

CONVEGNI

FRANCESCO TRAPELLA

Introduzione ad una ricerca su indagini preliminari, intelligenza artificiale e *bias* del ragionamento*

In un contesto caratterizzato da una sempre crescente riflessione sul potenziale impatto dell'intelligenza artificiale nell'accertamento del reato, risulta opportuno soffermarsi sulle possibili distorsioni cognitive che possono compromettere l'affidabilità dell'esplorazione. Considerato che l'*output* generato da un algoritmo dipende dai dati forniti da colui che ne cura l'addestramento, al fine di promuovere un'interazione virtuosa tra la macchina e l'operatore, appare essenziale acquisire adeguata consapevolezza circa i difetti cognitivi che influenzano il giudizio umano e che, se non adeguatamente sorvegliati, potrebbero riflettersi negativamente anche sull'attività dell'intelligenza artificiale impiegata a supporto delle decisioni.

Introducing a research perspective on preliminary investigations, Artificial Intelligence and cognitive biases.

In a context increasingly marked by reflection on the potential impact of artificial intelligence in criminal proceedings, it is appropriate to consider the possible cognitive distortions that may undermine the reliability of such exploratory processes. Given that the output generated by an algorithm depends on the data provided by those responsible for its training, and with the aim of fostering a virtuous interaction between the machine and the human operator, it is essential to develop a clear awareness of the cognitive biases that influence human judgment and that, if not properly monitored, may adversely affect the performance of artificial intelligence systems employed in decision-making support.

SOMMARIO: 1. Intelligenza artificiale e distorsioni del ragionamento umano: le ragioni di una ricerca. - 2. Segreto investigativo, monopolio delle indagini, WYSIATI. - 3. Il canone della proporzionalità investigativa nell'era dell'intelligenza artificiale. - 4. Prime, necessariamente provvisorie conclusioni.

1. Intelligenza artificiale e distorsioni del ragionamento umano: le ragioni di una ricerca. È sempre più vivace, nella letteratura processualpenalistica, il dibattito sull'impiego possibile dell'intelligenza artificiale nell'accertamento del reato e nella valutazione delle corrispondenti responsabilità: alla curiosità per l'imminente cambio di passo¹ si aggiungono le riflessioni attorno all'ipotesi di

^{1*} Questo lavoro risulta dalle relazioni - riviste e annotate - tenute dallo scrivente nei seminari *La funzione sociale dell'intelligenza artificiale: sorveglianza, biometria passiva e tutela dei diritti fondamentali della persona* (24 giugno 2025 - organizzazione scientifica: Prof.ssa Annarita Ricci) e *Indagini preliminari, intelligenza artificiale e "parità delle armi"* (*Notte europea dei ricercatori* - 26 settembre 2025). Esso risente, inoltre, della prima analisi dei dati raccolti nel laboratorio *Bias cognitivi, ricerca della prova e accertamento del reato*. I seminari e il laboratorio sono stati organizzati nell'ambito delle attività del Dipartimento di Scienze giuridiche e sociali dell'Università "G. d'Annunzio" di Chieti-Pescara.

Come evidenzia LORUSSO, *La sfida dell'intelligenza artificiale al processo penale nell'era digitale*, in *Sist. pen.*, 27 marzo 2024, 1, si palesa «una nuova era dai tratti incerti, che come sempre accade in

una “macchina pensante” che riesca a sostituirsi al giudice o agli investigatori² e le perplessità derivanti dai nuovi equilibri che la rivoluzione digitale (*rectius*: artificiale) imporrà, con correlata necessità di ridisegnare le garanzie individuali³.

Un rilievo si impone preliminarmente, dal momento che il sistema di apprendimento automatico eseguito dall’algoritmo (*machine learning*) viene addestrato sulla base di dati selezionati, elaborati e ricevuti dalla mano umana⁴.

Se l’*input* è estraneo alla macchina, ma da esso dipendono i risultati che produce, ne deriva che, almeno allo stato attuale di conoscenze e di progresso scientifico-tecnologico, il responso del sistema risente inevitabilmente dell’impostazione di chi lo addestra e, con essa, delle eventuali distorsioni del percorso cognitivo seguito per l’inserimento delle informazioni utili all’operazione algoritmica⁵.

Qualora si ipotizzasse di fare ricorso all’intelligenza artificiale nello svolgimento dell’attività decisoria o, prim’ancora, di quella investigativa, non si potrebbe sottacere la premessa, né, quindi, sarebbe possibile ignorare il rischio che il

questi casi genera paure e atteggiamenti protettivi, alimenta tabù e pregiudizi, quando non ispira crociate. O, al contrario, viene osannata con entusiasmi smisurati, accompagnata da roboanti peana, o persino acclamata come un Eden ritrovato».

² Dovendosi senz’altro accogliere la considerazione di BELVINI, *Intelligenza artificiale e circuito investigativo*, Bari, 2025, 9 che rileva «l’impossibilità, allo stato attuale, dell’I.A. di replicare i processi cognitivi e di affrontare in maniera indipendente situazioni articolate (secondo l’archetipo della c.d. *strong A.I.*)» e sottolinea «l’intento perseguito ... di realizzare sofisticati algoritmi appositamente programmati per rispondere a quesiti specifici in base agli *input* forniti (c.d. *weak A.I.*)».

³ O. MAZZA, *Distopia del processo artificiale*, in *Arch. pen. web*, 2025, 1, 3 inquadra il fenomeno nella temperie attuale. «Lo scenario degli anni Venti» – afferma – «ha mostrato finora una netta regressione della cultura processuale. Non solo il definitivo tramonto dell’illusione di un processo accusatorio garantista che non ha mai realmente attecchito nei costumi italiani, ma anche l’affiorare di una nuova tendenza alla gestione economico-aziendalistica degli affari penali». Preconizza, perciò, un’«efficienza artificiale» esiziale per le garanzie, di cui sarà necessario attestare il definitivo tramonto ove si trasferisca «al computer il pensiero», con il risultato che «nessun essere umano potrà mai competere con le capacità cognitive e intellettive di una macchina in grado non solo di gestire ogni informazione, ma anche di decidere con discernimento giudiziale».

⁴ Rilevano, in modo condivisibile, VALENTINI-DI MATTEO, *L’IA e la “scoperta” dei bias cognitivi nel processo penale*, in corso di pubblicazione su un volume collettaneo diretto da Triggiani-Incampo: «la vera preoccupazione che deve nutrire le nostre riflessioni sull’IA è in effetti proprio questa: l’IA si nutre di dati storici, in realtà proprio come un cervello umano, con l’evidente conseguenza che i *bias* cognitivi contenuti nei dati di addestramento andranno ad inquinare la decisione dell’IA già in partenza».

⁵ Analogamente, con riferimento a fatti notori e a massime di esperienza, e al loro impiego nel percorso argomentativo giudiziale, FALATO, *Il giudice come utente euristico nel ragionamento induttivo basato sul senso comune. Soluzioni di debiasing per assicurare l’igiene decisionale*, in *Arch. pen. web*, 2025, 2, 3-4.

software impiegato a fini giudiziali o di indagine risentisse delle fallacie, delle scorciatoie mentali o delle deduzioni affrettate commesse da chi avesse allenato il modello⁶.

Un paio di esempi chiarirà questo, che appare il solo preambolo utile per impostare correttamente qualsiasi previsione sull'uso, nei prossimi tempi, dell'intelligenza artificiale nel giudizio penale.

Il primo richiama il caso *Loomis*, definito dalla Corte suprema del Wisconsin nel 2016⁷. L'imputato era stato condannato a sei anni di reclusione per alcuni reati contro il patrimonio e contro l'ordine pubblico; nel calcolo della pena, il tribunale aveva tenuto conto delle indicazioni del *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS), un programma per il calcolo di un tasso di pericolosità individuale che aveva identificato l'uomo come soggetto ad alta probabilità di recidiva.

Il *risk assessment* si basava sulla combinazione di fattori variegati: alcuni, riferiti all'appartenenza dell'interessato a gruppi sociali ed etnici – valutata, di per sé, come elemento potenzialmente foriero di rischio – e altri, derivanti dalle informazioni provenienti dallo stesso imputato, tra cui quelle inerenti ai suoi precedenti contatti con la giustizia⁸. Il punto era il modo in cui questi elementi venivano apprezzati e, di conseguenza, valutati; complice, anche, il segreto commerciale che assisteva al funzionamento del programma informatico, alle parti non era consentito conoscere il valore attribuito a ciascun parametro, il metodo di calcolo utilizzato o la ragione per cui si sceglieva di considerare

⁶ Il giudice – sostiene SIMON, *Theories of bounded rationality*, in *Decision and organization*, a cura di McGuire-Radner, Amsterdam, 1972, 161-176 – è un soggetto «a razionalità limitata», al pari di qualsiasi altro essere umano. Non si può, quindi, ignorare il pericolo di trasferire questa sua caratteristica alla macchina, influenzando, in tal modo, la risposta dell'algoritmo.

⁷ Case no. 2015AP157-CR, 5 aprile – 13 luglio 2016, *State of Wisconsin vs. Eric L. Loomis*.

⁸ Tecnicamente, «COMPAS broadens the theoretical coverage to include key constructs from low self-control theory, strain theory or social exclusion, social control theory (bonding), routine activities-opportunity theory, sub cultural or social learning theories, and a strengths or good lives perspective»; «thus, COMPAS includes both theoretically relevant factors and the critical eight criminogenic predictive factors that emerged from recent meta-analytic studies»; da ultimo, «COMPAS includes a number of strength and protective factors that have shown empirical support for potential risk reduction and protecting offenders from the full impact of criminogenic needs. These include job and educational skills, history of successful employment, adequate finances, safe housing, family bonds, social and emotional support, noncriminal parents and friends, and so on». I passaggi sono di BRENNAN-DIETERICH-EHRET, *Evaluating the predictive validity of the COMPAS risk and need assessment system*, in *Criminal Justice and Behaviour*, 2009, 23-24.

alcune circostanze – come il genere – e di impiegarle, così, nel vaglio sulla criminogenicità dell'accusato⁹.

Il caso *Loomis* è interessante proprio perché svela un profilo inedito dell'applicazione della scienza al processo: se, da sempre, la questione controversa è l'intelligibilità delle teorie che giustificano i dati prodotti, ora il problema è la conoscibilità dei meccanismi che governano il processo computazionale da cui esita la risposta agli interrogativi della ricerca¹⁰.

Il nodo si scioglie a seconda della preferenza che si intenda accordare all'algoritmo o all'intuizione umana.

⁹ Per una puntuale disamina delle doglianze difensive, QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale "predittiva"*, in *Cass. pen.*, 2019, 1751-1752. Come sottolinea l'Autrice, al centro delle censure mosse dalla parte privata vi era la considerazione che «tra i parametri presi in considerazione dal *software* vi sarebbe il genere, con un grave profilo di incostituzionalità denunciato dalla difesa: tale dato sarebbe impiegato dallo strumento di valutazione come fattore criminogenico, sebbene, sempre per le ragioni sopra riportate, non sussistano riscontri chiari sui parametri di sua valutazione» (1752). Di fatto, parrebbe evincersi che un uomo delinqua più di una donna (o viceversa): una conclusione che, con evidenza, può essere smentita da molti studi empirici e che, perciò, meriterebbe di essere approfondita per poterne testare la validità, semmai in combinazione con altri fattori che, però, il silenzio sul concreto funzionamento del COMPAS impediva di apprezzare.

¹⁰ Ciò che BELVINI, *Intelligenza artificiale e circuito investigativo*, cit., 35 definisce come la principale insidia della macchina, ossia l'opacità: «constatata l'estrema complessità e, talvolta, l'impossibilità di illustrare con esattezza le modalità attraverso le quali la macchina restituisce l'*output*, essa è talvolta paragonata a una sorta di "*black box*", atteso che in diverse ipotesi, pur essendo noto il *set* di dati somministrato in partenza al sistema, non sempre si può comprendere fino in fondo come esso produce uno specifico esito». In argomento, si vedano anche PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge, 2016, 131; CONTISSA-LASAGNI-SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Dir. int.*, 2019, 619; PALMIOTTO, *The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings*, in *Algorithmic Governance and Governance of Algorithms*, a cura di Ebers-Canero Gamito, Cham, 2020, 49; WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box*, in *Regulating Artificial Intelligence*, a cura di Wischmeyer-Rademacher, Cham, 2020, 80-81; ARDUINI, *La "scatola nera" della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in *BioLaw Journal*, 2021, 453.

Rispetto al caso *Loomis* QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, cit., 1753 rileva che la «Corte suprema statale, nel confermare la sentenza, non sembra cogliere chiaramente questo punto, limitandosi ad escludere la violazione del *due process*, data la possibilità per l'imputato di confrontare i dati individuali di partenza (*input*) e le valutazioni di rischio finali (*output*) sulla base del manuale d'uso dello strumento, potendo confutarne l'attendibilità». Epperò, almeno le teorie applicate per addivenire ai risultati erano note al momento della sentenza, in quanto diversi studi sull'applicativo circolavano nella comunità da qualche anno; ciò che non era noto – e, sul punto, la difesa avrebbe probabilmente potuto fare maggiore chiarezza – era la «codificazione digitale con cui l[e] teori[e] [sono] stat[e] tradott[e] nel *software*» e, di lì, elaborate fino all'*output* (1752).

Da una parte, infatti, non manca chi sostiene la maggiore affidabilità del primo, per la sua minore incoerenza e per la sua estraneità alla tendenza di complicare le caratteristiche del caso con dati ed esperienze, ancora una volta, ricavati da sensazioni estemporanee o da non sempre ponderate associazioni di pensieri¹¹. D'altro canto, però, sarebbe insensato liquidare il giudizio intuitivo; piuttosto, conviene implementare i due metodi di valutazione, ad esempio elaborando una griglia di domande e una serie di criteri per l'assegnazione di un valore numerico alle risposte, oppure facendo seguire alla procedura algoritmica un momento per la formulazione di una stima svincolata dall'analisi statistica e dal riferimento a strumenti digitali¹².

Il cenno agli studi di psicologia cognitiva è utile ad inquadrare il passaggio più critico del caso *Loomis*. È persino consigliabile guidare la disamina di una situazione concreta propedeutica alla formulazione di una prognosi relativa ad un certo fatto sociale attraverso l'analisi di predittori definiti aprioristicamente¹³; tuttavia, l'attribuzione di un peso specifico ad ognuna delle voci è una

¹¹ Esattamente in questi termini, MEEHL, *Causes and Effects of My Disturbing Little Book*, in *Journal of Personality Assessment*, 1986, 370 ss., che richiama i risultati del suo precedente studio, *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, del 1954, e il dibattito che ne è scaturito nella comunità scientifica tra i sostenitori del modello proposto dall'Autore, che eleva la statistica a criterio privilegiato per la previsione di fatti sociali, e l'opinione opposta, che, invece, ricorre all'elaborazione umana e, quindi, all'intuizione. Sull'incoerenza che, secondo Meehl, contraddistingue il ragionamento analogico, si citano, a titolo di esempio, gli studi di HOFFMAN-SLOVCO-RORER, *An Analysis-of-Variance Model for the Assessment of Configurational Cue Utilization in Clinical Judgment*, in *Psychological Bulletin*, 1968, 338 ss. – che registrarono contraddizioni nel 20% delle diagnosi offerte da radiologi ai quali veniva presentata la stessa immagine, in momenti diversi – o quelli di BROWN, *Independent Auditor Judgment in the Evaluation of Internal Audit Functions*, in *Journal of Accounting Research*, 1983, 444 ss. che, su 101 revisori dei conti, rilevò un analogo tasso di incoerenza nella valutazione di attendibilità dei conti interni di alcune imprese.

¹² KAHNEMAN, *Pensieri lenti e veloci*, Milano, rist. 2025, 309-310: durante la valutazione di alcune reclute dell'esercito israeliano, pensò di concentrarsi «su domande standardizzate e fattuali, spera[ndo] di combattere l'effetto alone, in base al quale le prime impressioni favorevoli influenzano i giudizi successivi. Come ulteriore precauzione contro gli alone, ordin[ò] agli intervistatori di vagliare le sei caratteristiche sempre nella stessa sequenza, valutando ciascuna su una scala da uno a cinque prima di passare a quella successiva» (309). Rivolse, poi, ai propri collaboratori questo invito: «quando avrete finito, fate come desiderate voi: chiudete gli occhi, cercate di immaginare la recluta come un soldato e assegnatele un punteggio su una scala da uno a cinque». Il risultato fu sorprendente: «anche il giudizio intuitivo che gli intervistatori formulavano alla fine, chiudendo gli occhi, funzionava molto bene, anzi, altrettanto bene della somma dei sei punteggi specifici [...] L'intuizione aggiunge valore anche a colloqui di selezione ritenuti giustamente risibili, ma solo dopo una serie disciplinata di informazioni obiettive e una classificazione disciplinata di singoli tratti della personalità».

¹³ Così, DAWES, *The Robust Beauty of Improper Linear Models in Decision Making*, in *American Psychologist*, 1979, 571 ss. e ID., *The Superiority of Simple Alternatives to Regression for Social*

questione risolvibile solo sulla base di una scelta umana, di colui, cioè, che imposta i parametri di valutazione¹⁴.

Anche, poi, quando si volesse elaborare uno schema di *equal-weighting*, occorrerebbe fissare una regola di valenza identitaria dei fattori in gioco¹⁵; tale regola, però, si giustificerebbe, ancora una volta, sulla scorta di un'opzione estranea alla macchina e di cui la formula matematica in uso sarebbe una semplice e diretta conseguenza. Meglio detto, sarebbe il ricognitore a decidere che gli elementi si equivalgono, per operare, di seguito, secondo questa premessa.

Tornando al caso *Loomis*, nella comunità scientifica non si registrava un pieno consenso quanto all'attendibilità del metodo utilizzato, né era dato comprendere la rilevanza criminogena di alcuni dei predittori impiegati dal modello¹⁶; ancora, nonostante l'affermazione della Corte statale, che escludeva la violazione del diritto di difesa, avendo potuto l'imputato confrontare gli *input* e gli *output* seguendo le indicazioni del manuale d'uso dello strumento, residuavano incertezze circa la capacità del *software* di non limitarsi ad una valu-

Science Predictions, in *Journal of Educational and Behavioral Statistics*, 2004, 317, nel quale viene affermata la maggiore affidabilità dei sistemi che attribuiscono valori uguali a tutti i predittori, perché in grado di evitare incidenti da campionamento.

¹⁴ Proprio riprendendo le conclusioni della Corte statale in *Loomis*, l'esito di questa operazione, irrimediabilmente intriso di umanità, dovrà essere apprezzato dal giudice, che, certo, non potrà dirsi obbligato ad utilizzarlo come fattore determinante, quasi copiando e incollando in sentenza i risultati raggiunti. *Amplius*, CARRER, *Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in *Giur. pen.* (web), 24 aprile 2019; GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont.* (web), 29 maggio 2019, 6; G. MAZZA-TOGNAZZI, *La valutazione della pericolosità ante e post delictum*, in *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, a cura di Baccari-Felicioni, Milano, 2023, 194-195.

¹⁵ Riprendendo l'esempio di Dawes, il fatto che la stabilità coniugale sia prevedibile secondo la formula *frequenza del sesso meno frequenza dei litigi* – maggiore è la differenza, più forte è il legame nella coppia – impone, a monte, una duplice scelta, umana, di equiparare il peso dei due predittori e di escludere l'incidenza di ulteriori fattori. Lo stesso KAHNEMAN, *Pensieri*, cit., 303-304 recupera la vicenda di Virginia Apgar, che elaborò un sistema a cinque variabili per valutare la vitalità del neonato; attribuendo un punteggio da 0 a 2 a ciascun parametro, si conclude che l'infante che totalizza almeno 8 punti è vitale, mentre una situazione di gravità è quella del bambino che consegue un totale pari o minore di 4. Sia la specificazione delle variabili, sia l'assegnazione dei risultati parziali esita da un'operazione umana.

¹⁶ Critici verso il COMPAS, incapace di predire la recidivanza degli imputati, oltre ad essere in grado di generare decisioni discriminatorie, FASS-HEILBRUN-DEMATTEO-FRETZ, *The LSI-R and the COMPAS: Validation Data on Two Risk-Need Tools*, in *Criminal Justice and Behaviour*, 2008, 1095 ss.

tazione riferibile ai gruppi sociali di appartenenza dell'interessato e di spingersi, invece, ad un giudizio individualizzato.

I (troppi) interrogativi attorno al concreto ed esatto funzionamento del COMPAS e, prim'ancora, al modo in cui era stato istruito hanno portato a dubitare dell'attitudine della soluzione algoritmica di dare obiettività ad un ambito tradizionalmente sfuggente quale è l'indagine sulla pericolosità sociale. Il problema, però, non era tanto nell'accertare la procedura meccanica che aveva consentito di trasformare le informazioni iniziali nel *risk assessment* – ciò che, dice la Corte, era possibile consultando il manuale – quanto, piuttosto, nelle difficoltà di comprendere quali scelte fossero state seguite nella selezione dei fattori di rischio e di verificare che un tale percorso logico fosse rimasto immune da distorsioni cognitive¹⁷.

Si ritorna, così, al punto di partenza: ogni discorso sull'intelligenza artificiale non può prescindere da una riflessione sulla linearità logica, la coerenza e la non contraddittorietà del ragionamento seguito da chi ha istruito la macchina. Il caso *Loomis* offre un ottimo esempio al riguardo, calando la premessa in una dinamica tipica dell'accertamento penale: il giudizio sull'autore – e, quindi, anche sulle sue possibili condotte future – propedeutico al calcolo della pena applicabile.

Ulteriori conferme provengono dal secondo dei due esempi proposti, riconducibile ad una situazione piuttosto comune nel corso delle indagini preliminari, quando gli inquirenti chiedono alla persona che ha assistito ai fatti quanto sia sicura di ciò che sta riferendo.

La dinamica può essere rappresentata nei seguenti termini, attraverso un semplice caso¹⁸. Un taxi è coinvolto in un incidente stradale; in quella città i taxi hanno solo due colori: blu – che costituiscono il 15% del totale – e verde,

¹⁷ HILDEBRANDT, *Algorithmic Regulation and the Rule of Law*, in *Philosophical Transactions of the Royal Society*, 2018, 3 distingue, in modo del tutto condivisibile, tra *spiegazione* e *giustificazione* di un output. «next to the problems raised by a new type of discretion that is inherent in the design of ALI, these systems also generate problems for their interpretability. It is important to note that the output of an ALI system can be explained at different levels: (1) at the level of the research design, explaining the choices made and the trade-offs they generate in terms of reliability, generalizability and explainability; (2) at the level of generalized output, consisting of inferred rules that supposedly determine argumentation lines or court decisions; and (3) at the level of individual predictions, targeting a specific case based on algorithms trained on preceding case law. Though explainability at all levels is important, we need to keep in mind that, in the end, an explanation is not the same as a justification; knowing how the algorithm came to its conclusion does not imply that the conclusion is 'in accordance with the law'».

¹⁸ KAHNEMAN, *Pensieri lenti e veloci*, cit., 222-223.

pari al restante 85%. Un testimone ha identificato la vettura responsabile dell'incidente come blu e ha dichiarato di essere sicuro all'80% della propria percezione¹⁹.

Intuitivamente, la maggior parte delle persone – e, di conseguenza, anche gli operatori incaricati della ricostruzione dei fatti – sarebbe portata a concludere che vi sia una probabilità dell'80% che il veicolo coinvolto nello scontro fosse effettivamente di colore blu.

In realtà, ci si trova di fronte ad un classico problema di inferenza bayesiana. Esistono, infatti, due gruppi di dati – quelli relativi alla distribuzione del colore delle vetture e il grado di attendibilità del testimone – che interagiscono tra loro e devono essere considerati congiuntamente.

Analizzando il caso da questa prospettiva, si isolano i dati rilevanti:

- la probabilità a priori che il taxi sia blu è del 15%, quindi $P(B) = 0,15$;
- la probabilità a priori che il taxi sia verde è dell'85%, quindi $P(V) = 0,85$;
- la probabilità che il testimone dichiari che il taxi è blu, quando lo è effettivamente, è dell'80%, quindi $P(T_B | B) = 0,80$;
- la probabilità che il testimone affermi che il taxi è verde, quando in realtà è blu, è del 20%, quindi $P(T_V | B) = 0,20$.

L'obiettivo è determinare la probabilità che il taxi fosse blu, *dato che* il testimone ha affermato che lo era, ossia il valore di $P(B | T_B)$.

Il problema si risolve applicando il Teorema di Bayes, che consente di calcolare la probabilità condizionata di un evento rispetto ad un altro, a patto di conoscere le probabilità a priori dei due eventi e la probabilità condizionata del secondo evento dato il primo²⁰:

¹⁹ Nell'esempio di Kahneman, «il tribunale ha verificato l'attendibilità del testimone ponendolo nelle stesse condizioni della notte dell'incidente e ha constatato che egli ha riconosciuto ciascuno dei due colori l'80 per cento delle volte e non lo ha fatto il 20 per cento delle volte» (KAHNEMAN, *Pensieri lenti e veloci*, cit., 222). Nella prassi nostrana, più degli esperimenti giudiziali sono tipiche le interlocazioni tra gli investigatori e la persona informata sui fatti, rivolte ad ottenere un'autovalutazione dell'attendibilità dei dati esposti. Da qui, la leggera modifica del narrato che, però, non inficia la parte terminale del ragionamento proposto.

²⁰ Il teorema di Bayes è stato definito come quella «funzione logico-probabilistica che descrive la procedura corretta per revisionare la fiducia verso un'ipotesi alla luce di un insieme di prove» da CHERUBINI, *Trappole cognitive nel ragionamento probatorio*, in *Cass. pen.*, 2007, 1363. Per alcune disamine sull'uso della regola nel procedimento penale GARBOLINO, *Il teorema di Bayes applicato alla prova dattiloscopica trova cittadinanza nel tribunale di Milano*, in *Cass. pen.*, 2016, 3463; DELLA

$$P(B/T_B) = (P(T_B/B) \times P(B))/P(T_B)$$

Il denominatore $P(T_B)$ – ossia, la probabilità che il testimone affermi che l'auto era blu – si calcola considerando sia:

- i casi in cui il taxi è blu e il testimone rende una dichiarazione corretta (cioè l'80% del 15% - $P(T_B|B) \times P(B)$);
- i casi in cui il taxi è verde, ma il testimone rende una dichiarazione errata (cioè il 20% dell'85% - $P(T_V|B) \times P(V)$)

Quindi:

$$P(T_B) = 0,80 \times 0,15 + 0,20 \times 0,85 = 0,12 + 0,17 = 0,29$$

Sostituendo i valori numerici nella formula di Bayes:

$$(0,80 \times 0,15) / 0,29 = 0,41$$

La probabilità che il taxi fosse blu, dato quanto riferito dal testimone, è pari al 41% e non all'80% come, invece, suggerirebbe una valutazione intuitiva.

A seconda dell'approccio, si giunge a soluzioni molto diverse, le quali, ovviamente, non possono essere considerate egualmente esatte. Una ponderazione attenta del caso suggerisce già l'incongruenza di una stima all'80%, che deriverebbe, in modo acritico, dal semplice recepimento dell'attestazione di attendibilità del *loquens*, senza tenere in conto la composizione del parco macchine.

Tuttavia, poiché il ricorso al teorema di Bayes – e, più in generale, ai modelli statistici – non è unanimemente condiviso tra gli operatori²¹, potrebbero

TORRE, *Il "teorema di Bayes" fa capolino al tribunale di Milano*, in *Dir. pen. cont.* (web), 21 ottobre 2015.

Per alcune riflessioni sulla regola di Bayes, sempre da una prospettiva utile alla comprensione delle dinamiche dell'accertamento penale, CATALANO, *Prova indiziaria, probabilistic evidence e modelli matematici di valutazione*, in *Riv. dir. proc.*, 1996, 518 ss.; EAD., *Logica della prova, statistical evidence e applicazione della teoria delle probabilità nel processo penale*, in *Dir. pen. cont.* (Riv. trim.), 2013, 4, 132 ss.; GARBOLINO, *Probabilità e prova in un'ottica operativa*, in *Dir. pen. proc.*, 1995, 998 ss.; STELLA-GALAVOTTI, *«L'oltre il ragionevole dubbio» come standard probatorio. Le infondate divagazioni dell'epistemologo Laudan*, in *Riv. it. dir. proc. pen.*, 2005, 926 ss.; MURA, *Teorema di Bayes e valutazione della prova*, in *Cass. pen.*, 2004, 1808 ss.; ID., *Toeria bayesiana della decisione e ragionevole durata del processo*, ivi, 2007, 3104 ss.

Da una posizione – diremmo – di prudenza sull'applicazione di regole formali nei procedimenti penali, LUPARIA DONATI, *Trial by probabilities. Qualche annotazione "eretica"*, in *L'uso della prova scientifica nel processo penale*, a cura di Cucci-Gennari-Gentilomo, Rimini, 2012, 96, con ampi riferimenti allo studio di TRIBE, *Trial by mathematics: precision and ritual in the legal process*, in *Harvard Law Review*, 1971, 1329 ss., nettamente contrario all'impiego di modelli statistico-algebrici in ambito giudiziale.

²¹ Ad esempio, secondo TRIBE, *Trial by mathematics*, cit., 1360, «the statistical information in these cases will, if given to the jury, create a high probability assessment of civil or criminal liability - and

emergere diversi orientamenti nell'istruzione di un ipotetico algoritmo, il quale, ancora una volta, risentirebbe dei dati in ingresso, ereditandone eventualmente l'impostazione errata.

Tanto il secondo esempio – in una prospettiva futura possibile – quanto il primo – riferito ad un caso reale – ribadiscono la solidità dell'impostazione teorica e del metodo adottato per questa ricerca: prima di interrogarsi sui *bias* di un ipotetico – e, ad oggi, improbabile – giudice o investigatore digitale, è opportuno soffermarsi sulle distorsioni cognitive degli operatori umani e su come queste incidano sui nodi più critici del sistema, in rapporto ai suoi principi fondamentali.

2. *Segreto investigativo, monopolio delle indagini, WYSIATI*. I tratti distintivi della fase preliminare sono la sua segretezza e il ruolo predominante del pubblico ministero, tradizionalmente considerato il *dominus* delle indagini²². Senza soffermarsi su profili ampiamente noti della disciplina, occorre ora interrogarsi su come un simile contesto possa prestarsi ad errori cognitivi in grado di influenzare l'esito stesso dell'attività investigativa e, conseguentemente, i presupposti del giudizio.

L'unilateralità della prospettiva inquirente – e, dunque, la conseguente assenza di un confronto con altre possibili ricostruzioni del fatto – comporta il rischio concreto che l'investigatore, una volta avviata la raccolta degli elementi di prova, elabori una teoria del caso, vi si affezioni e fatichi ad abbandonarla, con l'effetto di sottovalutare, se non addirittura tralasciare, elementi di segno opposto, favorendone, così, la definitiva dispersione²³.

there is no assurance that the jury, either with or without the aid of Bayes' Theorem, will be able to make all of the adjustments in that high prior assessment that will be called for by the other evidence (or lack of it) that the rest of the trial reveals. The problem of the overpowering number, that one hard piece of information, is that it may dwarf all efforts to put it into perspective with more impressionistic sorts of evidence. This problem of acceptably combining the mathematical with the non-mathematical evidence is not touched in these cases by the Bayesian approach». L'impossibilità di ricorrere al teorema di Bayes per combinare le informazioni derivate dal ricorso agli schemi matematici con i dati risultanti da altri esperimenti probatori impedisce, secondo l'Autore, di ritenere quello in argomento il più efficace tra i modelli possibili.

²² In particolare, sul segreto investigativo e sui nessi di sistema con gli altri istituti delle indagini preliminari si rinvia alle sempre attuali osservazioni di CIAMPI, *L'informazione dell'indagato nel procedimento penale*, Milano, 2010, *passim*.

²³ Conviene sempre rammentare che «la tempestività e la completezza di approccio al materiale probatorio sono parametri fondamentali per una ricostruzione affidabile del fatto storico, come recita il motto investigativo “il tempo che passa è verità che fugge”». Sono parole di VALENTINI, *The*

Ancora una volta, si delinea uno scenario che la psicologia cognitiva riconosce come tipico del funzionamento dell'intelletto umano. In prima approssimazione, è possibile affermare che la mente tende a privilegiare schemi cognitivi rapidi, poco dispendiosi e istintivi, rispetto a modelli più riflessivi e analitici, i quali richiedono un maggiore impiego di energie mentali²⁴. La visione a *tunnel* irrobustisce la premessa²⁵: risulta, infatti, più semplice costruire un ragionamento fondato su associazioni di idee che rinforzino l'intuizione iniziale, piuttosto che metterla continuamente in discussione sulla base di argomentazioni contrastanti. Questo già basta a dimostrare la naturale riluttanza del pensiero a reperire autonomamente elementi in grado di confutare l'ipotesi di partenza.

Sulla scorta di questo preambolo, risulta di particolare interesse un esperimento condotto da tre studiosi, volto a dimostrare gli effetti di una visione parziale – come quella del pubblico ministero che si limiti ad acquisire soltan-

untouchables: la fase delle indagini preliminari, l'ufficio del pubblico ministero e i loro misteri, in *Arch. pen. web*, 2022, 2, 18.

Si badi: non ci si riferisce tanto ad una «lacunosità delle indagini» dettata «dalla volontà del p.m. di non perseguire una determinata pista investigativa» (fenomeno approfondito da GIUNCHEDI, *I poteri istruttori del giudice*, in *Arch. pen. web*, 2014, 2, 5), ma di un meccanismo precedente alla definizione di qualsivoglia progetto investigativo; un meccanismo inconscio, che muove per giustapposizioni mentali di conferme dell'intuizione iniziale e per scarti automatici di ogni elemento di segno contrario.

²⁴ STANOVICH-WEST, *Individual Differences in Reasoning: Implications for the Rationality Debate*, in *Behavioral and Brain Sciences*, 2000, 645 ss. distinguono tra un *Sistema 1*, che coincide con il pensiero intuitivo, e il *Sistema 2*, maggiormente incline all'analisi ponderata. La mente umana tende ad assecondare l'intuizione e, quindi, il lavoro mentale del Sistema 1, capace di risposte più veloci ma, per questo, meno riflettute. Si rinvia agli studi di EVANS, *Dual-Processing Accounts of Reasoning, Judgment and Social Cognition*, in *Annual Review of Psychology*, 2008, 255 ss.; EVANS-FRANKISH, *In Two Minds: Dual Processes and Beyond*, New York, 2009; LAVIE, *Attention, Distraction and Cognitive Control Under Load*, in *Current Directions in Psychological Science*, 2010, 143.

²⁵ Il *bias* della conferma, che si pone alla base di queste associazioni mentali, è frutto di un Sistema 1 che opera mentre il Sistema 2 è indaffarato o è, semplicemente, pigro. L'individuo sarà, così, portato ad assecondare la propria intuizione, fidandosi delle sensazioni acquisite in precedenza in merito ad un certo fenomeno. In termini analoghi, già ASCH, *Forming Impressions of Personality*, in *Journal of Abnormal and Social Psychology*, 1946, 258 ss.

In letteratura, sulla c.d. *tunnel vision*, KOEHLER-BRENNER-GRIFFIN, *The calibration of expert judgment: Heuristics and biases beyond the laboratory*, in *Heuristics and Biases: The Psychology of Intuitive Judgment*, a cura di Gilovich-Griffin-Kahneman, New York, 2002, 686; KUMARAN-SEYMOUR-DOLAN, *Frames, biases, and rational decision-making in the human brain*, in *Science*, 2006, 684 ss.; FINDLEY-SCOTT, *The multiple dimensions of tunnel vision in criminal cases*, in *Wisconsin Law Review*, 2006, 2, 291; KASSIN-DORR-KUKUCKA, *The forensic confirmation bias: Problems, perspectives, and proposed solutions*, in *Journal of Applied Research in Memory and Cognition*, 2013, 2, 42. Nel nostro Paese, GULOTTA, *Trattato di psicologia giudiziaria nel sistema penale*, Milano, 1987, 1112.

to gli elementi a sostegno della sua teoria preliminare – nella ricostruzione del fatto storico portato in giudizio²⁶.

I ricercatori suddivisero un certo numero di volontari in tre gruppi e fornirono loro informazioni relative ad alcune controversie, pur con significative differenze: a ciascuno dei primi due gruppi (i *non giurati*) fu presentata la versione dell'una o dell'altra delle parti in causa, mentre il terzo gruppo, formato dai *giurati*, ebbe accesso alle argomentazioni e alle prove di entrambi i contendenti²⁷.

Ai partecipanti che avevano ricevuto i dati parziali fu spiegato che ai giurati erano state comunicate le posizioni di tutti gli interessati e fu chiesto loro di prevedere quanti membri, in un ipotetico collegio composto da venti decisori, si sarebbero schierati a favore dell'una o dell'altra parte. Nella formulazione di tale previsione, era loro consentito specificare un intervallo all'interno del quale ritenevano ragionevole si collocasse il numero esatto di giudici favorevoli. Ai giurati fu sottoposto un quesito analogo: dovevano indicare quanti dei venti decisori ipotetici si sarebbero espressi a favore dell'una o dell'altra parte, precisando, anche in questo caso, il margine di incertezza della loro stima.

Ne derivò un duplice dato.

Sotto un primo profilo, chi disponeva della sola versione dell'accusa tendeva a prevedere la condanna dell'inquisito con maggiore facilità di chi conosceva il teorema difensivo, ossia i giurati e il gruppo antagonista di non giurati. Per altro verso, i giurati mostravano maggiore incertezza nella loro valutazione rispetto ai membri degli altri due gruppi²⁸.

²⁶ I risultati dell'esperimento sono pubblicati in BRENNER-KOEHLER-TVERSKY, *On the Evaluation of One-Sided Evidence*, in *Journal of Behavioral Decision Making*, 1996, 59 ss.

²⁷ Uno dei casi sottoposti ai volontari è riportato da KAHNEMAN, *Pensieri lenti e veloci*, cit., 116-117, oltre che nell'appendice al lavoro di BRENNER-KOEHLER-TVERSKY, *On the Evaluation of One-Sided Evidence*, 69, e può essere riassunto come segue. Un rappresentante sindacale si reca in un *drugstore* per incontrare i dipendenti. Trattenendosi nell'area di vendita, viene invitato dal titolare ad attendere la pausa pranzo e a svolgere i colloqui nel retrobottega. Il rappresentante rifiuta, rivendicando il diritto di continuare il dialogo con i lavoratori, così come era consuetudine. Il titolare chiama la polizia, che procede all'arresto del sindacalista. Questi, in seguito, denuncia il negozio per arresto illegale. Nelle versioni fornite ai non giurati, l'accusa descrive l'arresto come un tentativo di intimidazione, in contrasto con i rapporti abitualmente intercorrenti tra il *drugstore* e il sindacato; la difesa, invece, interpreta l'iniziativa del rappresentante come un atto eversivo, ritenendo, dunque, legittima la reazione del titolare.

²⁸ Il dato conferma l'intuizione esposta già in KAHNEMAN-TVERSKY, *On the psychology of prediction*, in *Psychological Review*, 1973, 237 e ripresa da PETERSON-PRITZ, *Effects of amount of information on predictions of uncertain quantities*, in *Acta Psychologica*, 1986, 229 ss.: la sensazione di solidità di una

La provvista di evidenze “a senso unico” – com’era quella dei non giurati e come, normalmente, è quella del pubblico ministero – può esitare – e, di regola, esita – in un racconto in sé coerente e capace di convincere chi lo ascolta della sua fondatezza. Questo, però, non vale ad affermare la reale esattezza della ricostruzione che, come si è detto, non ha tenuto conto degli argomenti avversari, capaci, in ipotesi, di inficiare la teoria iniziale.

L’esperimento appena descritto attesta le differenze tra le prognosi di chi aveva una visione incompleta del caso, ma intrinsecamente attendibile, e quelle di chi disponeva di tutti i dati; rilevava, infine, la minore incertezza dei primi rispetto agli altri²⁹.

Il concetto di fondo è quello racchiuso nell’acronimo WYSIATI (*What you see is all there is* – Quello che vedi è tutto ciò che esiste), coniato da Daniel Kahneman³⁰ per spiegare come la mente umana accolga di buon grado un racconto dei fatti internamente consistente, benché basato su prove limitate. Di nuovo, è una scorciatoia dell’intelletto, per nulla faticosa, che permette di saltare dai (pochi) dati a disposizione a conclusioni ritenute affidabili.

Questi aspetti debbono essere tenuti in debita considerazione nel preconizzare l’affidamento di compiti decisionali ad un algoritmo. Il suo responso, infatti, risentirà dell’impostazione di chi inserisce l’*input*: ancor più che per l’essere umano, per la macchina opera il vincolo alle informazioni disponibili, non essendo capace di affrontare questioni articolate in modo indipendente³¹, né possedendo qualità tipiche dell’intelligenza umana, come la coscienza o la creatività³².

previsione è determinata dall’intrinseca coerenza degli argomenti a disposizione, e non dalla loro completezza.

²⁹ Per un’applicazione di questo metodo di studio alle distorsioni cognitive, pur relative ad un altro tipo di giudizio, quello arbitrale, e per conclusioni simili alle nostre, si rinvia a RISSE, *The WYSIATI-Effect and Biases Arbitrators*, in *ASA Bulletin*, 2021, 315.

³⁰ KAHNEMAN, *Pensieri*, cit., 116. Il termine, però, viene impiegato a più riprese nello sviluppo del saggio.

³¹ Ciò che, almeno secondo l’attuale livello di progresso scientifico – già s’è detto – non è possibile: la considerazione si unisce a quella della non replicabilità dei meccanismi della mente umana, molti ancora ignoti. In questi termini e per ulteriori considerazioni, GALETTA, *Human-stupidity-in-the-loop?* *Riflessioni (di un giurista) sulle potenzialità e i rischi dell’Intelligenza Artificiale*, in *federalismi.it*, 22 febbraio 2023.

³² LEUCADITO, *Il mondo come dato. La ricerca del giusto tra arte ermeneutica e calcolo algoritmico*, in *Dir. informaz. e informatica*, 2019, 1188; MORO, *Macchine come noi. Natura e limiti della soggettività robotica*, in *Intelligenza artificiale. Il diritto, i diritti, l’etica*, a cura di Ruffolo, Milano, 2020, 48 ss.;

Il WYSIATI influenza la ricerca investigativa e, di conseguenza, compromette le basi della predizione decisoria. Se si può condividere che, un domani, i pubblici ministeri potranno ricorrere a strumenti di prognosi fondati sull'intelligenza artificiale³³, il rischio non è tanto – o, meglio, non è solo – il loro appiattimento sui risultati elaborati dalla macchina, quanto, piuttosto, che il *computer* erediti la loro visione unilaterale e, sulla base di essa, sciogla il nodo dell'azione penale, risponda ad una qualunque istanza della difesa o suggerisca l'accesso ad un rito alternativo³⁴.

In altri termini, familiarizzare con i meccanismi mentali sintetizzati in queste pagine aiuta, già ora, a comprenderne i riverberi sull'economia complessiva dell'accertamento: il segreto investigativo e il monopolio del pubblico ministero sulla fase preliminare, escludendo la difesa, favoriscono una narrazione necessariamente frammentaria e, contestualmente, facilitano la consunzione degli elementi a discarico, dispersi perché ignorati dagli inquirenti e non conosciuti per tempo dall'indagato. Sono evidenti le ricadute sul principio di

TAMPIERI, *L'intelligenza artificiale e le sue evoluzioni*, Padova, 2022, 66 ss.; BELVINI, *Intelligenza artificiale*, cit., 2.

³³ La riflessione è di PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. web*, 2021, 21: «la predizione decisoria diventa quindi parte integrante del sistema processuale penale. Con l'effetto di innalzare notevolmente l'asticella predittiva: perché escludere ragionevolmente la condanna significa, nei fatti, escludere che si possa raggiungere in dibattimento la prova della colpevolezza oltre ogni ragionevole dubbio. Ecco dunque che la regola di giudizio dibattimentale (la quale, lo sappiamo, è il precipitato tecnico della presunzione di non colpevolezza) si espande, si proietta a ritroso, va a condizionare, nella direzione del *favor rei*, le scelte di pubblici ministeri e giudici per le indagini preliminari (i quali, per decidere tra azione ed inazione, e tra decreto di citazione a giudizio e sentenza di non luogo a procedere, dovranno rispettivamente affidarsi ad indagini complete *ex art.* 358 c.p.p., e a supplementi probatori solidi *ex art.* 422 c.p.p.). In questo complesso scenario, è plausibile ritenere che, soprattutto i pubblici ministeri, dovendo spesso gestire atti investigativi irripetibili (di cui occorrerà pronosticare la tenuta in dibattimento), possano ricorrere spesso a strumenti di predizione decisoria basati sull'AI. Con il rischio però, vista l'indubbia difficoltà della prognosi imposta, di appiattirsi troppo sui dati elaborati dalla macchina per sottrarsi (almeno in parte) al peso della responsabilità di una scelta (azione/inazione) che appare davvero difficile da sostenere (soprattutto nei confronti dell'opinione pubblica)».

³⁴ Per quest'ultimo aspetto si rimanda alle considerazioni di KOSTORIS, *Predizione decisoria, diversion processuale e archiviazione*, in *Sist. pen.*, 23 luglio 2021, in partic. 12 ss. L'Autore osserva come un sistema automatizzato impiegato per attività predittive finalizzate all'accesso ai riti speciali debba, in via ipotetica, valutare la convenienza della *diversion* in relazione alle posizioni delle parti che ne facciano richiesta. Per svolgere tale valutazione, tuttavia, sarà necessario confrontare la scelta del rito speciale con i possibili esiti del giudizio dibattimentale. Si ritorna, così, al nodo originario: il rischio che i *bias* di chi programma o istruisce la macchina finiscano per condizionare l'intero processo decisionale, pur rimesso all'algoritmo.

completezza e sulla struttura dell'imputazione, e, quindi – già si diceva – sui presupposti del contraddittorio processuale.

In una battuta, il WYSIATI rende illusoria l'indicazione dell'art. 358 c.p.p.³⁵, donde la necessità di una riflessione che colga come si stia guardando ad un fenomeno fisiologico della psiche umana, del quale occorre tenere conto nella disciplina delle indagini preliminari: come si è sostenuto, esso concerne gli inquirenti, esattamente come tutti gli altri esseri umani³⁶.

Ma, se questo è l'esistente, sovengono dubbi che un algoritmo, sinora incapace di rendersi autonomo da chi lo istruisce, possa ovviare al WYSIATI, offrendo al pubblico ministero quella visione a tutt'oggi che gli è naturalmente estranea³⁷.

Ne deriva una conferma: almeno per ora, i *bias* della macchina sono quelli del suo programmatore³⁸. Ancora una volta, quindi, il tema dell'intelligenza

³⁵ Il riferimento all'art. 358 c.p.p. non è certo casuale, muovendo dalla convinzione che completezza investigativa «significa (anche) rispetto dell'art. 358 c.p.p., e dunque investigazioni a 360 gradi, attente all'esistenza di prove a scarico» (VALENTINI, *The untouchables*, cit., 18).

³⁶ Da una prospettiva analoga si colloca PASTA, *L'attività integrativa d'indagine tra diritto di difesa, separazione delle fasi e pseudoprincipi*, in *Proc. pen. giust.*, 2022, 599-600, che paragona il pubblico ministero ad un ricercatore il quale, acquisita notizia del reato, elabora un'ipotesi accusatoria e lavora per confermarla. Come osserva, poi, CORDERO, *Procedura penale*, Milano, 2007, 1212, anche qualora volesse smentire la propria ipotesi iniziale, il pubblico ministero difficilmente vi riuscirebbe, risultando, comunque, influenzato dal lavoro già svolto. Spetta, dunque, alla difesa il compito di falsificare la tesi investigativa.

Lo ricorda efficacemente NOBILI, *Un «quarto potere»? in Recenti orientamenti in tema di pubblico ministero ed esercizio dell'azione penale. Atti del convegno, Modena, 27 aprile 1996*, a cura di Tirelli, Milano, 1998, 42: «il pubblico ministero non può e non deve fungere da “tutto fare”: proprio perché non lo vogliamo imbecille e con le unghie spuntate, al resto (garanzie, controlli, difesa, tutela pure di necessità eccezionali dell'accusa stessa) devono provvedere altri ruoli, ossia rispettivamente, quello della difesa e quello di un giudice».

Affinché la parte privata possa adempiere a questa funzione, è, però, necessario che sia posta nelle condizioni più idonee: condizione che oggi non sempre si realizza e alla quale l'eventuale impiego dell'intelligenza artificiale non sembra, almeno al momento, offrire un rimedio efficace.

³⁷ Nella seconda metà del diciannovesimo secolo, metteva in luce i rischi connessi ad una ricostruzione del fatto demandata esclusivamente all'accusa e, perciò stesso, incompleta o lenta (connotati, questi, potenzialmente esiziali per un qualunque ideale di giustizia), CARRARA, *Programma del corso di diritto criminale. Del giudizio criminale*, Lucca, 1867, p. 537, §893: «perciò la contestazione dell'accusa (affinché non si converta in una farisaica simulazione del rito) deve aver sempre due condizioni – 1° deve essere *completa* – 2° deve essere *tempestiva*. La scienza è indifferente sul modo diverso di tale contestazione, purché sia *fatta* e sia fatta *utilmente*».

³⁸ Per citare SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo*, in *Riv. dir. proc.*, 2020, 614, la macchina «è ontologicamente condizionat[a] dal sistema di valori e dalle intenzioni di chi ne commissiona la creazione e/o di chi [la] crea».

artificiale merita di essere affrontato solo dopo avere indagato gli errori cognitivi e le semplificazioni concettuali a cui è incline l'essere umano – e, così, pure colui che addestra l'algoritmo – e che possono trovare adeguato approfondimento, conciliando gli studi sul sistema penale con gli approdi della psicologia cognitiva.

3. *Il canone della proporzionalità investigativa nell'era dell'intelligenza artificiale.* La prospettiva di una delega all'intelligenza artificiale di porzioni dell'accertamento del reato sollecita una riflessione sul canone della proporzionalità, irrinunciabile punto di riferimento nello studio di un sistema dei rapporti tra l'individuo e l'autorità incaricata della ricognizione, improntato alle indicazioni delle Carte dei valori³⁹. Lo stesso caso *Loomis* offre uno spunto interessante e, nonostante le rassicurazioni della Corte statale circa il rispetto del diritto di difesa, consente di avvertire almeno un sentore di disagio di fronte alla prospettiva di un calcolo algoritmico i cui funzionamento e impianto risultano in larga parte opachi, ma che, ciononostante, viene utilizzato con disinvoltura per l'irrogazione della pena.

D'altro canto, è indubbio che, nella gestione delle istanze di politica criminale, si fa un ricorso sempre più esteso alle nuove tecnologie, con l'obiettivo non solo di esplorare la *scena criminis* e i pertinenti risvolti fattuali, ma anche – e, sotto un certo profilo, soprattutto – di raccogliere informazioni, catalogarle, archivarle e impiegarle in funzione prognostica, a scopi di sicurezza e, quindi, non solo di repressione, ma anche di prevenzione del reato⁴⁰. Si tratta di un ambito che, quasi naturalmente, si presta a sperimentazioni, anche ri-

³⁹ Il principio di proporzionalità viene inquadrato in questo modo, cioè a salvaguardia di un'equilibrata gestione dei rapporti tra l'autorità e l'individuo, da NEGRI, *Compressione dei diritti di libertà e principio di proporzionalità davanti alle sfide del processo penale contemporaneo*, in *Riv. it. dir. proc. pen.*, 2020, 8 che parla di struttura «univocamente bipolare», giustappunto per contrapporre l'ingerenza del pubblico potere e la sfera più intima delle posizioni personali. Ponendosi da un'analoga prospettiva, secondo BELVINI, *Principio di proporzionalità e attività investigativa*, Napoli, 2022, 10, «nonostante già nel pensiero filosofico classico la proporzionalità fosse qualificata un elemento legato in modo inscindibile al concetto stesso di "giustizia", solo in età illuministica maturarono le prime riflessioni per postularne l'impiego come strumento per limitare la potestà punitiva dello Stato».

⁴⁰ In questi termini, NEGRI, *Nuove tecnologie e compressione della libertà personale: la sorveglianza con dispositivi elettronici dell'imputato sottoposto a misure cautelari*, in *Rev. bras. dir. proc. pen.*, 2019, 1257 e, dello stesso, Autore, *La regressione della procedura penale ad arnese poliziesco (sia pure tecnologico)*, in *Arch. pen.*, 2016, 44 ss.

correndo all'intelligenza artificiale⁴¹.

Non è, quindi, insensato porsi un problema su cui, da tempo, si concentra l'attenzione di quanti affrontano i temi dell'atipicità investigativa, delle indagini mediante captazioni occulte e della contrapposizione tra il diritto dei cittadini all'intangibilità della propria *Intimsphäre*⁴² e le pretese di sicurezza – forse sarebbe più corretto dire: securitarie – della collettività.

D'altra parte, a meno di un mutamento di paradigma, poco probabile senza una ridefinizione delle relazioni tra i singoli e il pubblico potere, la disciplina di un'ipotetica indagine algoritmica dovrebbe allinearsi ai precetti di salvaguardia dei diritti fondamentali, sanciti a livello domestico e sovranazionale⁴³. In quest'ottica, non si potrà trascurare l'impostazione personalistica che traspare dalla convergenza delle fonti europee e della Costituzione, organizzate secondo il principio della maggiore tutela desumibile dagli artt. 52, §3 e 53 Cdfue e 53 C.E.D.U.⁴⁴. Ugualmente, non andrà sottovalutato l'indirizzo dato dall'art. 49, §3 Cdfue che, pur menzionando il rapporto tra le pene inflitte e il reato in sanzione, offre un parametro generale⁴⁵ utile ad orientare gli operatori impegnati nelle attività potenzialmente rivolte alla declaratoria di responsabilità dell'accusato. Difatti, un'indagine non sufficientemente attenta alle posizioni individuali si sostanzia in un atto indebitamente invasivo della sfera dell'interessato; la pena eventualmente applicata quale epilogo di un'inchiesta siffatta e del successivo processo si conferma come un'aggressione ingiustifica-

⁴¹ Quasi per assecondare la «potenziale deriva ubiquitaria del controllo», per usare le parole impiegate per descrivere un pericolo «oggi quanto mai concreto ed attuale» da NICOLICCHIA, *Il principio di proporzionalità nell'era del controllo tecnologico e le sue implicazioni processuali rispetto ai nuovi mezzi di ricerca della prova*, in *Dir. pen. cont.* (web), 8 gennaio 2018, 2.

⁴² Si prende a prestito il lessico della Corte costituzionale tedesca. Si deve a BVerfG, 16 luglio 1969, in BVerfGE 27, 1 l'elaborazione della *Sphärentheorie*, con distinzione tra le situazioni giuridiche individuali in base al grado di resistenza rispetto all'interferenza dell'autorità statale. Vengono specificate tre categorie: le posizioni che rientrano nella *Intimsphäre*, intoccabili, quelle che fanno parte della *Privatsphäre* e le ultime, della *Sozialsphäre*, queste poco a poco più erodibili per il "bene comune".

⁴³ Come ammonisce UBERTIS, *Intelligenza artificiale e giustizia predittiva*, in *Sist. pen.*, 13 ottobre 2023, 4.

⁴⁴ In rapporto al canone di proporzionalità, MCBRIDE, *Proportionality and the European Convention on Human Rights*, in *Proportionality and the Laws of Europe*, a cura di Ellis, Oxford, 1999, 23.

⁴⁵ Si deve questa intuizione a QUATTROCOLO, *Tenuità e irrilevanza sociale del fatto*, in ASPP, *L'obbligatorietà dell'azione penale*, Milano, 2021, 196-197. L'Autrice legge l'art. 49, §3 Cdfue in modo da farne un canone di ragionevolezza della fattispecie penale, intesa non solo nel momento di applicazione della pena, ma lungo tutto il percorso di accertamento del reato.

ta delle situazioni soggettive delle persone coinvolte⁴⁶.

Il richiamo alle categorie generali aiuta ad affrontare al meglio la parte terminale del ragionamento, che deve necessariamente riprendere sia il pericolo concreto che le parti, toccate dall'uso di applicativi nel corso del procedimento, incontrino difficoltà a comprendere la programmazione sottesa a quei dispositivi⁴⁷ (si parla, correttamente, di *black box*⁴⁸), sia la più volte ricordata impossibilità, per il *software*, di rendersi autonomo dalla mano umana che lo ha addestrato.

Senza arrivare all'immagine di un giudice-automa che evolve nel modo di pensare e di agire grazie all'autoapprendimento⁴⁹ – si tratta di uno scenario probabilmente non immediato⁵⁰, sul quale, per ora, non ci si vuole soffermare – è sufficiente interrogarsi sulle potenzialità defensionali di fronte ad una scatola nera⁵¹: *mutatis mutandis*, è lo stesso tema, a suo tempo, introdotto dai difensori di Eric Loomis nell'omonimo caso, pur declinato secondo le peculiarità dello schema nostrano delle indagini preliminari.

Appare certamente problematico parlare di parità delle armi nella fase preliminare, la cui segretezza rispetto all'indagato potrebbe prestarsi all'impiego di ulteriori tecniche investigative occulte, quali sarebbero quelle legate all'algoritmo. Si potrebbe archiviare la questione, concludendo che

⁴⁶ In questo senso, è possibile invocare il canone di proporzionalità quale precetto capace di «investire l'intera reazione dell'ordinamento all'asserita commissione del reato, calibrando non solo la risposta sanzionatoria, ma anche l'attivazione della macchina processuale». Così, QUATTROCOLO, *Tenuità e irrilevanza sociale del fatto*, cit., 197.

⁴⁷ Chiarissimo nell'ammonire circa questo rischio LUPÀRIA DONATI, *La promessa della giustizia tecnologica*, in *Sist. pen.*, 1 agosto 2024, 4.

⁴⁸ Oltre alla bibliografia già in nota, per utili approfondimenti, PALMIOTTO, *The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings*, in *Algorithmic governance and governance of algorithms*, a cura di Ebers-Cantero Gamito, Cham, 2021, 49 o, sul piano dei diritti di difesa e della parità tra le parti, LASAGNI, *Difendersi dall'Intelligenza artificiale o difendersi con l'intelligenza artificiale? Verso un cambio di paradigma*, in *Riv. it. dir. proc. pen.*, 2022, 1545 ss.

⁴⁹ Sul quale, però, si rinvia agli approfondimenti di LUPÀRIA DONATI, *La promessa della giustizia tecnologica*, cit., 4; DANAHER, *Algorithmic decision-making and the problem of opacity*, in *Comput. Law*, 2016, 8, 29 ss.; VUILLE-LUPÀRIA DONATI-TARONI, *Scientific evidence and the right to a fair trial under Article 6 ECHR*, in *Law, Probability and Risk*, 2017, 55 ss.

⁵⁰ Se non nei limiti in cui (sempre) la mano umana fornisce all'algoritmo i dati delle sue precedenti predizioni, così da poterli successivamente impiegare (O. MAZZA, *Distopia del processo artificiale*, cit., 6): anche questa è un'attività che richiede la necessaria interazione tra uomo e macchina.

⁵¹ O, più correttamente, dinanzi all'approccio meccanicizzato della scatola nera: in questi termini, FALATO, *L'inferenza generata dai sistemi esperti e dalle reti neurali nella logica giudiziale*, in *Arch. pen. web*, 2020, 2, 4-5.

l'inquirente artificiale sarebbe solo l'ultimo degli strumenti nella disponibilità dei pubblici ministeri per la migliore gestione della cosiddetta "fase che non conta e che non pesa"⁵². Tuttavia, un tale approdo non terrebbe conto di un aspetto ormai assodato e, di fatto, in grado di scardinare la sistematica del rito criminale, così come elaborata nel 1988: le indagini preliminari hanno assunto un rilievo tale – si potrebbe, anzi, dire: una centralità effettiva – da non potersi ignorare le ripercussioni sul successivo dibattimento⁵³. Sottovalutare le problematiche inerenti al diritto di difesa in quella sede significherebbe sminuire il dato che il giudizio viene costruito proprio sulle risultanze investigative: i vizi, *lato sensu* intesi, di queste ultime incidono direttamente sul modo in cui viene costruita l'imputazione, sul dialogo che si instaura tra le parti circa il teorema accusatorio, sull'efficacia della strategia difensiva e, infine, sulla decisione giudiziale.

Ecco, allora, che, se già si guarda ad un frangente sbilanciato, la cui compatibilità con i principi di correttezza procedurale meriterebbe oggi un'attenta verifica, anche in ragione della sua attitudine a proiettare questo squilibrio sul *sequitur*⁵⁴, immaginare un'ulteriore limitazione delle prerogative difensive connessa all'inintelligibilità, per la parte privata, dei meccanismi interni e nascosti dello strumentario in uso agli investigatori condurrebbe ad uno scenario di evidente divergenza rispetto ai dettami di equità e di giustizia del processo⁵⁵.

⁵² Per una bella panoramica degli argomenti che, al giorno d'oggi, non consentono più di definire quella delle indagini preliminari come la fase che non conta e che non pesa, si rimanda al saggio di CAMON, *La fase che "non conta e non pesa": indagini governate dalla legge?*, in *Dir. pen. proc.*, 2017, 426.

⁵³ Coglie e approfondisce la «travagliata metamorfosi» dell'inchiesta preliminare, che ormai «nasconde una palinogenesi istruttoria capace di influire sulle sorti del giudizio», SCALFATI, *Il fermento pre-investigativo*, in *Pre-investigazioni. Espedienti e prassi*, a cura di Id., Torino, 2020, 1.

⁵⁴ Sul potenziale svuotamento delle garanzie costituzionali per un uso scorretto dello strumentario tecnologico, CERQUA-LUPARIA DONATI, *La versione della Consulta sulla corrispondenza elettronica: un bouleversement in materia di prova digitale?*, in *Dir. inf. e informatica*, 2023, 679 ss.

⁵⁵ A tal proposito, O. MAZZA, *Distopia*, cit., 19-20 promuove l'*umanesimo processuale* ad unico rimedio all'avvento dell'algoritmo: «se realmente si vuole arginare l'impiego massiccio e crescente dei sistemi di AI in ambito giudiziario, impedendo la progressiva sostituzione dell'uomo con la macchina, non è sufficiente regolare e limitare l'uso degli algoritmi sulla base di precetti irrealistici e destinati ad essere travolti dalla prova dei fatti ossia dalla dimostrazione della straordinaria efficacia della macchina nell'affrontare il *problem solving* penale, sia nei giudizi storici sia in quelli prognostici, per non parlare della sistematizzazione dei dati investigativi».

Sul tema del rapporto tra parti pubblica e privata nell'ipotesi (distopica) di investigatori e giudici artificiali, LUPARIA DONATI, *La promessa della giustizia tecnologica*, cit., 5 parla di uno «squilibrio tra

Questi rilievi permettono di testare l'ipotesi del ricorso all'intelligenza artificiale nel volgere delle indagini preliminari attraverso i tre *step* del giudizio di proporzionalità, elaborati dalla giurisprudenza tedesca al fine di fissare dei limiti superiori all'influenza, nelle libertà dei singoli, delle autorità impegnate nella persecuzione dei reati (*Schranken-Schranken*)⁵⁶.

Molto probabilmente, lo strumento algoritmico supererebbe il *test* iniziale, di *idoneità* (*Geeignetheit*), in quanto teoricamente impiegabile in un contesto conforme all'assetto costituzionale⁵⁷: si immagina, senz'altro, il contrasto al reato e, quindi, un'azione a tutela dell'ordine comune e a protezione dei valori lesi dall'illecito.

Qualche dubbio sorge, invece, per il secondo passaggio, di *necessità* (*Erforderlichkeit*): occorrerebbe acclarare l'attitudine del *software* a realizzare l'obiettivo previsto dal legislatore con il minore dispendio possibile sul piano delle situazioni individuali⁵⁸. L'appesantimento della posizione difensiva, impossibilitata a contrastare la *black box* e a comprenderne il funzionamento, potrebbe rendere il programma informatico un'opzione da subordinare ad altre, parimenti efficaci, che garantissero, però, alla parte privata la possibilità di intervenire nell'attività investigativa.

Proseguendo nell'*iter* di verifica, lo squilibrio che si creerebbe tra l'indagato e l'operatore che istruisce il *computer* per governarne, poi, il funzionamento provocherebbe il fallimento dell'ultimo stadio: la prova di proporzionalità in

un Leviatano con molte artiglierie disponibili nel suo arsenale e un singolo cittadino con armi spuntate o desuete».

⁵⁶ Così, NEGRI, *Compressione dei diritti di libertà*, cit., 7; FALATO, *La proporzione innova il tradizionale approccio al tema della prova: luci ed ombre della nuova cultura probatoria promossa dall'ordine europeo di indagine penale*, in *Arch. pen.*, 2018, 1, 23. Entrambi gli Autori ripercorrono le radici germaniche del concetto e della soluzione procedurale per la verifica della proporzionalità degli assetti esaminati.

⁵⁷ Secondo KUNIG, *Das Verhältnismäßigkeitsprinzip im deutschen öffentlichen Recht*, in *Deutschland und Japan im rechtswissenschaftlichen Dialog*, a cura di Kunig-Nagata, Colonia, 2006, p. 71 questa verifica si traduce in un controllo di *non manifesta* incostituzionalità e della *non manifesta* ineffettività della funzione espletata dagli operatori.

⁵⁸ La scelta di necessità ha un'impronta politica, che richiama il canone efficientistico e, di lì, ove correttamente inteso, la logica di un processo che sappia unire l'obiettivo del contrasto al crimine alla protezione dei diritti individuali. Si scorge, così, una doppia valenza nel giudizio in parola: quella propria delle operazioni di *razionalità pratica* e quella tipica delle soluzioni di *opportunità politica*. In questi termini, ORLANDI, *Garanzie individuali ed esigenze repressive (ragionando intorno al diritto di difesa nei procedimenti di criminalità organizzata)*, in AA.VV., *Studi in ricordo di Giandomenico Pisapia*, II, Milano, 2000, 560.

senso stretto (*Verhältnismäßigkeit im engeren Sinne*⁵⁹).

Il nesso oggi inscindibile tra la macchina e colui che la addestra aggrava il divario, già ben chiaro durante la fase preliminare, con colui che è coinvolto nell'inchiesta, che ne ignora i contenuti, i progressi e il teorema accusatorio in fase di formazione, e che, in aggiunta, subirebbe l'incapacità di interagire con l'intelligenza artificiale adoperata dagli investigatori.

È, dunque, possibile raggiungere un punto fermo. Il modello su cui è disegnata la fase preliminare risulta, oggi, difficilmente conciliabile con l'obiettivo di parità che traluce dallo statuto costituzionale del giusto processo, con le conseguenze sul prosieguo già richiamate.

L'innesto di tecniche investigative informatiche – quella che alcune voci in dottrina⁶⁰ hanno definito *prova algoritmica* e che, probabilmente, si potrebbe ricondurre all'elaborazione teorica sulle prove e sulle metodiche di indagine atipiche – accentuerebbe questo quadro, con rinnovato pregiudizio per la difesa.

Conviene, pertanto, auspicare che la produzione normativa – da collocare a livello di fonte primaria e non di regolamento, né di *soft-law*, per un'irrinunciabile esigenza di legalità – vorrà muovere dai riferimenti costituzionali nella direzione di una struttura tendenzialmente paritaria tra i soggetti dell'indagine. Certo, questo richiederà una seria rimeditazione delle categorie di sistema e, probabilmente, una ristrutturazione dell'odierna impalcatura del segmento investigativo.

4. *Prime, necessariamente provvisorie conclusioni.* Resta da chiarire quale possa essere la ricaduta di queste riflessioni sulla pratica degli operatori, in particolar modo della fase preliminare.

Prima di tutto, occorre prendere contezza del rapporto tra l'uomo e la macchina, un rapporto che non può prescindere dal contributo del primo e che, almeno secondo le attuali indicazioni unioniste, che includono gli affari di

⁵⁹ Riprendendo le parole di Henne, *Mit Kanonen auf Spatzen schießen*, in *Deutsches Verwaltungsblatt*, 2002, 1094, il canone in parole è riassumibile nella terna «*Angemessenheit, Verhältnismäßigkeit im engeren Sinne, Zumutbarkeit*», ossia adeguatezza, proporzionalità in senso stretto, ragionevolezza.

⁶⁰ GIANGRECO, *Limiti, paradossi e potenzialità della prova algoritmica nella giustizia penale*, in *Leg. pen.*, 15 febbraio 2025, 1 la definisce come «da prova che accerta l'esistenza di un fatto attraverso informazioni contenute o elaborate da algoritmi per l'automazione o da algoritmi di intelligenza artificiale», non senza richiamare l'effetto *black box*, su cui rinvia ad ARDUINI, *La "scatola nera" della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in *BioLaw Journal*, 2021, 2, 453.

giustizia tra le attività ad alto rischio, dovrà rispettare la garanzia della *sorveglianza umana*, così come resa esplicita dall'art. 14 dell'*A.I. Act*⁶¹.

Questo primo risultato giustifica l'attenzione rivolta al ragionamento umano, alle sue lacune e alle sue distorsioni, ancor prima che ai meccanismi che governano il funzionamento dell'algoritmo⁶². È, dunque, prioritario interrogarsi anzitutto sul modo in cui la macchina venga istruita e su come possa ereditare le falle del processo mentale di chi la addestra; solo successivamente sarà opportuno affrontare il problema della *black box* e la possibilità che, in futuro, il *software* possa elaborare autonomamente le informazioni ricevute, seguendo percorsi non accessibili a chi le ha originariamente inserite⁶³.

Si giustifica, dunque, una ricerca approfondita sui *bias* che influenzano investigatori e decisori. In questa prospettiva, è raccomandabile una maggiore sensibilità verso i contributi provenienti dalle scienze extragiuridiche – *in primis*, come mostrano gli esempi proposti in queste pagine, dalle scienze comportamentali, dalla psicologia cognitiva e dalla statistica – in quanto utili ad illustrare le peculiarità dei comportamenti di coloro che partecipano all'accertamento del reato.

Su un piano eminentemente pratico, questa considerazione sollecita una maggiore consapevolezza dei rapporti interni agli uffici preposti alle indagini. Per come è strutturata la relazione tra il pubblico ministero e la polizia giudizi-

⁶¹ Secondo O. MAZZA, *Distopia*, cit., 16-17, questa resterebbe una falsa garanzia: in un sistema che già indulge alla «giustizia penale assistita o meglio delegata» – un sistema, nel quale le stesse *guidelines* del CSM ammettono la partecipazione di tirocinanti e addetti UPP alle funzioni preparatorie all'esercizio della funzione giurisdizionale – «ben presto si passerà dal controllo del giudice sull'operato dei suoi aiutanti umani a quello sulla macchina. Quanto questo controllo sarà "significativo" possiamo prevederlo sulla scorta dell'esperienza della lasca sorveglianza esercitata sulle decisioni predisposte dai "giudici senza toga". L'ipocrisia del controllo umano significativo è così svelata da quanto accade con l'impiego di questi coadiutori del giudice. Sostituendo i *clerk* con la AI il risultato sarà il medesimo in termini di deresponsabilizzazione del giudice togato, mentre cambierà notevolmente sul piano della quantità, ma anche, a ben vedere, della qualità del lavoro».

⁶² Si pensi alle linee guida del CSM in attuale fase di elaborazione (per una sintesi, si può consultare il pezzo *Pronte le linee guida del CSM sull'uso dell'AI* a firma di Giovanni Negri, edito su *Il Sole 24 ore* del 4 ottobre 2025); esse non prevedranno l'impiego dell'intelligenza artificiale per la ricerca su banche dati giurisprudenziali: esso, «sebbene riconducibile a compiti procedurali, può presentare profili di rischio elevati qualora l'output generato venga utilizzato come base esclusiva o prevalente della formazione del convincimento del giudice». Verrebbe, però, da chiedersi per quale ragione le medesime linee guida contemplino la ricerca dottrinale algoritmica, che presenta analoghi margini di rischio.

⁶³ Come ipotizzato da più parti; oltre agli Autori già in nota, ROSSI, *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?*, Milano, 2019, *passim*.

ziaria, non è concepibile che il primo trascuri di interrogarsi sui possibili errori cognitivi compiuti dagli agenti o dagli ufficiali nell'esecuzione di una direttiva o di una delega. Riprendendo l'esempio delle automobili blu e verdi, si è osservato quali conseguenze possa produrre il fraintendimento del dato sull'attendibilità del testimone: un'ipotetica relazione di polizia giudiziaria che affermi, con un margine di sicurezza dell'80%, che il taxi coinvolto nel sinistro era blu cederebbe al pubblico ministero un'informazione falsata. D'altra parte, lo stesso procuratore è tenuto ad una scrupolosa disamina degli esiti dell'indagine poliziesca, se del caso – e, anzi, auspicabilmente – servendosi di quei saperi che gli consentano di superare eventuali pregiudizi cognitivi.

Lo scenario opposto risulta, invece, particolarmente preoccupante. Se si considera la natura del procedimento penale come una sequenza di atti tra loro interconnessi, l'errore commesso da uno dei soggetti coinvolti rischia di ripercuotersi sul prosieguo e di ricadere, in modo amplificato, sulla decisione finale. È proprio in ragione di ciò che si è ritenuto opportuno sollecitare una maggiore attenzione verso i temi introdotti in queste pagine, anche viste le indicazioni del legislatore, che ha da poco delegato il governo a regolare l'utilizzo dei sistemi di intelligenza artificiale nelle indagini preliminari, nel rispetto delle garanzie inerenti al diritto di difesa e ai dati personali dei terzi, nonché dei principi di proporzionalità, non discriminazione e trasparenza (art. 24, co. 5, lett. e) L. 23 settembre 2025, n. 132).

Immaginando l'inserimento dello strumento algoritmico nella catena delle attività del procedimento, al pari degli altri soggetti coinvolti, esso finirebbe per recepire gli abbagli cognitivi di chi fosse intervenuto in precedenza. Il problema non si limiterebbe solo ad un *output* distorto, ma riguarderebbe – in questo caso, sì – la non piena conoscibilità del funzionamento della macchina⁶⁴.

Una tale circostanza inciderebbe, innanzitutto, sulla posizione della difesa: gli inquirenti, infatti, tenderebbero ad interpretare i risultati forniti dall'intelligenza artificiale come comuni elementi di prova, più o meno funzionali alla costruzione del teorema accusatorio; la parte privata, invece, al momento della *discovery*, non sarebbe in grado di ricostruire con esattezza il percorso logico che avrebbe condotto alla formulazione di un certo addebito.

⁶⁴ Come nel caso *Loomis*, poi, il problema si amplificherebbe ove si ponessero ulteriori questioni di proprietà intellettuale e di segreto industriale. Lo spunto è sviluppato da KOSTORIS, *Predizione decisoria*, diversion, cit., 4.

Il divario tra i soggetti del procedimento – come già evidenziato – aumenterebbe in modo significativo, determinando un’ingerenza sovrabbondante dell’autorità nella sfera individuale dell’indagato. In questo senso si è rivelato utile recuperare la disamina sul canone della proporzionalità, anche alla luce degli spunti offerti dalla dottrina tedesca, che, per prima, ha ricondotto questo precetto al valore indefettibile della dignità umana⁶⁵. Proprio quel valore risulterebbe irrimediabilmente compromesso, qualora si giungesse ad una conclamata incapacità del singolo di confrontarsi, in modo effettivo, con un’azione investigativa opaca e totalizzante, con tutte le conseguenze già rimarcate in relazione al contraddittorio processuale.

Al di là dell’indubbio fascino che l’intelligenza artificiale esercita sugli studiosi – talvolta attratti dalla possibilità di riscrivere le regole del gioco in base ai nuovi paradigmi introdotti dall’irruzione della macchina nelle aule di giustizia – è imprescindibile concentrarsi sui connotati di questi inediti archetipi, al fine di evitare che, nella prassi, essi si trasformino in alternative ai principi di giustizia processuale sanciti dalle fonti sovraordinate.

Lo sforzo richiesto tanto ai teorici, quanto agli operatori è quello di uno studio propedeutico, che metta in luce un difetto – potremmo dire – sistemico: l’immagine del pubblico ministero quale *dominus* delle indagini preliminari rischia, infatti, di degradare in una figura torbida e quasi onnipotente, alla quale viene attribuita, a torto, un’infallibilità che finisce per ripercuotersi negativamente sulla difesa, la quale subisce le conseguenze degli errori da altri commessi.

Una rilettura di concetti cardinali della materia – quali il segreto investigativo, il monopolio degli inquirenti nella fase preliminare e la correlata unilateralità della loro prospettiva nella ricostruzione del fatto – attraverso gli approcci delle neuroscienze o delle discipline comportamentali, consente di acquisire strumenti utili a prevenire una deriva delle basi accusatorie del sistema nel momento in cui le risorse digitali faranno definitivamente ingresso nel proce-

⁶⁵ Conviene riprendere un passaggio fondamentale della sentenza nota come *Elles Urteil* (BVerfG, 16 gennaio 1957, in *BVerfGE* 6, 32) che descrive la dignità umana come «*die Würde ... die im Grundgesetz der oberste Wert ist*», ossia il più elevato tra i valori proclamati dalla Carta fondamentale. Su questa pronuncia epocale e sui suoi riverberi sulla giurisprudenza costituzionale successiva, GRIMM, *Das Elles-Urteil*, in *Verfassungsge richtsbarkeit in der Bonner Republik*, a cura di Meinel, Tubinga, 2019, 155 ss.; H. BIRKENKÖTTER, *Art. 2 Abs. 1 GG als allgemeines Freiheitsrecht und Auffanggrundrecht: Das Elles-Urteil*, in *Vorbereiter-Nachbereiter? Studien zum Verhältnis von Verfassungsrechtsprechung und Verfassungsrechtswissenschaft*, a cura di Grimm, Tubinga, 2019, 9.

dimento penale.

È questo, dunque, il tempo per un ripensamento radicale, che indirizzi sia il lavoro del legislatore e della dottrina – nella preparazione di modelli capaci di salvaguardare l'autonomia e i risultati dell'indagine senza, però, sacrificare la partecipazione della difesa – sia l'azione degli operatori, guidata da una convinzione ormai matura: quella della necessità di un confronto sui contenuti e sulle modalità dell'esplorazione preliminare, così da azzerare – o, almeno, ridurre – il pericolo di errori destinati ad ingigantirsi nel corso del procedimento e ad incidere, in modo irreversibile, sull'esito del giudizio.