AI ACT E PROCESSO PENALE: SFIDE E OPPORTUNITÀ

di Giovanni Canzio

SOMMARIO: 1. L'applicazione dell'IA nella giurisdizione penale. – 2. Il Regolamento dell'Unione europea. Dal modello «*forte*» alla «*Hybrid AI*». – 3. Pratiche di IA vietate e deroghe ai divieti (Capo II, art. 5 Reg.). – 4. Sistemi di IA ad alto rischio: requisiti e deroghe (Capo III, art. 6 Reg.). – 5. L'uso di sistemi di IA nelle indagini e nel giudizio: condizioni e limiti. – 6. Una difficile transizione.

1. L'applicazione dell'IA nella giurisdizione penale.

L'intelligenza artificiale applicata al settore della giurisdizione (*Legal AI*) può definirsi un sistema che, con specifico riguardo all'ambiente della giustizia, acquisisce, ordina e rielabora una enorme quantità di informazioni (*Big Data*), di tipo giudiziario o giurisprudenziale, al fine di identificare, sulla base di calcoli algoritmici di tipo probabilistico, la soluzione ottimale della questione posta, formulandone la previsione (*giustizia predittiva*) o addirittura la stessa decisione.

Il modello «*forte*» di IA postula l'automazione del processo decisionale in luogo degli attori della giurisdizione (*machina sapiens*). Un modello, questo, conseguente all'impiego di schemi matematico-statistici nell'esercizio di quella che viene definita *giustizia predittiva*, indubbiamente inquietante e opaco, e però connotato da un'indubbia forza espansiva, a fronte della crisi di certezza, calcolabilità, uniformità e celerità delle procedure, che promette una risposta pronta e neutra alla domanda di giustizia, perciò deresponsabilizzante per il decisore, con l'ulteriore effetto negativo del conformismo e della sclerotizzazione del formante giurisprudenziale. E che comporterebbe, all'evidenza, un incipiente cambiamento della dimensione umana ed etica del tradizionale paradigma del dire/fare il diritto nel XXI secolo[1].

Ciò a maggior ragione con riguardo all'ambiente della giurisdizione penale, il cui statuto epistemologico s'innerva intorno ai concetti di ipotesi e fatti, indizi e prove, contraddittorio, verità o dubbio, conferma o falsificazione dell'ipotesi, giustificazione razionale della decisione, controllo endo- ed extra-processuale della stessa. Il percorso cognitivo e decisorio, in funzione dell'accertamento della corrispondenza o verosimiglianza dell'ipotesi prospettata rispetto al fatto realmente accaduto nel passato (*lost facts*), non postula che la soluzione decisoria sia «esatta» bensì che sia «giusta», in termini di qualificata probabilità logica e di elevata credibilità razionale.

Le operazioni giudiziali di valutazione delle prove, non immuni da distorsioni ed errori cognitivi a causa della razionalità limitata della mente umana, sono disciplinate da una fitta rete di regole epistemologiche e di legalità (artt. 187, 192, comma 1, 533,

[1] Cfr., volendo, G. CANZIO, *Intelligenza artificiale e processo penale*, in G. CANZIO – L. LUPARIA DONATI (a cura di), *Prova scientifica e diritto penale*, Walter Kluwer – CEDAM, Milano, II ed., 2022, pp. 903 ss.

Editore Associazione "Progetto giustizia penale" | via Festa del Perdono 7, 20122 Milano
c/o Università degli Studi di Milano, Dipartimento di Scienze Giuridiche "Cesare Beccaria"
redazione@sistemapenale.it

comma 1, 546, comma 1 lett. *e*, 606 lett. *e* cod. proc. pen.), che s'ispirano ai valori costituzionali del *giusto processo*, quali: la presunzione di innocenza dell'imputato e l'onere della prova a carico dell'accusa; il principio del contraddittorio come metodo dialettico di verifica delle prove e di ricerca della verità; il giudizio conclusivo di conferma o falsificazione dell'ipotesi, nel contesto di una motivazione coerente e argomentata e alla stregua del criterio dell' «*al di là di ogni ragionevole dubbio*»; il controllo impugnatorio della legalità e logicità del discorso giustificativo.

La consapevolezza del rischio di anomalie e distorsioni dello stesso algoritmo (*Bias Automation*) ha determinato pertanto il preoccupato intervento della comunità dei giuristi, per assicurare che il pur utile arricchimento delle fonti informative del giudice mediante l'utilizzo di tecnologie computazionali si coniughi con il nucleo delle garanzie del giusto processo e risponda comunque a criteri di sorveglianza e responsabilità dell'uomo.

Nel senso che si riconosce titolo ad accedere all'ambiente della giurisdizione penale solo allo standard «*debole*» o «*collaborativo*» della intelligenza artificiale che, nella complementarità uomo-macchina, consente comunque all'uomo di mantenere il controllo della macchina.

Secondo le linee guida tracciate dalla *Carta etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, adottata il 3 dicembre 2018 dalla Commissione europea per l'efficienza dei sistemi di giustizia (*CEPEJ*), la congruità del calcolo algoritmico va verificata in un processo d'integrazione fra le misurazioni quantitative da esso offerte con il percorso cognitivo e decisorio del giudice, nel rispetto dei metavalori dell'ordinamento.

Come, peraltro, già avvertiva la S.C. del Wisconsin nella sentenza pronunciata nel *leading case* «*Loomis*» (secondo cui il software COMPAS, uno strumento utilizzato nel sistema statunitense per misurare il rischio di recidivanza, «… *should be always constitute merely one tool available to a Court, that need to be confirmed by additional sound information*…»), viene in particolare rimarcato il criterio della non esclusività del dato algoritmico per la decisione, che dev'essere viceversa riscontrato - corroborato - da ulteriori e diversi elementi di prova. Assumono altresì rilievo gli ulteriori criteri di validazione in punto di garanzie dei diritti fondamentali della persona, di non discriminazione, trasparenza, imparzialità, equità e comprensibilità dei metodi di elaborazione dei dati informatici, controllabilità dei percorsi di calcolo, qualità e attendibilità scientifica del risultato (*fitness*, *discovery*, *transparency*, *corroboration*, *accountability*, *compliance, enforcement*), in un quadro di autonomia e responsabilità del decisore.

Nello stesso senso si muove la prescrizione dettata dall'art. 8 del d.lgs. 18/05/2018, n. 51, Attuazione della direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio del 27/04/2016, relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti, ai fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali[2].

---

[2] Art. 8: «1. Sono vietate le decisioni basate unicamente su un trattamento automizzato, compresa la

Anche il Rapporto dell'Alto Commissario delle Nazioni Unite per i Diritti Umani, Michelle Bachelet (*Il diritto alla privacy nell'era digitale*, Ginevra 15 settembre 2021), nell'evidenziare l'impatto dei sistemi di IA sul diritto alla *privacy* e su altri diritti umani associati in vari settori chiave, come la sicurezza nazionale e la giustizia penale, formulava per gli Stati e per le imprese una serie di raccomandazioni (par. 59-61) per affrontare le sfide del progresso tecnologico, lanciando un drammatico appello perché «*Artificial Intelligence risks to privacy demand urgent action*».

2. Il Regolamento dell'Unione europea. Dal modello «*forte*» alla «*Hybrid AI*».

In questo scenario di tumultuosa evoluzione dei sistemi di IA, nell'intento di definire i confini applicativi di un modello giuridicamente ed eticamente accettabile, è intervenuta prima la proposta di Regolamento elaborata dalla Commissione Europea il 21 aprile 2021 e infine il Regolamento (UE) del Parlamento europeo e del Consiglio 2024/1689 (*AI Act*). Un testo certamente non breve e puntigliosamente dettagliato, la cui architettura regolatoria postula in premessa l'esigenza di «*governance systems at Union and National level*», al quale ha fatto seguito il 5 settembre 2024 la formulazione a Vilnius della Convenzione quadro del Consiglio d'Europa su «*AI and Human Rights, Democracy and the Rule of Law*».

Fra i ben 180 *Considerando* che illustrano in premessa la portata delle disposizioni regolamentari vanno sinteticamente menzionati almeno i seguenti, per il rilievo diretto o indiretto che essi assumono nell'ambiente della giustizia penale.

(2) Il regolamento dev'essere applicato conformemente ai valori dell'Unione sanciti dalla Carta dei diritti fondamentali, agevolando la protezione delle persone fisiche, delle imprese, della democrazia e dello Stato di diritto, promuovendo l'innovazione e rendendo l'Unione un leader nell'adozione di un'IA «*antropocentrica e affidabile*».

(8)-(26) È necessaria l'adozione di un quadro giuridico dell'Unione, proporzionato ed efficace, che istituisca «*regole armonizzate*» e «*vincolanti*» in materia di IA, avvalendosi di un approccio basato sul rischio che può essere generato dal sistema di IA, definito in modo chiaro, e garantendo nel contempo un elevato livello di protezione degli interessi riconosciuti e tutelati dal diritto dell'Unione.

(27) Vanno ricordati i sette principi etici per una IA «*coerente, antropocentrica, affidabile ed eticamente valida*» elaborati dall'*AI HLEG* nominato dalla Commissione, nel senso che i sistemi di IA sono sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia personale, e funzionano in modo da poter essere adeguatamente controllati e sorvegliati dagli esseri umani.

---

profilazione, che producano effetti negativi per l'interessato, salvo che siano autorizzati dal diritto dell'Unione europea o da specifiche disposizioni di legge. 2. Le disposizioni di legge devono prevedere garanzie adeguate per i diritti e le libertà dell'interessato. In ogni caso è garantito il diritto di ottenere l'intervento umano da parte del titolare del trattamento».

(42) In linea con la presunzione d'innocenza la persona fisica dovrebbe sempre essere giudicata in base al suo comportamento effettivo, non al comportamento previsto dall'IA basato unicamente sulla profilazione, sui tratti della personalità o su caratteristiche personali, senza che vi sia un ragionevole sospetto che essa sia coinvolta in un'attività criminosa sulla base di fatti oggettivi e verificabili e senza una valutazione umana. Dovrebbero essere vietate le valutazioni del rischio intese a determinare la probabilità che una persona commetta un reato unicamente sulla base della sua profilazione o dei tratti della personalità o di sue caratteristiche.

(59) E' opportuno classificare ad alto rischio, nella misura in cui il loro uso è consentito, una serie di sistemi di IA destinati ad essere utilizzati nel contesto di azioni delle autorità di contrasto, in cui l'accuratezza, l'affidabilità e la trasparenza risultano particolarmente importanti per evitare impatti negativi sui diritti fondamentali e garantire la responsabilità, mezzi di ricorso efficaci, un giudice imparziale, la presunzione di innocenza e i diritti della difesa della persona indagata, attesa la difficoltà di ottenere informazioni significative sul funzionamento di tali sistemi e di confutarne i risultati in tribunale.

(61) Alcuni sistemi di IA destinati a essere utilizzati da un'autorità giudiziaria, per assisterla nelle attività di ricerca e interpretazione dei fatti e del diritto, dovrebbero essere classificati ad alto rischio per far fronte a potenziali distorsioni, errori e opacità. Il loro utilizzo può fornire sostegno al potere decisionale finale del giudice, ma non sostituirlo, dovendo esso rimanere un'attività «*a guida umana*»

(170)-(171) Il diritto dell'Unione e quello nazionale prevedono mezzi di ricorso efficaci per le persone sui cui diritti e libertà incide negativamente l'uso dei sistemi di IA. Inoltre, le persone interessate dovrebbero avere diritto di ottenere una spiegazione chiara qualora la decisione si basi principalmente sugli *output* di sistemi di IA ad alto rischio e incida significativamente sui diritti fondamentali.

Il Regolamento definisce, in generale, *sistema di IA* un sistema automatizzato progettato per funzionare con livelli di autonomia variabili, 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*» (art. 3.1).

Fin dal primo comma dell'art. 1 avverte che esso intende promuovere la diffusione di un'IA «*antropocentrica e affidabile*», garantendo nel contempo, mediante la fissazione di «*regole armonizzate*», un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dell'Unione, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente contro gli effetti nocivi dei sistemi di IA, promuovendo peraltro l'innovazione tecnologica.

A tal fine adotta un approccio metodologico *risk based*, identificando nella variabile *rischio* una scala graduale di livelli, obblighi e limiti di applicazione degli strumenti di IA: dal divieto di utilizzo all'utilizzo condizionato dall'esistenza di un alto rischio, fino al rischio moderato e a quello minimo e accettabile. Quel che conta, se il rischio è alto, è che in ogni caso siano assicurate la supervisione o sorveglianza umana (art. 14), la trasparenza del funzionamento, così da consentire l'interpretazione dell'*output* e l'utilizzo adeguato (art. 13), l'accuratezza, la robustezza e la cibersicurezza

(art. 15), la valutazione d'impatto sui diritti fondamentali (art. 27), il diritto alla spiegazione dei singoli processi decisionali (art. 86).

3. Pratiche di IA vietate e deroghe ai divieti (Capo II, art. 5 Reg.).

La pur dettagliata articolazione nell'art. 5 Reg. dei divieti di uso di talune pratiche di IA (applicabili a decorrere dal 2 agosto 2025, tenuto conto del rischio inaccettabile associato al loro uso) reca, tuttavia, una serie di deroghe ed eccezioni, che impegnano l'interprete in complesse operazioni di selezione delle regole applicabili nei singoli casi, ma ancor prima il legislatore interno nel gravoso e però doveroso e urgente impegno di armonizzazione del diritto nazionale rispetto alle disposizioni del diritto dell'Unione.

Sono vietate (art. 5, par. 1, lett. a-h) le pratiche di un sistema di IA: - che utilizza tecniche subliminali o manipolative o ingannevoli aventi l'effetto di distorcere materialmente il comportamento di una persona, anche sfruttandone la vulnerabilità; - che è mirato alla valutazione o classificazione delle persone sulla base del loro comportamento o di caratteristiche personali, inferendone un punteggio sociale pregiudizievole o sfavorevole (*social scoring*); - che effettua valutazioni o previsioni del rischio che una persona commetta un reato unicamente sulla base della profilazione della stessa (*profiling*), a meno che non sia utilizzato a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, basata «su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa»; - che crea o amplia banche dati di riconoscimento facciale mediante *scraping* di immagini facciali da internet o da filmati di telecamere; - che inferisce le emozioni di una persona nel luogo di lavoro, salvo che per motivi medici o di sicurezza; - che utilizza sistemi di categorizzazione biometrica che classificano le persone per trarne deduzioni in merito a razza, convinzioni politiche, sindacali, religiose, filosofiche, sessuali, a meno che il set di dati biometrici non sia acquisito legalmente come nel settore delle «attività di contrasto».

È infine vietato (lett. h) l'uso di «*sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico*», finalizzati alla identificazione della persona senza il suo coinvolgimento attivo e a distanza, mediante il confronto dei suoi dati biometrici con quelli contenuti una banca dati. A meno che e nella misura in cui tale uso non sia strettamente necessario per la ricerca mirata di specifiche vittime di sottrazione, tratta o sfruttamento di esseri umani o di persone scomparse, per la prevenzione di una minaccia specifica sostanziale e imminente per la vita o l'incolumità fisica delle persone o di una minaccia reale e attuale o prevedibile di un attacco terroristico, per la localizzazione o identificazione di una persona sospettata di avere commesso un reato ai fini dell'indagine o dell'esercizio dell'azione penale o dell'esecuzione di una sanzione penale per i gravi crimini elencati nell'Allegato II, punibile con una pena privativa della libertà della durata massima di almeno quattro anni.

Ai fini delle «*attività di contrasto*» (par. 5, comma 2), svolte dalle «*autorità di contrasto*» – qualsiasi autorità pubblica competente in materia di prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia e la prevenzione contro le minacce alla sicurezza pubblica -, l'uso è tuttavia

consentito solo per confermare «*l'identità*» della persona interessata, dovendosi peraltro tenere conto degli elementi attinenti alla gravità, probabilità e entità del danno che sarebbe causato in caso di mancato uso del sistema e delle conseguenze dell'uso del sistema per i diritti e le libertà delle persone interessate. Inoltre, vanno rispettate le tutele e le condizioni «*necessarie e proporzionate*» in relazione all'uso, che è autorizzato solo se l'autorità di contrasto ha completato una «*valutazione d'impatto sui diritti fondamentali*» e ha registrato il sistema nella banca dati UE; solo «in situazioni d'urgenza debitamente giustificate», è possibile usare tali sistemi senza la registrazione, che dovrà essere completata senza indebito ritardo.

L'uso di un siffatto sistema a fini di contrasto (par. 3) «*è subordinato a un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente*», la cui decisione è «vincolante» ed è rilasciata su richiesta motivata e in conformità alle regole dettagliate del diritto nazionale; l'autorità competente rilascia l'autorizzazione solo se ha accertato, sulla base di prove oggettive o indicazioni chiare, che l'uso è «necessario e proporzionato» al conseguimento degli obiettivi consentiti, restando «limitato a quanto strettamente necessario»; tuttavia, in una situazione d'urgenza debitamente giustificata, è possibile usare il sistema senza autorizzazione che va richiesta al più tardi entro 24 ore, ma, se l'autorizzazione non è concessa, l'uso è interrotto e gli *output* sono eliminati.

Spetta allo Stato membro prevedere di autorizzare l'uso di sistemi di identificazione biometrica remota in tempo reale in spazi accessibili al pubblico a fini di attività di contrasto, entro i limiti e alle condizioni di cui ai par. 1 lett. h), 2 e 3, stabilendo «*regole dettagliate*» per la richiesta, il rilascio, l'esercizio e il controllo delle autorizzazioni, potendo peraltro introdurre, in conformità del diritto dell'Unione, disposizioni più restrittive al relativo uso (par. 5).

4. Sistemi di IA ad alto rischio: requisiti e deroghe (Capo III, art. 6 Reg.).

4.1 Secondo l'art. 6, par. 2 Reg. (applicabile a decorrere dal 2 agosto 2026), sono considerati ad alto rischio (il rischio è definito «la combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso») i sistemi di IA elencati in uno dei settori indicati dall'Allegato III. E cioè: 1. Biometria; 2. Infrastrutture critiche; 3. Istruzione e formazione professionale; 4. Occupazione, gestione dei lavoratori e accesso al lavoro autonomo; 5. Accesso e fruizione di servizi privati e pubblici essenziali; 6. Attività di contrasto; 7. Migrazione, asilo e gestione del controllo delle frontiere; 8. Amministrazione della giustizia e processi democratici.

Per lo specifico settore delle «*attività di contrasto*» di cui al citato punto 6, sono considerati i sistemi di IA destinati ad essere utilizzati dalle autorità di contrasto: a) per determinare il rischio per una persona di diventare vittima di reati; b) come poligrafi e strumenti analoghi; c) per valutare l'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati; d) per determinare il rischio di commissione del reato o di recidiva in relazione a una persona, non solo sulla base della profilazione o per valutare i tratti della personalità o il comportamento criminale pregresso; e) per

effettuare la profilazione delle persone nel corso delle indagini, dell'accertamento e del perseguimento di reati.

Per il settore della «*amministrazione della giustizia*» di cui al citato punto 8, sono considerati i sistemi di IA destinati ad essere usati da un'autorità giudiziaria per assistenza nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti o nella risoluzione alternativa delle controversie.

4.2 Ai sensi dell'art. 6, par. 3 Reg. un sistema di IA di cui all'allegato III non è considerato ad alto rischio se non presenta un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone, anche nel senso di non influenzare materialmente il risultato del processo decisionale. Il che si verifica quando il sistema di IA è destinato: a) ad eseguire un compito procedurale limitato; b) a migliorare il risultato di un'attività umana completata; c) a rilevare schemi decisionali o deviazioni da schemi decisionali precedenti e non è finalizzato a sostituire o influenzare la valutazione umana precedentemente completata senza un'adeguata revisione umana; d) a eseguire un compito preparatorio per una valutazione pertinente ai fini dei casi d'uso di cui all'allegato III. Un sistema di IA di cui all'allegato III è sempre considerato ad alto rischio qualora effettui la profilazione di persone.

4.3 I sistemi di gestione ad alto rischio debbono rispettare una serie di condizioni di esercizio e requisiti, non solo di tipo tecnico o amministrativo, per gli aspetti attinenti alla garanzia di conformità, alla gestione dei rischi, alla *governance* dei dati, alla documentazione tecnica, alla conservazione delle registrazioni di tracciabilità del funzionamento, all'accuratezza, robustezza e cibersicurezza, alla responsabilità lungo la catena del valore dell'IA.

Assumono speciale rilievo, nella specifica materia delle attività di contrasto da parte delle competenti autorità, i requisiti della trasparenza e fornitura di informazioni ai *deployer* (art. 13), della sorveglianza umana (art. 14) e della valutazione d'impatto sui diritti fondamentali (art. 27).

In particolare, il funzionamento del sistema di IA ad alto rischio dev'essere sufficientemente trasparente da consentire ai *deployer* di interpretare l'*output* e utilizzarlo adeguatamente mediante informazioni concise, complete, corrette e chiare, che siano accessibili e comprensibili.

Per il profilo della sorveglianza umana, il sistema di IA ad alto rischio è progettato e sviluppato anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da una persona durante il periodo in cui sono in uso. La persona alla quale è affidata la sorveglianza deve avere la possibilità di comprenderne capacità e limiti di funzionamento per affrontare eventuali disfunzioni o prestazioni inattese ed essere consapevole della tendenza a fare eccessivo affidamento sull'*output* prodotto, in particolare per i sistemi utilizzati per fornire informazioni o raccomandazioni per le decisioni che sono prese da persone, nonché di interpretare correttamente l'*output* e decidere di non usare il sistema, ignorare o ribaltare l'*output*, anche interrompendone il funzionamento mediante una procedura di arresto.

Va infine sottolineato come, da un lato, qualsiasi persona che abbia motivo di ritenere che vi sia stata una violazione delle disposizioni del regolamento può presentare – almeno – un reclamo alla competente autorità di vigilanza del mercato (art. 85) e, dall'altro, qualsiasi persona oggetto di una decisione adottata dal *deployer* sulla base dell'*output* di un sistema di IA ad alto rischio elencato nell'allegato III, ad eccezione di quelli al punto 2, e che incida significativamente su tale persona per avere un impatto negativo sulla sua salute, sicurezza o sui diritti fondamentali, ha il diritto di ottenere dal *deployer* spiegazioni chiare e significative sul ruolo del sistema nella procedura decisionale e sui principali elementi della decisione adottata (art. 86).

5. L'uso di sistemi di IA nelle indagini e nel giudizio: condizioni e limiti.

Dalla lettura sistematica e coordinata delle menzionate disposizioni dell'*AI Act* emerge il profondo intreccio fra gli aspetti tecnico-giuridici e quelli di tipo etico della regolamentazione sovranazionale. La previsione di clausole aperte e di deroghe pure per le pratiche vietate e le prescrizioni generali dettate in materia di «*attività di contrasto*» comportano il necessario rinvio al puntuale disciplinamento del fenomeno ad opera del diritto armonizzato dell'Unione o di quello nazionale da adeguare e armonizzare.

Si deve registrare, a ben vedere, una sfida inedita e di straordinaria portata storica sia per il legislatore interno che per i giuristi e in particolare per l'interprete[3].

Con particolare riguardo al processo penale[4], i valori costituzionali del giusto processo (presunzione di innocenza, parità delle armi, contraddittorio, motivazione, criterio BARD, controllo impugnatorio) e di tutela dei diritti fondamentali della persona pretendono la fissazione di principi e regole chiare circa la dislocazione e l'equilibrio dei poteri fra polizia giudiziaria, pubblico ministero e giudice nella fase delle indagini preliminari, nonché circa la validità e utilizzabilità probatoria o meno degli atti investigativi supportati da sistemi di IA, l'ammissibilità e la valutazione della relativa prova nel giudizio, l'effettività della garanzia di partecipazione dialettica di tutte le parti prima alla perimetrazione del ragionamento decisorio e poi al sindacato impugnatorio.

Dunque, come organizzare correttamente l'accesso di questo peculiare tipo di prova tecnologica nell'ambiente del processo penale, al fine di implementare la qualità delle *performance* cognitive e decisionali del giudicante, assicurandone nel contempo la legalità e l'utilizzabilità? Come garantire il legittimo spazio per l'esercizio del diritto di difesa, attraverso il confronto dialettico, la confutazione, la prova contraria, il dubbio, e

---

[3] Fra i tanti contributi dottrinali sull'argomento, cons. F. PETRELLI, *Intelligenza artificiale e processo penale*, in *Diritto di difesa*, 6 giugno 2024; A. BALSAMO, *L'impatto dell'IA nel settore della giustizia*, in questa *Rivista*, 22 maggio 2024; G. UBERTIS, *Perizia, prova scientifica e IA nel processo penale*, *ivi*, 3 giugno 2024; A. GIOVENE, *AI Act. Tra perfettibilità e compromessi*, in *federalismi.it*, 17 aprile 2024.

[4] Per le prime, sperimentali elaborazioni metodologiche in materia di sistemi intelligenti e diritto alle prove, cons. L. LUPARIA DONATI, *Introduzione. Prova giudiziaria e ragionamento artificiale: alcune possibili chiavi di lettura*, in J. SALLANTIN – J.J. SZEZECINIARZ (a cura di), *Il concetto di prova alla luce dell'intelligenza artificiale*, Milano, Giuffré, 2005, p. VII ss.

perciò il contraddittorio effettivo fra le parti «*per*» e «*sulla*» prova, in funzione della validazione scientifica del risultato e contro la deriva algoritmica della giurisdizione?

Si apre, all'evidenza, uno scenario inedito nel quale il legislatore nazionale, per primo, è chiamato a dislocare, con «*regole dettagliate*», i poteri e le funzioni di accertamento, selezione e decisione della competente autorità giudiziaria, nellin materia di 'esercizio di una discrezionalità tecnica, che sia guidata dal sapiente e prudente ricorso a categorie valoriali e clausole aperte.

Ad esempio, rispetto alla definizione della stretta necessità e proporzione dell'uso del sistema di IA in linea di principio vietato, alla valutazione d'impatto sui diritti fondamentali e sulle libertà delle persone, al governo della procedura autorizzatoria, alla debita giustificazione della eventuale urgenza, alla probabilità e all'entità del danno dal mancato uso ecc., sembrano invero insufficienti i criteri enunciati dalla Corte Suprema statunitense nella nota sentenza *Daubert*[5], in base ai quali il giudice deve vagliare l'effettiva affidabilità di una teoria o un metodo e di una *expert witness's scientific testimony*, ai fini della loro ammissibilità come prova scientifica nel processo: la controllabilità mediante esperimenti; la falsificabilità mediante test di smentita con esito negativo; la *peer review* della comunità scientifica di riferimento; la conoscenza della percentuale di errore dei risultati; infine, il criterio subordinato e ausiliario della generale accettazione da parte della comunità degli esperti. Criteri, tutti, sostanzialmente condivisi e anzi arricchiti dalla Corte di cassazione italiana[6].

5.1. Con riguardo alla fase delle indagini preliminari, che com'è ormai diffusamente percepito dagli osservatori, è assurta a indiscusso baricentro anche mediatico del processo, sembra opportuno prendere le mosse dalla saggia scelta del legislatore della recente riforma *Cartabia* di riportare in (un seppure ancora parziale) equilibrio i rapporti fra il pubblico ministero e il giudice, mediante il riconoscimento di una serie di rimedi di tipo ordinatorio affidati al giudice per le indagini o della udienza preliminare - le cd. «*finestre di giurisdizione*» - in taluni momenti topici delle indagini.

Non sembra dubbio che, soprattutto laddove l'azione investigativa di contrasto si concreti nell'utilizzo di strumenti di IA vietati o ad alto rischio, potenzialmente pervasivi dei diritti fondamentali e delle libertà individuali delle persone, la previsione di clausole aperte e deroghe rinvii ad operazioni logiche di verifica e valutazione discrezionale di fatti e circostanze e di bilanciamento di valori più propriamente pertinenti alla figura del giudice, anziché a quella del pubblico ministero, che è parte benché pubblica.

5.2 Con riguardo alla fase del giudizio, risulta davvero efficace rinviare la verifica di coerenza e attendibilità di questa speciale prova scientifica e tecnologica al contraddittorio fra le parti «*sulla*» prova, cioè quando essa sia stata già ammessa e

---

[5] Daubert v. Merrel Dow Pharmaceuticals, Inc., 509 US 579 (1993).

[6] Cass., Sez. IV, 17/09/2010, n. 43786, Cozzini, sull'attendibilità di una teoria relativa all'eziologia del cancro dovuta ad esposizione all'amianto; Sez. Un., 11/09/2002, n. 30328, Franzese, sulle leggi scientifiche di copertura del nesso causale tra la condotta omissiva del medico e l'evento lesivo in danno del paziente.

acquisita? Non sarebbe più utile costruire, nell'organizzazione degli snodi procedimentali, un filtro di accesso, preventivo e a maglie strette, e un agile contraddittorio «*per*» la prova, al fine di escludere addirittura che entrino nel patrimonio probatorio informazioni non sorrette da legittima validazione di attendibilità e utilizzabilità? L'irruzione nel giudizio di una siffatta, potente e incisiva, prova tecnologica giustificherebbe quindi la previsione di uno spazio dialettico preliminare già nel momento e in funzione dell'ammissione, prima ancora che della valutazione del risultato probatorio.

In questa direzione potrebbe rinvenirsi nel sistema processuale vigente la disposizione dell'art. 189 cod. proc. pen., per cui l'apprezzamento di rilevanza, non superfluità e concreta idoneità della prova «*ad assicurare l'accertamento dei fatti*» – senza che ne resti pregiudicata «*la libertà morale delle persone*» per il divieto di perizia criminologica *ex* art. 220, comma - è rimesso al vaglio preliminare del giudice. Questi, anche al fine di garantire l'anticipata conoscenza circa le metodologie che saranno applicate nell'accertamento, dopo avere sentito le parti sulle modalità di assunzione della prova, provvede all'ammissione con ordinanza, fissando le regole per la corretta applicazione dei metodi e delle procedure tecniche di acquisizione della stessa. Va segnalato, in proposito, un passo della Relazione al Progetto preliminare del nuovo codice di procedura penale del 1989 (p. 60), riguardante la portata dell'art. 189: «*È sembrato che una norma così articolata possa evitare eccessive restrizioni ai fini dell'accertamento della verità, tenuto conto del continuo sviluppo tecnologico che estende le frontiere dell'investigazione, senza mettere in pericolo le garanzie difensive*». Norma, questa, che nell'*intentio legis* mirava ad assicurare, con l'apporto della scienza e della tecnologia nella ricerca della verità, l'opportuna flessibilità del sistema processuale. Come si vede, un filtro, questo dell'art. 189, a maglie ben più strette rispetto a quello previsto dall'art. 190, comma 1, che, ai fini dell'ammissione della prova in genere, si limita a selezionare negativamente solo «*le prove vietate dalla legge e quelle che manifestamente sono superflue o irrilevanti*», e inoltre assistito da un significativo rafforzamento del contraddittorio anticipato, «*per*» la prova, ancor prima che «*sulla*» prova.

6. Una difficile transizione.

In una equilibrata interazione tra giurisdizione e sistemi di intelligenza artificiale, secondo il metodo della *Hybrid AI*, giuristi e algoritmica sono chiamati a costruire un modello di conoscenza condiviso e raffinato, nel solco dei valori etici - di stampo europeo - di spiegabilità, trasparenza, controllo umano significativo e autonomia del decisore[7].

Il disciplinamento, etico e giuridico, del fenomeno ad opera del Regolamento europeo consente di reggere l'urto impetuoso e intrigante, nel contesto del processo penale in particolare, dei pervasivi strumenti dell'intelligenza artificiale, senza che ne risulti sconvolto – almeno finora - il tradizionale paradigma razionalistico della struttura

[7] N. LIPARI, *Il diritto del nuovo millennio tra giurisdizionalizzazione ed algoritmo*, in *Accademia*, 2024, p. 9 ss.

e della funzione della giurisdizione penale, insieme con l'equità, l'efficacia e le garanzie del modello del giudizio *reasonig under uncertainty* e *by probabilities*.

Il diritto penale del fatto è incentrato sulla persona umana e il metodo dialettico dell'accertamento resta affidato, in ultima istanza, all'opera valutativa e decisoria dell'uomo, fallibile ma sindacabile secondo legge e ragione.

Gli esiti della sfida lanciata dal progresso della scienza, tuttavia, non sono affatto scontati. Al bivio tra tecnologia e tecnocrazia, essa si sposta sul terreno della concreta efficacia e qualità della giurisdizione, che, sulla base di regole puntuali e chiare dettate dal legislatore, dovrebbe essere esercitata secondo una dinamica interazione fra i saperi e le operazioni logiche tradizionalmente affidate al giudice e alle parti e le evidenze della prova o del calcolo di un algoritmo, in un orizzonte culturale di intesa su obiettivi e valori responsabilmente condivisi.

I principi di spiegabilità, trasparenza e sorveglianza umana, racchiusi in norme a rilevanza sovranazionale, hanno bisogno di essere concretamente attuati. Non sono certo poche le questioni controverse circa la portata e l'effettiva tenuta dei nuovi principi e regole, per la cui soluzione occorrerà prestare la doverosa attenzione allo sviluppo delle prassi applicative e interpretative.

Non sembra tuttavia all'altezza delle aspettative di armonizzazione del sistema processuale penale interno l'impianto del disegno di legge governativo del 23 aprile 2024, recante disposizioni e delega al Governo in materia di intelligenza artificiale, sebbene in premessa si avverta che «*le disposizioni della presente legge si applicano conformemente al diritto dell'Unione europea*»[8].

Per un verso, l'art. 14, nei suoi due concisi commi, si limita a prevedere che l'IA sia utilizzata nell'attività giudiziaria «*esclusivamente*» per l'organizzazione e la semplificazione del lavoro giudiziario e per la ricerca giurisprudenziale e dottrinale, riservando al Ministero della giustizia la disciplina del relativo impiego da parte della sola magistratura ordinaria e riconoscendo - ovviamente - al magistrato la decisione sulla interpretazione della legge, sulla valutazione dei fatti e delle prove e sulla adozione di ogni provvedimento. Nulla è detto o è prescritto, viceversa, circa i poteri, i divieti, le condizioni, i limiti e le procedure dirette ad acquisire, utilizzare e valutare informazioni probatorie generate in via deduttiva dai sistemi di IA.

Per altro verso, l'art. 22, pur rinviando a uno o più decreti legislativi «*per l'adeguamento della normativa nazionale al Regolamento europeo sull'intelligenza artificiale*», individua una serie di principi e criteri direttivi affatto generici ed eccentrici rispetto all'obiettivo di armonizzazione del sistema in materia di indagini e giudizio penale.

A questo punto, nella consapevolezza che la frammentazione e la miope lentezza del processo decisionale, insieme con l'incoerenza, la scarsa chiarezza o anche l'eccessiva severità della regolamentazione dello sviluppo e degli utilizzi dei sistemi di IA, possono costituire un serio rischio che l'Italia perda terreno, anche culturalmente, nella evoluzione e nell'addestramento di generazioni di modelli di IA, appare doveroso

[8] UCPI, *Prime brevi riflessioni sull'impianto del DDL governativo in materia di IA e giustizia penale*, 6 maggio 2024.

rimarcare che «*il Regolamento europeo è solo un prologo*»[9] di una nuova e lunga fase che riguarderà il mondo scientifico ed economico e che metterà a dura prova anche quello giuridico su diritti, garanzie, rischi e sanzioni. E, per una *governance* efficace e credibile del fenomeno, «*se non ci si mette subito al lavoro due anni non basteranno*».

[9] Così si esprime A. CORRADO, in *Corriere della sera*, 12/9/2024, p. 28.