazione.⁴⁵⁷ Ancora più significativo è il rischio che tali organismi siano assoggettati ad interessi privati (*regulatory capture*). Infatti, l'Unione europea delega tale potere a imprese private, prive della legittimità democratica propria delle istituzioni pubbliche⁴⁵⁸.

Tra i vari portatori di interesse coinvolti nel procedimento decisionale adottato da tali organismi di normazione, le realtà più modeste sono spesso sottorappresentate⁴⁵⁹. Ad esempio, le organizzazioni della società civile a tutela dei diritti fondamentali potrebbero avere delle difficoltà ad incidere in maniera significativa in tale procedimento decisionale.

Un'altra ragione per cui il regolamento potrebbe risultare insufficiente riguarda la scelta di adottare il cosiddetto *product safety approach*, con riferimento ad un settore in cui rilevano i diritti fondamentali. Con *product safety*, si fa riferimento a quella parte del diritto dei consumatori funzionale a garantire la sicurezza dei prodotti tramite prescrizioni normative. L'obiettivo consiste nel mitigare le asimmetrie informative, a causa delle quali un consumatore potrebbe acquistare un prodotto che risulta lesivo per la propria sicurezza⁴⁶⁰. L'AI Act adotta il *product safety approach*, in quanto mira a garantire la sicurezza dei prodotti, da un lato proibendo pratiche inaccettabili (art. 5) e dall'altro, prevedendo degli obblighi per i sistemi ad alto rischio, volti a mitigare i rischi legati alla sicurezza di tali prodotti (artt. 8-27).

⁴⁵⁵ H. POUGET -K. HOLTMAN -T. EMBORG, *EU Artificial Intelligence Act, Standard Setting*, 2025, <https://artificialintelligenceact.eu/standard-setting-overview/>, consultato il 27 ottobre 2025.

⁴⁵⁶ S.DE VRIES -O. KANEVSKAIA -R.DE JAGER, *Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation*, in *European Papers*, vol. 8, n. 2, 2023, p. 596.

⁴⁵⁷ *Ibid*, p. 602.

⁴⁵⁸ M. ALMADA -N. PETIT, *The EU AI Act: between the rock of product safety and the hard place of fundamental rights*, in *Common market law review*, vol. 62, 2025, pp. 1114-1118.

⁴⁵⁹ S.DE VRIES -O. KANEVSKAIA -R.DE JAGER, *op. cit.*, p. 602.

⁴⁶⁰ J. RUOHONEN, *A Review of Product Safety Regulations in the European Union*, in *International Cybersecurity Law Review*, 2022, pp. 345-347.

Tale approccio risulta difficile da conciliare con la protezione dei diritti fondamentali, per diverse ragioni. In primo luogo, coerentemente con il *product safety approach*, il regolamento fornisce una definizione di rischio che presuppone la misurabilità del danno: “*il rischio è dato dalla combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso*” (art. 3, par. 2). Il danno conseguente alla violazione di diritti fondamentali risulta spesso difficile da misurare in termini quantitativi⁴⁶¹. Ad esempio, il danno conseguente alla lesione del diritto alla difesa, dovuto all’opacità del sistema IA impiegato nel processo penale, può risultare difficile da quantificare. La mitigazione di un rischio relativo ad un possibile danno non pienamente misurabile, come quello relativo a diritti fondamentali, potrebbe quindi non risultare agevole.

L’approccio basato sulla sicurezza del prodotto non è poi adatto a proteggere i diritti fondamentali, perché si fonda su una logica di soddisfacimento minimo, in quanto, affinché i sistemi IA siano conformi alla normativa, risulta sufficiente il soddisfacimento di determinati standard. La conformità al regolamento non viene invece pregiudicata dalla circostanza che sia possibile adottare delle misure aventi una minore incidenza sui diritti fondamentali. Tale impostazione non è compatibile con il principio di necessità desumibile dall’art. 52, par. 1, CDFUE che prevede che le limitazioni dei diritti fondamentali devono essere proporzionate. Secondo la CGUE, la proporzionalità implica che la limitazione dei diritti fondamentali deve essere necessaria: tra le opzioni disponibili, occorre privilegiare quella che comporta una minore incidenza sui diritti fondamentali⁴⁶². La logica del soddisfacimento minimo non è compatibile con tale principio, in quanto risulta sufficiente soddisfare lo standard previsto dal regolamento, mentre non vi è alcun obbligo ulteriore di adottare, tra le opzioni disponibili, le misure che comportino il minore impatto sui diritti fondamentali⁴⁶³.

Un ulteriore profilo critico si rinviene nella legge 23 settembre 2025, n. 132, con cui il legislatore italiano ha definito le autorità competenti in materia di AI Act in Italia. Il legislatore italiano ha scelto di selezionare l’Autorità Nazionale per la Cybersicurezza (ACN) quale autorità di vigilanza sul mercato. Tale autorità “*è responsabile per la vigilanza, ivi incluse le attività ispettive e sanzionatorie, dei sistemi di intelligenza artificiale, secondo quanto previsto dalla normativa nazionale e dell’Unione europea*” (art. 20, comma 1, lett.b), della suddetta legge). Tale scelta è stata criticata da una parte della dottrina in quanto si è dubitato che tale autorità eserciterà il proprio potere in maniera indipendente e imparziale, secondo quanto richiesto dall’art. 70, par. 1, dell’AI Act⁴⁶⁴. L’ACN non è un’autorità amministrativa indipendente ma un’agenzia governativa, a differenza, ad esempio, dell’Autorità Garante per la Protezione dei Dati Personali a cui è affidata la *governance* del GDPR⁴⁶⁵. Se l’ACN non dovesse soddisfare le garanzie di indipendenza necessarie, le norme a tutela dei diritti fondamentali, previste nel regolamento, potrebbero risultare prive di un’attuazione sufficientemente incisiva.

⁴⁶¹ M. ALMADA -N. PETIT, *op. cit.*, pp. 106-107.

⁴⁶² Corte di giustizia dell’Unione europea, *La Regina contro il Ministro dell’Agricoltura, della Pesca e dell’Alimentazione e il Segretario di Stato per la Salute, su ricorso di Fedesa e altri*, C-381/88, 13 novembre 1990, par. 113.

⁴⁶³ M. ALMADA -N. PETIT, *op. cit.*, p. 109.

⁴⁶⁴ A. PAJNO, *La governance dell’IA tra regolamento europeo e disciplina nazionale*, in *AStrid rassegna*, n.13/24, 2024.

⁴⁶⁵ M. RABITTI -F. BASSAN, *L’applicazione dell’AI Act in Italia e la tutela del consumatore. Il ruolo delle autorità indipendenti*, in *FEDERALISMI.IT*, fasc. 30, 2024, pp. 250-265.

In conclusione, occorre tenere presente che l'AI Act potrebbe presentare dei limiti nella tutela dei diritti fondamentali. L'entità di tale rischio non è ancora pienamente verificabile, in quanto la progressiva entrata in vigore del regolamento, a norma dell'art. 113, non è ancora terminata. Il regolamento è infatti soltanto parzialmente entrato in vigore, mentre sarà completamente applicabile in tutta l'Unione europea soltanto a partire dal 2 agosto 2026 (art. 113), salvo eventuali proroghe.

3.5) Dal diritto penale del fatto al diritto penale d'autore?

3.5.1) Introduzione

Un altro rischio che occorre considerare concerne la possibilità di una deriva verso il diritto penale d'autore, di natura illiberale, contrapposto al diritto penale del fatto, di stampo liberale. Con diritto penale del fatto, si intende che il reato corrisponde ad un fatto dannoso, cioè a un comportamento che mette in pericolo o lede beni individuali o collettivi. Storicamente, con l'affermarsi del diritto penale del fatto, si è superata la precedente corrispondenza tra reato e peccato, intesa come la repressione di comportamenti o anche meri atteggiamenti interiori, puniti non in quanto dannosi, bensì in quanto contrastanti con la legge divina. Il diritto penale del fatto ha assunto un ruolo predominante soprattutto grazie alla cosiddetta Scuola Classica a partire dall'Ottocento ed è stato fatto proprio dal legislatore sia nella codificazione del 1889 che del 1930⁴⁶⁶.

Tra la fine dell'Ottocento e l'inizio del Novecento, si è successivamente affermata la Scuola Positiva, secondo cui il diritto penale avrebbe primariamente ad oggetto persone socialmente pericolose (ad esempio, i delinquenti abituali), anziché fatti di reato. Ne conseguirebbe che non sarebbe neppure necessario che il legislatore individui tassativamente i fatti tipici costituenti reato, essendo piuttosto sufficiente individuare i tipi d'autore meritevoli di sanzione penale. Tale tesi è ampiamente superata sia dalla dottrina che dal legislatore, in ragione dei poteri incontrollabili che verrebbero affidati al giudice penale nel limitare la libertà personale senza dover accertare la sussistenza di un fatto tipico⁴⁶⁷.

Questa concezione risulta chiaramente contraria al principio del *nullum crimen sine lege, nulla poena sine lege*, secondo cui una persona può essere punita soltanto in presenza dei presupposti fissati dalla legge e soltanto nei limiti previsti dalla legge. Il principio in esame trova oggi espresso riconoscimento nell'art. 25, comma 2, Cost., secondo cui “nessuno può essere punito se non in forza di una legge che sia entrata in vigore prima del fatto commesso”⁴⁶⁸.

Nel vigente codice penale italiano del 1930, anche quando la legge attribuisce rilievo alla pericolosità sociale, il fatto di reato rimane un presupposto indefettibile. Ad esempio, ai sensi degli artt. 202 e 203 c.p., la mera pericolosità sociale non è sufficiente a fondare una misura di sicurezza, essendo altresì necessaria la commissione di un fatto, previsto dalla legge come reato o come quasi reato (ossia quelle condotte umane,

⁴⁶⁶ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022, pp. 5-7.

⁴⁶⁷ *Ibid*, pp. 7-9.

⁴⁶⁸ *Ibid*.

che, pur non essendo esplicitamente qualificate come reato, il legislatore considera idonee a giustificare una misura di sicurezza)⁴⁶⁹.

Con riferimento al ricorso all'intelligenza artificiale nel processo penale, un'autorevole dottrina ha evidenziato il rischio che l'utilizzo dei *risk assessment tool* apra la strada ad una forma di inaccettabile determinismo penale, secondo cui il comportamento dell'imputato è determinato da cause naturali, sociali, o psicologiche e non è, quindi, espressione del libero arbitrio. La conseguenza sarebbe che dal diritto penale del fatto, costituzionalmente fondato nell'art. 25, comma 2 Cost., si passi ad un inaccettabile diritto penale del profilo d'autore, in cui la pericolosità del soggetto viene desunta in via esclusiva dagli schemi comportamentali del soggetto, anziché dal fatto commesso⁴⁷⁰.

Occorre inoltre considerare che gli algoritmi predittivi della recidiva si basano su fattori quali, ad esempio, la storia criminale, l'età, il sesso, lo stato civile o l'abuso di sostanze stupefacenti. Tali fattori possono essere utili a comprendere la personalità dell'imputato, ma non riguardano specificamente il fatto oggetto dell'imputazione. Per tale ragione, l'imputato potrebbe essere giudicato in base a chi è e non in base al fatto commesso. Un simile risultato sarebbe in contrasto non soltanto con il *nullum crimen sine lege, nulla poena sine lege*, ma anche con il principio di materialità (*nullum crimen sine actione*), secondo cui soltanto un comportamento effettivamente realizzato può costituire reato, e non anche mere intenzioni astratte o stati mentali⁴⁷¹.

In definitiva, tramite i *risk assessment tools* verrebbero indirettamente reintrodotte alcune delle (ormai superate) tesi della Scuola Positiva. Per contrastare questo pericolo, occorre evitare ogni forma di automatismo, per cui dall'accertamento della pericolosità sociale da parte di un sistema IA non può conseguire automaticamente la limitazione della libertà personale. Risulta dunque essenziale che l'uso dell'IA non faccia venir meno la centralità che il fatto ha assunto nel diritto penale moderno.

3.5.2) IA e pericolosità sociale: misure di sicurezza e misure di prevenzione

In questo paragrafo, verrà analizzato il pericolo di una deriva verso il diritto penale d'autore, con specifico riferimento alla centralità che la pericolosità sociale assume nelle misure di sicurezza e nelle misure di prevenzione.

Con riferimento alle misure di sicurezza, occorre ripercorrere le controverse originarie finalità politico-criminali di tali misure. Nelle intenzioni del legislatore del codice penale del 1930, le misure di sicurezza si configuravano come un ulteriore e più penetrante strumento di contrasto della criminalità, nei casi in cui la

⁴⁶⁹ *Ibid.* V. par. 2.5.1.

⁴⁷⁰ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, <https://archiviodypc.dirittopenaleuomo.org/d/6702-quando-la-giustizia-penale-incontra-l-intelligenza-artificiale-luci-e-ombre-dei-risk-assessment-too>, consultato il 3 agosto 2025, p. 21; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia.*, in U. RUFFOLO, *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Milano, Giuffrè Francis Lefebvre, 2020, <https://discrimen.it/oracolo-algoritmico-e-la-giustizia-penale-al-bivio-tra-tecnologia-e-tecnocrazia/>, consultato il 6 ottobre 2025, pp. 13-14.

⁴⁷¹ L. NOTARO, *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021, <https://arpi.unipi.it/handle/11568/1147326>, consultato il 13 luglio 2025, pp. 65-66.

sola pena si fosse rivelata inefficace. Tali misure erano infatti state concepite come una sanzione svincolata dalle garanzie proprie della pena, con particolare riferimento ai limiti di durata, e al principio di colpevolezza (infatti, le misure di sicurezza si basano sulla pericolosità sociale)⁴⁷². Questa impostazione trova conferma anche nella Relazione del Guardasigilli al progetto definitivo del codice penale⁴⁷³. Dalle intenzioni del legislatore del 1930 emerge, quindi, il rischio che le misure di sicurezza diventino uno strumento attraverso cui eludere le garanzie proprie delle pene.

Questo pericolo sarebbe particolarmente accentuato nel caso in cui i presupposti per applicare una misura di sicurezza venissero accertati mediante l'ausilio dell'intelligenza artificiale. Le misure di sicurezza si basano sulla pericolosità sociale (art. 203, comma 1, c.p.), con la conseguenza che il ricorso all'IA potrebbe spostare il baricentro sulla personalità dell'imputato a discapito del fatto compiuto⁴⁷⁴. L'intelligenza artificiale potrebbe così essere impiegata per eludere la centralità che il fatto assume in materia di pene, dando, invece, maggiore risalto alla pericolosità sociale.

Allo stesso tempo, è interessante notare come questo rischio appaia mitigato dal presupposto oggettivo delle misure di sicurezza di cui all'art. 202 c.p.: l'aver commesso un reato o un fatto in forza del quale la legge penale prevede che possa essere applicata una misura di sicurezza. Questa garanzia mitiga il rischio di un passaggio dalla tipicità del fatto alla tipicità d'autore, in quanto limita l'applicabilità di una misura di sicurezza al compimento di un determinato fatto. La mera pericolosità sociale, accertata mediante l'intelligenza artificiale, non potrebbe dunque di per sé legittimare l'adozione di una misura di sicurezza.

Per quanto riguarda invece le misure di prevenzione, occorre premettere che anche tali misure si basano sulla pericolosità sociale. Se da un lato il legislatore ha ormai accolto la tesi secondo cui il diritto penale ha primariamente oggetto dei fatti tipici e non la pericolosità sociale, dall'altro lo stesso legislatore ha previsto le misure di prevenzione come strumento di repressione della criminalità maggiormente penetrante, in quanto non richiede una previa condanna penale per un fatto specifico⁴⁷⁵.

Specificamente, le misure di prevenzione possono essere applicate ai seguenti soggetti: *“a) coloro che debbano ritenersi, sulla base di elementi di fatto, abitualmente dediti a traffici delittuosi; b) coloro che per la condotta ed il tenore di vita debba ritenersi, sulla base di elementi di fatto, che vivono abitualmente, anche in parte, con i proventi di attività delittuose; c) coloro che per il loro comportamento debba ritenersi, sulla base di elementi di fatto, (comprese le reiterate violazioni del foglio di via obbligatorio di cui all'articolo 2, nonché dei divieti di frequentazione di determinati luoghi previsti dalla vigente normativa), che sono dediti alla commissione di reati che offendono o mettono in pericolo l'integrità fisica o morale dei minorenni, la sanità, la sicurezza o la tranquillità pubblica”* (art. 1, lett. c), d.lgs. 6 settembre 2011, noto come codice antimafia).

⁴⁷² G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, pp. 884-885.

⁴⁷³ *Codice penale e Codice di procedura penale*, Lav. Prep., vol. V, pt. 1, 1929.

⁴⁷⁴ M. DI FLORIO, *Calculate Criminal Law? Criticità nell'uso degli algoritmi di pericolosità sociale* in *La legislazione penale*, 2023, <https://www.la legislazione penale.eu/calculate-criminal-law-criticita-nelluso-degli-algoritmi-di-pericolosita-sociale-mattia-di-florio/>, consultato il 10 settembre 2025, pp. 7-9.

⁴⁷⁵ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*

La Cassazione ha chiarito che il giudizio di prevenzione, basato sulla disposizione sopra riportata, non si esaurisce in un giudizio prognostico in merito alla pericolosità sociale. Tale accertamento comprende altresì l'apprezzamento di fatti storici indicanti la possibilità di iscrivere il soggetto in una delle categorie criminologiche individuate dalla legge⁴⁷⁶. Vi è quindi una prima fase constatativa, in cui si verifica l'idoneità degli elementi presenti a rappresentare la sussistenza di condotte antisociali, e una seconda fase prognostica, consistente nel valutare la probabilità che le condotte antisociali si ripetano.

La rilevanza di una fase constatativa permette di mitigare tale rischio, seppur non di escluderlo. Qualora il giudizio di prevenzione presentasse esclusivamente una componente prognostica, i fatti compiuti potrebbero assumere un ruolo secondario, con la conseguenza che il soggetto verrebbe giudicato sostanzialmente per ciò che è, anziché per quello che ha fatto. L'orientamento giurisprudenziale consolidatosi implica che, anche nel caso in cui si dovesse far ricorso all'IA, il giudice sarebbe comunque obbligato a valutare la sussistenza di elementi idonei a rappresentare condotte contrarie alla convivenza sociale. La mera pericolosità sociale stimata dal modello predittivo non sarebbe dunque sufficiente a fondare una misura di prevenzione.

È opportuno evidenziare che è lo stesso legislatore a prescrivere la non sufficienza della mera pericolosità, imponendo che la valutazione di cui alle lett. a), b), c), dell'art. 1 del d.lgs. 159/2011 debba essere compiuta sulla base di elementi di fatto.

Il rischio di un diritto penale d'autore risulta comunque maggiormente accentuato rispetto alle misure di sicurezza, relativamente alle quali l'art. 202 c.p. richiede che il destinatario di una misura di sicurezza abbia commesso un reato, oppure un quasi reato. La sussistenza del fatto di reato o di quasi reato, deve quindi essere provata con il medesimo standard probatorio idoneo a fondare la responsabilità penale: l'oltre ogni ragionevole dubbio (art. 533, comma uno, c.p.p.). La formulazione di cui all'art. 1 d.lgs. 159/2011 induce invece a pensare che si tratti di una soluzione intermedia tra il mero sospetto e l'elemento in grado di fondare una pronuncia di responsabilità penale⁴⁷⁷. Tale scelta è stata criticata in dottrina, evidenziando come gli unici presupposti di fatto a cui un giudice possa riferirsi sono quelli idonei a fondare la responsabilità penale. La prescrizione legislativa si presterebbe soltanto ad essere utilizzata strumentalmente per riproporre, di fatto, una repressione del mero sospetto⁴⁷⁸.

In definitiva, il presupposto oggettivo delle misure di sicurezza, di cui all'art. 202 c. p., mitiga in maniera maggiormente significativa il rischio di un diritto penale d'autore, rendendo irrilevante la mera pericolosità sociale qualora il fatto di reato o di quasi reato non sia provato, secondo il medesimo standard probatorio previsto ai fini della responsabilità penale. Nelle misure di prevenzione, non si può escludere che, in presenza di un mero sospetto, l'elevato livello di rischio valutato da un sistema IA risulti sufficiente per fondare una misura di prevenzione limitativa della libertà personale.

⁴⁷⁶ Cass. pen., Sez. I, 11 febbraio 2014, n. 23641.

⁴⁷⁷ A. GAITO -B. ROMANO -M. RONCO *et al.*, *Misure di prevenzione*, 1ª edizione, UTET Giuridica, Milanofiori Assago (MI), 2013, p. 99.

⁴⁷⁸ D. PETRINI, *La prevenzione inutile. Illegittimità delle misure praeter delictum*, Giovene, Napoli, 1996, p.22.

3.5.3) Diritto penale d'autore nella determinazione della pena?

Per quanto concerne la commisurazione della pena, ricorre un rischio differente rispetto alle misure di prevenzione. Il giudice può ovviamente commisurare la pena soltanto dopo aver accertato la sussistenza del fatto di reato, secondo lo standard probatorio dell'oltre ogni ragionevole dubbio, ai sensi dell'art. 533, comma 1, c.p.p. Vi sono quindi delle garanzie maggiormente significative quanto alla centralità del fatto, rispetto alle misure di prevenzione.

Rimane però il pericolo che, in sede di commisurazione della pena, la capacità a delinquere del reo assuma un ruolo preminente a discapito della gravità del reato. L'accertamento della capacità a delinquere tramite l'IA potrebbe influenzare indebitamente il giudice, con l'effetto di non tenere sufficientemente conto della gravità del reato. L'imputato, nella sostanza, verrebbe giudicato per ciò che è, con riferimento alla capacità a delinquere (art. 133, comma 2, c.p.) e non per quello che ha fatto, relativamente alla gravità del reato (art. 133, comma 1, c.p.).

Si possono allora suggerire alcune cautele. Si potrebbe stabilire, come è stato sostenuto in dottrina, che è la gravità del reato, e non la capacità a delinquere, il criterio principale di commisurazione. Secondo questa tesi, dal tenore dell'art. 133 c.p. emergerebbe che il giudice deve, innanzitutto, tenere conto della gravità del reato, nelle sue componenti soggettive ed oggettive. A tale prima valutazione, occorrerebbe poi aggiungere il giudizio sulla capacità a delinquere⁴⁷⁹. Nel caso in cui l'IA stimi come elevato il rischio di recidiva (e dunque la capacità a delinquere), sarebbe comunque il giudizio sulla gravità del reato ad assumere un ruolo preminente. Si garantirebbe così il primato della condotta del reo rispetto alla sua pericolosità sociale, impedendo derive verso un diritto penale d'autore.

In secondo luogo, si potrebbe sancire la rilevanza esclusivamente *in bonam partem* della capacità a delinquere. Secondo autorevole dottrina, la finalità rieducativa della pena, ai sensi dell'art. 27, comma 3, Cost., preclude la rilevanza della capacità a delinquere per l'applicazione di pene eccedenti la misura della gravità del fatto colpevole. La capacità a delinquere potrebbe soltanto suggerire l'applicazione di pene inferiori. Una pena inflitta in misura sproporzionata alla gravità del fatto non sarebbe in grado di favorire la collaborazione necessaria per un programma di risocializzazione a cui tende la finalità rieducativa. Una pena siffatta favorirebbe piuttosto gli effetti desocializzanti propri di ogni giudizio che viene percepito come ingiusto⁴⁸⁰.

Sancendo la rilevanza esclusivamente *in bonam partem* della capacità a delinquere, il ruolo dell'IA ai fini della commisurazione della pena sarebbe limitato a quei casi in cui, alla luce della scarsa pericolosità del soggetto, pare opportuno evitare di irrogare una pena particolarmente elevata che rischierebbe piuttosto di avere un effetto criminogeno, anziché rieducativo. Sarebbe invece esclusa la rilevanza dell'IA nel caso opposto in cui

⁴⁷⁹ M. RONCO – B. ROMANO, *Codice penale commentato*, UTET Giuridica, Torino, 2012, disponibile su OneLegale, consultato il 25 settembre 2025, *sub art. 133 c.p.*

⁴⁸⁰ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.*, p. 804.

il fatto non risulti particolarmente grave nelle sue componenti oggettive, ma il *risk assessment tool* ha stimato un rischio di recidiva particolarmente elevato.

3.6) Possibili violazioni del principio di offensività

Un altro rischio, connesso al diritto penale d'autore, concerne la possibilità che il ricorso all'IA risulti in contrasto con il principio di offensività, secondo il quale non vi può essere reato senza offesa a un bene giuridico determinato⁴⁸¹. La Corte costituzionale ha sancito che tale principio vincola sia il legislatore che il giudice. Il principio di offensività opera in astratto, in quanto vincola il legislatore che è tenuto a qualificare come reato soltanto fatti che, nella loro configurazione astratta, presentino attitudine lesiva. Ancora, l'offensività opera altresì in concreto, vincolando il giudice a un'indagine circa la sussistenza dell'offesa nel fatto di reato così come concretamente verificatosi⁴⁸². Ad esempio, in giurisprudenza, si è affermato che il reato di coltivazione di stupefacenti (di cui all'art. 73 del d.p.r. 9 ottobre 1990, n. 309) non è configurabile qualora l'attività di coltivazione sia di minime dimensioni, venga svolta in forma domestica e appaia destinata in via esclusiva all'uso personale del coltivatore⁴⁸³.

Il principio in esame muove dall'odierna concezione di reato come fatto dannoso per la società, secondo cui occorre reprimere soltanto i comportamenti che ledono o mettono in pericolo beni individuali o collettivi. Non è dunque sufficiente limitarsi a punire fatti previsti dalla legge come reato, in virtù del diritto penale del fatto, ma è altresì necessario punire soltanto quei fatti che risultino offensivi di determinati beni giuridici. Anche sotto tale profilo, vi è un netto contrasto con le tesi della Scuola Positiva, secondo cui, nel diritto penale, dovrebbero essere posti in primo piano tipi di persone socialmente pericolose, indipendentemente dall'offensività del fatto commesso⁴⁸⁴.

In dottrina, si è evidenziato che giudicare una persona sulla base di un *risk assessment tool* rischia di tradursi in un giudizio su ciò che la persona è, anziché su ciò che la persona ha fatto. Sulla base di tale considerazione, è stato evidenziato un possibile contrasto con il principio di danno (*harm principle* o *nullum crimen sine iniuria*), secondo cui un determinato comportamento può essere considerato reato soltanto se provoca un danno effettivo a un bene giuridico altrui⁴⁸⁵.

⁴⁸¹ G. MARINUCCI -E. DOLCINI -G. L. GATTA, *op. cit.*, pp. 10-11.

⁴⁸² Corte cost., 20 dicembre 2019, n. 278; Corte cost., 7 giugno 2019, n. 141; Corte cost., 20 maggio 2016, n. 109; Corte cost., 20 giugno 2008, n. 255.

⁴⁸³ Sez. Un., 19 dicembre 2019, n. 12348, par. 7: “Il reato di coltivazione di stupefacenti è configurabile indipendentemente dalla quantità di principio attivo ricavabile nell'immediatezza, essendo sufficienti la conformità della pianta al tipo botanico previsto e la sua attitudine, anche per le modalità di coltivazione, a giungere a maturazione e a produrre sostanza stupefacente; devono però ritenersi escluse, in quanto non riconducibili all'ambito di applicazione della norma penale, le attività di coltivazione di minime dimensioni svolte in forma domestica che, per le rudimentali tecniche utilizzate, lo scarso numero di piante, il modestissimo quantitativo di prodotto ricavabile, la mancanza di ulteriori indici di un loro inserimento nell'ambito del mercato degli stupefacenti, appaiono destinate in via esclusiva all'uso personale del coltivatore”; R.GIOVAGNOLI, *Manuale di diritto penale. Parte generale*, V edizione, Ita edizioni, 2025, Torino, pp. 222-229.

⁴⁸⁴ G. MARINUCCI -E. DOLCINI -G. L. GATTA, *op. cit.*, pp. 5-9;

⁴⁸⁵ L. NOTARO, *op. cit.*, p. 66.

Il giudice potrebbe attribuire rilevanza penale a fatti che, in concreto, non risultano realmente offensivi del bene giuridico, in quanto potrebbe essere influenzato dalla pericolosità sociale del soggetto accertata tramite l'IA, con la conseguenza che non verrebbe prestata la dovuta attenzione alla reale attitudine lesiva della condotta in esame. La circostanza che un soggetto sia valutato con un *risk score* che ha previsto un rischio di recidiva particolarmente elevato potrebbe ingenerare nella collettività un allarme sociale, dinanzi al quale, per il giudice, potrebbe essere difficile non punire anche qualora dovesse ricorrere l'inoffensività in concreto. In conclusione, l'utilizzo dell'IA nel processo penale pone delle sfide anche sotto il profilo dell'offensività in concreto. Anche rispetto a tale profilo, l'IA potrebbe far riemergere, sotto altra veste, le ormai superate teorie della Scuola Positiva, a discapito dell'odierna concezione di reato come fatto lesivo di beni giuridici.

3.7) Possibili violazioni della presunzione di innocenza

Un'altra possibile criticità è che l'introduzione dell'IA come prova in merito alla personalità dell'imputato potrebbe contaminare il giudizio sulla colpevolezza. Tale situazione si porrebbe in contrasto con la presunzione di innocenza⁴⁸⁶.

Rileva l'art. 27, comma 2, Cost., a norma del quale l'imputato "*non è considerato colpevole sino alla condanna definitiva.*" Data l'ambiguità della formulazione normativa, la disposizione è stata interpretata dalla dottrina tradizionale con l'espressione "*presunzione di non colpevolezza*", nel senso che ogni persona non è considerata colpevole fino alla condanna definitiva. Tale interpretazione assume peraltro un significato riduttivo⁴⁸⁷, in quanto la dottrina più recente ha sostenuto che dalla disposizione costituzionale deve essere desunta una vera e propria presunzione di innocenza, nel senso che la persona deve considerarsi innocente (e non soltanto non colpevole) sino alla condanna definitiva⁴⁸⁸. La stessa Corte costituzionale, sebbene inizialmente avesse aderito alla tesi della presunzione di non colpevolezza⁴⁸⁹, ha ormai riconosciuto che l'art. 27, comma 2, Cost. sancisce la presunzione di innocenza⁴⁹⁰.

La ragione dell'ambiguità normativa discende dalla circostanza che l'Assemblea costituente ha voluto unire, con un'unica espressione, una regola di trattamento e una regola probatoria⁴⁹¹. La regola di trattamento vuole che l'imputato non venga assimilato al colpevole, fintanto che non venga condannato con sentenza definitiva. Ad esempio, tale principio impone che le misure cautelari non possono consistere, nella sostanza, in un'anticipazione della pena⁴⁹².

La regola probatoria vuole che l'onere della prova quanto alla colpevolezza dell'imputato ricada sull'accusa, anziché sulla difesa. Occorre distinguere tra l'onere formale, consistente nell'introdurre il mezzo di prova che

⁴⁸⁶ *Ibid*, pp. 68-69.

⁴⁸⁷ P. TONINI -C. CONTI, *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023, p.272.

⁴⁸⁸ G. ILLUMINATI, *La presunzione di innocenza dell'imputato*, Bologna, 1979, p. 28.

⁴⁸⁹ Corte costituzionale, 22 giugno 1972, n. 124.

⁴⁹⁰ Corte costituzionale, 22 ottobre 2007, n. 348; Corte costituzionale, 22 ottobre 2007, n. 349.

⁴⁹¹ A. MALINVERNI, *Principi di diritto penale tributario*, Milano, Giuffrè, 1984, p. 474.

⁴⁹² P. TONINI -C. CONTI, *op. cit.*, p. 272.

permetterà il formarsi dell'elemento di prova e l'onere sostanziale, relativo al dovere di convincere il giudice dell'esistenza del fatto di reato⁴⁹³.

Sul versante del diritto sovranazionale, rilevano gli artt. 6, par. 2 della Convenzione europea dei diritti dell'uomo (CEDU) e l'art. 48, par. 1 della Carta dei diritti fondamentali dell'Unione europea (CDFUE). Entrambe le disposizioni sanciscono in maniera inequivoca una presunzione di innocenza. La stessa Corte costituzionale ha riconosciuto la sussistenza della presunzione di innocenza nell'ordinamento italiano, in virtù dell'integrazione delle garanzie costituzionali con quelle sovranazionali. Infatti, se da un lato la Costituzione non sancisce espressamente la presunzione di innocenza, dall'altro la CEDU e la CDFUE tutelano l'accusato fino all'accertamento legale della colpevolezza, e non fino alla sentenza definitiva. Dall'integrazione delle tutele costituzionali e sovranazionali, discende la presunzione di innocenza (anziché di non colpevolezza) che tutela l'imputato fino al passaggio in giudicato della sentenza di condanna (anziché fino all'accertamento legale della colpevolezza)⁴⁹⁴.

La presunzione di innocenza, nel diritto dell'Unione europea, rileva altresì come regola di trattamento dell'imputato al di fuori del processo, con riferimento ai pericoli del cosiddetto processo mediatico. Tale profilo è oggi disciplinato dalla direttiva 2016/343/UE *“sul rafforzamento di alcuni aspetti della presunzione di innocenza e del diritto di presenziare al processo nei procedimenti penali.”* Tale direttiva, ad esempio, mira ad evitare che le autorità pubbliche presentino l'imputato come colpevole.

Qualora l'IA venisse implementata nel processo penale, i principali pericoli legati alla presunzione di innocenza riguarderebbero soprattutto gli ambiti della commisurazione della pena e della sospensione condizionale della pena. In tali sedi, il ricorso all'IA per valutare la personalità dell'imputato potrebbe indebitamente contaminare la valutazione giudiziale sulla colpevolezza.

Si tratterebbe di una violazione sul piano sostanziale e non formale, in quanto, naturalmente, il giudice motiverebbe la colpevolezza dell'imputato sulla base del fatto di reato e non sulla base dell'*output* algoritmico⁴⁹⁵. In concreto, il giudice potrebbe però convincersi della colpevolezza dell'imputato sulla base del *risk score* prodotto dall'IA, nonostante tale elemento non sia chiaramente idoneo a vincere la presunzione di innocenza, in quanto estraneo al fatto specificamente contestato. Ne conseguirebbe una violazione della presunzione di innocenza come regola di giudizio nel processo.

Per il PM potrebbe risultare più agevole dimostrare la colpevolezza dell'imputato se il giudice si è già indebitamente convinto in ragione dell'*output* algoritmico. L'onere sostanziale della prova, consistente nel dovere di convincere il giudice dell'esistenza del reato, ne risulterebbe dunque altrettanto attenuato.

Per contrastare tale pericolo, l'art. 5, par. 1, lett. d), dell'AI Act vieta *“l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazioni del rischio relative a*

⁴⁹³ *Ibid*, pp. 272-276.

⁴⁹⁴ Corte costituzionale, 22 ottobre 2007, n. 348; Corte costituzionale, 22 ottobre 2007, n. 349.

⁴⁹⁵ Il giudice potrebbe in realtà motivare sulla base dell'*output* algoritmico nel caso in cui l'IA venisse utilizzata come mezzo di prova quanto alla sussistenza del fatto compiuto. Tuttavia, in questo paragrafo ci si riferisce specificamente all'uso dell'IA nella commisurazione della pena o sospensione condizionale della pena, che quindi non rileva ai fini dell'accertamento del fatto.

persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità.” Nel considerando 42 del medesimo regolamento, si chiarisce che la *ratio* di tale divieto consiste nel tutelare la presunzione di innocenza, in quanto le persone fisiche dovrebbero essere giudicate sulla base del loro comportamento effettivo, anziché sulla base di un comportamento previsto dall’IA basato unicamente sulla profilazione.

Oltre ad esserci rischi legati alla presunzione di innocenza come regola di giudizio, vi sono altresì dei pericoli relativi alla presunzione di innocenza come regola di trattamento dell’imputato al di fuori del processo, con riferimento al cosiddetto processo mediatico. Ad esempio, potrebbe divenire pubblico un *output* algoritmico che attesti un rischio di recidiva particolarmente significativo. Tale *output* potrebbe ledere indebitamente la reputazione del soggetto, in quanto questi potrebbe apparire come colpevole dinanzi alla collettività, senza che sia stato celebrato un processo nei suoi confronti. Sotto quest’ultimo profilo, non sarebbe sufficiente formare i giudici, ma sarebbe anche necessario che la collettività sia consapevole che la previsione algoritmica sul rischio di recidiva non dice nulla sulla colpevolezza in merito al fatto di reato specificamente contestato.

In conclusione, è essenziale che l’accertamento della pericolosità sociale non influenzi indebitamente il giudizio sulla colpevolezza dell’imputato, al fine di garantire il rispetto della presunzione di innocenza, intesa come regola di trattamento dell’imputato fuori e dentro il processo.

3.8) Compatibilità con l’individualizzazione della pena e del trattamento cautelare

Il ricorso all’IA nel processo penale potrebbe porsi altresì in contrasto con il principio di individualizzazione del trattamento sanzionatorio, quando impiegata in sede di commisurazione della pena, e di individualizzazione del trattamento cautelare, in caso di applicazione ai fini delle misure cautelari, come sostenuto in dottrina⁴⁹⁶.

Tale rischio risulta analogo rispetto a quanto già emerso nel caso *Loomis*, in cui la difesa aveva lamentato la violazione del diritto ad una pena individualizzata, poiché l’imputato era stato giudicato in quanto appartenente ad un determinato gruppo di persone, anziché come singolo⁴⁹⁷. La Corte Suprema del Wisconsin aveva, invece, considerato legittimo il ricorso al *risk assessment tool*, in quanto non era stato impiegato come fattore determinante, bensì come mera informazione aggiuntiva, da valutare congiuntamente con altri fattori⁴⁹⁸.

In sede di commisurazione della pena ex art. 133 c.p., tanto il giudice quanto il legislatore devono tenere conto dei principi costituzionali, con particolare riferimento alla finalità rieducativa della pena, ex art. 27, comma 3,

⁴⁹⁶ M. GIALUZ, *Quando la giustizia penale incontra l’intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, <https://archiviodypc.dirittopenaleuomo.org/d/6702-quando-la-justizia-penale-incontra-l-intelligenza-artificiale-luci-e-ombre-dei-risk-assessment-too>, consultato il 3 agosto 2025, p. 21.

⁴⁹⁷ Corte Suprema del Wisconsin, *State v. Loomis*, n. 881 N.W.2d 749, 2016, par. 34.

⁴⁹⁸ *Ibid.*, par. 72.

Cost⁴⁹⁹. La Corte costituzionale ha sancito il principio di individualizzazione della pena, definito come il “*tenere conto dell’effettiva entità e delle specifiche esigenze dei singoli casi*”. Tale principio “*si pone come naturale attuazione e sviluppo di principi costituzionali, tanto di ordine generale (principio d’uguaglianza) quanto attinenti direttamente alla materia penale*”. Alla base dell’individualizzazione della pena, vi è il principio di proporzionalità, in quanto la pena deve essere proporzionalmente adeguata alle personali responsabilità e alle personali esigenze di risposta.

L’individualizzazione della pena trova inoltre fondamento nell’art. 27, comma 1, Cost., in forza del quale “*la responsabilità penale è personale*”, in quanto tenere conto dell’effettiva entità del fatto e delle specifiche esigenze dei singoli casi mira a rendere la responsabilità penale quanto più possibile personale. Tale principio risulta altresì strumentale alla finalità rieducativa della pena, di cui all’art. 27, comma 3, Cost., in quanto permette di adeguare il trattamento sanzionatorio alle specifiche esigenze di rieducazione del condannato. Specificamente, la Corte ha sancito tale principio per censurare previsioni sanzionatorie rigide che non permettono di adeguare la pena all’effettiva entità e alle specifiche esigenze dei singoli casi⁵⁰⁰.

Nell’ipotesi dell’impiego dell’IA ai fini della commisurazione della pena, il principio di individualizzazione del trattamento sanzionatorio assumerebbe rilievo sotto un altro profilo differente rispetto alla rigidità delle previsioni sanzionatorie. Siccome il sistema di IA stima un determinato livello di rischio sulla base dell’appartenenza ad un *pattern*, l’imputato verrebbe giudicato in quanto appartenente ad una determinata categoria e non in relazione all’effettiva entità e alle specifiche esigenze del singolo caso, come richiesto dalla Corte costituzionale. Ad esempio, la pena potrebbe essere determinata sulla base dell’etnia dell’imputato, sebbene tale elemento non dica nulla sull’effettiva entità del reato nel singolo caso di specie⁵⁰¹.

Nell’ambito delle misure cautelari, potrebbe rilevare il canone di individualizzazione del trattamento cautelare. Per comprendere tale principio, occorre svolgere delle premesse in merito alla legittimità costituzionale delle misure cautelari. Si potrebbe sostenere che tali misure siano in contrasto con la presunzione di innocenza *ex* art. 27, comma 2, Cost., in quanto le misure cautelari vengono disposte prima della condanna definitiva. La questione si pone con particolare riferimento ai gravi indizi che, a norma dell’art. 274, comma 1, c.p.p., costituiscono una delle condizioni generali di applicabilità delle misure cautelari. Il giudice è infatti tenuto a formulare un giudizio prognostico in merito alla colpevolezza dell’imputato sulla base degli elementi esistenti allo stato degli atti.

La Corte costituzionale ha invece chiarito che si tratta di un’antinomia soltanto apparente, in quanto le misure cautelari, affinché siano costituzionalmente legittime, devono assumere connotati differenti rispetto alle pene.

⁴⁹⁹ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *op. cit.* p. 798; A. CADOPPI -S. CANESTRARI -A. MANNA *et al.*, *Parte generale III, la punibilità e le conseguenze del reato*, in *Trattato di diritto penale*, UTET giuridica, Milanofiori Assago (MI), 2014, pp. 202-208.

⁵⁰⁰ Corte cost., 2 aprile 1980, n. 50, parr. 3-4, considerato in diritto.

⁵⁰¹ Chiaramente, in tale circostanza, vi sarebbe altresì una violazione del principio di uguaglianza (art. 3 Cost.) e del principio di non discriminazione (art. 14 CEDU e art. 21 CDFUE), come argomentato nel par. 3.1.

Ne consegue che l'applicazione delle misure cautelari non può consistere in un giudizio anticipato di colpevolezza, né può tendere direttamente o indirettamente a finalità proprie della sanzione penale.

Le misure cautelari sono poi soggette al criterio del minor sacrificio necessario secondo cui la compressione della libertà personale dell'indagato o dell'imputato deve essere contenuta entro i limiti indispensabili a soddisfare le esigenze cautelari riconoscibili nel caso concreto. Da un lato, tale principio trova fondamento nella libertà personale, ai sensi dell'art. 13, comma 1, Cost., in quanto mira a ridurre al minimo la limitazione di tale libertà. Allo stesso tempo, tale criterio si ricollega alla presunzione di innocenza, in quanto una limitazione della libertà personale che non sia limitata a quanto strettamente necessario per soddisfare le esigenze cautelari avrebbe i connotati propri di una pena.

Il criterio del minor sacrificio implica che il ricorso alle forme di restrizione della libertà personale maggiormente intense possa ritenersi consentito soltanto se le esigenze cautelari non possono essere soddisfatte mediante misure di minore incisività. Tale criterio comporta altresì che il legislatore sia tenuto a predisporre una pluralità di misure, tra loro alternative, caratterizzate da differenti gradi di incidenza sulla libertà personale⁵⁰².

La Corte costituzionale ha sancito che il criterio del minore sacrificio è funzionale a realizzare una piena individualizzazione della coercizione cautelare⁵⁰³. Tale principio *“impegna il legislatore a prefigurare meccanismi individualizzati del trattamento cautelare, parametrati sulle esigenze configurabili nelle singole fattispecie concrete”*⁵⁰⁴.

Nel caso del ricorso all'intelligenza artificiale al fine di valutare la sussistenza delle esigenze cautelari *ex art.* 274 c.p.p., il trattamento cautelare potrebbe non essere sufficientemente individualizzato, perché l'interessato potrebbe essere profilato dal sistema di IA sulla base del gruppo sociale di appartenenza, in maniera analoga a quanto già osservato con riferimento al principio di individualizzazione della pena. I giudici potrebbero quindi non essere in grado di adeguare il trattamento cautelare alle esigenze configurabili nelle singole fattispecie concrete. Il modello predittivo potrebbe attribuire rilevanza al gruppo sociale di appartenenza, anziché a tali esigenze.

In definitiva, è essenziale che sia il giudizio sulla commisurazione della pena sia la valutazione delle esigenze cautelari restino focalizzati sul singolo individuo e non sulla categoria di appartenenza, affinché l'uso dell'IA non contrasti con i principi costituzionali di individualizzazione del trattamento sanzionatorio e di quello cautelare.

3.9) Compatibilità con l'obbligatorietà dell'azione penale

⁵⁰² P. TONINI -C. CONTI, *op. cit.*, p. 471; A. CAMON -M. DANIELE -D. NEGRI, *Fondamenti di procedura penale*, 4° edizione, Wolters Kluwer, CEDAM, Milano, 2023, pp. 101-104.

⁵⁰³ Corte costituzionale, 7 luglio 2010, n. 265, par. 7.

⁵⁰⁴ *Ibid*, par. 6.

Con specifico riferimento al ricorso all'IA nella selezione delle notizie di reato da trattare con precedenza (art. 1, comma 6, lett. b), d.lgs. 20 febbraio 2006, n. 106), nonché nella scelta spettante al PM tra archiviazione e esercizio dell'azione penale (artt. 50, comma 1 e 408, comma 1, c.p.p.), occorre valutare la compatibilità con l'obbligatorietà dell'azione penale (art. 112 Cost.).

Ai sensi di tale disposizione, “*il pubblico ministero ha l'obbligo di esercitare l'azione penale*”. Nel significato attribuitovi dal legislatore costituzionale, l'obbligatorietà impone al PM di valutare la fondatezza della *notitia criminis* e di svolgere le “*indagini necessarie per le determinazioni inerenti all'esercizio dell'azione penale*”⁵⁰⁵. L'art. 112 Cost. deve leggersi in combinato disposto con l'art. 3 Cost. sul principio di uguaglianza. Il reato deve infatti essere perseguito anche se la persona offesa non ha possibilità economiche. Impedire che la valutazione sull'esercizio dell'azione penale sia rimessa alla volontà del PM mira poi a garantire che la repressione penale si espliciti nei confronti di tutti nello stesso modo.

In secondo luogo, l'obbligatorietà dell'azione penale è connessa con il principio di legalità di cui all'art. 25, comma 2, Cost: deve essere la legge e non il singolo pubblico ministero a determinare chi deve essere punito e chi deve essere esente da pena⁵⁰⁶.

L'art. 112 Cost. si raccorda altresì con l'art. 108, comma 2, Cost. sull'indipendenza del pubblico ministero⁵⁰⁷: se il PM non fosse pienamente indipendente rispetto all'esecutivo, potrebbe selezionare quali reati perseguire in base a pressioni esterne, violando così il precetto costituzionale.

Tale principio potrebbe risultare in contrasto con il ricorso all'IA nella selezione delle notizie di reato da trattare con precedenza o nella scelta tra archiviazione e obbligatorietà dell'azione penale, in quanto il sistema di IA potrebbe compiere, di fatto, una scelta di priorità, potenzialmente in contrasto con l'obbligo del PM di perseguire indistintamente tutti i reati.

Peraltro, in dottrina è stata sostenuta la relatività di tale principio, in contrapposizione all'assolutezza vigente nel sistema pre-costituzionale. In particolare, prima dell'avvento della Costituzione, l'obbligatorietà dell'azione penale esprimeva un valore assoluto, a tutela della potestà sovrana dello Stato sull'individuo⁵⁰⁸. L'entrata in vigore della Costituzione ha determinato un mutamento di tale valore, da assoluto a relativo. Il principio di obbligatorietà dell'azione penale necessita infatti di un bilanciamento con altri diritti costituzionalmente garantiti, non esistendo diritti “tiranni” rispetto agli altri⁵⁰⁹. L'obbligatorietà dell'azione penale può risultare in tensione con il principio di buon andamento della pubblica amministrazione, ai sensi dell'art. 97, comma 2, Cost., esteso agli uffici giudiziari del pubblico ministero. L'art. 112 Cost, deve anche

⁵⁰⁵ Si veda il disposto dell'art. 326 c.p.p. sulle finalità delle indagini preliminari.

⁵⁰⁶ P. TONINI -C. CONTI, *op. cit.*, pp. 672-673.

⁵⁰⁷ M. CHIAVARIO, *L'obbligatorietà dell'azione penale: il principio e la realtà*, in *Cass. pen.*, 1993, p. 2658.

⁵⁰⁸ O. DOMINIONI, *Interesse alla persecuzione penale e irrilevanza sociale del fatto nel prisma dell'efficienza giudiziaria*, in *Studi in onore di M. Pisani*, La Tribuna, 2010, p. 323; O. DOMINIONI, *Azione penale*, in *Dig. discipl. pen.*, vol. I, Torino, 1987, p. 409.

⁵⁰⁹ V. ZAGREBELSKY, *Stabilire le priorità nell'esercizio obbligatorio dell'azione penale*, in *AA.VV., Il pubblico ministero oggi*, Milano, 1994, p. 10; G. SILVESTRI, *Relazione*, in *AA.VV., Pubblico ministero e riforma dell'ordinamento giudiziario*, Associazione tra gli studiosi del processo penale, Milano, 2006, p. 227.

essere bilanciato con l'art. 111, comma 2, Cost., sulla ragionevole durata del processo⁵¹⁰. La ragione di tale tensione è che l'obbligatorietà dell'azione penale, se interpretata come valore assoluto, rischia di compromettere l'interesse all'efficienza giudiziaria, desumibile sia dal principio di buon andamento che dal principio di ragionevole durata del processo.

In dottrina, è stata espressa la necessità di un intervento normativo nei casi in cui risulti utile affievolire l'interesse alla persecuzione penale di cui dall'art. 112 Cost., al fine di perseguire l'interesse all'efficienza giudiziaria. Si evidenzia che soltanto una scelta legislativa possa fondare le scelte concretamente idonee a bilanciare, in concreto, tali contrastanti esigenze costituzionali⁵¹¹.

Il legislatore ha recepito tale esigenza con la Riforma Cartabia. Sono stati modificati i commi 6 e 7 dell'art. 1 del d.lgs. 20 febbraio 2006, n. 106, prevedendo che il procuratore della Repubblica disponga i criteri organizzativi dell'ufficio (i quali rilevano ai fini della selezione delle notizie di reato da trattare con precedenza rispetto alle altre) “*nell'ambito di criteri generali indicati dal Parlamento con legge*”. Vengono così predeterminati dei criteri generali dai rappresentanti dei cittadini, al fine di garantirne la legittimità democratica⁵¹².

Una parte della dottrina ha accolto con favore tale intervento normativo, argomentando che l'introduzione di criteri di priorità non contraddice il dettato di cui all'art. 112 Cost., alla luce della natura relativa e non assoluta di tale principio. La scelta del legislatore sarebbe dunque espressione di un legittimo bilanciamento tra l'art. 112 Cost. e l'interesse all'efficienza giudiziaria quale presupposto della ragionevole durata del processo⁵¹³.

Questa conclusione non gode però di consenso unanime in dottrina. Si è sostenuto che, i criteri di priorità, avendo natura selettiva, costituiscono una delle caratteristiche tipiche dell'azione penale facoltativa. Nel sistema statunitense, dove l'azione penale è facoltativa, le decisioni dei procuratori si basano proprio su dei criteri di priorità, preannunciati dagli stessi procuratori in campagna elettorale (nel sistema statunitense, i procuratori vengono spesso nominati sulla base di elezioni democratiche)⁵¹⁴.

Nel caso del ricorso all'IA in questo settore, sussiste un'analoga tensione tra l'interesse all'efficienza processuale e l'interesse alla persecuzione penale, in quanto l'IA potrebbe rendere maggiormente efficiente il processo penale. Impiegare l'IA nella selezione delle notizie di reato da trattare con precedenza potrebbe permettere ai PM di concentrarsi sulle notizie di reato che presentano maggiori probabilità di portare ad una

⁵¹⁰ N. GALANTINI, *Il principio di obbligatorietà dell'azione penale, tra interesse alla persecuzione penale e interesse all'efficienza giudiziaria*, in *Diritto penale contemporaneo*, 2019, <https://archiviopcdpc.dirittopenaleuomo.org/d/6788-il-principio-di-obbligatorieta-dell-azione-penale-tra-interesse-alla-persecuzione-penale-e-interess>, consultato il 27 ottobre 2025, p. 2.

⁵¹¹ M. DI NAPOLI, *Tutela dei diritti ed efficace esercizio dell'azione penale: il punto di vista di una procura della repubblica*, in *L'azione penale tra obbligatorietà e discrezionalità*, Milano, 2009, p. 45.

⁵¹² M. GIALUZ, *Per un processo penale più efficiente e giusto. Guida alla lettura della riforma Cartabia*, in *Sistema penale*, 2022, <https://www.sistemapenale.it/it/scheda/gialuz-per-un-processo-piu-efficiente-e-giusto-guida-alla-lettura-della-riforma-cartabia>, consultato il 26 settembre 2025, pp. 42-43.

⁵¹³ V. GRAMUGLIA, *Obbligatorietà dell'azione penale e ruolo dell'IA nelle scelte del pubblico ministero*, in *Archivio Penale*, Giuffrè Francis Lefebvre, Milano, 2024, <https://archiviopenale.it/obbligatorieta-dellazione-penale-e-ruolo-dellia-nelle-scelte-del-pubblico-ministero/articoli/46820>, consultato il 25 settembre 2025, p. 4.

⁵¹⁴ R. E. KOSTORIS *Per un'obbligatorietà temperata dell'azione penale*, in *Rivista di Diritto Processuale*, 2007, <https://www.research.unipd.it/handle/11577/2438042>, consultato il 27 ottobre 2025.

condanna definitiva, in modo tale da evitare l'inutile dispendio di risorse processuali.⁵¹⁵ In sede di esercizio dell'azione penale, l'IA potrebbe fornire al PM una base statistica più affidabile in base alla quale formulare una ragionevole previsione di condanna. Si eviterebbe così di utilizzare risorse per processi che, statisticamente, sono destinati a concludersi con un'assoluzione. Se si condivide la tesi della natura relativa dell'interesse alla persecuzione penale, da bilanciare con l'interesse all'efficienza processuale, il ricorso all'IA potrebbe configurarsi come un legittimo bilanciamento tra tali esigenze, in maniera analoga a quanto già sostenuto da parte della dottrina con riferimento ai criteri di priorità introdotti dalla Riforma Cartabia.

Va sottolineato che il ricorso all'intelligenza artificiale come supporto alle scelte del PM non può però essere pienamente equiparato ai criteri di priorità introdotti dalla Riforma Cartabia. In tale ipotesi, è il Parlamento, dotato di legittimità democratica, a disporre i criteri generali in base ai quali andranno individuati i criteri di priorità contenuti nel progetto organizzativo dell'ufficio. Nel caso di ricorso all'intelligenza artificiale, il rischio è che tali criteri non siano più determinati da organi dotati di legittimità democratica, bensì dagli sviluppatori del sistema IA i quali, ovviamente, non sono stati eletti democraticamente. Si potrebbe così eludere il principio di legalità *ex art. 25, comma 2, Cost.*, ineludibilmente collegato all'obbligatorietà dell'azione penale (in quanto quest'ultima mira a garantire che sia la legge e non il PM a determinare chi deve essere perseguito e chi deve essere esente da pena). In dottrina, è stato infatti sottolineato il rischio che l'IA tenda sostituire la legge⁵¹⁶, con conseguente pericolo di svalutare la riserva di legge di cui all'art. 25, comma 2, Cost.⁵¹⁷.

In conclusione, il ricorso all'IA potrebbe rappresentare un legittimo strumento di bilanciamento tra l'interesse all'efficienza processuale e quello alla persecuzione penale. Allo stesso tempo, resta da verificare il rischio che tale impiego possa indebolire il principio di legalità a cui l'obbligatorietà dell'azione penale è funzionalmente connessa.

⁵¹⁵ V., par. 2.8.2.

⁵¹⁶ F. SGUBBI, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Bologna, Il Mulino, 2019, p. 41.

⁵¹⁷ V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia.*, in U. RUFFOLO, *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Milano, Giuffrè Francis Lefebvre, 2020, <https://discrimen.it/oracolo-algoritmico-e-la-justizia-penale-al-bivio-tra-tecnologia-e-tecnocrazia/>, consultato il 6 ottobre 2025, pp. 11-13.

CAPITOLO QUARTO:

LIMITI LEGALI

4.1) Il divieto di perizia criminologica (art. 220, comma 2, c.p.p.)

4.1.1) Introduzione: il divieto, le deroghe e il giudizio sulla personalità

Un primo possibile limite legale all'introduzione dell'IA predittiva concerne il divieto di perizia criminologica di cui all'art. 220, comma 2, c.p.p. Si tratta di valutare se tale divieto possa estendersi all'introduzione di *risk assessment tools* funzionali a valutare la personalità dell'imputato.

La perizia è uno dei mezzi di prova tipici previsti dal Titolo II del Libro terzo del codice di procedura penale, dedicato alla disciplina dei singoli mezzi di prova. Di regola, *“la perizia è ammessa quando occorre svolgere indagini o acquisire dati o valutazioni che richiedono specifiche competenze tecniche, scientifiche o artistiche”* (art. 220, comma 2, c.p.p.). Il comma successivo prevede che *“salvo quanto previsto ai fini dell'esecuzione della pena o della misura di sicurezza, non sono ammesse perizie per stabilire l'abitudine o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche”*.

La perizia criminologica funzionale ad accertare la personalità dell'imputato non è quindi ammessa qualora tale accertamento non abbia ad oggetto patologie (le quali possono essere rilevanti, ad esempio, ai fini del giudizio sulla capacità di intendere e di volere dell'imputato). Per espressa previsione normativa, tale divieto non sussiste in sede di esecuzione penale a norma dell'art. 678 c.p.p. e in tema di misure di sicurezza, ex art. 679 c. p. p.

L'art. 220, comma 2, c.p.p. trova poi un'ulteriore eccezione nel processo penale a carico di imputati minorenni in cui il giudice e il pubblico ministero possono sempre *“sentire il parere di esperti, anche senza alcuna formalità”* al fine di *“accertarne l'imputabilità e il grado di responsabilità, valutare la rilevanza sociale del fatto nonché disporre le adeguate misure penali e adottare gli eventuali provvedimenti civili”* (art. 9, commi 1 e 2 del decreto del Presidente della Repubblica 22 settembre 1988, n. 448). Il riferimento al parere di esperti, specie quando concerne il giudizio sulla personalità dell'imputato minorenne, sembrerebbe infatti derogare il divieto in esame⁵¹⁸.

Per comprendere la norma in esame, occorre ricordare che, nel codice di procedura penale, non mancano norme che legittimano la valutazione giudiziale sulla personalità dell'imputato. L'art. 194, comma 1, prevede che il *“testimone non può deporre sulla moralità dell'imputato salvo che si tratti di fatti specifici, idonei a qualificare la personalità in relazione al reato e alla pericolosità sociale”*⁵¹⁹.

L'art. 236, comma 1, c.p.p. ammette che vengano acquisite le sentenze definitive e i certificati del casellario giudiziale *“ai fini del giudizio sulla personalità dell'imputato o della persona offesa dal reato, se il fatto per*

⁵¹⁸ F. GIANFROTTA, *Perizia*, in M. CHIAVARIO, *Commento al nuovo Codice di procedura penale*, Utet, Torino, 1990, pp. 578-79.

⁵¹⁹ L. NOTARO, *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021, <https://arpi.unipi.it/handle/11568/1147326>, consultato il 13 luglio 2025, p. 76.

il quale si procede deve essere valutato in relazione al comportamento o alle qualità morali di questa”. Tuttavia, la giurisprudenza limita il ricorso a tale documentazione alle valutazioni in merito alle circostanze attenuanti generiche e alla sospensione condizionale⁵²⁰.

4.1.2) Analisi comparatistica e storica di un divieto peculiare

Occorre poi notare che l’art. 220, comma 2, c.p.p. rappresenta, sul piano del diritto comparato, una peculiarità dell’ordinamento italiano. Innanzitutto, emerge una differenza netta rispetto agli Stati Uniti. Come si è già visto, vi sono degli Stati in cui la perizia criminologica e i *risk assessment tools*, in sede di determinazione della pena (*sentencing*), sono ammessi, come nel caso della Louisiana. In altri Stati, come l’Arizona, il ricorso a tali strumenti è obbligatorio, con la conseguenza che il giudice deve tenere in considerazione il *risk score* prodotto da tali strumenti di valutazione del rischio⁵²¹.

L’art. 220, comma 2, c.p.p. si discosta altresì dai principali ordinamenti europei in cui non sembra esserci un divieto di perizia criminologica analogo a quello italiano⁵²². Ad esempio, l’ordinamento francese prevede che possano essere disposte delle c.d. “inchieste rapide” (*enquêtes sociales rapides*) sulla situazione familiare, materiale e sociale dell’imputato. Tali inchieste servono a verificare quali siano le misure sanzionatorie praticabili e a favorire l’inserimento sociale. I risultati di tale accertamento confluiscono in un “*dossier della personalità*” che viene messo a disposizione dell’autorità giudiziaria⁵²³. Il giudice può così accedere a valutazioni di esperti circa la personalità dell’imputato, in maniera differente rispetto a quanto previsto, nell’ordinamento italiano, dall’art. 220, comma 2, c.p.p.

Ancora, nell’ordinamento tedesco, si consente al PM di chiedere agli operatori del servizio sociale di costituire un fascicolo di “aiuto alla decisione” (*Gerichtshilfe*)⁵²⁴. In maniera analoga a quanto già visto in merito all’ordinamento francese, tale fascicolo deve contenere elementi relativi all’ambiente e alla situazione sociale, familiare e psicologica dell’accusato⁵²⁵.

Per quanto riguarda, invece, l’evoluzione storica di tale divieto, il codice di procedura penale del 1988 (il Codice Vassalli) ha ribadito il tradizionale scetticismo verso le scienze psico-criminologiche che era già

⁵²⁰ *Ibid*; Cass. pen. Sez. VI, 24 settembre 2013, n. 42823.

⁵²¹ D. KEHL -P. GUO -S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2017, <https://dash.harvard.edu/entities/publication/73120379-010a-6bd4-e053-0100007fdf3b>, consultato il 26 settembre 2025, p. 16.

⁵²² L. NOTARO, *Intelligenza artificiale e giustizia penale*, op. cit., p. 147.

⁵²³ M. CHIAVARIO, *Procedure penali d’Europa. Belgio, Francia, Germania, Inghilterra, Italia. Sintesi nazionali e analisi comparatistiche*, CEDAM, Padova, 2020; Si riporta di seguito il testo dell’art. D16, c.p.p. francese: “*L’enquête sur la personnalité des personnes mises en examen ainsi que sur leur situation matérielle, familiale ou sociale prévue à l’article 81, alinéa 6, du code de procédure pénale et les examens, notamment médical et médico-psychologique, mentionnés à l’alinéa 7 du dit article, constituent le dossier de personnalité de la personne mise en examen. Ce dossier a pour objet de fournir à l’autorité judiciaire, sous une forme objective et sans en tirer de conclusion touchant à l’affaire en cours, des éléments d’appréciation sur le mode de vie passé et présent de la personne mise en examen. Il ne saurait avoir pour but la recherche des preuves de la culpabilité*”.

⁵²⁴ Si riporta di seguito il § 160, par. 3 del c.p.p. tedesco, in materia di indagini da parte della procura: “*Die Ermittlungen der Staatsanwaltschaft sollen sich auch auf die Umstände erstrecken, die für die Bestimmung der Rechtsfolgen der Tat von Bedeutung sind. Dazu kann sie sich der Gerichtshilfe bedienen*”.

⁵²⁵ M. CHIAVARIO, *Procedure penali d’Europa*, op. cit., p. 209.

emerso nel codice del 1930 (il Codice Rocco). Quest'ultimo, all'art. 314, comma 2, vietava, senza deroghe, le perizie funzionali a “*stabilire l'abitudine o la professionalità nel reato, la tendenza a delinquere, il carattere e la personalità dell'imputato, e in genere le qualità psichiche indipendenti da cause patologiche*”.

Tale norma era stata oggetto di forti critiche da parte di dottrina autorevole che l'aveva giudicata anacronistica già a partire dagli anni '60. Si argomentava che fosse contraddittorio attribuire rilievo alla personalità dell'imputato, per poi precludere al giudice il ricorso agli strumenti necessari per compiere tale accertamento⁵²⁶. In questa fase, anche la Corte costituzionale, pur senza dichiarare l'illegittimità costituzionale di tale norma, non esclude che la diffidenza verso la perizia psicologica fosse discutibile alla luce degli sviluppi odierni sulla psiche ed auspicò che la normativa venisse aggiornata⁵²⁷.

In questo contesto, venne approvata la legge 3 aprile 1974, n. 108, contenente “*la delega legislativa al Governo della Repubblica per l'emanazione del nuovo codice di procedura penale*”. In questa prima delega al Governo (la quale, beninteso, non ha portato all'emanazione del vigente codice di procedura penale), si richiedeva “*il riordinamento dell'istituto della perizia, con particolare riferimento alla perizia medico-legale, psichiatrica e criminologica*” (art. 2 n. 10). Dal riferimento alla perizia criminologica si poteva desumere la volontà del legislatore di rivedere, almeno in parte, il divieto assoluto di perizia criminologica di cui all'art. 314, comma 2 del Codice Rocco. Tale norma è stata però oggetto di forti critiche in dottrina, in quanto avrebbe riaperto allo scientismo (cioè, la tendenza a considerare la scienza come unico valido strumento di conoscenza) predicato dalla Scuola Positiva, secondo cui il diritto penale dovrebbe avere ad oggetto in primo luogo persone socialmente pericolose, anziché fatti di reato⁵²⁸.

Ad una fase caratterizzata da una progressiva affermazione di istanze favorevoli all'introduzione nel processo penale della perizia criminologica è seguita, a partire dagli anni '80, una stagione di ripensamento sull'opportunità di tale scelta⁵²⁹. Non stupisce allora che il legislatore, nell'art. 2, n. 10 della legge 16 febbraio 1987, avente ad oggetto la delega legislativa al Governo in forza della quale è stato poi adottato il Codice Vassalli, abbia fatto genericamente riferimento al “*riordinamento dell'istituto della perizia*”, senza richiedere espressamente che venisse introdotta la perizia criminologica.

Nel Codice Vassalli, da un lato il legislatore appare maggiormente aperto al possibile contributo da parte di discipline extragiuridiche, in quanto il ricorso alla perizia, ex art. 220, comma 1, c.p.p., appare non più legato ad una situazione di necessità o al carattere accentuatamente specialistico dell'indagine che si intende compiere.⁵³⁰ Dall'altro, il legislatore recepisce l'orientamento dottrinale allora prevalente che guardava con forte perplessità, se non con decisa contrarietà, all'accertamento della personalità con le forme della perizia⁵³¹. Il divieto previsto nel codice del 1988 non è più assoluto, ma viene esplicitamente derogato dalla stessa

⁵²⁶ F. BRICOLA, *La discrezionalità nel diritto penale*, Giuffrè, Milano, 1965.

⁵²⁷ Corte cost., 24 giugno 1970, n. 124.

⁵²⁸ F. CORDERO, *Procedura penale*, 9° ed., Giuffrè, Milano, 2012, pp. 786-787.

⁵²⁹ F. GIANFROTTA, *op. cit.*, p. 577.

⁵³⁰ *Ibid.*, pp. 572-573.

⁵³¹ *Ibid.*, pp. 577-578.

disposizione con riferimento alle misure di sicurezza e all'esecuzione della pena (la deroga in materia di processo minorile era invece già presente in forza dell'allora vigente art. 11 del regio decreto 20 luglio 1934, n. 1404).

Dall'analisi storica e comparatistica emerge la peculiarità del divieto in esame che, da un lato sembra non trovare particolari riscontri in altri ordinamenti e, dall'altro, è rimasto in vigore nonostante i dubbi, sorti fin dagli anni '60, sulla presunzione assoluta di inattendibilità della perizia criminologica

4.1.3) Le *rationes* alla base del divieto e l'assimilabilità all'IA predittiva

Un'autorevole dottrina ha sostenuto che l'art. 220, comma 2, c.p.p. sarebbe posto a tutela della presunzione di innocenza, ex art. 27, comma 2, Cost., in quanto la perizia ivi prevista potrebbe avere un effetto sfavorevole sul convincimento del giudice nel giudizio sulla responsabilità⁵³². Dall'accertamento peritale di una personalità negativa, potrebbe dunque conseguire un convincimento giudiziale circa la colpevolezza dell'imputato.

A conferma di questa tesi, si può osservare che, nel Codice Vassalli, a differenza del Codice Rocco, la perizia criminologica, di regola, non è ammissibile nel processo di cognizione in cui la presunzione di innocenza assume un ruolo centrale. La perizia criminologica è invece ammissibile nel processo di esecuzione dove la presunzione di innocenza non assume rilievo, perché il giudice ha già accertato la colpevolezza dell'imputato con sentenza definitiva. Mentre, nel Codice Rocco, la norma era espressione di una generale sfiducia verso le teorie psico-criminologiche, con la conseguenza che il divieto era assoluto, nel Codice Vassalli, la norma tutela invece la presunzione di innocenza, con la conseguenza che il divieto viene meno nel processo esecutivo ove questa esigenza di tutela non si rinviene.

In senso contrario, un'altra parte della dottrina nega che, anche nel Codice Vassalli, il divieto sia riconducibile alla presunzione di innocenza. Si argomenta che, qualora il legislatore avesse voluto tutelare la presunzione di innocenza, non avrebbe ammesso giudizi prognostici sulla pericolosità sociale circa soggetti presunti innocenti, per evitare che tale giudizio finisca per influenzare il convincimento sulla responsabilità penale dell'imputato. Al contrario, il legislatore permette tali giudizi sulla pericolosità sociale, anche quando il soggetto non è stato condannato in via definitiva, sul presupposto che tale accertamento non risulta contrario alla presunzione di innocenza⁵³³.

Per questa ragione, un'altra parte della dottrina ha individuato, come *ratio* della norma, la sfiducia del legislatore nella scienza psicologica criminologica come strumento per accertare la sfera interna dell'imputato⁵³⁴. Il divieto in esame andrebbe allora interpretato in combinato disposto con l'art. 188 c.p.p. il quale prevede che “*non possono essere utilizzati, neppure con il consenso della persona interessata, metodi o*

⁵³² G. VASSALLI, *Il potere discrezionale del giudice nella commisurazione della pena*, in G. VASSALLI, *Scritti giuridici*, vol. I, Milano, 1997, p. 1334.

⁵³³ M. GIALUZ, *op. cit.*, p. 20.

⁵³⁴ F. BRICOLA, *La discrezionalità nel diritto penale. Nozione ed aspetti costituzionali*, Milano, Giuffrè, 1965, p. 116.

tecniche idonei a influire sulla libertà di autodeterminazione o ad alterare la capacità di ricordare e di valutare i fatti”⁵³⁵. La perizia criminologica sarebbe quindi vietata in quanto lesiva della libertà di autodeterminazione.

Esaminate le *rationes* dell’art. 220, comma 2, c.p.p., occorre verificare se tale disposizione si possa applicare anche all’impiego dell’IA predittiva nel processo penale.

Occorre premettere che, per espressa previsione normativa, l’art. 220, comma 2, c.p.p. non può, in ogni caso, essere invocato per escludere il ricorso all’IA predittiva nell’esecuzione della pena o nelle misure di sicurezza⁵³⁶. Pare altrettanto difficile argomentare *ex art.* 220, comma 2, c.p.p. nel senso di escludere il ricorso all’IA predittiva nel processo penale minorile, in quanto, in tale processo, si ammette il parere di esperti per finalità quali l’accertamento dell’imputabilità e del grado di responsabilità, *ex art.* 9, commi 1 e 2 del d. p.r. 448/1988. Anche nei settori in cui l’art. 220, comma 2, c.p.p. non è applicabile, il ricorso all’IA predittiva potrebbe comunque essere vietato da altre norme, come si avrà modo di evidenziare nei paragrafi successivi. Ciò premesso, in dottrina, si possono distinguere due orientamenti in merito alla possibilità o meno di applicare tale norma per escludere il ricorso all’IA predittiva nel processo penale. Secondo un primo orientamento, l’insormontabile divieto di perizia criminologica precluderebbe il ricorso a *risk assessment tools* nel processo penale⁵³⁷. Si evidenzia che il divieto di cui all’art. 220, comma 2, c.p.p. è funzionale ad evitare che il giudice possa rimanere condizionato dalle valutazioni sul carattere dell’imputato, trascurando l’accertamento sul fatto oggetto di giudizio⁵³⁸. Analizzando le domande che vengono sottoposte all’interessato in sede di utilizzo di *risk assessment tools*, è evidente come vengano in rilievo informazioni sul carattere e sulla personalità del soggetto nei cui confronti vengono poste le domande⁵³⁹. Ad esempio, *COMPAS* si basa su dati personali relativi all’impiego del tempo libero, alle inclinazioni personali o alla fiducia nella giustizia⁵⁴⁰. Vi è dunque il rischio che tramite tali strumenti le informazioni sulla personalità inducano il giudice a soffermarsi più sul carattere dell’interessato che sul fatto nella sua oggettività, con conseguente elusione del divieto di perizia criminologica⁵⁴¹.

Nello stesso senso, si è osservato che gli strumenti di valutazione del rischio rappresenterebbero una valutazione tecnico-scientifica assimilabile alla perizia, in quanto, in fase di programmazione del sistema, vengono utilizzate teorie psico-criminologiche⁵⁴². Lo stesso *COMPAS* è stato sviluppato sulla base di teorie

⁵³⁵ M. GIALUZ, *op. cit.*

⁵³⁶ Quest’ultima ipotesi è stata considerata nei parr. 2.5.1-3 (misure di sicurezza in generale) e 2.5.4 sulla pericolosità sociale del minore.

⁵³⁷ S. QUATTROCCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale “predittiva”*, in *Cassazione penale*, 2019, p. 1762; L. MALDONATO, *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Diritto penale contemporaneo*, pp. 411-412.

⁵³⁸ L. MALDONATO, *op. cit.*, p. 411.

⁵³⁹ *Ibid.*, p. 412.

⁵⁴⁰ *Sample-COMPAS-Risk-Assessment-COMPAS-“CORE”*, <https://embed.documentcloud.org/documents/2702103-Sample-Risk-AssessmentCOMPAS-CORE/>, consultato il 16 dicembre 2025.

⁵⁴¹ L. MALDONATO, *op. cit.*, p. 412.

⁵⁴² S. QUATTROCCOLO, *op. cit.*, p. 1762.

psico-criminologiche, tra cui gli studi a cura di Marie Van Nostrand, una criminologa e statistica statunitense che ha svolto delle ricerche circa i fattori predittivi del fallimento preprocessuale⁵⁴³.

La tesi secondo cui il divieto di perizia criminologica preclude il ricorso al *risk assessment tool* è coerente con l'orientamento, avallato da autorevole dottrina, secondo cui l'art. 220, comma 2, c.p.p. sarebbe posto a tutela della presunzione di innocenza⁵⁴⁴. L'esigenza di evitare che il giudice si convinca indebitamente della colpevolezza dell'imputato sulla base di elementi che afferiscono invece alla sua personalità emerge sia nella perizia criminologica che nel ricorso all'IA predittiva, in quanto, in entrambi i casi, ricorrono valutazioni inerenti alla personalità.

In dottrina, è emerso un altro orientamento secondo cui l'assimilazione dei *risk assessment tools* al divieto di perizia criminologica non sarebbe del tutto scontata⁵⁴⁵. Si argomenta che l'art. 220, comma, 2, c.p.p. non sarebbe funzionale a tutelare la presunzione di innocenza, ma sarebbe espressione della sfiducia del legislatore nei confronti delle scienze criminologiche. Di conseguenza, non sarebbe scontata l'estensione del divieto anche ai *risk assessment tools*, perché tali strumenti potrebbero non essere viziati dalle medesime criticità che il legislatore rinviene nelle perizie criminologiche.

Si osserva poi che i *risk assessment tools* sembrerebbero non essere specificamente funzionali ad esplorare la sfera interiore dell'interessato al fine di accertare “*il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche*”⁵⁴⁶.

Nello stesso senso, si è evidenziato che l'IA predittiva potrebbe essere uno strumento per supplire alla mancanza di strumenti che il giudice ha a disposizione per accertare la personalità dell'imputato⁵⁴⁷. In dottrina, è emersa l'espressione “*crisi della pericolosità sociale*” per indicare la crescente sfiducia in tale categoria⁵⁴⁸. Il tema è stato già affrontato con specifico riferimento alle misure di sicurezza, dove il legislatore fa espresso riferimento alla pericolosità sociale (artt. 202 e 203 c.p.p.).⁵⁴⁹ In generale, il problema ricorre ogni volta che il processo penale richieda una prognosi circa il futuro compimento di reati, in quanto la pericolosità sociale consiste nella probabilità che la persona commetta nuovi fatti preveduti dalla legge come reati (art. 203, comma 1, c.p.p.). Alla base della crisi di tale categoria, vi è una forte sfiducia quanto all'affidabilità e alla validità delle prognosi criminologiche.

⁵⁴³ NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015, <https://embed.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core/>, consultato il 25 settembre 2025; M. VAN NOSTRAND, *Assessing Risk among Pretrial Defendants in Virginia: The Virginia Pretrial Risk Assessment Instrument*, Virginia Department of Criminal Services, Richmond, Virginia, 2007, <https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/assessing-risk-among-pretrial-defendants-virginia-virginia-pretrial-risk-assessment-instrument.pdf>, consultato il 25 settembre 2025; Cfr. Par. 2.2.2.

⁵⁴⁴ Cfr. G. VASSALLI, *Il potere discrezionale del giudice nella commisurazione della pena*, in G. VASSALLI, *Scritti giuridici*, vol. I, Milano, 1997, p. 1334.

⁵⁴⁵ M. GIALUZ, *op. cit.*, p. 20.

⁵⁴⁶ *Ibid.*

⁵⁴⁷ M. CAIANIELLO, *potenzialità e rischi derivanti dall'interazione tra I. A. e giustizia penale preventiva* in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, p. 273.

⁵⁴⁸ G. MARINUCCI -E. DOLCINI -G.L. GATTA, *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022, p. 891. V. par. 2.5.1.

⁵⁴⁹ V. par. 2.5.2.

La sfiducia è dovuta anche alla circostanza che il giudice, a causa dell'art. 220, comma 2, c.p.p., è privato dell'ausilio di periti nell'effettuare tali prognosi, con la conseguenza che potrebbe non avere le competenze necessarie, in quanto il percorso formativo dei magistrati riguarda essenzialmente l'ambito legale.

Tale valutazione rischia poi di risultare fortemente incerta, in quanto, in assenza di un parametro statistico comune, vi è il rischio che giudici diversi abbiano metri di giudizio differenti. Un simile risultato sarebbe in contrasto con il principio di uguaglianza (art. 3 Cost.), in quanto, a parità di grado di pericolosità sociale, persone diverse potrebbero essere giudicate diversamente.

L'IA potrebbe allora supplire alla mancanza di competenze in ambito criminologico che i giudici potrebbero non avere, senza ricorrere a perizie criminologiche. Vi sarebbe poi il vantaggio di rendere le prognosi sulla pericolosità sociale meno incerte e maggiormente uniformi, con un impatto positivo anche sotto il profilo del principio di uguaglianza (art. 3 Cost.).

Come si è visto nel paragrafo antecedente, l'art. 220, comma 2, c.p.p. risulta anche peculiare dal punto di vista comparatistico e storico, essendo stato giudicato come anacronistico sin da risalente dottrina. Queste considerazioni suggeriscono quanto meno cautela nell'estendere l'ambito di applicazione di un divieto che, già di per sé, è stato oggetto di forti critiche e che non trova riscontro in ordinamenti simili a quello italiano, come a quello francese e tedesco.

In conclusione, a seconda della *ratio* che si attribuisce all'art. 220, comma 2, c.p.p. (presunzione di innocenza o sfiducia nelle scienze psico-criminologiche), si perviene ad una diversa conclusione in merito all'assimilabilità dell'IA predittiva al divieto di perizia criminologia. Pare preferibile la tesi che rinviene nel divieto in esame la sfiducia nelle scienze criminologiche e che ritiene quanto meno non scontata l'assimilabilità della perizia criminologica all'IA predittiva, anche in ragione delle potenzialità che l'IA presenta come strumento per superare la crisi della pericolosità sociale.

Peraltro, la questione potrebbe non essere dirimente alla luce dei significativi limiti che, in tale ambito, sembrano porre l'art. 5, par. 1, lett. d), del regolamento 2024/1689/UE (AI Act) e l'art. 15, comma 1 della legge 23 settembre 2025, n. 132. Come si vedrà, tali divieti rischiano di compromettere fortemente la possibilità di fare ricorso all'IA nel processo penale, privando di rilevanza concreta alla questione esaminata in questo paragrafo.

4.2) I limiti imposti dalla direttiva 2016/680/UE

4.2.1) Il divieto di processo decisionale automatizzato: introduzione

Un ulteriore limite legale all'introduzione dell'IA nel processo penale concerne l'art. 11 della direttiva 2016/680/UE (la c.d. Law Enforcement Directive o direttiva LED). Nel primo paragrafo si prevede che: “*Gli Stati membri dispongono che una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato sia vietata [...]*”.

La disposizione è stata attuata nell'ordinamento italiano dall'art. 8, comma 1, del decreto legislativo 18 maggio 2018, n. 51, il quale dispone che “*Sono vietate le decisioni basate unicamente su un trattamento automatizzato,*

compresa la profilazione, che producono effetti negativi nei confronti dell'interessato, salvo che siano autorizzate dal diritto dell'Unione europea o da specifiche disposizioni di legge". Di seguito, verrà analizzata direttamente l'art. 11 della direttiva LED, evidenziando, dove necessario, le differenze tra le due disposizioni. Il divieto in esame è simile a quanto previsto nell'art. 22, par. 1, del regolamento 2016/679/UE (il c.d. GDPR), con la precisazione che la direttiva LED si applica specificamente al trattamento dei dati personali da parte delle autorità competenti ai fini di *"prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia contro minacce alla sicurezza pubblica e la prevenzione delle stesse"* (art.1, par. 1, direttiva 2016/680/UE). L'art. 22, par. 1 del GDPR prevede che *"l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona"*. Risulta utile analizzare congiuntamente l'art. 22, par. 1, del GDPR e l'art. 11 della direttiva LED, in quanto molti dei principi elaborati con riferimento al GDPR non possono che applicarsi anche alla direttiva LED, data la somiglianza (seppur non equivalenza) tra le due norme.

4.2.2) Il divieto di processo decisionale automatizzato: i singoli requisiti

Ai sensi dell'art. 11, par. 1, della direttiva LED, affinché sussista il relativo divieto devono essere soddisfatti i seguenti requisiti: deve sussistere una decisione; tale decisione deve basarsi *"unicamente"* su di un trattamento automatizzato, compresa la profilazione; la decisione deve produrre effetti giuridici significativi o incidere negativamente sull'interessato.

Sia l'art. 22, par. 1, del GDPR, che l'art. 11, par. 1, della direttiva LED utilizzano il termine *"decisione"*. La Corte di giustizia dell'Unione europea ne ha dato un'interpretazione particolarmente ampia. Tale termine non si applica esclusivamente ad una decisione amministrativa o giudiziale. La Corte ha invece riconosciuto che anche la valutazione della solvibilità finanziaria di una persona può costituire una decisione rilevante ai fini di tale divieto, qualora abbia un ruolo decisivo nella concessione o del diniego del credito o comunque al fine di stipulare un contratto economicamente rilevante⁵⁵⁰.

Oltre alla sussistenza di una decisione, si richiede altresì che questa si basi *"unicamente"* su di un trattamento automatizzato, compresa la profilazione, quest'ultima definita come *"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"* (art. 3 n.4 della direttiva LED). Secondo le linee guida del Gruppo di lavoro ex art. 29, l'organo deputato all'applicazione uniforme del diritto europeo in materia di protezione dei dati personali, oggi sostituito dal Comitato europeo per la protezione dei dati personali, la profilazione richiede tre elementi: un trattamento automatizzato, i dati personali e l'obiettivo di valutare aspetti personali di una persona.

⁵⁵⁰ Corte di giustizia dell'Unione europea, *OQ c. Land Hessen, Schufa Holding AG*, C-634-2021, 7 dicembre 2023, parr. 44-45.

Circa il primo requisito, si è chiarito che il coinvolgimento di una persona fisica non esclude che il trattamento costituisca una profilazione, in quanto, a differenza dell'art. 22, par. 1 del GDPR, la definizione di profilazione non richiede che quest'ultima sia basata “unicamente” su di un trattamento automatizzato. Riguardo al secondo requisito, si rinvia alla definizione di dati personali di cui all'art. 4, par. 1 del GDPR: “*qualsiasi informazione riguardante una persona fisica identificata o identificabile*”. In merito al terzo requisito, si chiarisce che una mera classificazione di individui sulla base di caratteristiche note non costituisce una profilazione, bensì una mera classificazione. L'uso della parola “*valutare*” suggerisce infatti che la profilazione implica una qualche forma di giudizio su quella persona⁵⁵¹.

Sul significato dell'avverbio “unicamente”, sussiste un ampio dibattito rispetto al grado di intervento umano idoneo ad evitare che il trattamento ricada all'interno di tale divieto. Secondo una lettura restrittiva, il divieto, da interpretarsi in senso assoluto, sarebbe circoscritto alle decisioni adottate interamente dalla macchina⁵⁵². I sostenitori di questa interpretazione precisano che l'intervento umano debba comunque riguardare la sostanza della decisione e non possa esaurirsi in un vuoto passaggio procedimentale. L'essere umano può utilizzare la macchina soltanto come supporto decisionale, in quanto la decisione finale non può che rimanere a guida umana⁵⁵³. Il Gruppo di lavoro *ex art. 29* sembra aver accolto questa tesi, avendo sostenuto che, ai fini dell'applicabilità di tale divieto, è necessario che non vi sia coinvolgimento umano. Ciò non implica che, secondo il Gruppo di lavoro *ex art. 29*, sia possibile eludere tale divieto mediante un coinvolgimento umano fittizio e non effettivo. Si è infatti richiesto che la sorveglianza umana sia significativa e non soltanto un gesto simbolico e che venga effettuata da un soggetto che abbia “*l'autorità e la competenza di ribaltare la decisione*”⁵⁵⁴.

Il riferimento alla competenza e all'autorità, come requisiti necessari per garantire che l'intervento umano sia effettivo, è stato ripreso anche dal considerando 73 del regolamento 2024/1689/UE (AI Act), in materia di sorveglianza umana per i sistemi di IA ad alto rischio. Nello specifico, si richiede che “*le persone fisiche alle quali è stata affidata la sorveglianza umana dispongano delle competenze, della formazione e dell'autorità necessarie per svolgere tale ruolo*”.

In dottrina è stata poi avanzata un'interpretazione secondo cui l'avverbio “unicamente” debba, in realtà, essere interpretato in senso relativo e non assoluto. Seguendo questa linea interpretativa, dovrebbero essere oggetto

⁵⁵¹ GRUPPO DI LAVORO ART. 29, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, adottata il 3 ottobre 2017, revisionata e adottata il 6 febbraio 2018, pp. 6-8. Il Gruppo di lavoro era un organismo previsto dagli artt. 29 e 30 della direttiva 95/46/CE, *relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati*. Era composto dalle autorità di controllo di ciascuno Stato membro e la sua principale funzione era “*esaminare ogni questione attinente all'applicazione delle norme nazionali di attuazione della presente direttiva per contribuire alla loro applicazione omogenea*” (art. 30, par. 1, lett. a) della suddetta direttiva). Con l'entrata in vigore del GDPR, tale organismo è stato sostituito dal Comitato europeo per la protezione dei dati, il quale svolge essenzialmente la stessa funzione (artt. 68-76 del GDPR).

⁵⁵² L. NOTARO, *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021, <https://arpi.unipi.it/handle/11568/1147326>, consultato il 13 luglio 2025, p. 142.

⁵⁵³ M. BRKAN, *Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond*, in *Electronic Journal*, 2017, p. 10.

⁵⁵⁴ GRUPPO DI LAVORO ART. 29, *op. cit.*, pp. 20-21.

di tale divieto anche le decisioni, formalmente adottate da un essere umano, ma basate su di un procedimento automatizzato i cui risultati non siano stati attivamente verificati⁵⁵⁵. Si argomenta che, altrimenti, si applicherebbe tale divieto soltanto in casi particolarmente rari. Non sarebbe poi possibile fornire tutela nelle ipotesi in cui le discriminazioni e i danni cagionati da decisioni basate su procedimenti automatizzati risultano maggiormente frequenti. Ad esempio, non verrebbe compreso il cosiddetto *credit scoring* che le banche utilizzano per valutare la solvibilità finanziaria di chi richiede un finanziamento⁵⁵⁶. Si evidenzia anche che, per assurdo, qualora una ratifica meramente formale fosse sufficiente a eludere il divieto di cui all'art. 22, tale compito potrebbe essere svolto persino da un animale addestrato e non necessariamente da un essere umano⁵⁵⁷. Con specifico riferimento all'introduzione dell'IA nel processo penale, è stata avanzata una proposta interpretativa maggiormente esigente, secondo cui, per garantire che l'intervento umano sia effettivo, sarebbe necessario altresì che la decisione non si basi in via esclusiva sul procedimento automatizzato⁵⁵⁸. In termini penalistici, si potrebbe introdurre una regola di valutazione in forza della quale l'*output* prodotto dall'IA dovrebbe essere considerato come un mero indizio, da corroborare insieme ad altri elementi di prova⁵⁵⁹. Le stesse Linee guida sul processo decisionale automatizzato, adottate dal Gruppo di lavoro *ex art. 29*, esplicitano che il tenere conto di altri fattori, nel riesaminare il risultato del processo automatizzato, comporta che tale decisione non sarà basata “unicamente” sul trattamento automatizzato⁵⁶⁰. Tale interpretazione sarebbe doverosa anche alla luce del diritto di accesso al giudice, previsto, in caso di limitazione della libertà personale, dagli artt. 5, par. 2 e 3, della CDFUE, e 5, par. 3 e 4, della CEDU. Il “macro” diritto di accesso al giudice sarebbe quindi arricchito di un ulteriore “micro” diritto: l'interessato deve essere giudicato da un giudice persona che tenga conto di elementi di prova ulteriori rispetto al risultato prodotto dalla macchina⁵⁶¹. La necessità che la decisione non sia basata “unicamente” su di un trattamento automatizzato si ricollega alle considerazioni già svolte con riferimento all' *automation bias*⁵⁶². Se un soggetto, a causa di tale pregiudizio cognitivo, accoglie in maniera acritica il suggerimento offerto dalla macchina, la decisione sarebbe, nella sostanza, basata “unicamente” su di un trattamento automatizzato, secondo l'interpretazione relativa dell'avverbio “unicamente”⁵⁶³. La circostanza che l'*output* prodotto dall'IA venga corroborato da altri

⁵⁵⁵ L. A. BYGRAVE, *Automated Profiling – Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, in *Computer Law & Security Review*, vol. 17, no.1, 2001, pp. 17-24.

⁵⁵⁶ G. MALGIERI -G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, vol. 7, 2017, pp. 13-15.

⁵⁵⁷ *Ibid.*

⁵⁵⁸ M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019, p. 17; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia.*, in U. RUFFOLO, *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Milano, Giuffrè Francis Lefebvre, 2020, p. 20.

⁵⁵⁹ M. GIALUZ, *op. cit.*

⁵⁶⁰ GRUPPO DI LAVORO ART. 29, *op. cit.*, pp. 21-25.

⁵⁶¹ M. GIALUZ, *op. cit.*, p. 18.

⁵⁶² Cfr. par. 3.3.

⁵⁶³ G. MALGIERI -G. COMANDÈ, *op. cit.*, pp. 13-15.

elementi di prova, può contribuire a mitigare il rischio di *automation bias*, perché impone di considerare elementi ulteriori rispetto al suggerimento algoritmico.

Si noti poi l'affinità tra l'art. 11, par. 1, della direttiva LED e l'art. 14 dell'AI Act che, anche al fine di mitigare il rischio di *automation bias*⁵⁶⁴, richiede che i sistemi ad alto rischio siano oggetto di sorveglianza umana. Entrambe le disposizioni sono espressione del principio eurounitario dello *human in command*, secondo cui le decisioni adottate da sistemi automatizzati o dall'intelligenza artificiale devono rimanere sotto il controllo umano.

Tornando all'analisi dell'art. 11, par. 1, della direttiva LED, un ulteriore requisito necessario ai fini dell'applicabilità del divieto è che la decisione “*produca effetti giuridici negativi o incida significativamente sull'interessato*”.

Nel confronto con il corrispondente art. 22, par. 1 del GDPR, occorre notare che, nella direttiva LED, si prevede un divieto applicabile soltanto agli effetti giuridici negativi, mentre, nel GDPR, vi è un divieto generalizzato applicabile a qualsiasi effetto giuridico, dunque anche in *bonam partem*. Un tipico effetto giuridico negativo potrebbe essere il rafforzamento delle misure di sicurezza o di sorveglianza da parte delle autorità competenti⁵⁶⁵.

Particolarmente complessa è l'interpretazione della locuzione “*incida significativamente sull'interessato*”. Il Gruppo di lavoro ex art.29 ha richiesto che l'effetto sia sufficientemente rilevante da meritare attenzione ed incidere sull'interessato, come nel caso di un soggetto che non sia ammesso a bordo perché inserito in una lista nera⁵⁶⁶.

L'art. 22, par.1, del GDPR è formulato in maniera differente anche sotto tale profilo, in quanto il divieto si applica quando la decisione incide “*in modo analogo*” significativamente sulla persona, mentre, nell'art. 11, par. 1, della direttiva LED, la locuzione “*in modo analogo*” non compare. Sulla base dell'espressione “*in modo analogo*”, il Gruppo di lavoro ex art. 29 ha chiarito che la decisione incide significativamente sull'interessato se produce effetti simili, in termini di impatto, agli effetti che si avrebbero qualora la decisione avesse effetti giuridici. La soglia di importanza deve quindi essere analoga a quella di una decisione che produca un effetto giuridico⁵⁶⁷. Tale interpretazione è stata però adottata con riferimento al GDPR, mentre non è chiaro se l'effetto giuridico debba essere considerato come termine comparativo circa l'importanza della decisione anche con riferimento all'art. 11, par. 1, della direttiva LED.

In merito alla locuzione “*incida significativamente sull'interessato*”, la CGUE, nel caso *Schufa Holding*, ha sancito che il calcolo della solvibilità finanziaria dell'interessato rientra nell'ambito di applicazione dell'art. 22, par. 1, del GDPR se da tale tasso di probabilità dipende in modo decisivo la stipula, l'esecuzione o la

⁵⁶⁴ Si veda, in proposito, l'art. 14, par. 4, lett. b, del regolamento UE/2024/1689 che, nella versione inglese menziona espressamente l'*automation bias*, mentre, nella versione italiana, il concetto è reso con il termine “*distorsione dell'automazione*”.

⁵⁶⁵ GRUPPO DI LAVORO ART. 29, *Parere su alcune questioni fondamentali della direttiva (UE) 2016/680 sulla protezione dei dati nelle attività di polizia e giustizia*, adottato il 29 novembre 2017, pp. 12-13.

⁵⁶⁶ *Ibid*, p. 13.

⁵⁶⁷ GRUPPO DI LAVORO ART. 29, *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, Adottate il 3 ottobre 2017 e riviste il 6 febbraio 2018, p. 24.

cessazione di un rapporto contrattuale⁵⁶⁸. Nel caso di specie, il calcolo della solvibilità finanziaria era funzionale a valutare le richieste di finanziamento da parte dei consumatori nei confronti della banca. Era stato accertato che, quando veniva calcolata una scarsa probabilità di adempimento, ne conseguiva, nella quasi totalità dei casi, il rifiuto da parte della banca di concedere il finanziamento⁵⁶⁹.

Nell'art. 8, comma 1, del d.lgs. 51 del 2018, con cui il legislatore italiano ha recepito la direttiva LED, si rinviene una formulazione diversa sia rispetto al GDPR che rispetto alla direttiva LED. Si prevede che il divieto sussiste qualora le decisioni completamente automatizzate “*producano effetti negativi nei confronti nell'interessato*”. Rispetto alla direttiva LED, non si richiede esplicitamente né che tali effetti siano giuridici, né che, in alternativa, incidano significativamente sull'interessato. Adottando un'interpretazione conforme alla stessa direttiva, pare comunque preferibile interpretare la locuzione “*effetti negativi*” nel senso di effetti giuridici negativi o che incidono significativamente sull'interessato, come previsto dal diritto dell'Unione europea.

4.2.3) Le garanzie spettanti all'interessato qualora il divieto sia derogato

Dopo aver sancito il divieto precedentemente esaminato, l'art. 11, par. 1, della direttiva LED fa salva l'ipotesi in cui tale decisione “*sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento e che preveda garanzie adeguate per i diritti e le libertà dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento*”. L'art. 11, par. 1, prevede come garanzia minima, nell'ipotesi in cui tale divieto sia derogato, il diritto ad ottenere l'intervento umano.

Anche l'art. 8, comma 1, del d.lgs. n. 51/2018, legittima il processo decisionale automatizzato quando viene autorizzato dal diritto dell'Unione europea o da specifiche disposizioni di legge. Si prevede poi come garanzia minima il diritto ad ottenere l'intervento umano (art. 8, comma 2).

Anche l'art. 22, parr. 2 e 3, del GDPR prevede delle eccezioni al divieto di processo decisionale automatizzato, in presenza delle quali devono comunque essere soddisfatte determinate garanzie: il diritto ad ottenere l'intervento umano, di esprimere la propria opinione e di contestare la decisione. L'art. 11, par. 1, della direttiva LED, prevede espressamente soltanto la prima di queste garanzie, non esplicitando il diritto ad esprimere la propria opinione e di contestare la decisione.

Le eccezioni al divieto di processo decisionale automatizzato ai sensi dell'art. 22, par. 2, del GDPR sono meno stringenti rispetto all'art. 11, par. 1, della direttiva LED, in quanto nel GDPR il divieto può essere derogato non soltanto “*quando la decisione sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato*” (lett. b), ma anche quando “*sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento*” (lett. a) o “*si basi sul consenso esplicito dell'interessato*” (lett. c). Il divieto previsto dalla direttiva LED, a differenza di quello previsto dal GDPR, non

⁵⁶⁸ Corte di giustizia dell'Unione europea, *OQ c. Land Hessen, Schufa Holding AG*, C-634-2021, 7 dicembre 2023, par. 73.

⁵⁶⁹ *Ibid*, par. 48.

ammette deroghe basate sulla volontà delle parti. Si tratta di una differenza fondamentale in quanto, con riferimento al GDPR, si è notato che l'estensione delle eccezioni sembra quasi svuotare di qualsiasi utilità l'art. 22, sebbene, in quel caso, l'equilibrio venga garantito dall'applicazione delle misure a tutela dei diritti dell'interessati, ai sensi dell'art. 22, par. 3⁵⁷⁰.

Secondo il Gruppo di lavoro ex art. 29, tra le garanzie spettanti all'interessato, in caso di deroga al divieto, ai sensi della direttiva LED, rileva altresì il considerando 38 della stessa direttiva, il quale prevede il “*diritto di esprimere la propria opinione e il diritto di ottenere una spiegazione della decisione raggiunta dopo tale valutazione e di impugnare la decisione*”. Lo stesso Gruppo di lavoro ne individua la *ratio*: l'intervento umano serve ad evitare che l'interessato sia sottoposto a decisioni automatizzate indecifrabili che potrebbero essere viziate da errori o distorsioni. Il Gruppo di lavoro ribadisce poi l'importanza che l'intervento umano venga effettuato da un soggetto che disponga dell'autorità e competenza necessaria per modificare la decisione⁵⁷¹.

Di particolare interesse è il “*diritto di impugnare la decisione*”, anch'esso previsto dal considerando 38 della suddetta direttiva. Tale diritto rischia però di essere compromesso, qualora, per via dell'opacità algoritmica, l'interessato non sia pienamente a conoscenza delle ragioni a fondamento del risultato prodotto dalla macchina su cui si basa la sentenza impugnata⁵⁷².

Nel GDPR, viene invece previsto il diritto a contestare la decisione e non anche il diritto di impugnare la stessa (art. 22, par. 3 e considerando 71). Per comprendere questa differenza di disciplina, occorre ricordare che la direttiva LED si applica al trattamento di dati personali da parte delle forze di polizia e dell'autorità giudiziaria, dove l'impatto sulle libertà fondamentali è spesso diretto. Nel GDPR, il legislatore ha invece voluto introdurre un mero diritto a esprimere la propria opinione, in quanto l'impatto delle relative decisioni sui diritti fondamentali non è altrettanto diretto.

4.2.4) Decisioni e profilazioni discriminatorie

Uno degli obiettivi perseguiti dall'art. 11 consiste nell'evitare che il processo decisionale automatizzato conduca a delle discriminazioni vietate dal diritto dell'Unione. Le discriminazioni possono rientrare nell'ambito di applicazione dell'art. 11, par.1 della direttiva: “*Gli Stati membri dispongono che una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato sia vietata [...]*”⁵⁷³. La discriminazione può infatti comportare una decisione che incida significativamente sull'interessato.

L'art. 11, par. 2 prevede che, di regola, le decisioni basate unicamente su un trattamento automatizzato “*non si basano sulle categorie particolari di dati personali di cui all'articolo 10*” della stessa direttiva. In

⁵⁷⁰ G. CIACCI, *Macchine intelligenti e protezione dei dati personali*, in G. BUONOMO -G. CIACCI -A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, p. 258.

⁵⁷¹ GRUPPO DI LAVORO ART. 29, *Parere su alcune questioni fondamentali della direttiva (UE) 2016/680 sulla protezione dei dati nelle attività di polizia e giustizia*, adottato il 29 novembre 2017, p. 12.

⁵⁷² V. par. 3.2.1 e 3.2.2.

⁵⁷³ GRUPPO DI LAVORO ART. 29 *Parere su alcune questioni fondamentali della direttiva (UE) 2016/680 sulla protezione dei dati nelle attività di polizia e giustizia*, adottato il 29 novembre 2017, pp. 13-14.

particolare, l'art. 10 fa riferimento ai *“dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche o l'appartenenza sindacale, e il trattamento di dati genetici, di dati biometrici intesi a identificare in modo univoco una persona fisica o di dati relativi alla salute o di dati relativi alla vita sessuale della persona fisica o all'orientamento sessuale”*. I dati in questione meritano una specifica protezione in quanto particolarmente sensibili sotto il profilo dei diritti e delle libertà fondamentali (considerando 37)⁵⁷⁴.

Il divieto di processo decisionale automatizzato basato su particolari categorie di dati personali è stato recepito nell'ordinamento italiano dall'art. 8, comma 3, del d.lgs. 51 del 2018 che replica, nella sostanza, il divieto di cui all'art. 11, par. 2, della direttiva.

Il divieto non è assoluto ma può essere derogato a condizione che vi *“siano misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato”* (art. 11, par.2 della direttiva LED). Il processo decisionale automatizzato può quindi eccezionalmente essere basato su dati sensibili, a condizione che vi sia una base giuridica ai sensi del diritto dell'Unione o dello Stato Membro che assicuri idonee garanzie⁵⁷⁵.

L'art. 11, par. 3, della direttiva LED prevede che *“la profilazione che porta alla discriminazione di persone fisiche sulla base di categorie particolari di dati personali di cui all'articolo 10 è vietata, conformemente al diritto dell'Unione”*. Si tratta di un divieto significativo, sebbene sia meno stringente del divieto assoluto suggerito dal Parlamento europeo in prima lettura ove si era proposto di vietare, in tutti i casi, la profilazione che, intenzionalmente o meno, desse luogo a discriminazioni o che comportasse misure aventi effetti discriminatori⁵⁷⁶.

Il considerando 38 individua come fondamento di tale norma il divieto di discriminazioni di cui all'art. 21 della CDFUE. Lo stesso considerando rinvia altresì all'art. 52 della CDFUE a norma del quale eventuali limitazioni dei diritti previsti dalla Carta ne devono rispettare il contenuto essenziale, devono essere previste dalla legge e devono essere conformi al principio di proporzionalità.

Il divieto di cui all'art. 11, par. 3 della direttiva LED è stato recepito nell'ordinamento italiano dall'art. 8, comma 4 del d.lgs. 51/2018, il quale prevede che *“fermo il divieto di cui all'articolo 21 della Carta dei diritti fondamentali dell'Unione europea, è vietata la profilazione finalizzata alla discriminazione di persone fisiche sulla base di categorie particolari di dati personali di cui all'articolo 9 del regolamento UE”*. Tale disposizione, richiamando l'art. 21 della CDFUE, ribadisce ulteriormente che il fondamento della norma si rinviene nel divieto di discriminazione previsto nella Carta dei diritti fondamentali dell'Unione europea.

4.3) Il regolamento 2024/1689/UE tra pratiche vietate e sistemi ad alto rischio

⁵⁷⁴ Cfr. l'art. 9 del regolamento 2016/679UE, il quale, similmente, limita il trattamento di categorie particolari di dati personali.

⁵⁷⁵ GRUPPO DI LAVORO ART. 29, *Parere su alcune questioni fondamentali della direttiva (UE) 2016/680 sulla protezione dei dati nelle attività di polizia e giustizia*, adottato il 29 novembre 2017, pp. 13-14.

⁵⁷⁶ Parlamento Europeo, Risoluzione legislativa del 12 marzo 2014, *sulla proposta di direttiva del Parlamento europeo e del Consiglio concernente la tutela delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, e la libera circolazione di tali dati*, art. 9, par. 2-ter; M. GIALUZ, *op. cit.*, p. 18.

4.3.1) Il divieto di cui all'art. 5, par. 1, lett. d)

L'impiego dell'IA nel processo penale incontra un ulteriore limite nell'art. 5, par. 1, lett. d) del regolamento 2024/1689/UE (AI Act). La disposizione vieta: “*l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazione del rischio relative a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità*”.

Circa la *ratio* del divieto, il considerando 42 si riferisce espressamente alla tutela della presunzione di innocenza, in quanto “*le persone fisiche dell'Unione dovrebbero essere sempre giudicate sulla base del loro comportamento effettivo*”. Il riferimento al comportamento effettivo sembrerebbe riferirsi al già menzionato pericolo di deriva verso un diritto penale d'autore in cui la personalità del soggetto assume un ruolo centrale, a discapito del fatto compiuto.

Uno degli aspetti maggiormente problematici della disposizione riguarda il significato della generica espressione “*valutare o prevedere il rischio che una persona fisica commetta un reato*”. Autorevole dottrina ha interpretato la norma con riferimento alla cosiddetta polizia predittiva (c.d. *predictive policing*). Con questa espressione si fa solitamente riferimento ad una varietà di tecnologie che, utilizzando una grande quantità di dati, combinati tra loro applicando teorie criminologiche, sono in grado di prevedere i crimini potenziali e le aree geografiche in cui essi si manifestano con maggiore frequenza. Sulla base di tale previsione algoritmica vengono orientate le azioni e le strategie adottate dalle forze dell'ordine per contrastare la criminalità. La dottrina ha poi distinto tra la polizia predittiva basata sulla posizione (c.d. *location-based*) e la polizia predittiva basata sul singolo (c.d. *individual-based*). Il divieto si riferirebbe solamente alla seconda ipotesi, in quanto nella disposizione si fa riferimento al rischio che una determinata persona fisica commetta un reato. Un sistema di IA che preveda soltanto il luogo in cui un reato venga commesso senza riferirsi ad uno specifico individuo sembrerebbe invece non essere compreso⁵⁷⁷.

Sembrerebbe però che tale divieto presenti un ambito di applicazione più ampio della sola polizia predittiva. Si è osservato che l'art.5, par. 1, lett. d), si riferisce altresì a strumenti psico-criminologici volti a prevedere il rischio di recidiva e di comportamento violento da parte di uno specifico soggetto.⁵⁷⁸

A sostegno di questa interpretazione, va sottolineato come il divieto è stato formulato nella versione in inglese che, per quanto formalmente presenti lo stesso valore delle altre versioni nelle lingue ufficiali dell'Unione europea, è la lingua in cui il regolamento è stato negoziato e risulta quindi particolarmente utile a cogliere le intenzioni del legislatore⁵⁷⁹. L'italiano “*per effettuare la valutazione del rischio*” in inglese è resa con il termine “*for making risk assessments*”. Quest'ultima espressione sembra richiamare i *risk assessment tools* (in italiano, strumenti di valutazione del rischio) impiegati negli Stati Uniti da parte dell'autorità giudiziaria per stimare la

⁵⁷⁷ C. N. PEHLIVAN -N. FORGÒ -P. VALCKE, *The EU Artificial Intelligence (AI) ACT a Commentary*, Kluwer Law International B.V., Olanda, 2025, pp. 152-153.

⁵⁷⁸ S. QUATTROCCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, in *Diritto di difesa, la rivista dell'Unione delle Camere Penali Italiane*, Giuffrè Francis Lefebvre, 2025, pp. 4-5

⁵⁷⁹ *Ibid.*

probabilità che un individuo commetta nuovi reati. Sembrerebbe quindi che il legislatore intendesse vietare proprio gli stessi *risk assessment tools* adottati nel processo penale americano.

La previsione del rischio di commissione del reato, a norma dello stesso art. 5, par. 1, lett. d), è vietata soltanto se si basa unicamente sulla profilazione o sulla valutazione dei tratti e delle caratteristiche della personalità. Il concetto di profilazione non è una novità nel diritto dell'Unione europea. La profilazione viene definita allo stesso modo sia all'art. 4, par.4, del regolamento 2016/679/UE (GDPR) che all'art. 3, par.4, della direttiva 2016/680/UE (direttiva LED), nonché all'art. 3, par. 52, dell'AI Act: *“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”*.

Oltre che alla profilazione, si fa riferimento ai sistemi basati unicamente sulla valutazione dei tratti e delle caratteristiche della personalità. Il considerando 42 dell'AI Act fornisce un elenco esemplificativo di tratti che potrebbero rilevare in tal senso, tra cui la nazionalità, il luogo di nascita, il luogo di residenza o la situazione finanziaria. Lo stesso considerando fornisce degli esempi di casi in cui non si rientra nel divieto, in quanto l'analisi del rischio non si basa sulla profilazione né sulla valutazione di tratti o di caratteristiche della personalità. Si menzionano i sistemi funzionali a valutare il pericolo di frode finanziaria da parte delle imprese sulla base di transazioni sospette e viene fatto l'esempio di sistemi di IA funzionali a prevedere la localizzazione di stupefacenti da parte delle autorità doganali sulla base di rotte conosciute.

Particolarmente delicato è il significato dell'avverbio *“unicamente”*. Si tratta dello stesso termine già utilizzato dal legislatore europeo nell'art. 11, par. 1 della direttiva 2016/680/UE per vietare *“la decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato”*. A differenza di quest'ultima disposizione, nell'art. 5, par.1, lett. d), dell'AI Act, l'avverbio *“unicamente”* non si riferisce al processo decisionale. Si riferisce invece alla necessità, ai fini dell'applicabilità del divieto, che il sistema di IA si basi in via esclusiva sulla profilazione o sulla valutazione dei tratti e delle caratteristiche della personalità. Il divieto in esame può quindi essere integrato anche qualora il procedimento decisionale non sia totalmente automatizzato, a differenza del divieto di cui all'art. 11, par. 1 della direttiva LED. Ne consegue che la norma non si limita a evitare il rischio di un'eccessiva automazione, come sembrerebbe fare la Direttiva LED, ma introduce un divieto dalla portata più ampia.

I potenziali campi applicativi dell'IA nel processo penale che sono stati analizzati nel capitolo secondo della presente disamina si riferiscono a ipotesi in cui l'IA verrebbe utilizzata per supportare la p.g. o l'autorità giudiziaria e non per sostituirla. La circostanza che in questi casi l'IA avrebbe una funzione di mero supporto e non di sostituzione rispetto all'attività umana non potrebbe essere invocata nel senso di escludere tali ipotesi dall'ambito di applicazione del divieto in esame.

Lo stesso art. 5, par. 1, lett. d), dell'AI Act esclude dall'ambito di applicazione della norma il ricorso a *“sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa”*. La disposizione

non deve essere interpretata nel senso di escludere l'applicabilità del divieto per il semplice fatto che l'IA sia utilizzata a supporto di una decisione adottata da una persona fisica, anziché per sostituirla. Si ritiene piuttosto che l'AI Act vieti l'utilizzo di un software di previsione del rischio di commissione di reati sulla base di fattori quali precedenti penali o il luogo di residenza, in quanto tale previsione si basa soltanto sulla profilazione e non su fatti oggettivi e verificabili⁵⁸⁰. La circostanza che tale previsione non sia giuridicamente vincolante (ad esempio, perché il giudice se ne può discostare) non è quindi sufficiente ad escludere l'applicabilità del divieto. In definitiva, è stato introdotto un divieto dalla portata particolarmente ampia in quanto prescinde dal carattere non vincolante e dalla funzione meramente ausiliaria dell'intelligenza artificiale rispetto alla decisione umana.

4.3.2) I sistemi ad alto rischio rilevanti nel processo penale

Qualora non rientrino nel divieto di cui all'art. 5, molti dei sistemi di IA destinati ad essere impiegati nel processo penale sono comunque classificati come ad alto rischio, secondo l'elenco di cui all'allegato III del medesimo regolamento. Al n. 6 di tale allegato, si prevedono sistemi ad alto rischio nelle attività di contrasto, quest'ultima definita dall'art. 3, par. 46, come *“attività di prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, ivi incluse la salvaguardia contro le minacce alla sicurezza pubblica e la prevenzione delle stesse”*. È evidente come nell'attività di contrasto possano rientrare dei sistemi riconducibili al concetto di IA predittiva. Ad esempio, si fa riferimento ai sistemi destinati ad essere utilizzati *“per determinare il rischio per una persona fisica di diventare vittima di reati”*, oppure per *“determinare il rischio di commissione del reato o di recidiva in relazione a una persona fisica”*, nonché *“per valutare i tratti e le caratteristiche della personalità o il comportamento criminale pregresso di persone fisiche o gruppi”* (allegato III, n. 6, lett. a) e d). Il considerando 59 chiarisce che non rientrano in quest'ultima ipotesi i sistemi di IA destinati ad essere usati dalle autorità fiscali e doganali che svolgono analisi in conformità alle norme in materia di antiriciclaggio.

In altri casi, sempre riconducibili alle attività di contrasto, vi rientrano sistemi di IA che esulano dalla nozione di IA predittiva e che, pertanto, non sono stati specificamente analizzati nel capitolo secondo. Ad esempio, sono compresi sistemi destinati ad essere usati dalle autorità di contrasto come poligrafi o, più in generale, funzionali a valutare l'affidabilità degli elementi probatori (allegato III, n. 6, lett. b e c).

Sono altresì sistemi ad alto rischio quelli destinati ad essere usati dalle autorità di contrasto *“per effettuare la profilazione delle persone fisiche, nel corso di indagine, accertamento e perseguimento di reati”*. Come si è già visto, la profilazione può rientrare nel divieto di cui all'art. 5, par. 1 lett. d)⁵⁸¹. Un sistema non può essere vietato e, allo stesso tempo, essere ad alto rischio, con la conseguenza che, se ricorrono i requisiti previsti di cui all'art. 5, par. 1, lett. d), il sistema è vietato, altrimenti, viene classificato come ad alto rischio in via residuale.

⁵⁸⁰ P. VOIGT -N. HULLEN, *The EU AI Act. Answers to Frequently Asked Questions*, C.H. Beck, Monaco di Baviera 2024, pp. 42-43.

⁵⁸¹ V. p. 141.

Secondo il considerando 59, le *rationes* alla base della scelta del legislatore di classificare tali sistemi nell'ambito delle attività di contrasto come ad alto rischio sono riconducibili a diversi diritti fondamentali, tra cui il diritto alla difesa, il diritto a un ricorso effettivo (con particolare riferimento all'ipotesi in cui il sistema IA non sia trasparente), alla presunzione di innocenza, nonché al pericolo che l'utilizzo dell'IA sia un fattore di disuguaglianza⁵⁸².

L'allegato III, n. 8, prevede sistemi ad alto rischio nell'ambito dell'amministrazione della giustizia e dei processi democratici. Nella lett. a), vengono inclusi tra i sistemi ad alto rischio *“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”*. Viene compresa anche l'ipotesi, di minore rilievo ai fini del processo penale, in cui tali sistemi vengano utilizzati per la risoluzione stragiudiziale delle controversie.

Non rientrano invece all'interno di tale categoria i sistemi di IA inerenti ad attività meramente accessorie e non incidenti sulle singole decisioni giudiziarie come i sistemi funzionali alla pseudonimizzazione di documenti giudiziali (considerando 61).

Non si fa neppure espresso riferimento ai sistemi impiegati al fine di profilare i giudici. Si tratta di una questione che ha portato ad un dibattito molto acceso in diversi Stati Membri. Ad esempio, il legislatore francese ha introdotto un divieto esplicito in tal senso⁵⁸³. La legge francese 23 marzo 2019, n. 222, art. 33, comma 2, prevede che i dati identificativi dei magistrati e dei membri della cancelleria non possono essere riutilizzati con lo scopo o con l'effetto di valutare, analizzare, confrontare o prevedere le loro pratiche professionali, reali o presunte⁵⁸⁴.

Anche con riferimento ai sistemi ad alto rischio di cui all'allegato III n.8, la *ratio* è riconducibile a rischi già trattati nel capitolo terzo. Ad esempio, nel considerando 61, si fa riferimento a potenziali distorsioni o opacità di cui il sistema può essere affetto, nonché a possibili violazioni del diritto a un ricorso effettivo e a un giudice imparziale. Per mitigare tali rischi, lo stesso considerando 61 sancisce un principio fondamentale: *“L'utilizzo di strumenti di IA può fornire sostegno al potere decisionale dei giudici o all'indipendenza del potere giudiziario, ma non dovrebbe sostituirlo: il processo decisionale finale deve rimanere un'attività a guida umana”*.

Nel complesso, molti dei potenziali campi applicativi dell'intelligenza artificiale nel processo penale considerati nel secondo capitolo rientrano tra i sistemi ad alto rischio, con conseguente applicazione degli obblighi di cui agli artt. 8-27. Come si è avuto modo di notare nel terzo capitolo, l'AI Act fornisce spesso degli strumenti funzionali a tutelare i principali rischi che l'IA pone per i diritti fondamentali. Ad esempio, si è visto

⁵⁸² Si tratta, in gran parte, di rischi ampiamente discussi nel capitolo terzo, a cui si rinvia.

⁵⁸³ S. QUATTROCCOLO, *op. cit.*, pp. 10-11.

⁵⁸⁴ In originale: *“Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées. La violation de cette interdiction est punie des peines prévues aux articles 226 18, 226-24 et 226-31 du code pénal, sans préjudice des mesures et sanctions prévues par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés”*.

che il requisito della rappresentatività dei dati di cui all'art. 10, par. 4, del suddetto regolamento potrebbe mitigare il rischio di discriminazione algoritmica, in osservanza della tutela predisposta dall'art. 21 CDFUE. Sembra quindi che la disciplina dei sistemi ad alto rischio potrebbe essere lo strumento migliore per garantire una tutela effettiva per i diritti fondamentali, senza compromettere l'innovazione, anche in conformità al c.d. *risk based approach*, secondo cui le regole devono essere proporzionate al rischio che un sistema di IA presenta. Allo stesso tempo, occorre superare le criticità evidenziate nel par. 3.4. Ad esempio, non si può escludere che gli organismi europei di normazione (che ai sensi dell'art. 40, par. 1, hanno il compito di individuare gli standard la cui osservanza comporta la presunzione circa la conformità al regolamento) risultino eccessivamente condizionati da interessi privati.

4.3.3) Le difficoltà ermeneutiche nel coordinare l'art. 5 con l'allegato III

I sistemi predittivi del rischio che una persona fisica commetta un reato sembrerebbero, allo stesso tempo, essere vietati dall'art. 5 ed essere classificati come meramente ad alto rischio dall'allegato III.

In primo luogo, l'allegato III, n. 6, lett. d) classifica come ad alto rischio i sistemi IA funzionali a “*determinare il rischio di commissione del reato o di recidiva in relazione a una persona fisica non solo sulla base della profilazione delle persone fisiche*”. Il *discrimen* tra sistemi ad alto rischio e pratiche vietate riguarda l'impiego esclusivo della profilazione. Se il sistema si fonda esclusivamente sulla profilazione, la sua immissione nel mercato costituisce pratica vietata. Se il sistema non si basa in via esclusiva sulla profilazione, viene invece classificato come meramente ad alto rischio.

La *ratio* di tale distinzione sarebbe che, qualora un sistema si basi su esclusivamente su di elementi meramente soggettivi come avviene nel caso di una profilazione, si favorirebbe la deriva del c.d. diritto penale d'autore⁵⁸⁵. Si pensi, ad esempio, ad un sistema basato in via esclusiva sui precedenti penali: l'utilizzo di un simile strumento da parte dell'autorità giudiziaria o da parte della polizia giudiziaria potrebbe favorire una deriva in cui la personalità dell'agente assume un rilievo primario e il fatto di reato oggetto dell'imputazione in esame presenta un ruolo secondario.

La questione è più complessa per quanto concerne i sistemi funzionali a “*valutare i tratti e le caratteristiche della personalità*”, a cui si fa riferimento nella parte finale dello stesso allegato III, n.6, lett. d). Anche tali sistemi potrebbero rientrare nei sistemi predittivi del rischio di commissione del reato, oggetto del divieto di cui all'art. 5. Tuttavia, con riferimento a tale ipotesi, la circostanza che questi sistemi si basino o meno in via esclusiva sulla profilazione non viene indicata dall'allegato III, n. 6, lett. d) come *discrimen* tra le pratiche vietate e i sistemi ad alto rischio, con la conseguenza che la distinzione non risulta agevole.

Per risolvere questa (forse apparente) contraddizione, si è notato che l'art. 5, par. 1, lett. d) vieta l'uso di strumenti di valutazione del rischio per determinare il rischio che una (singola) persona fisica commetta un

⁵⁸⁵ M. GIANGRECO, *Gli spazi applicativi dell'intelligenza artificiale nella giustizia penale: riflessioni alla luce dell'AI act dell'Ue e del blueprint statunitense*, in *Cammino e diritto*, 2025, pp. 5-7; E. STANIG, *Il nuovo diritto penale d'autore*, in *Scuola Positiva e sistema penale: quale eredità?*, Pittaro, Trieste, 2012, pp. 45 ss.

reato. Tale norma vieterebbe quindi l'uso dei *risk assessment* nell'ottica di prevenzione nei confronti di una persona singola senza precludere l'applicazione con finalità di prevenzione relativamente ad un'intera comunità. Questa interpretazione sarebbe coerente con l'allegato III, il quale, nel legittimare gli strumenti di valutazione del rischio funzionali a valutare i tratti e le caratteristiche della personalità, sembrerebbe classificare come ad alto rischio l'applicazione di tali sistemi nei confronti di più soggetti⁵⁸⁶.

Secondo un'altra ricostruzione, non essendoci soluzioni interpretative che permettano di superare la contraddizione alla luce del significato letterale delle disposizioni in esame, occorre risolvere tale antinomia facendo prevalere l'art. 5 sull'allegato III. Tale esito interpretativo si giustificerebbe alla luce dell'obiettivo generale del regolamento concernente la tutela dei diritti fondamentali. Se si ritiene che il ricorso al *risk assessment* costituisca un rischio così elevato da dover essere inserito tra le pratiche vietate, risulta allora ragionevole interpretare l'allegato III in senso recessivo, cioè non in grado di rovesciare il divieto generale⁵⁸⁷. Secondo questa ricostruzione, i sistemi predittivi del rischio di commissione del reato, quando si basano in via esclusiva sulla profilazione, sono vietati in ogni caso. Non rileva invece la circostanza che essi possano rientrare nei sistemi funzionali a valutare i tratti e le caratteristiche della personalità che l'allegato III, n.6, lett. d) classifica come meramente ad alto rischio, perché, con riferimento a tale ipotesi, l'allegato III deve interpretarsi in senso recessivo rispetto all'art. 5.

4.4) L'art. 15 della legge 132/2025

4.4.1) L'art. 15, comma 1: la riserva al magistrato nelle attività incidenti nel merito

Tra i limiti legali all'introduzione dell'IA nel processo penale italiano, occorre, da ultimo, analizzare l'art. 15 della legge 23 settembre 2025, n. 132, con la quale il legislatore italiano ha disciplinato l'intelligenza artificiale in diversi settori, introducendo norme di complemento rispetto all'AI Act. L'art. 15, comma 1 prevede che *“nei casi di impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti.”*

La disposizione in esame è collocata subito dopo l'art. 14 della stessa legge n. 132 del 2025, dedicato all'uso dell'IA nella pubblica amministrazione (PA). Si tratta di una scelta non casuale, in quanto evidenzia le specificità del settore giudiziario rispetto ad altri settori di natura pubblicistica. In ambito giudiziario, l'esigenza di mitigare il rischio di un'eccessiva automatizzazione è maggiormente avvertita rispetto ad altri settori della PA⁵⁸⁸. Questa differenza emerge soprattutto se si confronta l'art. 14, comma 2, con l'art. 15, comma 1 della suddetta legge. Mentre in materia di PA il legislatore si limita a richiedere che l'utilizzo dell'IA avvenga *“in funzione strumentale e di supporto all'attività provvedimentale, nel rispetto dell'autonomia e del*

⁵⁸⁶ M. GIANGRECO, *op. cit.*, pp. 5-7.

⁵⁸⁷ S. QUATTROCCOLO, *op. cit.*, pp. 7-8.

⁵⁸⁸ M. IASELLI, *Commentario alla legge sull'intelligenza artificiale: Principi, governance e applicazioni settoriali - legge 23 settembre 2025, n. 132*, EPC Editore, Milano, 2025, pp. 61-64.

potere decisionale della persona” (art. 14, comma 2), nel settore della giustizia si prescrive che sia “*sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti*” (art. 15, comma 1).

Il requisito dell'autonomia e del potere decisionale della persona, afferente genericamente a tutta la PA, richiede che in ogni procedimento amministrativo che utilizzi l'IA sia sempre possibile individuare una persona fisica che ne sia responsabile e che sia dotata non soltanto della competenza tecnica, ma anche dell'autorità amministrativa necessaria per analizzare in maniera critica gli *output* prodotti dall'IA e, eventualmente, discostarsene⁵⁸⁹. Si tratta quindi di un requisito funzionale ad evitare il pericolo di un'automatizzazione eccessiva, in continuità con quanto già previsto dall'art. 14 dell'AI Act in materia di sorveglianza umana. Il requisito mira a evitare che si faccia eccessivo affidamento sull'*output* prodotto dall'IA, finendo per ricadere nel c.d. *automation bias*⁵⁹⁰.

Invece, nel disciplinare l'utilizzo dell'IA in ambito giudiziario, il legislatore non si limita a mitigare il rischio di un'automatizzazione eccessiva, ma prevede l'esclusiva competenza umana per le attività che caratterizzano il nucleo essenziale della giurisdizione⁵⁹¹. Sembrerebbe che la locuzione “*è sempre riservata al magistrato*” (art. 15, comma 1) non significhi che al magistrato debba spettare la decisione finale, sia pur supportata dall'ausilio dell'intelligenza artificiale, come pure si potrebbe pensare ad una prima analisi della disposizione in esame. Piuttosto, significherebbe che viene esclusa in radice la possibilità che il magistrato si possa avvalere dell'IA per fondare la propria decisione.

Altrimenti, non si comprende per quale motivo non prevedere, anche in ambito giudiziario, il principio dell'autonomia della persona rispetto all'intelligenza artificiale (previsto invece in materia di PA dall'art. 14). Pare dunque evidente, anche confrontando l'art. 15 con l'art. 14, che in ambito giudiziario, il legislatore introduce un limite maggiormente stringente rispetto a quanto previsto in merito alla PA. Con riferimento a quest'ultima, il limite sembra consistere nel garantire che la decisione finale spetti ad una persona fisica non soltanto nella forma ma anche nella sostanza. Invece, in ambito giudiziario, si esclude che il magistrato si possa avvalere dell'IA per fondare la decisione.

In secondo luogo, l'intenzione del legislatore di vietare l'uso dell'IA nell'attività decisionale del magistrato emerge chiaramente in un *Dossier* parlamentare in cui si legge che per le decisioni “*che costituiscono il nucleo fondamentale e più sensibile nell'esercizio della giurisdizione, viene esclusa qualsiasi possibilità di fare ricorso all'intelligenza artificiale*”. Nello stesso documento, si legge altresì che la disposizione in esame non consente “*l'impiego dei sistemi di AI riconducibili alla c.d. giustizia predittiva*”⁵⁹².

Inoltre, le ipotesi in cui, ai sensi dell'art. 15, comma 2, l'utilizzo dell'IA è legittimo afferiscono ad attività che non incidono direttamente sul merito delle decisioni giurisdizionali: organizzazione dei servizi relativi alla

⁵⁸⁹ *Ibid*, p. 60.

⁵⁹⁰ Cfr. par. 3.3.4.

⁵⁹¹ M. IASELLI, *op. cit.*, p. 61.

⁵⁹² Camera dei deputati, *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, Servizio Studi, Roma, Dossier 19 giugno 2025, p. 54.

giustizia, semplificazione del lavoro giudiziario e attività amministrative accessorie. Pertanto, non si può invocare il secondo comma dell'art. 15 per sostenere che i magistrati possano utilizzare l'IA per supportare valutazioni incidenti nel merito delle decisioni.

Le attività che costituiscono il nucleo essenziale della giurisdizione, rispetto alle quali sussiste tale divieto, sono “*l'interpretazione e l'applicazione della legge, la valutazione dei fatti e delle prove e l'adozione dei provvedimenti*”. Per quanto concerne l'interpretazione della legge, in dottrina si è sostenuto che la ratio sottesa al divieto sarebbe rinvenibile nel fatto che si tratterebbe di un'attività ineludibilmente umana, poiché non presuppone soltanto la conoscenza del diritto, ma anche la capacità di applicare le norme in relazione ai principi generali, di coordinare le norme tra loro, di adattare in relazione all'evoluzione del contesto sociale e tecnologico, nonché di bilanciare valori e interessi reciprocamente in contrasto⁵⁹³.

Tuttavia, in dottrina non mancano tesi secondo cui l'incapacità della macchina di censire valori e principi o di adattarsi al contesto sociale e tecnologico sia in realtà sopravvalutata⁵⁹⁴. Ad esempio, mediante il ricorso a tecniche di elaborazione del linguaggio naturale (*Natural Language Processing, NLP*), è stato possibile prevedere l'esito delle sentenze della Corte EDU con un'accuratezza particolarmente elevata (79%)⁵⁹⁵. Tale successo non può che mettere in discussione l'asserita incapacità della macchina di censire principi o di adattarsi al contesto. Infatti, le sentenze della Corte EDU, il cui esito, nell'esempio sopra citato, è stato possibile stimare, sono fortemente caratterizzate dall'applicazione di principi generali.

Per quanto riguarda l'incapacità di adattarsi al contesto sociale, si è poi evidenziato che la macchina presenta una minore capacità di giurisprudenza evolutiva, ma non un'assoluta incapacità, essendo in grado di elaborare molti dati sociali come la letteratura (anche non giuridica), le notizie o le banche dati⁵⁹⁶. Impiegando tali dati di natura sociale, la macchina può quindi mutare il proprio “orientamento” alla luce del contesto di riferimento. Un altro ambito oggetto del divieto in esame concerne la valutazione dei fatti e delle prove. Come possibile *ratio*, è stato evidenziato che tale attività richieda necessariamente l'intervento umano, in quanto è necessario valutare l'attendibilità delle fonti, la coerenza delle diverse ricostruzioni, nonché il peso da attribuire ad elementi probatori in contrasto tra loro⁵⁹⁷. Ad esempio, si è sostenuto che il giudice umano potrebbe avere una maggiore sensibilità nel valutare la genuinità delle dichiarazioni dei testimoni rispetto alla macchina.⁵⁹⁸ Peraltro, si tratta di considerazioni che chiariscono ulteriormente perché tali valutazioni non possono essere devolute interamente alla macchina, ma che non spiegano la scelta del legislatore di impedire altresì che il magistrato possa avvalersi dell'IA come mero supporto decisionale. In altri termini, la circostanza che il magistrato umano abbia delle abilità che la macchina potrebbe non essere in grado di replicare pienamente

⁵⁹³ M. IASELLI, *op. cit.*, p. 62.

⁵⁹⁴ U. RUFFOLO, *La macchina sapiens come «avvocato generale» ed il primato del giudice umano*, in RUFFOLO U., *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021. pp. 214-215.

⁵⁹⁵ N. ALETRAS -D. TSARAPANIS -D. PREOTIUC *et al.* *Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective*, in *PeerJ Computer Science* 2, 2016, <https://doi.org/10.7717/peerj-cs.93>, consultato il 25 settembre 2025, pp. 1-3.

⁵⁹⁶ U. RUFFOLO, *op. cit.*, p. 215.

⁵⁹⁷ M. IASELLI, *op. cit.*, p. 62.

⁵⁹⁸ *Ibid.*

non spiega perché non ricorrere ad una virtuosa interazione uomo-macchina al fine di beneficiare sia dei pregi del magistrato che dell'intelligenza artificiale.

Pertanto, il divieto di cui all'art. 15 non trova una *ratio* convincente nell'asserita infungibilità della macchina rispetto alle attività compiute dal magistrato umano. Piuttosto, il divieto pare comprensibile alla luce dei rischi analizzati nel Capitolo Terzo della presente disamina, con particolare riferimento al pericolo di discriminazione, di opacità e di *automation bias*. Si noti che la norma si applica ad ogni provvedimento giurisdizionale adottato da un magistrato, non soltanto nell'ambito del processo penale. Con riferimento a quest'ultima ipotesi emergono poi altri pericoli, anch'essi analizzati nel Capitolo Terzo, come, ad esempio, la deriva verso un diritto penale d'autore.

Rispetto all'AI Act, il legislatore italiano vieta ai magistrati di fare ricorso, nell'esercizio dell'attività giurisdizionale, a dei sistemi che, di regola, vengono classificati dal diritto dell'Unione come sistemi ad alto rischio e non come pratiche vietate. Infatti, l'allegato III, n. 8, lett. a), classifica come sistema ad alto rischio *“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”*. Peraltro, anche in assenza dell'art. 15 della legge n. 132 del 2025, molti sistemi impiegati nel processo penale rientrerebbero comunque nel divieto avente ad oggetto sistemi IA funzionali alla *“valutazione del rischio relativa a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità”*, di cui all'art. 5, par. 1, lett. d), dell'AI Act.

4.4.2) Spazi residui di ricorso all'IA nelle attività non incidenti nel merito

L'art. 15 della suddetta legge sembrerebbe non lasciare particolare spazio alla maggior parte delle ipotesi di impiego dell'IA predittiva nel processo penale italiano considerate nel Capitolo Secondo della presente disamina. In particolare, in un *Dossier* della Camera dei deputati si chiarisce che la disposizione non consente l'impiego di sistemi di giustizia predittiva, intesi come sistemi IA che, sulla base di un sistema statistico elaborato in autonomia tramite l'analisi di un'elevata quantità di atti giuridici, sono in grado di formulare una previsione, quale l'esito di un giudizio⁵⁹⁹.

Allo stesso tempo, vi sono delle possibili applicazioni dell'IA (non necessariamente predittiva) nel processo giudiziario che sia l'AI Act che la legge italiana sull'intelligenza artificiale fanno salve. In particolare, l'art. 15, comma 2, della legge n. 132 del 2025 legittima *“gli impieghi dei sistemi di intelligenza artificiale per l'organizzazione dei servizi relativi alla giustizia, per la semplificazione del lavoro giudiziario e per le attività amministrative accessorie”*. Pertanto, vengono fatte salve le attività che non incidono direttamente sul merito delle decisioni giurisdizionali⁶⁰⁰. Tale scelta è coerente con lo stesso AI Act che non soltanto non inserisce tra le pratiche vietate, ma neppure classifica come ad alto rischio *“quei sistemi destinati ad attività amministrative*

⁵⁹⁹ Camera dei deputati, *Dossier*, op. cit., p. 54.

⁶⁰⁰ M. IASELLI, op. cit., p. 62.

puramente accessorie, che non incidono sull'effettiva amministrazione della giustizia nei singoli casi” (considerando 61 dell’AI Act).

Con “*organizzazione dei servizi relativi alla giustizia*”, l’art. 15, comma 2, della legge n. 132 del 2025 fa riferimento, ad esempio, alla gestione delle risorse, alla programmazione delle udienze, nonché all’allocazione del personale amministrativo. Con “*semplificazione del lavoro giudiziario*”, la disposizione in parola mira a legittimare l’utilizzo di sistemi IA che automatizzano compiti ripetitivi e standardizzati, come, ad esempio, la predisposizione di atti processuali ricorrenti o l’estrazione di informazioni rilevanti da una grande quantità di documenti. Infine, con “*attività amministrative puramente accessorie*”, vengono comprese quelle attività amministrative che, pur essendo necessarie per il buon funzionamento dell’amministrazione giudiziaria, non sono tuttavia inerenti al merito delle decisioni giurisdizionali. Si pensi, ad esempio, all’uso dell’IA per l’archiviazione o la catalogazione dei fascicoli, oppure per redigere statistiche sull’attività giudiziaria⁶⁰¹. Con riferimento ai sistemi sopra elencati, la disposizione in esame delega al Ministero della giustizia la relativa disciplina.

Si noti che, in una precedente formulazione dell’art. 15, comma 1, erano compresi nel divieto anche i sistemi per le attività di ricerca giurisprudenziale e dottrinale. In un *dossier* della Camera dei deputati, si legge che per tali attività dovrebbe applicarsi la disciplina di cui all’art. 15, comma 2⁶⁰².

In generale, i sistemi disciplinati dal secondo comma dell’art. 15 mirano a rendere maggiormente efficiente l’attività giurisdizionale, spesso delegando alla macchina compiti particolarmente ripetitivi, lasciando così all’autorità giudiziaria e al personale amministrativo le mansioni maggiormente complesse. In questo senso, tali impieghi dell’IA in ambito giudiziario sono funzionali a soddisfare l’interesse all’efficienza processuale desumibile dall’art. 111, comma 2, Cost., in merito alla ragionevole durata del processo.

In conclusione, lo spazio residuo per il ricorso all’IA in ambito giudiziario sembra essere limitato ad attività non incidenti direttamente sul merito delle decisioni giurisdizionali. Nel complesso, l’approccio del legislatore italiano è certamente più prudente rispetto al legislatore europeo.

⁶⁰¹ *Ibid.*

⁶⁰² Camera dei deputati, *Dossier*, op. cit., p. 54.

CONCLUSIONI

Nel presente lavoro sono state analizzate le potenzialità, i rischi e i limiti legali dell'eventuale introduzione dell'intelligenza artificiale predittiva al fine di supportare l'attività della polizia giudiziaria, dei PM e dei giudici nella formulazione di giudizi prognostici nel processo penale italiano.

Per quanto riguarda le potenzialità di questa nuova tecnologia, dalla ricerca sono emerse diverse ragioni a sostegno del ricorso all'IA predittiva nel processo penale.

In primo luogo, in alcuni ambiti, come la commisurazione della pena, l'IA potrebbe rendere la valutazione giudiziale maggiormente standardizzata, riducendo l'influenza di fattori soggettivi.⁶⁰³ Inoltre, si avrebbe una maggiore uniformità di giudizio, in quanto giudici diversi dovrebbero fare riferimento al medesimo modello predittivo⁶⁰⁴. Ancora, negli accertamenti in cui assume rilievo la pericolosità sociale, l'IA potrebbe supplire alla sfiducia del legislatore quanto alla capacità delle scienze criminologiche di accertare la personalità dell'imputato⁶⁰⁵. In alcuni campi applicativi, come, ad esempio, nella scelta del PM tra archiviazione ed esercizio dell'azione penale (artt. 50, comma 1 e 408 c.p.p.), il ricorso all'IA sarebbe altresì funzionale a soddisfare l'interesse all'efficienza processuale, desumibile dal principio della ragionevole durata del processo (art. 111, comma 2, Cost.).

Per quanto riguarda, invece, i potenziali campi applicativi dell'IA predittiva nel processo penale, è emerso che in alcuni casi è possibile fare affidamento sull'esperienza di ordinamenti stranieri per individuarne le potenzialità e i limiti. Ad esempio, per quanto concerne le esigenze cautelari del pericolo di reiterazione del reato e del pericolo di fuga (art. 274, lett. b) e c) c.p.p.), si può tenere conto della previsione del rischio di fallimento preprocessuale da parte di *COMPAS* e *PSA*. Allo stesso modo, relativamente all'impiego dell'IA nella determinazione della pena, non si può non considerare il dibattito statunitense in merito all'utilizzo dei *risk assessment tools* nel c.d. *sentencing*.

In altri casi, il ricorso all'IA sembra avere un carattere maggiormente speculativo, come, ad esempio, nel caso dell'esigenza cautelare dell'inquinamento probatorio (art. 274, lett.a), c.p.p.), nell'accertamento della pericolosità sociale (artt. 202 e 203 c.p.), nelle misure di prevenzione (d.lgs. 6 settembre 2011, n. 159) o nel supporto alla scelta del PM tra archiviazione o esercizio dell'azione penale (artt. 50, comma 1 e 408 c.p.p.). Nei casi in cui non è possibile fare pienamente affidamento sull'esperienza di ordinamenti stranieri, occorre maggiore cautela, in quanto i rischi e i benefici potrebbero essere meno chiari, non essendoci ancora stata una sperimentazione in tal senso.

Occorre poi segnalare che, se le norme impongono esplicitamente al giudice di basare il giudizio su elementi oggettivi, il ricorso all'IA presenta minori criticità. Ad esempio, la disciplina legislativa sulla pericolosità sociale del minore (art. 37, comma 2, d.P.R. 22 settembre 1988 n. 448) richiede che tale accertamento avvenga sulla base delle specifiche modalità e circostanze del fatto, mentre la pericolosità sociale *ex* artt. 202 e 203 c.p.

⁶⁰³ Cfr. parr. 2.1 e 2.7.1.

⁶⁰⁴ Cfr. par. 2.1

⁶⁰⁵ Cfr., parr. 2.1 e 2.5.2.

risulta maggiormente incerta⁶⁰⁶. Quando le norme non impongono esplicitamente al giudice di basare il giudizio su elementi oggettivi, il giudice potrebbe limitarsi a rinviare all'output algoritmico, con l'ulteriore conseguenza negativa di favorire indirettamente un diritto penale d'autore.

Per quanto concerne i rischi analizzati nel capitolo terzo, alcune questioni sono comuni ad ogni ipotesi di impiego dell'IA a supporto delle decisioni umane, come il pericolo di discriminazione, l'opacità algoritmica e i *bias* cognitivi. In questo caso, anche l'introduzione dell'IA nel processo penale deve tenere conto del più ampio dibattito europeo sulla concezione antropocentrica dell'intelligenza artificiale, secondo cui è l'IA che deve essere a servizio dell'uomo e non il contrario⁶⁰⁷.

Nel complesso, vi è il pericolo che si passi da un diritto penale del fatto, in cui vige la presunzione d'innocenza, a un diritto penale d'autore, in cui l'accusa può limitarsi a provare la pericolosità sociale, finendo per ridare nuova linfa alle ormai superate tesi predicate dalla Scuola Positiva. Per evitare che tale pericolo si concretizzi, è essenziale che le valutazioni giudiziali, pur supportate dall'IA, rimangano ancorate ad elementi di fatto.

Occorre poi notare che, alla luce dei rischi esaminati, il ricorso all'IA nel processo penale rischia di svalutare numerosi principi e diritti costituzionali e sovranazionali, come ad esempio, il principio di uguaglianza (art. 3 Cost.), con riferimento al pericolo che gli algoritmi siano discriminatori, il diritto alla difesa (art. 25, comma 2, Cost.), nell'ipotesi in cui non sia possibile comprendere il funzionamento del sistema di IA, e la presunzione di innocenza (art. 27, comma 2, Cost.), in caso di indebito convincimento del giudice circa la colpevolezza dell'imputato sulla base dell'output algoritmico e non in ragione del fatto contestato.

In alcuni casi il diritto vigente offre già delle garanzie potenzialmente idonee quanto meno a mitigare tali pericoli. In particolare, molte soluzioni sono offerte dall'AI Act. Si pensi all'art. 10 di tale regolamento che, richiedendo la rappresentatività dei dati, mira ad evitare che dati di scarsa qualità conducano a delle distorsioni foriere di discriminazioni. Ancora, l'art. 14, par. 4, al fine di prevenire l'*automation bias*, prevede che coloro i quali sono deputati alla sorveglianza dei sistemi IA siano consapevoli del rischio di fare eccessivo affidamento sul risultato prodotto dall'intelligenza artificiale. In questo senso, è essenziale che le garanzie presenti nell'AI Act non rimangano sulla carta, ma vengano implementate adeguatamente nell'ordinamento italiano. In particolare, sarà decisivo il ruolo svolto dall'Agenzia per la Cybersicurezza Nazionale (ACN), la quale, in questo settore, svolgerà la funzione di autorità di sorveglianza del mercato.

Vi sono poi dei casi in cui i rischi analizzati nel capitolo terzo necessitano non soltanto di rimedi legali, ma anche di soluzioni tecniche. Ad esempio, l'*Explainable AI*, cioè l'insieme delle tecniche per comprendere il funzionamento dei sistemi IA, potrebbe essere una valida soluzione, per quanto in via di sviluppo, per risolvere il problema della *black box*.

Occorre poi notare la natura fortemente multidisciplinare della materia. Infatti, l'analisi delle sfide presentate dall'IA richiede di tenere presente studi che spaziano dalla statistica (si pensi ai criteri di equità algoritmica)⁶⁰⁸,

⁶⁰⁶ Cfr. par. 2.5.4.

⁶⁰⁷ Cfr. 1.3.3.

⁶⁰⁸ Cfr. par. 3.1.2.

all'informatica (come nel caso del problema della *Black box* o all'*Explainable AI*). Sul versante prettamente giuridico, è fondamentale tenere conto sia della normativa penale interna che di quella europea in materia di intelligenza artificiale o di protezione dei dati personali. A tal fine, è fondamentale la collaborazione tra studiosi di discipline differenti, in modo tale che le risposte alle sfide poste dall'IA si basino su di una visione complessiva del fenomeno.

Per quanto concerne i limiti posti dal diritto vigente, analizzati nel capitolo quarto, una garanzia significativa, nel senso di evitare un'eccessiva automazione, è posta dalla direttiva LED la quale, all'art. 11, par. 1, vieta, a determinate condizioni, il processo decisionale automatizzato.

Il legislatore europeo e italiano, sembrano invece essere andati oltre l'esigenza di evitare un'eccessiva automazione, introducendo due divieti dall'ambito di applicazione particolarmente ampio: l'art. 5, par. 1, lett. d dell'AI Act e l'art. 15, comma 1, della legge 23 settembre 2025, n. 132. Alla luce di tali divieti, il dibattito sulla possibilità di assimilare l'IA al divieto di perizia criminologica *ex art.* 220, comma 2, c.p. potrebbe perdere la sua rilevanza in concreto, in quanto il ricorso ai *risk assessment tools* potrebbe, in ogni caso, ricadere nei divieti previsti dall'AI Act e dalla legge 23 settembre 2025, n. 132.

Infine, resta da valutare la coerenza di tali divieti con il c.d. *risk based approach*, adottato dall'AI Act, secondo cui le regole devono essere proporzionate al livello di rischio che un determinato sistema o una determinata pratica presenta⁶⁰⁹.

Da un lato, indubbiamente, i pericoli che l'IA pone rispetto a un'ampia gamma di diritti fondamentali permettono quanto meno di comprendere le ragioni alla base della prudenza che sembra emergere sia dalla disciplina nazionale che da quella sovranazionale. Allo stesso tempo, si sarebbe anche potuto consentire, senza divieti, il ricorso all'IA nel processo penale, per poi mitigare i rischi che l'IA pone per i diritti fondamentali prevedendo, per ciascun rischio emerso, puntuali obblighi normativi. D'altronde, lo stesso AI Act prevede, per i sistemi ad alto rischio, diversi obblighi (ai sensi degli artt. 8 e ss.) funzionali ad evitare che l'IA risulti lesiva per i diritti fondamentali. Classificare l'utilizzo dell'IA nel processo penale come sistema ad alto rischio, senza prevedere divieti di ampia portata né nello stesso AI act, né nella disciplina nazionale potrebbe essere, dunque, il miglior punto di equilibrio tra l'esigenza di tutelare i diritti fondamentali e le innegabili opportunità che l'IA presenta.

In definitiva, l'introduzione dell'IA predittiva nel processo penale presenta opportunità significative, come la standardizzazione delle valutazioni giudiziali e una maggiore efficienza, ma comporta anche rischi concreti per i diritti fondamentali, come l'uguaglianza, il diritto alla difesa e la presunzione di innocenza. La sfida principale, in conclusione, consiste nel trovare un equilibrio ponderato, che consenta di sfruttare i benefici dell'IA senza compromettere le garanzie costituzionali e sovranazionali.

⁶⁰⁹ Cfr. par. 1.3.3.

Bibliografia

- AGUDO U. -LIBERAL K. G. -ARRESE M., *et al. The impact of AI errors in a Human-in-the-loop Process*, in *Cognitive Research: Principles and Implications*, 2024.
- ALBRIGHT A., *If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions* in *University of Toronto Law Journal*, vol. 73, 2023.
- ALETRAS N. -TSARAPANIS D. -PREOTIUC D., *et al. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, in *PeerJ Computer Science* 2, 2016.
- ALMADA M.-PETIT N., *The EU AI Act: between the rock of product safety and the hard place of fundamental rights*, in *Common market law review*, vol. 62, 2025, pp. 1114-1118.
- ALON-BARKAT S. -BUSUIOC M., *Human–AI Interactions in Public Sector Decision Making: ‘Automation Bias’ and ‘Selective Adherence’ to Algorithmic Advice*, in *Journal of Public Administration Research and Theory*, vol. 33, 2023, pp. 513-524.
- AMATO G., *Individuo e autorità nella disciplina della libertà personale*, Giuffrè, Milano, 1967.
- ANGWIN J. -LARSON J. -MATTU S., *et al., Machine Bias: Risk Assessments in Criminal Sentencing*, ProPublica, 2016.
- ANTOLISEI F., *La capacità a delinquere* in *Rivista italiana di diritto penale (RIDP)*, CEDAM, Padova, 1934, p. 68.
- ASHLEY K. D., *Introducing AI & Law and Its Role in Future Legal Practice*, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, Cambridge, 2017, pp. 1-6.
- ASHWORTH A. -ZEDNER L., *Preventive Justice*, Oxford University Press, Oxford, 2014.
- BARREDO ARRIETA A., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*, in *Information Fusion*, 2020, p. 283.
- BAR HILLER M., *The base rate fallacy in probability judgements*, *Acta psychologica*, 1980.
- BARTOLI R. -PELLISSERO M. -SEMINARA S., *Diritto penale. Lineamenti di parte speciale*, seconda edizione, Giappichelli, Torino, 2022.
- BASILE F., *Esiste una nozione ontologicamente unitaria di pericolosità sociale? Spunti di riflessione con particolare riferimento alle misure di sicurezza e di prevenzione*, in *La pena, ancora, fra attualità e tradizione*, Giuffrè, 2018, pp. 975-978.
- BASILE F., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in G. BALBI G. -DE SIMONE F. -ESPOSITO A. *et al. Diritto penale e intelligenza artificiale: nuovi scenari*, Torino, Giappichelli Editore, 2022, p. 22.
- BELLE V.-PAPANTONIS I., *Principles and Practice of Explainable Machine Learning*, in *Frontiers in Big Data*, 2021, pp. 6-8.

- BENCH CAPON T. -SARTOR G., *A Model of Legal Reasoning with Cases Incorporating Theories and Values in Artificial Intelligence*, Amsterdam, Paesi Bassi, 2003, pp. 1-6.
- BERK R. -HEIDARI H. -JABBARI S. *et al.* *Fairness in Criminal Justice Risk Assessments: The State of the Art*, in *Sociological Methods & Research*, 2018, pp. 13-15.
- BETTIOL G., *Diritto penale: parte generale*, 11°ed., CEDAM, Padova, 1982.
- BIN R. – PITRUZZELLA G., *Diritto costituzionale*, Giappichelli, Torino, 2025, pp. 468-488.
- BLAIOTTA R., *Giustizia, errore, intelligenza artificiale*, in *Sistema penale*, 2023, pp. 14-17.
- BLOMBERG T., -BALES W. -MANN K., *Validation of the Compas Risk Assessment Classification Instrument*, Center for Criminology and Public Policy Research College of Criminology and Criminal Justice Florida State University, Tallahassee, Florida, 2010.
- BONTA J. -ANDREWS D. A., *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, Public Safety Canada, Carleton University, 2007.
- BRICOLA F., *La discrezionalità nel diritto penale. Nozione ed aspetti costituzionali*, vol.1, Giuffrè, 1965.
- BRKAN M., *Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond*, in *Electronic Journal*, 2017, p. 10.
- BROUWER E., *Challenging Bias and Discrimination in Automated Border Decisions, Ligue des droits humains and the Right to Effective Judicial Protection*, in *Verfassungsblog*, 2023.
- BROUWER E., *Ligue des droits humains and the validity of the PNR Directive: Balancing individual rights and State powers in times of new technologies*, in *Common Market Review*, 2023, pp. 855-857.
- BROŽEK B., *The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making*, in *Artificial Intelligence and Law*, 2024, pp. 429-430.
- BYGRAVE L.A., *Automated Profiling – Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, in *Computer Law & Security Review*, vol 17, no.1, 2001, pp. 17-24.
- BUONOMO G., *Macchine intelligenti, processo induttivo e apprendimento autonomo*, in BUONOMO G. -CIACCI G.- COSTANZO A., *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, pp. 3-36.
- CADOPPI A. -CANESTRARI S.-MANNA A., *et al*, *Parte generale III, la punibilità e le conseguenze del reato*, in *Trattato di diritto penale*, UTET giuridica, Milanofiori Assago (MI), 2014, pp. 202-208.
- CAIANIELLO M., *potenzialità e rischi derivanti dall'interazione tra I. A. e giustizia penale preventiva* in RUFFOLO U., *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2025, pp. 265-279.
- CAMERA DEI DEPUTATI, *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, Servizio Studi, Roma, Dossier 19 giugno 2025, pp. 54 e ss.
- CAMON A. -DANIELE M. -NEGRI D., *Fondamenti di procedura penale*, 4° ed., Wolters Kluwer, CEDAM, Milano, 2023, pp. 101-104.
- CAMPOBASSO G. F., *Diritto commerciale, vol. 1, Diritto dell'impresa*, UTET giuridica, Milano, 2021.

- CHEN S., *Chinese Scientists Create 'World First' AI Prosecutor that Can Press its own Charges* in *South China Morning Post*, 2021.
- CHIAVARIO M. *Commentario al nuovo codice di procedura penale*, UTET, Torino, 1990.
- CHIAVARIO M., *L'obbligatorietà dell'azione penale: il principio e la realtà*, in *Cass. pen.*, 1993.
- Codice penale e codice di procedura penale*, *Lav. Prep.*, vol. V, pt. 1, 1929, p. 2658.
- CHIAVARIO M., *Procedure penali d'Europa. Belgio, Francia, Germania, Inghilterra, Italia. Sintesi nazionali e analisi comparatistiche*, CEDAM, Padova, 2020.
- CIACCI G., *Macchine intelligenti e protezione dei dati personali*, in G. BUONOMO -G. CIACCI -A. COSTANZO, *Macchine Intelligenti e diritto*, Giappichelli, Torino, 2025, pp. 223-260.
- COMMISSIONE EUROPEA, *Comunicazione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni – Creare fiducia nell'intelligenza artificiale antropocentrica*, COM (2019) 168 final, Bruxelles, 2019.
- COMMISSIONE PER L'EFFICIENZA DELLA GIUSTIZIA DEL CONSIGLIO D'EUROPA(CEPEJ), *Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, Consiglio d'Europa, 2018.
- CONSO G. -GREVI V.-BARGIS M., *Compendio di procedura penale*, 11° edizione, CEDAM, Milano, 2023, pp. 368-370 e 518-520.
- COPPOLA F., *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo*, in *Archivio Penale*, 2023.
- CORDERO F., *Le situazioni soggettive nel processo penale*, Utet, Torino, 1956, pp. 158 e ss.
- CORDERO F., *Procedura penale*, 9° ed., Giuffrè, Milano, 2012, pp. 786-787.
- DE VRIES S. -KANEVSKAIA O. -DE JAGER R., *Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation*, in *European Papers*, vol. 8, n. 2, 2023, pp. 594-602.
- DIETERICH W. -MENDOZA C.-BRENNAN T., *COMPAS Risk Scales: Demonstrating Accuracy, Equity and Predictive Parity*, Northpointe Inc., 2016, pp. 1-10.
- DI FLORIO M., *Calculate Criminal Law? Criticità nell'uso degli algoritmi di pericolosità sociale*, in *La Legislazione Penale*, 2023.
- DI NAPOLI M., *Tutela dei diritti ed efficace esercizio dell'azione penale: il punto di vista di una procura della repubblica*, in *L'azione penale tra obbligatorietà e discrezionalità*, Milano, 2009, p. 45.
- DOMINIONI O., *Azione obbligatoria penale e efficienza giudiziaria*, in *Diritto penale e procedura penale*, 2020, p. 4.
- DOMINIONI O., *Interesse alla persecuzione penale e irrilevanza sociale del fatto nel prisma dell'efficienza giudiziaria*, in *La giustizia penale*, 2010, p. 323.

EDWARDS L. -VEALE M., *Slave to the Algorithm? Why 'a Right to an Explanation' is Probably not the Remedy you are Looking for*, in *Duke Law & Technology Review*, vol. 16, 2017, pp. 51-56.

AGENZIA DELL'UNIONE EUROPEA PER I DIRITTI FONDAMENTALI, *Getting the Future Right*, 2021, Lussemburgo, 2020, pp. 60-81.

EVAS T., *The EU Artificial Intelligence Act: Advancing Innovation for Trustworthy AI*, in *Journal of AI Law and Regulation*, vol. 1, n. 1, 2024, pp. 99-100.

FAIR TRIALS, *FOI reveals over 12,000 people profiled by flawed Durham police predictive AI tool*, in *Fair Trials*, 2022.

FARID H. -DRESSEL J., *The Accuracy, Fairness, and Limits of Predicting Recidivism*, in *Science Advances*, vol. 4, n.1, American Association for the Advancement of Science, 2018.

FERRARA E., *The Butterfly Effect in Artificial Intelligence Systems: Implications for AI Bias and Fairness*, in *Machine Learning with Applications*, 2024.

FIANDACA G. -MUSCO E., *Manuale di diritto penale. Parte generale*, 9^a edizione, Bologna, Zanichelli, 2024.

FINK M., *Human Oversight under art. 14 of the EU AI Act*, Leiden University, Europa Institute, Católica Global School of Law, Central European University (CEU) - Department of Legal Studies, Leiden, Olanda, 2025, pp. 11-12.

FORNARI L., *Misure di sicurezza e doppio binario: un declino inarrestabile?* in *Rivista italiana di diritto e procedura penale*, Milano, 1993, p. 637.

FRAGASSO B., *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, in *Diritto penale contemporaneo*, 2023, pp. 33-34.

GAITO A -BARGI A. *et al.*, *Codice di procedura penale commentato*, UTET Giuridica, Torino, 2012, *sub art.* 274, 282-ter, 384.

GAITO A. -ROMANO B., RONCO M. *et al.*, *Misure di prevenzione*, UTET Giuridica, Torino, 2013, pp. 99.

GALANTINI N., *Il principio di obbligatorietà dell'azione penale, tra interesse alla persecuzione penale e interesse all'efficienza giudiziaria*, in *Diritto penale contemporaneo*, 2019, p. 2.

GATTA G.L., *Riforma della giustizia penale. Contesto, obiettivi e linee di fondo della "legge Cartabia"*, in *Sistema Penale*.

GEIGER C., *The Social Function of Intellectual Property Rights, or How Ethics can Influence the Shape and Use of IP Law*, Max Planck Institute for Intellectual Property and Competition Law, in G.B. DINWOODIE *Intellectual Property Law: Methods and Perspectives*, Cheltenham, Regno Unito 2013, pp. 153-176.

GIALUZ M., *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 2019.

- GIALUZ M., *Per un processo penale più efficiente e giusto. Guida alla lettura della riforma Cartabia*. In *Sistema penale*, 2025, pp. 42-25.
- GIANFROTTA F., *Perizia*, in M. CHIAVARIO, *Commento al nuovo Codice di procedura penale*, Utet, Torino, 1990, *sub art.* 220.
- GIANGRECO M., *Gli spazi applicativi dell'intelligenza artificiale nella giustizia penale: riflessioni alla luce dell'AI act dell'UE e del blueprint statunitense*, in *Cammino e diritto*, 2025, pp. 5-7.
- GIARDA A. - SPANGHER G. *Codice di procedura penale commentato*, 4° edizione, Ipsoa, Milano, 2010, *sub art.* 381.
- GIOVAGNOLI R., *Manuale di diritto penale. Parte generale*, V edizione, Ita edizioni, 2025, Torino, pp. 222-229.
- GIOVAGNOLI R., *Manuale di diritto penale. Parte speciale*, IV edizione, Ita edizioni, 2025, Torino.
- GRAMUGLIA V., *Obbligatorietà dell'azione penale e ruolo dell'IA nelle scelte del pubblico ministero*, in *Archivio Penale*, Giuffrè Francis Lefebvre, Milano, 2024, pp. 1-29.
- GRAVETT W.H., *Judicial Decision-Making in the Age of Artificial Intelligence* in SOUSE ANTUNES H., *Multidisciplinary Perspectives on Artificial Intelligence and the Law, Law, Governance and Technology Series*, Springer International Publishing, Cham, 2023, p. 291.
- GRAY R., *Probation: An Exploration in Meaning*, National Criminal Justice Reference Service, Washington D.C., 1988, p. 26.
- GRUPPO DI LAVORO ART. 29 PROTEZIONE DEI DATI, *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, Adottate il 3 ottobre 2017 e riviste il 6 febbraio 2018.
- GRUPPO DI LAVORO ART. 29 PROTEZIONE DEI DATI, *Parere su alcune questioni fondamentali della direttiva (UE) 2016/680 sulla protezione dei dati nelle attività di polizia e giustizia*, adottato il 29 novembre 2017.
- GRUPPO DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'IA affidabile*, Commissione europea, Bruxelles, 2019, pp. 15-23.
- GULOTTA G. - EGNOLETTI P. - NICCOLAI B. *et. al.*, *Tendenze generali e personali ai bias cognitivi e la loro ricaduta in campo forense: fondamenti e rimedi*, in *Sistema Penale*, 2021, p. 6.
- HARDT M. - PRICE E. - SREBRO N. *Equality of Opportunity in Supervised Learning*, in LEE D.D., SUGIYAMA V., LUXBORG U.V. *et al.* *Equality of Opportunity in Supervised Learning. Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems*, Barcellona, Spagna, 2016, pp. 3315-3323.
- IASELLI M., *Commentario alla legge sull'intelligenza artificiale: Principi, governance e applicazioni settoriali - legge 23 settembre 2025, n. 132*, EPC Editore, 2025, pp. 61-64.

- IGUALADA R., *Domestic Violence and Family Crises: A perspective from Spanish Law in Familia. Il diritto della famiglia e delle successioni in Europa*, 2018, pp. 389-391.
- KATZ D.M. -BOMMARITO M.J. -BLACKMAN J. *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, in *PLOS ONE* 12, 2017.
- KAZIM T. -TOMLINSON J., *Automation Bias and the Principles of Judicial Review in Judicial Review*, 2023, p 9.
- KEHL D.- GUO P. -KESSLER S., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2017, pp. 16-25.
- KLEINBERG G.-MULLAINATHAN S. -RAGHAVAN A.M., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, in *arXiv preprint*, 2019, p. 17.
- KLINGELE C., *The Promises and Perils of Evidence-Based Corrections*, in *Notre Dame Law Review*, 2015.
- KOSTORIS R.E., *Per un'obbligatorietà temperata dell'azione penale*, in *Rivista di Diritto Processuale*, 2007, p. 577.
- KOVIC M. -KRISTIANSEN S., *The Gambler's Fallacy Fallacy (Fallacy)*, in *Journal of risk research*, Routledge, Taylor & Francis Group, 2017, p. 1.
- LA GIOIA F. -SARTOR G., *Il sistema compas: algoritmi, previsioni, iniquità*, in U. RUFFOLO XXVI *Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 226-241.
- LARSON J. -MATTU S- KIRCHNER L., *et. al. How We Analyzed the Compas Recidivism Algorithm*, ProPublica, 2016.
- LATTANZI G.-LUPO E., *Codice penale. Rassegna di giurisprudenza e dottrina*, Giuffrè, Milano, 2010, p.9.
- LEVY K. -CHASALOW K. -RILEY S., *Criminal Justice profiling and EU Data Protection Law: Precarious Protection from Predictive Policing*, in *International Journal of Law in Context*, Cambridge University Press, 2022.
- MACRI L., *I primi passi dell'Italia verso l'impiego dell'IA nel processo penale e il calcolo del rischio di recidiva* in *Giurisprudenza penale web*, 2025, p. 41.
- MAHONY S.O., *Medicine and the McNamara fallacy*, in *National Library of Medicine*, Edinburgo, Scozia, 2017.
- MALDONATO L., *Algoritmi predittivi e discrezionalità del giudice: una nuova sfida per la giustizia penale*, in *Diritto penale contemporaneo*, 2019, pp. 411-412.
- MALGIERI G., COMANDÈ G., *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, vol. 7, 2017, pp. 13-15.
- MALINVERNI A., *Principi di diritto penale tributario*, Milano, Giuffrè, 1984, p. 474.

- MALONEY K., KILEEN J., MALONEY A., *The Use of HemoSpat to Include Bloodstains Located on Nonorthogonal Surfaces in Area-of-Origin Calculations*, in *Journal of Forensic Identification*, Vol. 59, 2009, pp. 513-524.
- MANES V., *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia.*, in U. RUFFOLO, *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Milano, Giuffrè Francis Lefebvre, 2020, p. 20.
- MARINUCCI G. -DOLCINI E. GATTA G. L., *Manuale di diritto penale. Parte generale*, 11° edizione, Milano, Giuffrè Francis Lefebvre, 2022.
- MARTINI A., *La pena sospesa in Itinerari di diritto penale*, Giappichelli, 2001, p. 208.
- MOSIER K. L. -SKITKA L. J., *Automation Use and Automation Bias*, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, n. 3, 1999, pp. 344-348.
- MUGARI I -OBIOHA E., *Predictive Policing and Crime Control in The United States of America and Europe: Trends in a Decade of Research and the Future of Predictive Policing*, in *Social Sciences*, 2021.
- NARAYANAN A. -KAPOOR S., *AI Snake Oil*, Princeton University Press, Princeton, 2024, pp. 43-49.
- NORTHPOINTE INC., *Practitioners Guide to COMPAS Core*, Northpointe Institute for Public Management, 2015.
- NOTARO L., *Intelligenza artificiale e giustizia penale*, in MASSARO A., *Intelligenza artificiale e Giustizia Penale*, Università degli Studi "Roma Tre", Roma, 2020, pp. 93-149.
- NOTARO L., *Predictive Algorithms and Criminal Justice: A Synthetic Overview from an Italian and European Perspective*, Università di Pisa, 2021.
- NUVOLONE P., *La capacità a delinquere nel sistema del diritto penale*, in *Trent'anni di diritto e procedura penale: studi*, vol. I, CEDAM, Padova, 1969, p. 578.
- OSWALD M.-GRACE J.-URWIN S.-BARNES G.C., *Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality*, in *Information & Communications Technology Law*, Taylor & Francis, 2018, p. 227.
- OXFORD INSTITUTE OF TECHNOLOGY AND JUSTICE, *Tech and Justice Atlas*, Oxford Institute of Technology and Justice – Blavatnik School of Government, 2025.
- PAULESU P., *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Archivio penale*, 2022, pp. 1-16.
- PARLAMENTO EUROPEO, Comitato LIBE, *Studio sull'impatto dell'intelligenza artificiale sui diritti fondamentali nella U.E.*, 2020.
- PEHLIVAN C.L. -FORGÒ N.-VALCKE P., *The EU Artificial Intelligence (AI) ACT: a Commentary*, Kluwer Law International B.V, Olanda, 2025, pp. 152-153 e 265-298.
- PETRINI D., *La prevenzione inutile. Illegittimità delle misure praeter delictum*, Giovene, Napoli, 1996, p. 22.
- PETROCELLI B., *La funzione della pena* In *Saggi di diritto penale*, CEDAM, Padova, 1952, p. 403.

- PIETROCARLO E., *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali in Sistema penale*, 2023, pp. 1-2.
- PILA J.-TORREMANS P., *European Intellectual Property Law*, Oxford University Press, Oxford, 2019, pp. 513-515.
- POUGET H, -HOLTMAN K. -EMBORG T., *EU Artificial Intelligence Act, Standard Setting*, 2025, <https://artificialintelligenceact.eu/standard-setting-overview/>.
- QUATTROCOLO, S. *Intelligenza artificiale e processo penale: le novità dell'AI ACT*, in *Diritto di Difesa*, 2023, pp. 4-11-
- QUATTROCCOLO S., *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, 2019, p. 1762.
- QUIJANO SANCHEZ L., LIBERATORE F., CAMACHO COLLADOS J., *Applying Automatic Text-Based Detection of Deceptive Language to Police Reports: Extracting Behavioral Patterns from a Multi-Step Classification Model to Understand how we Lie to the Police*, in *Knowledge-Based Systems*, vol. 149, 2018, pp. 155–168.
- ROMANO M. -GRASSO G., *Commentario sistematico del Codice penale. Parte generale*, vol. II, Giuffrè, Milano, 2012, p. 355.
- RONCO M.– ROMANO B., *Codice penale commentato*, UTET Giuridica, Torino, 2012, *sub artt.* 133, 202 e 203 c.p.
- ROVATTI R., *Il processo di apprendimento algoritmico e le applicazioni nel settore legale*, in U. RUFFOLO, *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 31-40.
- RUFFOLO U., *La macchina sapiens come "avvocato generale" ed il primato del giudice umano*, in RUFFOLO U., *XXVI Lezioni di diritto dell'intelligenza artificiale*, Giappichelli, Torino, 2021, pp. 205-223.
- RUOHONEN J., *A Review of Product Safety Regulations in the European Union*, in *International Cybersecurity Law Review*, 2022, pp. 345-347.
- RUSSELL S. -NORVIG P., *Artificial Intelligence: A Modern Approach*, 3^a ed., Prentice Hall, Upper Saddle River (NJ), 2010, pp. 1-98 e 802-829.
- SANTUOSSO A. –SARTOR G., *La giustizia predittiva: una visione realistica*, in *Giurisprudenza italiana*, 2022, p. 1760.
- SCHABAS W.A. *The European Convention on Human Rights: A Commentary*, Oxford University Press Incorporated, Oxford, 2015, pp. 555-587.
- SEARLE J. R., *Minds, Brains, and Programs*, in *Behavioral and Brain Sciences*, vol. III, n. 3, 1980, pp. 417-457.

- SGUBBI F., *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Bologna, Il Mulino, 2019, p. 41.
- SILVEIRA A.-COELHO L. A. -COSTA M. I., et al. *The Charter of Fundamental Rights of the European Union: a commentary*, JusGov, UMinho Law School, Escola de Direito da Universidade do Minho, Braga, Portugal, 2024, pp. 212-216 e 439-440.
- SIMONCINI A., *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, in *Associazione italiana dei costituzionalisti (AIC)*, 2023.
- SIMONCINI A., *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, in *Rivista di diritto processuale*, 2024.
- SPINDLER G. *Algorithms, Credit Scoring, and the New Proposals of the EU for an AI Act and on a Consumer Credit Directive*, in *Law and Financial Markets Review*, 2023, pp. 239-240.
- STANIG E., *Il nuovo diritto penale d'autore*, in *Scuola Positiva e sistema penale: quale eredità?*, Pittaro, Trieste, 2012, pp. 45 e ss.
- STEPONENAITE V. -VALCKE P., *Judicial Analytics on Trial: An Assessment of Legal Analytics in Judicial Systems in the Light of the Right to a Fair Trial*, in *Maastricht Journal of European and Comparative Law*, Vol. 27, n. 6, SAGE Publications, 2020, pp. 776-771.
- STEVENSON M.T. -MAISON S.G., *Pretrial Detention and Bail*, in E. LUNA, *Reforming Criminal Justice: A Report by the Academy for Justice*, Arizona State University, Phoenix, Arizona, 2017, pp. 31-34.
- STRAUSS N., *Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment*, in *Columbia Law Review*, vol. 120, 2020, pp. 1682-1685.
- SURDEN H., *Artificial Intelligence and Law*, Cambridge University Press, Cambridge, 2019, pp. 1306-1337.
- TOLAN S.-MIRON M. -GOMEZ. E., et al., *Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in Catalonia*, in *Proceedings of the 17th International Conference on Artificial Intelligence and Law (ICAIL '19)*, Montreal, Canada, Association for Computing Machinery, 2019, parr. 1-5
- TONINI P.-CONTI C., *Manuale di procedura penale*, 24° Edizione, Giuffrè Francis Lefebvre, Milano, 2023.
- TURING A. M., *Intelligent Machinery and Intelligence*, in *Mind*, pp. 433-460.
- ULENAERS J., *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?* in *Asian Journal of Law and Economics*, Vol.11, n.2, p. 33.
- UNITED STATES SENTENCING COMMISSION, *Federal Sentencing: The Basics*, Washington, 2020, [Federal Sentencing: The Basics | United States Sentencing Commission](#).

VALDIVIA A. -HYDE VAAMONDE C. -GARCÍA MARCOS J., *Judging the algorithm: A case study on the risk assessment tool for gender-based violence implemented in the Basque country*, in *AI & Society*, vol. 40 n. 4, 2025.

VALENTINI C., *Obbligatorietà dell'azione penale, patologie della prassi e mancanza di controlli*, in *Rivista di diritto processuale*, 2020, pp. 1023 e ss.

VAN NOSTRAND M., *Assessing Risk among Pretrial Defendants in Virginia: The Virginia Pretrial Risk Assessment Instrument*, Virginia Department of Criminal Services, Richmond, Virginia, 2007.

VARRASO G., *La legge Cartabia e il rapporto dei procedimenti speciali al recupero dell'efficienza processuale in Sistema Penale*, 2022, pp. 29-32.

VASSALLI G., *Il potere discrezionale del giudice nella commisurazione della pena*, in G. VASSALLI, *Scritti giuridici*, vol. I, Milano, 1997, p. 1334.

VASSALLI G., *Misure di prevenzione e diritto penale*, in *Studi in onore di Biagio Petrocelli*, vol. III, Giuffrè, 1972.

VENEZIANI P., *La punibilità e le conseguenze giuridiche del reato* in C. F. GROSSO -T. PADOVANI -A. PAGLIARO, *Trattato di diritto penale*, Giuffrè, Milano, 2014, p. 347.

VOIGT P. -HULLEN N., *The EU AI Act. Answers to Frequently Asked Questions*, C.H. Beck, Monaco di Baviera 2024, pp. 42-43.

WAGGONER P.-MACMILLEN A., *Pursuing Open-Source Development of Predictive Algorithms: The Case of Criminal Sentencing Algorithms*, in *Journal of Computational Social Science*, vol. 5, no. 1, Springer, 2022, pp. 4-5.

WATT L., *Chinese Scientists Create 'World First' AI Prosecutor that Will Decide People's Fate*, in *The Telegraph*, 2021, [Chinese scientists create 'world first' AI prosecutor that will decide people's fate](#), consultato il 20 settembre 2025.

WOOD J., *The Bail Reform Act of 1984, 4th Edition*, Federal Judicial Center, Washington (DC), 2022, pp. 18-35.

ZAGREBELSKY V., *Stabilire le priorità nell'esercizio obbligatorio dell'azione penale*, in AA.VV., *Il pubblico ministero oggi*, Milano, 1994, p. 10.