

## ORIENTAMENTI

---

**MARIANO MENNA**

### **Intelligenza artificiale e controllo contenutistico delle “forme” di produzione dei suoi *outputs* conoscitivi**

Se non si è mai riusciti a cogliere i meccanismi più interni di funzionamento della intelligenza umana - anche per il contributo ancora modesto in proposito delle neuroscienze - al fine di testare i risultati conoscitivi del cervello dell'uomo, si può sostenere che gli *outputs* conoscitivi della intelligenza artificiale possano sindacarsi anche indipendentemente dalla conoscenza dei meccanismi interni di elaborazione dei prodotti gnoseologici della IA. All'uopo possono essere utilizzati, per il riscontro, i due modelli di ragionamento impiegati nel processo penale: quello basato sul calcolo unilaterale delle probabilità e l'altro che sfrutta il contesto in cui si esternalizzano e dialettizzano le versioni sulla realtà dinanzi ad un decisore che faccia uso - per stabilire i significati - esclusivamente della logica formale. In futuro si potrà tentare di sindacare meglio anche gli stati computazionali interni delle forme di produzione conoscitiva dell'intelligenza artificiale mediante sensori che restituiscano la fotografia delle connessioni delle reti neurali dell'IA in associazione alle progressive configurazioni conoscitive che si creino nel processo di elaborazione degli *outputs* della IA.

*Artificial intelligence and content control of the "forms" of production of its cognitive outputs*

*If it has never been possible to grasp the most internal mechanisms of functioning of human intelligence - also due to the still modest contribution of neuroscience in this regard - in order to test the cognitive results of the human brain, it can be argued that the outputs knowledge of artificial intelligence can also be considered independently of the knowledge of the internal mechanisms of processing of the epistemological products of AI. For this purpose, the two reasoning models used in criminal proceedings can be used for verification: based on the unilateral calculation of probabilities and the other which exploits the context in which the versions of reality are externalized and dialectized before a decision maker who uses - to establish the meanings - exclusively formal logic. In the future, it will be possible to attempt to better syndicate the internal computational states of the forms of cognitive production of artificial intelligence through sensors that return the photograph of the connections of the AI's neural networks in association with the progressive cognitive configurations that are created in the process of processing of AI outputs.*

**SOMMARIO:** 1. Dall'ingovernabilità *ex ante* dell'elaborazione conoscitiva dell'intelligenza artificiale generativa al problema del riscontro *ex post* dei suoi risultati di apprendimento. - 2. I modelli di razionalità del processo penale italiano quali esempi di moduli generali di conoscenza utili ad approfondire gli *outputs* della AI. - 3. Dall'esemplificazione processuale lo spunto per il controllo contenutistico sulle “forme” degli argomenti conoscitivi dell'AI e la ripercorribilità *ex post* del percorso conoscitivo operato dalla intelligenza artificiale generativa.

1. *Dall'ingovernabilità ex ante dell'elaborazione conoscitiva dell'intelligenza artificiale generativa al problema del riscontro ex post dei suoi risultati di apprendimento.* La valutazione del materiale decisionale nel processo penale mediante sistemi di intelligenza artificiale è sicuramente un'operazione ad alto rischio, ma le necessarie cautele in proposito - imposte anche dalla recente

normativa europea<sup>1</sup> - non devono indurre ad abbandonare l'impiego della AI per il vaglio delle prove per come singolarmente considerate o inserite nel contesto di tutte le risultanze acquisite<sup>2</sup>.

In merito, si pone indubbiamente il problema del controllo dei prodotti della AI ma lo stesso è stato analizzato fino ad oggi con prevalente attenzione alla possibilità di vagliare l'algoritmo posto a base del procedimento valutativo<sup>3</sup>, per tacere degli altri fattori di contesto di utilizzo della AI da sindacare, come, in particolare, per ciò che concerne l'intelligenza artificiale generativa, la serie di dati che allenano le macchine intelligenti nelle operazioni di autoapprendimento<sup>4</sup> o la tipologia di operatori che presiedono alla selezione dei dati da fornire al sistema operativo di turno per generare gli *outputs* conoscitivi.

Si giunge a contrapporre l'uso di un sistema operativo all'altro nell'approfondimento di un unico problema conoscitivo<sup>5</sup> per dialettizzare i risultati attraverso la diversità dei dati che allenano le differenti macchine intelligenti o mediante la distinta programmazione che condiziona - senza però dominare totalmente - gli *outputs* conoscitivi dei sistemi messi a confronto.

Il controllo significativo che in tal modo l'uomo vuole operare sulle macchine, nondimeno, non comporta la possibilità di riprodurre totalmente attraverso la mente umana il processo di confezionamento dei risultati dei sistemi operativi intelligenti quando i medesimi siano di tipo generativo<sup>6</sup>.

In altri termini, il *modus procedendi* in tema di intelligenza artificiale generativa è tale per cui si cerca di imitare l'installazione nelle macchine intelligenti di elementi simili alle reti neurali del cervello umano e, quindi, senza go-

---

<sup>1</sup> Sulle basi tecniche e sulla recente normativa europea in tema di AI, cfr., per tutti, ZICCARDI-ROSSI, *Guida all'intelligenza artificiale. Dalle basi tecniche alla lettura dell'AI Act*, Milano, 2024, *passim*.

<sup>2</sup> Sul tema si interrogano, tra gli altri, UBERTIS, *Intelligenza artificiale, giustizia penale, controllo significativo umano*, in *Dir.pen. contemp.* 2020, 75 ss. e KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Discrimen* 2024, *passim*.

<sup>3</sup> Vedi, per esempio, in proposito SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv.it.dir. proc. pen.* 2020, 605 ss.

<sup>4</sup> In merito, ponendo l'accento sulla immensa mole di dati che i sistemi di intelligenza artificiale anche generativa utilizzano, SORBELLO afferma che "le nuove tecnologie mettono a disposizione dell'investigatore un ampio patrimonio informativo disponibile per orientare l'attività operativa in maniera selettiva e proficua" (*Banche dati, attività informativa e predittività. La garanzia di un diritto penale del fatto*), in *Dir. pen. contemp.* 2019, 374 ss.)

<sup>5</sup> Al confronto tra più sistemi operativi di intelligenza artificiale si accenna, per esempio, in NATALE, *Intelligenza artificiale, neuroscienze, algoritmi*, Pisa, 2024, *passim*.

<sup>6</sup> Per i problemi anche etici che l'ingovernabilità del percorso conoscitivo della AI suscita, cfr., per tutti, FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, *passim*.

vernare, ossia senza indicare al sistema operativo tutti i passaggi da porre in essere per processare ogni dato utilizzabile nel rispondere ad un quesito conoscitivo, si lascia alla macchina l'autonomia anche metodologica per risolvere il dubbio posto all'origine della interrogazione della macchina<sup>7</sup>.

Rispetto al procedimento generativo, si può allora controllare la capacità di trattamento dei dati della macchina (ed in particolare il margine di errore nel calcolo che possa essere specificamente previsto) e la correttezza metodologica delle istruzioni date al sistema operativo in funzione del miglioramento del risultato conoscitivo - in ciò potendo apprezzare anche la superiorità di un metodo oggetto di programmazione rispetto ad altro tipico di un altro genere di macchina intelligente che venga utilizzata per risolvere problemi analoghi -, si può insomma sindacare anche la qualità metodologica dell'algoritmo che impieghi un sistema operativo, ma, data l'autonomia - sia pure condizionata dalle istruzioni del programmatore - della generazione degli *outputs* conoscitivi, non si potrà prevedere il percorso logico seguito dalla macchina, né si potranno capire i tipi di connessione posti in essere per restituire una certa conoscenza, ed, allora, non si potrà essere sicuri del fatto che la macchina a cui si ponga un quesito conoscitivo, non si indirizzi in una prospettiva ricostruttiva dei fatti che non abbia la stessa tipologia selettiva di dati di chi avanzi il quesito o della comunità dei destinatari del messaggio conoscitivo del sistema operativo o degli esseri che siano stati protagonisti delle azioni da ricostruire.

Questo significa che la macchina, in ultima analisi, non essendo dominata nel suo percorso ricostruttivo dal soggetto che la interroghi, né subordinata al metodo di lettura degli avvenimenti dei destinatari del messaggio del sistema operativo, potrebbe essere autrice di una anomalia ricostruttiva che non restituisca all'interprete la realtà dei fatti da rivisitare.

Ed, allora, se non si può oggi come oggi governare *ex ante* l'*iter* conoscitivo della intelligenza artificiale generativa, si può almeno controllare *ex post* l'operazione gnoseologica posta in essere dalla macchina per essere sicuri che sia affidabile il percorso ricostruttivo battuto dalla medesima?

In merito, non è senza rilievo il fatto che, rispetto alla mente umana, non si è mai ritenuto necessario il governo *ex ante* del percorso conoscitivo della medesima con l'introduzione di meccanismi che disponendo della possibilità di

---

<sup>7</sup> In merito, cfr., tra gli altri, QUINTARELLI, *Intelligenza artificiale. Cos'è davvero, come funziona, che effetti avrà*, Torino, 2020, *passim*.

apprendere anticipatamente i tipi di connessione neurale sfruttati dall'uomo, nonché il tipo più interno di selezione metodologica dei percorsi cognitivi posti in essere dall'intelligenza umana, insomma disponendo della possibilità di accedere in anticipo al modo profondo di funzionamento del cervello rispetto a tutte le tappe - anche quelle minime - di elaborazione della conoscenza, permettano di prevedere come ed in quali direzioni si muoverà l'essere umano per sviluppare i suoi esiti conoscitivi.

Solo ora, attraverso le neuroscienze<sup>8</sup> - che peraltro stanno muovendo ancora i primi passi - si sta indagando sul funzionamento del cervello, ma è un dato il fatto che senza conoscere sostanzialmente il medesimo, bensì solamente ponderando le forme di condizionamento dell'elaborazione mentale umana, si è proceduto a verificare l'attendibilità degli *outputs* conoscitivi dell'uomo in altro modo e, cioè, riscontrando *aliunde* e, quindi, mediante un controllo *ex post*, la corrispondenza o anche solo la capacità di corrispondenza alla verità del percorso gnoseologico umano.

Sotto questo profilo, in caso di ripetibilità dei fenomeni, si controlla la conoscenza in prospettiva scientifica mediante la sperimentazione e nell'ipotesi, invece, di irripetibilità degli eventi, l'apprendimento degli avvenimenti passati<sup>9</sup> - come anche, poi, la predizione di quelli futuri non legati ad un'attività sperimentale - si opera con modi che non passano - o almeno non passano ancora - attraverso la conoscenza del funzionamento del cervello umano, bensì con verifiche che in ultima analisi, anche quando intendano avvalersi di metodologie di controllo categoriale e centrato sulla credibilità della operazione conoscitiva, hanno sempre bisogno di un riscontro contenutistico di dati esterni al più intimo modo di operare - apprezzato in termini generali - del processo gnoseologico messo in opera dal cervello umano.

Per la conoscenza degli eventi passati e la prognosi di quelli futuri non legati in prevalenza alle possibilità di sperimentazione, una dimensione favorevole all'approfondimento dei modi della conoscenza umana è quella del processo penale in cui, al netto delle limitazioni legali del processo inventivo motivate dalla tutela di interessi extraprocessuali che si contrappongono al poterdovere di approfondimento conoscitivo, sia la guida legale del processo di

---

<sup>8</sup> Per una sintesi da cui emerge lo stato attuale delle neuroscienze cfr., tra gli altri, KANDEL, KOESTER, MACK, SIEGELBAUM, *Principi di neuroscienze*, Milano, 2023.

<sup>9</sup> Si accenna alla particolarità del processo di apprendimento di avvenimenti del passato irripetibili in FERRUA, *Il concetto di "provare" e di "provato"*, in *Dir. pen. e processo*, 2021, 5, 557.

verifica del fatto addebitato, sia l'interpretazione che si fornisca in termini deduttivi dalla normativa di riferimento sono utili per indirizzare il riscontro di fenomeni che cercano di imitare la conoscenza umana di eventi passati e la predizione di avvenimenti futuri come l'elaborazione degli *outputs* conoscitivi su siffatto genere di accadimenti operata dai sistemi di intelligenza artificiale.

*2. I modelli di razionalità del processo penale italiano quali esempi di moduli generale di conoscenza utili ad approfondire gli outputs della AI*

Nel processo penale italiano abbiamo a che fare con due modelli di razionalità per l'approfondimento del fatto addebitato<sup>10</sup>.

In una prima direzione, la probabilità di conoscenza dell'evento passato e la prognosi di accadimenti futuri non legati solo alla sperimentazione, diventa forte - fino a divenire evidenza accettata da tutti anche nell'ipotesi in cui si prenda atto dell'insuperabilità del dubbio sulla descrizione degli eventi<sup>11</sup> - quando viene criticata in un processo di dialettica conoscitiva intersoggettiva che sia garantita da un punto di vista imparziale che senza avere la pretesa di salvaguardare l'accesso ad una verità assoluta consenta di esaltare la credibilità di una delle tesi a confronto - facendola prevalere sulle altre - attraverso la negazione della logicità delle versioni scartate nel giudizio di prevalenza. Questo modo di procedere comporta un *self restraint* nel sindacato finale di preferenza e di conseguente attestazione di maggiore credibilità di una versione sull'altra, quando tutte le ricostruzioni a confronto presentino delle crepe logiche per cui bisogna prendere atto del fatto che nessuna delle medesime sia credibile.

Solo se in mancanza del suindicato *self restraint* si dovesse stabilire e decidere comunque, quale delle versioni comparate sia da prendere per buona al fine di renderla termine di confronto di ulteriori azioni umane, si adotterebbe - come nel processo civile - il criterio del "più probabile che non" per cui si

---

<sup>10</sup> In proposito, cfr. MENNA, *Il ragionamento probabilistico dei contendenti e non del giudice dibattimentale*, in *Arch. Pen.* 2022, *passim*.

In prospettiva diversa da quella che affida al modello sillogistico e, quindi, logico-formale in senso deduttivo una perdurante funzione di accertamento della verità e non solo la caratteristica di schema generale dell'argomentare, cfr., per tutti, CECCHI, *Il sillogismo giudiziale: "quel che resta, quel che se ne va"*, in *Dir. proc. pen.* 2020, 1525 ss.

<sup>11</sup> Il rapporto tra il concetto di evidenza e, quindi, di insuperabilità del dubbio nell'ottica di una cognizione "quantitativa" è approfondito anche in MENNA, *La salvezza dell'interpretazione non discrezionale del giudice dibattimentale attraverso l'introduzione della "giuria elettiva di esperti"*, in *Dir. pen. e proc.* 2021, *passim*.

passerebbe dal giudizio falsificante - che impone il sindacato *self restraint* qualora non si superi il ragionevole dubbio mediante la rilevazione delle inconciliabilità logiche delle ricostruzioni a confronto e vieta di stabilire il peso probabilistico (rispetto al giudizio di corrispondenza alla verità) di una versione rispetto ad altra al netto della incongruenza logica rilevata nella medesima - al riscontro (incentrato sul calcolo delle probabilità) del grado di verosimiglianza di una tesi rispetto ad altra anche nel caso in cui tutte le tesi a confronto presentino delle crepe logiche.

In proposito, non deve trarre in inganno il fatto che anche a fronte del ragionamento probabilistico di cui si è detto dovrebbe valere il parametro del superamento di ogni ragionevole dubbio che sembra richiedere per la definizione della ipotetica questione il raggiungimento di una probabilità più alta di quella favorita dal ragionare sul solo “più probabile che non”<sup>12</sup>

In effetti, se ci si pone sul livello del ragionamento probabilistico, e, quindi, del grado di verosimiglianza di una tesi rispetto ad altra, il carattere solipsistico e, come si dirà, l'incertezza del punto di arresto del processo di formazione delle ipotesi ai fini della valutazione anche statistica della preferibilità di una versione rispetto ad altra, rende sempre soggettivo e, quindi, sempre variabile quanto alla percentuale di riferimento, il sindacato complessivo del coefficiente di probabilità della corrispondenza al vero dell'alternativa prescelta, per cui non sussiste reale differenza - in punto di prospettiva probabilistica e di grado di verosimiglianza - tra il livello di conoscenza fornito dal criterio del “più probabile che non” rispetto all'altro prodotto dall'“al di là di ogni ragionevole dubbio”-

Più specificamente, poi, in prospettiva diversa da quella segnalata come probabilistica, il sindacato logico può vertere sia sull'analisi delle “forme” di confezionamento degli argomenti conoscitivi, prendendo spunto dall'analisi della credibilità dei medesimi (connessa al riconoscimento dei fattori di condizionamento del modo di elaborare gli argomenti), sia sul vaglio dei “contenuti” degli argomenti attraverso l'individuazione di riscontri contenutistici desunti dal confronto con altri avvenimenti passati o futuri<sup>13</sup>.

---

<sup>12</sup> In proposito, soprattutto sul rapporto tra un ragionamento probabilistico come quello delle ordinanze cautelari ed il criterio dell'“al di là di ogni ragionevole dubbio” cfr. FALATO, *Principio del ragionevole dubbio e libero convincimento del giudice cautelare*, in *Arch. Pen.*, 2024, *passim*.

<sup>13</sup> Spesso la segnalata distinzione attiene alla differenza tra il sindacato di attendibilità della prova rappresentativa ed il riscontro della prova critico-indiziaria, ma la differenziazione segnalata nel testo non si

Tanto quando l'analisi logica sia contenutistica, quanto se la medesima verta sulla credibilità delle forme di elaborazione degli argomenti - che pure può dirsi in ultima analisi contenutistica, ma non nel senso del confronto con altri eventi passati o futuri, bensì nella direzione della necessità di scrutinare comunque dati esterni alla conoscenza del metodo di funzionamento intrinseco del modo di ragionare dell'interprete (vale a dire in un'ottica differente dal come su un piano di assoluta generalizzazione elabori le conoscenze il sistema intelligente che produca il risultato conoscitivo), il riscontro logico non dipende dalla conoscenza del meccanismo di funzionamento del cervello che opera l'apprendimento (quel che si può estendere anche al sistema cerebrale di conoscenza delle fonti umane di prova del fatto da approfondire come quello dei dichiaranti nel processo penale), bensì - lo si ribadisce - da dati esterni a siffatto sistema.

Di certo, se si conoscesse nel profondo il modo di operare del processo di apprendimento umano, questo fattore atterrebbe - attraverso il controllo dei segnalati dati (che pure hanno un "contenuto" apprezzabile come indice del metodo di lavoro del sistema intelligente) - all'analisi delle "forme" di confezionamento degli argomenti conoscitivi e restituirebbe un importante momento di verifica della credibilità formativa dei passaggi argomentativi. Ma, come si è detto, solo alcuni profili del funzionamento del cervello, adattati peraltro al caso concreto del soggetto del cui apprendimento e del cui apparato comunicativo si discute, si conoscono e di certo - si pensi all'ingresso nel sistema probatorio processuale dell'approccio neuroscientifico - oggi la considerazione di questo tipo di dati ha una capacità di effetto falsificante, all'occorrenza, nel giudizio logico che si opera nel processo.

Di altra natura è il secondo modello di razionalità che opera in sede processuale<sup>14</sup>, basato solo sul riscontro probabilistico di un unico soggetto conoscen-

---

sovrappone perfettamente alla selezione tradizionale tra ciò che tradizionalmente è rappresentazione e quanto è ricostruzione indiziaria.

Sulla valutazione della prova critico-indiziaria - anche rispetto alla segnalata differenza - cfr., per tutti, GAITO, *Il procedimento probatorio*, in *Teoria e prassi della prova. Profili processual-filosofici*, a cura di Bargi, Gaito, Sagnotti, Milano, 2009, 113 e FERRUA, *Teorie della prova: dialogo con Franco Cordero*, in *www.discrimen.it*, 16 dicembre 2020, 4 ss.

<sup>14</sup> Su siffatto modello razionale di giudizio e di motivazione - che ha il suo prototipo nell'argomentare *ex art.292 c.p.p.* in tema di cautele - per come distinto dal modulo logico-formale - che ha il suo prototipo nel ragionamento su cui fonda la sentenza dibattimentale - , cfr. già MENNA, *La motivazione del giudizio penale*, Napoli, 2000, *passim*, e, più di recente, MENNA, *Il ragionamento probabilistico dei*

te che opera nel suo foro interno il giudizio logico basato sul gioco della posizione delle ipotesi e del controllo delle medesime alla luce di possibili obiezioni logiche e, poi, supera siffatta operazione logica perché secondo quello che si è definito il criterio del “più probabile che non” va a stabilire il “grado” di verosimiglianza di un’alternativa ricostruttiva della realtà anche quando dovesse individuare in essa delle crepe logiche perché semplicemente scotomizza alcuni profili della supposta tesi e ne giudica la preferibilità nel confronto con altre versioni dei fatti pur se ne rivela alcune caratteristiche deficitarie. E’ chiaro, peraltro, che non si avrebbe a che fare con il calcolo di semplici probabilità se non permanessero delle ragioni di incertezza in una determinata prospettiva ricostruttiva e, quindi, invece che di semplice probabilità si potesse avere a che fare con una verità indiscutibilmente assoluta.

E’ palese il fatto che nemmeno nel modello logico basato sul confronto dialettico e critico dinanzi ad un punto di vista imparziale si può mai raggiungere la verità assoluta e si ha quindi sempre a che fare con un risultato conoscitivo che su un piano non relativo al contesto intersoggettivo dialettico è solo probabile. La differenza rispetto all’operazione conoscitiva unilaterale sta nella garanzia di un punto di vista imparziale che ha l’ardire di arrestare il suo processo conoscitivo quando scorga una crepa logica in tutte le versioni poste a confronto e sancisca il permanere di incertezze, mentre il soggetto che conosce unilateralmente, anche se fosse dotato di un punto di vista imparziale, dopo aver provato a tipizzare le alternative ricostruttive per controllarne le falle logiche, non può conseguentemente scartarle ed, in ipotesi, sancire il suddetto *self restraint* circa l’*output* conoscitivo qualora ravvisasse delle crepe logiche in tutte le versioni dei fatti immaginate nel suo foro interno; e ciò perché, essendo le versioni immaginate nel suo foro interno intrise di giudizi assiologici, egli non può arrestare il processo di formazione delle questioni relative al processo ricostruttivo in un punto corrispondente all’esatto limite concettuale della vicenda approfondita ed allora lo fa arbitrariamente secondo la sua opinione. Quindi, sapendo di non poter operare una *probatio diabolica* rispetto a tutte le infinite questioni che si paleserebbero se non delimitasse arbitrariamente il novero delle problematiche da affrontare, non può corrispondentemente rinunciare a produrre comunque un *output* conoscitivo in senso positivo o negativo di un fatto e deve, perciò, bandire il *self restraint*

---

*contendenti e non del giudice dibattimentale*, cit., loc. cit., passim e MINAFRA, *Proscioglimenti anticipati e moduli ragionali di giudizio*, Bari, 2024, passim.



pur quando evidenzi delle discrasie logiche in tutte le alternative ricostruttive immaginate, pena la paralisi che sempre nell'eventuale processo infinito di approfondimento toccherebbe al suo processo decisionale. Siffatta paralisi sarebbe dovuta al fatto che l'arbitraria fissazione del punto di arresto della progressiva formazione delle ipotesi soggettivizza inevitabilmente ogni esito conoscitivo e, quindi, anche l'individuazione delle crepe logiche in tutte le versioni elaborate sarebbe un fatto soggettivo e non tipico di uno stato di dubbio oggettivamente insuperabile a differenza della situazione in cui il contesto delle versioni a confronto sia, come si dirà, oggettivo perché tipizzato dall'esterno rispetto all'intervento del soggetto che debba decidere la soluzione delle questioni dubbie. Il decisore, per il fatto che egli stesso è conscio della imprescindibilità della sua opinione, non essendo sicuro mai della stessa insuperabilità del dubbio, non può essere costretto soggettivamente a ritenere sempre permanente l'incertezza (che in astratto sarebbe dovuta sempre al carattere opinabile di ogni versione degli avvenimenti) e, perciò, al fine di non paralizzare il giudizio in funzione appunto di questa sua opinabilità, dovrà sempre accedere ad una versione positiva o negativa dei fatti superando in ogni caso le incertezze relative. Di qui, l'abbinamento della formazione di una sua rappresentazione dei fatti (che trova luogo in ogni caso e si allontana dal mero sancire la permanenza di un dubbio insuperabile), con un giudizio inevitabile anche del "grado" - calcolato probabilisticamente - di corrispondenza alla verità di quella opinione.

Al contrario, se si esternalizza il processo di formazione della versione ricostruttiva rispetto a colui che deve stabilire la soluzione conoscitiva finale, come avviene nel processo attraverso la distinzione tra l'attività probatoria svolta in contraddittorio ed il momento finale di decisione e di motivazione della medesima, la versione ricostruttiva di parte solleva il giudice dal compito di fissare arbitrariamente il limite concettuale del fatto da ricostruire e, quindi, anche se questo confine è comunque apposto arbitrariamente dalle parti in ragione delle infinite questioni assiologiche che si potrebbero in ogni caso sollevare durante l'attività probatoria in contraddittorio, il giudice, facendo leva sulla delimitazione del contesto che la situazione dell'onere probatorio di parte ha creato, può nella versione di parte delimitare anche il novero delle problematiche più strettamente assiologiche restringendolo a quello dipendente dai profili - che, per il giudice, essendo a lui esterni, sono tipizzabili e, cioè, descrivibili in termini oggettivi -, a cui le parti hanno arrestato la sequen-

za di approfondimento del fatto processuale. Si può, allora, distinguendo gli aspetti tipici dalle questioni assiologiche, procedere al solo controllo logico falsificante dei primi, limitando la risoluzione delle questioni più prettamente valutative all'uso della analogia. Questa, dal canto suo, prescindendo dalle situazioni in cui non si ha a che fare con situazioni pur di opinione che siano comunque definibili in termini di evidenza e, quindi, di riscontro logico del parere soggettivo da assumere<sup>15</sup>, si associa - nel caso di esternalizzazione dei punti di vista ricostruttivi rispetto al decisore - ad un ragionamento pur sempre unilateralmente probabilistico, ma ciò nondimeno veicolabile in un ambito limitato, in ragione del confine concettuale che le parti hanno conferito alla vicenda da giudicare, e non impedisce, allora, al decisore di produrre un *output* conoscitivo prevalentemente, anche se non esclusivamente, incentrato sull'uso falsificante della logica formale e, quindi, più certo rispetto ai termini sollevati dal contesto di conoscenza.

3. *Dall'esemplificazione processuale lo spunto per il controllo contenutistico sulle "forme" degli argomenti conoscitivi dell'AI e la ripercorribilità ex post del percorso conoscitivo operato dalla intelligenza artificiale generativa*

Quale insegnamento trarre allora dal segnalato duplice modello razionale di controllo per operare il riscontro degli *outputs* conoscitivi di una forma intelligente come quella dei sistemi di intelligenza artificiale generativa per i quali, almeno oggi, non si può controllare l'intimo modo di formazione degli argomenti conoscitivi? E rispetto alla situazione dell'intelligenza umana si può progredire nelle tipologie di controllo dell'intelligenza artificiale generativa dal momento che comunque abbiamo a che fare, rispetto ad essa, con un suo processo conoscitivo che per quanto autonomo è pur sempre espresso da un sistema elettro-meccanico?

Nella prima prospettiva, dal doppio modo di controllare la rispondenza alla realtà degli argomenti conoscitivi elaborati dalla intelligenza umana, va da sé che quelli che si sono definiti condizionamenti del modo di elaborazione delle conoscenze - oggetto della verifica delle "forme" di elaborazione dei ragio-

---

<sup>15</sup> Ci si riferisce, qui, alle situazioni come quelle che, sulla base di indici accettati da tutti i consociati, fanno definire, per esempio, un delitto come "grave" e trasformano, quindi, in evidenza la valutazione di tipo assiologico che, se, invece, si associassero a zone grigie di definizione come "grave" del delitto, non potrebbero, a differenza delle prime, essere oggetto di una cognizione che da qualitativa divenga "quantitativa" ed oggettiva.

In proposito, cfr. MENNA, *Il ragionamento probabilistico dei contendenti e non del giudice dibattimentale*, cit., loc. cit., passim.

namenti - si associano alle “forme” di produzione degli *outputs* conoscitivi della intelligenza artificiale nel novero ancora ridotto delle medesime.

In queste rientrano controlli basati sempre su dati esterni e definibili anch’essi di tipo “contenutistico” - rispetto a quelli relativi alla conoscenza del percorso logico internamente operato dalla intelligenza artificiale generativa - come le parziali informazioni sul funzionamento della macchina (compreso il margine di errore del modo di elaborazione della medesima e le istruzioni date alla stessa per operare nel caso specifico di confronto con il problema conoscitivo da risolvere) e la conoscenza dei dati archiviati utilizzati dal sistema operativo, compresi quelli che alimentano l’autoapprendimento della macchina.

Al controllo delle “forme” di elaborazione si aggiunge il sindacato sui “contenuti” più tradizionalmente intesi come tali - nel senso di dati messi a confronto con eventi esterni a quelli inerenti ai dati metodologici di elaborazione dei risultati conoscitivi della macchina. A questo proposito, siffatti *outputs*, al pari di risultati accertativi ottenuti nel processo penale, vanno testati nella loro credibilità desunta dall’inserimento nel contesto di tutte le altre risultanze di cui si dispone per risolvere il problema da approfondire. Ed, allora, correttamente si è detto, che, innanzi tutto, bisogna capire quali “ipotesi” di soluzione della questione da affrontare quegli *outputs* sorreggono. Ciò risponde all’individuazione delle inferenze gnoseologiche che bisogna sviluppare rispetto alla versione complessiva di un fatto da elaborare, per come desumibile da tutto il novero del materiale conoscitivo anche diverso dalle conoscenze fornite dalla AI. Quindi, bisogna stabilire le eventuali incompatibilità logiche dei supposti risultati con le altre risultanze di contesto<sup>16</sup>.

Queste segnalate operazioni possono essere effettuate senza conoscere - lo si ribadisce - il funzionamento dell’elaborazione logica interna del sistema di intelligenza artificiale generativa e, per riprendere l’aggancio ai due modelli di razionalità di cui si è detto relativamente al processo penale, va annotato che esse possono avere luogo sia rispetto ad un contesto di proposizione di alternative ricostruttive esterno rispetto al foro di colui che dovrà decidere la solu-

---

<sup>16</sup> Nella stessa direzione, quanto al modo di testare nel contesto di tutte le risultanze i prodotti della intelligenza artificiale, la credibilità dei suoi *outputs* conoscitivi (che, però, come si vede, è centrata sul confronto con avvenimenti anche probatori esterni rispetto alle risultanze del sistema intelligente considerato, più che sulla analisi delle “forme” di condizionamento della produzione degli *outputs* che pure bisogna operare), UBERTIS, , *Intelligenza artificiale, giustizia penale, controllo significativo umano*, cit., loc. cit., 75 ss.

zione, sia a fronte di un contesto di versioni ricostruito solipsisticamente dall'ipotetico decisore.

Per quanto si è detto, solo in questa seconda prospettiva, il carattere solamente solipsistico e probabilistico del ragionamento dell'interprete indurrà il bisogno di stabilire, secondo un calcolo delle probabilità, in misura totale il grado di verosimiglianza - rispetto ad altre alternative ricostruttive - dell'*output* conoscitivo della macchina che impieghi l'intelligenza artificiale.

Quando il decisore non è anche proponente il contesto, egli potrà in prevalenza operare ragionamenti meramente deduttivi e per i profili tipici del risultato conoscitivo limitarsi a stabilire la prevalenza di una delle tesi a confronto utilizzando esclusivamente un ragionamento oggettivamente falsificante fondato sulla logica formale. Solamente quando il risultato conoscitivo della macchina comporterà la risoluzione di quesiti di carattere assiologico, l'interprete opererà in proposito ragionamenti meno certi, di tipo probabilistico, che, però, fonderanno su applicazioni analogiche che non avranno carattere prevalente rispetto alle vere e proprie "deduzioni" perché limitate dai confini tipici della vicenda complessiva affrontata e tipizzata nel suo contesto esaltando punti di vista esterni rispetto a quello del decisore.

In queste considerazioni è la sintesi dell'approccio di qualsiasi interprete a del materiale conoscitivo. In proposito, il processo penale offre solo lo spunto per effettuare un'analisi che trascende la vicenda giudiziaria.

A quanto detto bisogna solo aggiungere che, come nel processo, anche in generale lo scrutinio e, poi, l'impiego prammatico degli *outputs* conoscitivi è limitato da paletti metodologici che può darsi da solo l'interprete o che provengono dall'esterno, come, nel caso del processo, dalle norme scritte; e siffatte limitazioni, essendo proprie dell'approccio ermeneutico dell'interprete umano, sono ulteriori rispetto a quelli che si sono definiti "condizionamenti" della macchina intelligente, i quali in fin dei conti incidono sempre sul metodo di ricostruzione del fatto. Tra questi si collocano in bella vista le istruzioni del programma che fornisce l'operatore all'elaboratore degli *outputs* conoscitivi.

A questo punto, però, si può ulteriormente progredire nel riscontro delle conoscenze fornite da un sistema operativo che sfrutti l'intelligenza artificiale perché, oltre al riscontro contenutistico *ex post* dei risultati della macchina di cui si è detto, si può provare a penetrare qualche apparente segreto ulteriore

del percorso conoscitivo operato dal sistema operativo di intelligenza artificiale.

Con questo non si vuole dire che svelando questo ulteriore profilo nascosto – a cui si accennerà subito di seguito – si potrà sempre prevedere in anticipo il tipo di connessioni che decida di privilegiare la macchina, se non altro, perché di fronte all’infinita varietà di casi concreti avvicinati con l’intelligenza artificiale, il percorso di connessioni inevitabilmente cambia e non solo i collegamenti logico-formali, ma pure le analogie operate dal sistema operativo possono essere tali e tante da non risultare governabili in anticipo.

Ci si vuole, invece, riferire ad una possibilità di riscontro che potrebbe avvenire sempre *ex post* sul percorso conoscitivo operato dalla macchina e che, nondimeno, rispetto all’analisi tradizionale dei “condizionamenti” formali del processo di produzione degli *outputs* conoscitivi (tra cui lo studio dell’algoritmo), ed a fronte pure del riscontro che l’interprete può dare secondo il modello monologico unilaterale di approccio ai risultati conoscitivi oppure secondo lo schema insieme discorsivo e logico-formale di ragionamento sui medesimi che sfrutta l’esternalizzazione e la dialettica dei punti di vista rispetto a colui che legge la corrispondenza a verità del prodotto conoscitivo, fornisce qualche opportunità di considerazione maggiore circa il metodo in concreto seguito al suo interno dalla macchina per elaborare le conoscenze in questione.

Più specificamente, bisogna ritenere che il fatto che la macchina espressiva di intelligenza artificiale generativa processi un alto numero di dati significa che la stessa segua un percorso e, quindi, una progressione – per quanto rapidissima – nella generazione di concetti che, poi, portano al risultato conoscitivo finale.

Ebbene, siccome anche le neuroscienze stanno sviluppando dei sistemi di *neuroimaging* che associano a progressive fotografie dello stato del cervello umano la produzione di distinti concetti<sup>17</sup>, a maggior ragione, quando ci si confronta con un elaboratore fondato sulla intelligenza artificiale, si potrebbero installare nella macchina chiamata a generare conoscenza, dei sensori in grado di fotografare gli stati progressivi delle connessioni prescelte dal sistema operativo per formare la conoscenza.

---

<sup>17</sup> Se ne è dato conto in una trasmissione televisiva del 2024 della rubrica Tgr Leonardo, in [www.youtube.com/leonardo.archivio.it](http://www.youtube.com/leonardo.archivio.it).

Evidentemente, un'operazione del genere non può che essere fatta con l'ausilio di *computers* che in collegamento con i supposti sensori siano in grado di restituire le immagini dei progressivi stati delle connessioni elettromeccaniche formati per generare un certo risultato conoscitivo nel sistema di intelligenza artificiale, associando, per di più, i distinti stati delle accennate relazioni connettive a tipi di concetti.

Studiando, allora, la progressione dei concetti evolutisi nella macchina di intelligenza artificiale, l'interprete potrà avere a disposizione una notevole messe di dati simile a quella astratta sia dalla riproduzione fotografica degli stati di connessione, sia dalla associazione di questi ultimi a concetti progressivi.

Siffatti dati, pur essendo solo una traccia del *modus operandi* della macchina che non consente di capire la maniera di operare della medesima rispetto a casi reali nuovi, potrà consentire di interpretare meglio - senza mai la pretesa di giungere alla verità assoluta - il funzionamento della macchina a fronte di casi che nella tipologia appaiano analoghi a quelli già processati dal sistema di intelligenza artificiale generativa e siano già stati scrutinati rispetto alla progressione di concetti che abbia portato al risultato conoscitivo esaminato.

In siffatta prospettiva, grazie alla segnalata interazione tra l'intelligenza artificiale generativa e la capacità di calcolo degli elaboratori tradizionali - che potrebbe, tra l'altro, essere di qui a poco enormemente potenziata in relazione all'elaborazione informatica di tipo quantistico - si potrebbe arrivare a prevedere meglio e, quindi, a governare la produzione degli *outputs* conoscitivi che si riferiscano evidentemente a casi appartenenti alla medesima tipologia di altri già processati.

Quando, invece, si abbia a che fare con tipologie di casi "nuovi" avvicinati con l'intelligenza artificiale generativa, le parziali somiglianze che comunque possano stabilirsi con altri tipi di vicende potrebbero rientrare solo nella dimensione di quei "condizionamenti" che, al pari dello studio dell'algoritmo, aiutano l'interprete a fornire una patente di credibilità o meno agli argomenti gnoseologici forniti, ma senza gettare una luce risolutiva sul metodo specifico seguito dalla macchina per elaborare gli ipotizzati risultati conoscitivi.

In definitiva, il riscontro di dati *ex post* del tipo di quello accennato è comunque di tipo contenutistico perché non può prescindere dallo studio dei concetti elaborati progressivamente dalla macchina, anche se questi poi vengano utilizzati per studiare le "forme" della elaborazione delle conoscenze di intelligenza artificiale.

Va ribadito, nondimeno, che il siffatto tipo di sindacato non può mai gettare luce sul modo di creare gli *outputs* conoscitivi dell'intelligenza artificiale generativa che valga non solo per i casi analoghi già scrutinati, bensì anche per fattispecie nuove.