

INTERVENTO CONTRO LA STUPIDA ILLUSIONE DI UNA MACCHINA «INTELLIGENTE»*

Fabio Coppola**

SOMMARIO: 1.- Premessa: dalla “stupidologia” all’intelligenza artificiale; 2.- I social media e l’assuefazione al pensiero veloce; 3.- L’illusione dell’intelligenza artificiale; 4.- Dall’utente critico al megafono della macchina; 5.- Una lezione per la formazione giuridica.

1.- Premessa: dalla “stupidologia” all’intelligenza artificiale

In queste aule, ospite del professor Andrea R. Castaldo, il professor Ferrando Mantovani presentava, nel 2016, il suo ultimo lavoro monografico «Stupidi si nasce o si diventa? Compendio di stupidologia». Nell’opera, l’Autore individuava uno dei tratti ricorrenti della stupidità umana nella «demonizzazione» del pensiero altrui e nella speculare «sde-monizzazione» del proprio, aprioristicamente e tirannicamente ritenuto meritevole di preferenza¹.

All’epoca non era forse ancora possibile immaginare sino a che punto i *social media* e, da ultimo, l’intelligenza artificiale avrebbero offerto una risposta, sia pure approssimativa, al quesito posto da Mantovani, dissipando almeno un dubbio: stupidi si può nascere, certamente; ma oggi appare altrettanto concreto il rischio che, in assenza di un’adeguata struttura critica e della necessaria consapevolezza, lo si possa diventare anche per effetto dell’impiego dei sistemi di IA.

2.- I social media e l’assuefazione al pensiero veloce

Occorre muovere dai *social media*. Essi ci hanno progressivamente abituati alla rapidità del pensiero e all’immediatezza delle decisioni, sempre più spesso fondate sulla mera

* Il testo corrisponde all’intervento tenuto durante la Prolusione del Prof. Andrea R. Castaldo in occasione dell’inizio dei Corsi di Diritto Penale e Diritto Penale e Intelligenza Artificiale.

** Ricercatore in Tenure Track di Diritto Penale – Dipartimento di Scienze Giuridiche dell’Università degli Studi di Salerno.

¹ Cfr., oltre al manuale citato nel testo, F. MANTOVANI, *Suicidio assistito: aiuto al suicidio od omicidio del consenziente?*, in *La Giustizia Penale*, 1 (2017), 37.

approssimazione e sulla verosimiglianza. La realtà, al contrario, presenta sfumature difficilmente riconducibili entro rigide griglie classificatorie. Queste ultime finiscono per sacrificare la sensibilità necessaria alla comprensione dell'altro, l'esercizio del dubbio e quell'assenza di pregiudizio che dovrebbe precedere ogni valutazione².

Tre indizi compongono, per l'osservatore più scettico, la fantomatica prova di quanto affermato. Meta ha deciso di chiudere negli Stati Uniti un contenzioso, nel quale era accusata di manipolare i minori e di creare dipendenza, mediante un accordo economico di notevole entità. La Commissione europea ha approvato la proposta di un «EU Kids Act», contenente limitazioni all'accesso dei minori ai social network. Quanto ai maggiorenni, basti pensare alla familiarità ormai acquisita con il *dissing* oppure alla tendenza a riconoscersi negli stereotipi veicolati dai *social* attraverso rappresentazioni semplificate: nel linguaggio giovanile sono note le *red*, *green* e *white flags*, alle quali ciascuno di noi sarebbe agevolmente riconducibile sulla base di semplicistici indici comportamentali.

Il minimo comune denominatore è agevolmente individuabile: la riduzione della complessità personale e sociale a etichette immediate, rassicuranti proprio perché approssimative.

3.- *L'illusione dell'intelligenza artificiale*

E l'intelligenza artificiale? In virtù della capacità di elaborare e restituire risposte immediate, anch'esse fondate su meccanismi di approssimazione statistica, l'IA finisce per essere annoverata nell'olimpico degli "stupidi" pieni di risposte, proprio per questo ritenuti intelligenti.

L'assuefazione alla polarizzazione dei giudizi e delle emozioni, alimentata dai *social media*, ha contribuito a generare l'illusione che le *chatbots* siano effettivamente dotate di un'intelligenza superiore, in quanto capaci di offrire risposte verosimili pressoché su ogni argomento. Quante volte l'IA ha fornito il consiglio apparentemente giusto per le pene d'amore o un problema personale? Quante volte è sembrata un ascoltatore più attento dell'essere umano?

Tale illusione prende il nome di «epistemia»³: la convinzione che l'IA sia effettivamente intelligente, laddove essa restituisce soprattutto luoghi comuni, vale a dire risposte statisticamente ricorrenti all'interno del *dataset* sul quale è stata addestrata. Essa opera, in

² In argomento, è estremamente formativo leggere D. KAHNEMAN, *Pensieri lenti e veloci*, trad. it. a cura di L. Serra, Segrate 2016.

³ In Treccani, voce «Epistemia» (Neologismi 2026): «epistemia s. f.: La confortevole illusione di conoscenza prodotta dall'interazione con l'IA generativa dei grandi modelli linguistici (LLM), là dove la plausibilità simulativa del discorso fluente e la coerenza narrativa sostituiscono l'efficienza cognitiva e l'affidabilità dei dati. A parità di ignoranza, l'output generato da un LLM risulta più fluido, articolato, persuasivo di quello prodotto da un essere umano privo di competenze. Ma è un'efficienza simulativa, non cognitiva. E proprio per questo, più l'utente è inconsapevole dei limiti strutturali del modello, più è esposto al fascino

Interventi

altri termini, come una sorta di megafono delle convinzioni diffuse, senza che queste risultino necessariamente validate sul piano scientifico.

Un riscontro empirico di tale limite strutturale può essere rinvenuto nel comportamento di *ChatGPT* quando gli viene posta la domanda «Chi è Brian Hood?». Di tutta risposta, il sistema oppone un generico diniego, dichiarando di non essere in grado di fornire informazioni a riguardo. Il quadro fattuale è, tuttavia, più complesso: Brian Hood è un sindaco australiano che il modello ha erroneamente associato ad accuse diffamatorie di corruzione, sebbene egli avesse agito in qualità di *whistleblower*, denunciando fatti di corruzione cui aveva assistito.

È plausibile - sebbene non dimostrabile - che il sistema abbia privilegiato l'associazione, statisticamente dominante nei dati consultati, tra le parole «sindaco» e «corrotto» rispetto a una ricostruzione meno ricorrente, ma fattualmente corretta, dove il sindaco assume il ruolo di denunciante. Sotto questo profilo, il caso esemplifica la logica propria dell'epistemica, in forza della quale le regolarità linguistiche superficiali si sostituiscono a un ragionamento sensibile alle peculiarità del caso concreto.

L'episodio evidenzia altresì un *deficit* cruciale di quella che viene comunemente percepita come "intelligenza" della macchina, con particolare riguardo alla capacità di riconoscere e correggere l'errore. A causa dell'opacità algoritmica, che preclude una significativa ricostruzione del percorso valutativo interno da cui è scaturito l'*output* erroneo, OpenAI, anziché rettificare le informazioni restituite dal sistema, non ha potuto fare altro che impedirgli di fornire qualsiasi informazione su Brian Hood. Tale soluzione, lungi dal correggere l'errore, si limita a occultarne gli effetti e conferma l'incapacità del sistema di rivedere criticamente i propri risultati: un limite che ne certifica l'assenza di intelligenza e al quale si aggiungono i danni derivanti dalla censura sofferta dal sindaco australiano.

4.- *Dall'utente critico al megafono della macchina*

Il problema, tuttavia, non si esaurisce in tale profilo. L'essere umano, ove eserciti pienamente la propria capacità di giudizio e di discernimento, potrebbe trarre utilità dalla risposta 'veloce' della macchina - intesa come sintesi o primo *screening* - per poi sottoporla ad approfondimento personale, tentarne la falsificazione e pervenire, infine, alla formazione di un giudizio autonomo, agevolato, ma non sostituito, dal lavoro del sistema.

Si assiste, invece, a una pericolosa regressione verso l'acritica adesione ai giudizi approssimativi e trancianti della macchina, cui corrisponde la progressiva rinuncia allo sforzo

della forma e all'equivoco dell'affidabilità. Il risultato è una nuova dinamica sistemica, che potremmo chiamare epistemica: non più solo disinformazione (distorsione del contenuto) o infodemia (sovraccarico informativo), ma una condizione in cui la produzione del sapere è colonizzata dalla sua apparenza. Un'epidemia di testi verosimili che simulano razionalità, argomentazione, autorevolezza – senza alcuna garanzia di fondamento». Nella letteratura scientifica, cfr. E. LORU - J. NUDO - N. DI MARCO - A. SANTIROCCHI - R. ATZENI - M. CINELLI - V. CESTARI - C. ROSSI-ARNAUD - W. QUATTROCIOCCHI, *The simulation of judgment in LLMs*, in *PNAS* 122.42 (2025).

del discernimento. L'essere umano rischia così di ridursi a mero megafono del dato elaborato dal sistema. In altri termini, il rapporto tra essere umano e macchina rischia di subire un'inversione: lo strumento, concepito per potenziare le capacità umane, finisce per sostituirsi al momento propriamente umano della valutazione, trasformandosi da supporto al giudizio nel suo surrogato.

5.- Una lezione per la formazione giuridica

Proviamo, dunque, a trarre una lezione per il diritto, non soltanto penale, dinanzi all'imponente sfida posta dall'intelligenza artificiale. Per interrompere il circolo vizioso sin qui descritto e governare il corretto impiego di tali sistemi, tanto in ambito giudiziario⁴ quanto al di fuori di esso, occorre anzitutto infrangere il mito della loro asserita intelligenza e comprenderne a fondo il funzionamento, le potenzialità e, soprattutto, i limiti strutturali. Vi è, tuttavia, un'esigenza ancora più rilevante: coltivare il pensiero 'lento' attraverso le attività che preservano le peculiarità dell'essere umano: la lettura, la riflessione e la scrittura, le quali consentono di sottrarre il giudizio agli automatismi indotti dalla tecnologia. Per apprendere realmente il diritto penale, o qualsiasi altra materia che si intenda conoscere in profondità, non esistono scorciatoie, né riassunti elaborati dall'IA capaci di sostituire lo studio. Occorre riscoprire la riflessione critica, che richiede uno sforzo costante e faticoso. Al contempo, essa si pone quale unico anticorpo in grado di consentire all'essere umano di preservare la propria umanità dall'omologazione, dall'atrofia critica e dalla fallace convinzione che lo strumento creato mediante la tecnica possa essere migliore del suo creatore⁵.

In estrema sintesi, il rischio non risiede nella capacità della macchina di elaborare risposte, bensì nella disponibilità dell'essere umano a rinunciare a porsi domande, a coltivare il dubbio. Non si tratta, pertanto, di preservare una romantica contrapposizione tra essere umano e tecnica, ma di mantenere ferma la gerarchia tra mezzo e fine: l'intelligenza artificiale deve restare uno strumento di supporto al giudizio, senza trasformarsi nel suo autore.

La sfida, specie per il giurista, è dunque culturale prima ancora che tecnologica e giuridica. Comprendere la macchina significa sottrarla al mito; esercitare il pensiero lento significa impedire che il suo *output*, per abitudine o pigrizia, si sostituisca a quello umano. È in questa duplice consapevolezza che risiede la condizione minima, ma essenziale, per governare la tecnica senza rinunciare alla "grande bellezza" dell'imperfezione umana.

⁴ Non è questa la sede per approfondire, ma la sperimentazione giudiziaria è in una fase molto vivace (e delicata), come testimoniano le linee-guida del CSM e la prima giurisprudenza in materia.

⁵ Si fa riferimento alla «vergogna prometeica». Cfr. G. ANDERS, *L'uomo è antiquato*, Torino 2007.