RIVISTA ITALIANA
DI INFORMATICA E DIRITTO
Periodico internazionale di diritto, giustizia e tecnologie

ISSN 2704-7318 • n. 1/2026 • DOI 10.32091/RIID0267 • articolo sottoposto a peer review • pubblicato in anteprima il 16 mar. 2026

Vincenzo Calderonio

Causa ed effetto dell'intelligenza artificiale: sulla natura giuridica dell'output dell'IA e il suo impatto sul regime di responsabilità

Il contributo esplora criticamente la nozione di autonomia dell'intelligenza artificiale (IA) e il suo impatto sul piano giuridico, ponendo in dialogo informatica, giurisprudenza e recente normazione europea. Attraverso un'analisi comparata delle definizioni di IA e algoritmo, l'articolo tenta di chiarire le implicazioni della capacità adattiva dei sistemi d'intelligenza artificiale, interrogandosi sulla possibilità di attribuire valore giuridico agli output da essi generati. Particolare attenzione è riservata al paradigma causale sotteso ai modelli di machine learning, alla luce della giurisprudenza italiana (sentenza Franzese) e delle previsioni dell'AI Act, con lo scopo di evidenziare limiti strutturali dell'IA rispetto ai concetti di responsabilità, imputabilità e personalità giuridica. L'indagine si conclude con l'analisi di tre contesti applicativi – proprietà intellettuale, sistema giudiziario e ambito medico – per proporre un modello di responsabilità ancorato all'azione umana e coerente con la logica dell'accertamento *ex post*.

Intelligenza Artificiale – Causalità – XAI – Responsabilità

Cause and effect of artificial intelligence: on the legal nature of AI output and its impact on the liability regime

The contribution critically explores the notion of artificial intelligence (AI) autonomy and its impact on the legal level, placing theoretical computer science, jurisprudence and recent European standardisation in dialogue. Through a comparative analysis of the definitions of AI and algorithm, the article clarifies the implications of the adaptive capacity of artificial intelligence systems, questioning the possibility of attributing legal value to the generated outputs. Particular attention is paid to the causal paradigm underlying machine learning models, in the light of Italian case law (Franzese ruling) and the provisions of the AI Act, to highlight the structural limits of AI concerning the concepts of responsibility, imputability and legal personality. The survey concludes by analysing three application contexts – intellectual property, the judicial system and the medical field – to propose a liability model anchored to human action and consistent with the logic of ex post verification.

Artificial Intelligence – Causality – XAI– Responsibility

L'Autore è dottorando di ricerca presso il Dipartimento di Informatica dell'Università di Pisa e l'Istituto di Informatica Giuridica e Sistemi Giudiziari del Consiglio Nazionale delle Ricerche

1. Introduzione

L'intelligenza artificiale non possiede ancora oggi dei contorni ben definiti. Contemporaneamente alla recente diffusione di algoritmi di IA accessibili a chiunque, si sta sviluppando un dibattito intorno alla forma che una sua possibile definizione giuridica potrebbe assumere. In questo quadro, nonostante la crescente emanazione di atti giuridici aventi ad oggetto l'IA, il problema della comprensione dei suoi meccanismi di funzionamento permane. Il tema è stato ampiamente studiato in letteratura[1] e rappresenta ancora oggi uno dei maggiori ostacoli per il raggiungimento di una definizione coerente di intelligenza artificiale.

Lo scopo di quest'articolo è quello di analizzare questo problema con una metodologia comparatistica atta ad investigare le diverse definizioni di IA e algoritmo date dal diritto e dall'informatica. Attraverso questa ricognizione l'articolo si propone di focalizzare la sua attenzione sul problema dell'autonomia dell'IA. Mentre il dibattito prevalente sulla natura dell'intelligenza artificiale, nato dalla domanda di Turing "can machine think?"[2], si è sempre concentrato sulle possibilità che l'IA ha di raggiungere il pensiero umano[3], quest'articolo si propone di analizzare se la stessa abbia facoltà di produrre output giuridicamente rilevanti e se effettivamente agli stessi possa essere conferito un valore giuridico scisso dal rapporto di dipendenza rispetto all'agire umano.

La seconda parte del contributo analizzerà la questione attraverso la chiave interpretativa offerta dalla Corte di Cassazione con la sentenza Franzese (Sez. Unite Penali, n. 30328 del 10 luglio 2002). Laddove infatti l'output dell'IA verrà identificato come prodotto di un calcolo statistico, la sua influenza causale sulle dinamiche che lo coinvolgono dovrà essere controbilanciata da un adeguato valore giuridico che possa garantire le posizioni dei decisori umani che se ne avvalgono. A tal fine verranno analizzati nella parte conclusiva alcuni casi pratici per cercare di comprendere come questi sistemi potranno essere inseriti nel nostro ordinamento giuridico all'interno della cornice regolamentare istituita dall'AI Act, cercando di rispettare un modello di responsabilità improntato alla chiarezza.

2. Base giuridica per la definizione di intelligenza artificiale

Il primo paradigma giuridico per regolamentare l'IA emergente a livello mondiale è quello dell'Unione europea che sotto il nome di AI Act[4] riunisce

1. Si vedano Brennan 2023 e Bathaee 2018.
2. Turing 1950.
3. I lavori di Turing 1950 e Searle 1980 possono essere considerati i capostipiti delle due diverse visioni di IA che oggi sono presenti in letteratura: l'una che vede l'IA come un'entità al pari degli organismi viventi e l'altra come strumento incapace di comprendere o volere.
4. Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale (Artificial Intelligence Act).

una serie di atti giuridici a vocazione universale[5]. I quali comprendono questioni giuridiche complesse come quelle della definizione di IA, della gestione di questi sistemi nel mercato unico europeo e della regolamentazione della responsabilità per gli output prodotti. L'impatto di questa prima legislazione sul tema dell'inquadramento giuridico dell'intelligenza artificiale è stato determinante a livello mondiale. Infatti, frutto di un lavoro di investigazione scientifica e giuridica iniziato nel 2018[6], la terminologia e le forme del paradigma giuridico europeo hanno ispirato alcune proposte legislative comparse di recente sul tema[7], tra le quali è bene ricordare l'Executive order del Presidente degli Stati Uniti Biden[8], la Risoluzione ONU del marzo 2024[9], il disegno di legge brasiliano n. 2338/2023[10] e la legge italiana n.132/2025[11].

Il paradigma europeo, che ha istituito formalmente una definizione giuridicamente rilevante di intelligenza artificiale, non si è però ritrovato per primo ad affrontare l'argomento. Infatti in campo italiano un primo tentativo di definizione di intelligenza artificiale è stato prodotto dalla giurisprudenza amministrativa che, nella sentenza del Consiglio di Stato n. 7891/2021 ha tentato di dare una compiuta definizione di algoritmo e intelligenza artificiale, in funzione sostitutiva del legislatore, in un caso avente ad oggetto una decisione algoritmica effettuata illegittimamente. In questo caso il Consiglio di Stato ha inizialmente richiamato la definizione del giudice di prime cure che aveva definito l'algoritmo come: "Semplicemente una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato". Procedendo poi con una differenziazione tra algoritmo e intelligenza artificiale secondo la quale: "Cosa diversa è l'intelligenza artificiale. In questo caso l'algoritmo contempla meccanismi di machine learning e crea un sistema che non si limita solo ad applicare le regole software e i parametri preimpostati (come fa invece l'algoritmo 'tradizionale') ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico"[12].

Inquadrando così come punto determinante per la distinzione tra algoritmo e intelligenza artificiale l'autonomia, contenuta nella terminologia "elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti", che la seconda dimostra discostandosi in parte dai parametri preimpostati dal codice di programmazione.

D'altra parte, in campo comunitario, l'approccio dell'Unione europea alla regolamentazione ha prodotto nella versione definitiva del testo dell'AI

5. La terminologia AI Act oggi comprende il Regolamento orizzontale di recente approvazione in seno al Parlamento europeo fondato su numerosi atti giuridici tra i quali l'ormai ritirata AI Liability Directive, doc. COM (2022) 496, e due atti preliminari che hanno influito maggiormente sullo sviluppo della normativa: Commissione europea, *Libro bianco sull'intelligenza artificiale*, doc. COM(2020) 65; Commissione europea, *Relazione sulle implicazioni in materia di sicurezza e responsabilità dell'intelligenza artificiale*, doc. COM(2020) 64, 2020.
6. Si veda sul punto il lavoro preliminare di ricerca e inquadramento teorico della fattispecie effettuato dal gruppo di esperti indipendenti nominato dalla Commissione europea, che ha fortemente inciso sulla legislazione finale e sulla definizione della terminologia del dibattito regolatorio a livello internazionale. High-Level Expert Group on Artificial Intelligence 2020, 2019-a, 2019-b.
7. Per una rassegna articolata degli sviluppi legislativi sul tema si vedano i lavori di Taddei Elmi–Marchiafava segnalati nei riferimenti bibliografici.
8. United States, Executive Office of the President [Joseph Biden]. *Executive Order 14110 Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. 30 October 2023. Federal Register, vol. 88, no. 210, pp. 75191-75226.
9. United Nations General Assembly, Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, seventy-eight session A/RES/78/L.49, adopted 11 March 2024.
10. Brasil, *projeto de lei* n. 2338-2023, di cui si può trovare una trattazione con allegata una traduzione non ufficiale in Belli–Curzi–Gaspar 2023.
11. Legge 23 settembre 2025, n. 132, recante "Disposizioni e deleghe al Governo in materia di intelligenza artificiale".
12. Citazioni entrambe provenienti dalla sentenza del Consiglio di Stato n. 7891/2021. Si veda inoltre in argomento Pietrangelo–Nannipieri 2023, che sul tema inquadra con precisione i problemi della questione definitoria.

Act una definizione d'intelligenza artificiale basata sul rischio, nella quale l'IA viene definita in base a differenti livelli di autonomia associati a diversi livelli di rischio. L'articolo 3(1) dell'AI Act dà questa definizione di sistema d'intelligenza artificiale: "un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali"[13].

Un punto in comune che hanno le due definizioni del Consiglio di Stato e dell'Unione europea è il riferimento che fanno entrambe alle capacità di adattamento dei sistemi IA dopo lo sviluppo, caratteristica che li distingue dagli algoritmi che invece sono statici nel loro funzionamento[14]. Infatti, in letteratura sembra essere abbastanza affermata l'idea di definire l'IA in base all'autonomia che mostra[15] e contemporaneamente questo si spiega in virtù della criticità di questa caratteristica. È l'autonomia a segnare la differenza tra un algoritmo e un sistema IA, il quale, a differenza dell'algoritmo, sarà capace di apprendere autonomamente *pattern* di risposta grazie al machine learning[16]. Questo, nel quadro della legislazione europea, viene tradotto nella formula "progettato per operare con diversi livelli di autonomia", che pare essere impiegata strategicamente dalla Commissione europea al fine di classificare i sistemi di IA in base al rischio

La scelta regolamentare in questo senso pare efficace perché risponde ad un'esigenza reale, quella di definire diversi livelli di tutela a seconda del rischio potenziale di questi sistemi. La possibilità di associare a diversi livelli di autonomia corrispondenti obblighi normativi che crescono con l'autonomia del sistema si spiega in virtù del fatto che più un sistema di IA è autonomo, più sarà imprevedibile e di conseguenza aumenterà anche il rischio per un output non voluto. Ciononostante, il riferimento a livelli variabili di autonomia, lasciando ampi margini giuridici per l'inquadramento della fattispecie, rischia anche di omettere l'indicazione di un eventuale limite ad essa: esiste questo limite? E se vi è un limite all'autonomia dell'IA, dove va posto? Questi interrogativi sono fondamentali sia per una stringente definizione di sistema d'intelligenza artificiale[17] che per una chiara attribuzione della responsabilità degli output prodotti da esso. In questo senso può essere utile effettuare una prima comparazione tra IA e algoritmo per scoprire in quale punto del loro funzionamento appare questa caratteristica, l'autonomia, e, successivamente, verificarne l'estensione.

3. Distinzione tra intelligenza artificiale e algoritmo nella computer science

Algoritmo e intelligenza artificiale sono due concetti pratici della computer science. Entrambi possono essere fatti derivare dal concetto più ampio e primigenio di *computazione*, derivato da una

13. Art. 3(1) dell'AI Act, Regolamento (UE) 2024/1689 e ripresa dai principi dell'OECD in tema di AI, *OECD Principles on Artificial Intelligence*. OECD Publishing 2019.
14. Sulla centralità del ruolo dell'autonomia per la definizione di questi sistemi nel quadro regolamentare dell'AI Act si veda Fernández-Llorca–Gómez–Sánchez–Mazzini 2024, che propone un'interessante rassegna terminologica delle definizioni impiegate dall'AI Act, e le recenti linee guida della Commissione europea sul tema, *Orientamenti della Commissione sulla definizione di un sistema di intelligenza artificiale ai sensi del Regolamento (UE) 2024/1689 (AI Act)*, doc. C(2025) 924, che stabiliscono espressamente come "il livello di autonomia costituisce una condizione necessaria per determinare se un sistema possa qualificarsi come sistema di intelligenza artificiale".
15. Si vedano questi testi provenienti dal dibattito americano sulle potenzialità dell'IA: LeCun 2022, Bubeck–Chandrasekaran–Ronen et al. 2023 che hanno contribuito non poco anche grazie ai lavori svolti nel 2023 sugli LLM, in particolare OpenAI 2023, all'introduzione nel regolamento europeo della categoria dei *general-purpose AI systems* che ancora a oggi non sono stati chiaramente definiti dalla letteratura scientifica dell'area della computer science.
16. Su tutti si veda Russell–Norvig 2010.
17. Si veda Almada 2025 che lucidamente critica l'eccessiva ampiezza delle linee guida della Commissione in tema di definizione di sistema IA, le quali si prestano a interpretazioni estensive potenzialmente rischiose per la certezza del diritto.

categoria di numeri definiti computabili per la prima volta da Turing nel 1936[18]. È da questa categoria che è opportuno partire per ottenere una definizione di algoritmo, sorprendentemente assente sia nell'AI Act che nel Regolamento generale sulla protezione dei dati (GDPR)[19], nonostante il ruolo centrale che quest'architettura informatica ricopre per la costruzione di sistemi IA, nonché per il ruolo giuridicamente rilevante che ha in numerose previsioni delle citate normative.

Una più moderna e formale descrizione di cosa sia un algoritmo derivante dalla computer science la si deve a diversi tentativi di risolvere l'*Entscheidungsproblem* che fu proposto da David Hilbert nel 1928[20]. L'*Entscheidungsproblem*, o problema della decisione, originariamente aveva ad oggetto la domanda sull'esistenza di una procedura meccanica capace di risolvere ogni problema della matematica[21]. In particolare esso chiedeva se per una formula espressa nel linguaggio formale della logica del primo ordine esistesse una procedura in grado di dichiarare la solvibilità o meno della stessa[22]. Entrambe le soluzioni proposte da Alan Turing e Alonzo Church dimostrarono l'insolubilità del problema, determinando al contempo un mutamento della sua impostazione teorica. Turing, nell'elaborare l'omonima Macchina, riformulò infatti il problema in termini di computabilità, sostenendo che un problema è computabile, ossia risolvibile mediante una procedura meccanica, se la macchina, che incarna il metodo meccanico richiesto dal problema della decisione, termina il proprio calcolo[23]. La Macchina di Turing ha finito così col diventare la moderna formalizzazione di quello che oggi è un algoritmo.

Le due fondamentali proprietà dell'algoritmo nella teorizzazione di Turing sono la capacità di fornire una procedura meccanica che a partire da un input conduce a un output e la terminazione. Tuttavia, nel caso di un problema senza soluzione, la computazione della Macchina di Turing tende a essere infinita, ovvero a non terminare. In tali circostanze, non si potrebbe parlare propriamente di algoritmo, poiché verrebbe meno la seconda proprietà fondamentale: la capacità di giungere a un risultato finito. Qualunque procedimento meccanico è suscettibile di essere definito algoritmo. Ad esempio, la legge processuale e il diritto amministrativo sono a tutti gli effetti degli algoritmi e rientrano in quella parte del diritto che è da ritenersi computabile[24].

18. Turing 1936.
19. Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati).
20. In breve il problema della decisione chiede "[if] there is an absolutely mechanical procedure, that is an algorithm, for deciding the satisfiability or unsatisfiability of any given sentence" (Börger–Grädel–Gurevich 2001). Questo problema è stato scoperto essere irrisolvibile da Kurt Gödel in matematica con il primo teorema di incompletezza, e in informatica da Alan Turing e Alonzo Church indipendentemente. La scoperta dell'irrisolvibilità del problema ha dato origine alla distinzione tra funzioni computabili e non, secondo la quale per le funzioni computabili esiste un algoritmo che porta ad una soluzione univoca, fattispecie inesistente per le funzioni non computabili; sulla rilevanza del tema in campo giuridico si veda Brennan 2023, che propone un'interpretazione giuridicamente rilevante del problema per dimostrare l'impossibilità di conoscere *ex ante* se un'IA produrrà un output etico o meno, rientrando le questioni etiche nella sfera della non computabilità.
21. "The Entscheidungsproblem is solved when we know a procedure that allows for any given logical expression to decide by finitely many operations its validity or satisfiability [. . .] The Entscheidungsproblem must be considered the main problem of mathematical logic": Börger–Grädel–Gurevich 2001, p. 3.
22. *Ibidem*.
23. Turing 1936.
24. Si veda quanto affermato in Borruso 1995 "Tutte le volte nelle quali la legge è formulata in maniera così rigorosa da costituire una regola generale, astratta, completa, analitica, inequivoca […] la legge ha i requisiti di quello che in informatica (e in matematica) si chiama 'algoritmo' e, quindi, può ben essere applicata anche da un computer […] e in taluni casi – sia pur pochi – li ha, come, soprattutto, nel campo del diritto processuale e amministrativo".

Rispetto alla nozione di algoritmo la definizione di intelligenza artificiale non se ne discosta nelle sue architetture fondamentali, senonché, a differenza dell'algoritmo che è interamente programmato *ex ante* per raggiungere un determinato scopo, l'IA è basata sul machine learning, una tecnica che permette di apprendere autonomamente *pattern* di risposta senza che essi siano stati programmati interamente prima[25]. Diversamente rispetto a quanto avviene con un semplice algoritmo, nel caso dell'intelligenza artificiale non è il programmatore che crea integralmente l'IA. Il programmatore definisce l'architettura e le regole di apprendimento del sistema, selezionando specifici algoritmi o modelli di apprendimento automatico. Il comportamento del sistema non è tuttavia determinato direttamente dalle singole istruzioni, ma si struttura nel corso della fase di addestramento, in cui il modello si forma a partire dai dati e dai criteri di ottimizzazione stabiliti[26]. Una volta che le risposte fornite dal programma sono considerate soddisfacenti dagli sviluppatori, allora si può affermare di avere un modello di risposta. In definitiva, in ogni caso il modello non è altro che un *pattern* di risposte che il programma ha imparato a restituire a fronte di determinati input.

4. L'autonomia come segno distintivo dell'IA e la circoscrizione della sua area di operatività

Il vero e proprio *quid pluris* dell'IA è la sua autonomia, ovvero il come un dato modello di IA sviluppa una particolare prassi di risposta[27]. L'autonomia dell'IA sta nel fatto che essa apprende da sé le vie mediante le quali corrispondere la risposta esatta, corrispondente all'output atteso, senza che nella maggior parte dei casi gli sviluppatori siano in grado di stabilire il percorso logico in base al quale il programma è arrivato alla soluzione[28]. Sotto questo profilo la definizione di intelligenza artificiale proposta dal Consiglio di Stato potrebbe risultare troppo vaga. L'ambiguità sorge nella parte in cui si afferma che l'IA "elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico"[29]. Sul punto non è chiaro se si debba intendere che il processo di apprendimento automatico sia contemporaneo al funzionamento dell'IA o se esso avvenga in un'altra fase.

In questo senso la manualistica[30] spiega come i "criteri d'inferenza" non vengono elaborati costantemente, ma solamente nella fase di addestramento dove il modello IA apprende il *pattern* di risposte che poi adotterà in maniera più o meno uniforme nella fase di interazione con gli utenti. In essa, infatti, il modello rimarrà statico, come l'algoritmo, e per l'aggiornamento del modello di risposta sarà necessaria un'ulteriore fase di addestramento. Un sistema d'intelligenza artificiale dopo l'addestramento viene rilasciato in una determinata versione, ad esempio, GPT è stato rilasciato prima in versione 3.5 e successivamente in versione 4. Ogni versione è vincolata ad un particolare modello di risposta che non si auto aggiorna e per ogni innovazione delle capacità di risposta di un modello IA è necessario effettuare nuovamente una fase di training manualmente e aggiornare il precedente modello fornendogli

25. Per un'interpretazione giuridicamente rilevante del concetto di autonomia di un sistema tecnologico, anche relativamente alla nozione di machine learning, si veda Sartor–Omicini 2016.
26. Russell–Norvig 2010.
27. Un esempio nostrano di ricerca sulla *Explainable Artificial Intelligence* (XAI), branca dell'informatica dedita alla ricerca delle logiche per le quali un determinato sistema di IA corrisponde a determinati output piuttosto che altri, con ampia rilevanza in ambito internazionale è rappresentato da Guidotti–Monreale–Ruggieri et al. 2018; mentre, un altro approccio interessante proveniente dalla ricerca internazionale è quello proposto in Ribeiro–Singh–Guestrin 2016.
28. Questo è il fenomeno che dà origine al *black box problem*, la cui rilevanza giuridica è stata messa in luce da Bathaee 2018.
29. Sentenza del Consiglio di Stato n. 7891/2021.
30. Russell–Norvig 2010.

nuovi dati[31]. Ad oggi le tecniche di *continual learning*, che permetterebbero ai sistemi IA di apprendere in tempo reale, non sono integrabili nei modelli di machine learning per via della loro impraticabilità[32]. Tuttavia, anche se si ipotizzasse un futuro in cui l'IA fosse capace di adattarsi all'ambiente e all'informazione in tempo reale, ci si potrebbe domandare quanto questa forma d'interazione possa essere simile a quella propria delle forme d'intelligenza naturale[33].

In ogni caso queste brevi considerazioni sembrano porsi in antitesi con le definizioni di IA date dal Consiglio di Stato e dalla Commissione europea, nella parte in cui entrambe le definizioni affermano che l'intelligenza artificiale apprende e si auto aggiorna sulla base dei dati dell'ambiente circostante[34]. Ciò è vero, ma con la precisazione che questo tipo di comportamento adattivo non è qualcosa che la macchina compie spontaneamente, ma piuttosto essa necessita dell'intervento umano per ogni nuovo aggiornamento. Con la conseguenza che dopo l'aggiornamento il modello di risposta del sistema IA rimarrà cristallizzato fino al prossimo aggiornamento umano. Resta aperto il problema di come un modello di IA apprenda a corrispondere gli output esatti. È questo infatti il momento nel quale compare l'autonomia dell'intelligenza artificiale che la distingue da un mero algoritmo.

4.1. Il limite dell'autonomia dell'intelligenza artificiale

La questione è attualmente affrontata dal filone di ricerca sull'*explainable artificial intelligence* (XAI)[35], devota alla ricerca della risposta sul come i sistemi di intelligenza artificiale possano apprendere autonomamente, quali regole impiegano, se effettivamente vi è una loro autonomia e quanto essa sia estesa. È da questo campo di ricerca che sono nate le prime voci critiche sull'effettiva estensione dell'autonomia dell'intelligenza artificiale e dunque anche sulla natura degli output che produce.

31. Questa tra l'altro è una delle ragioni all'origine della pratica di raccolta e utilizzo dei dati degli utenti da parte di OpenAI che ha inizialmente causato il blocco da parte del Garante della Privacy, dati utili per aggiornare il modello di risposta, doc. web 10085455.
32. Per aver riscontro di quanto affermato si veda la sezione "On the path to more general artificial intelligence" al punto *continual learning* a p. 93 di Bubeck–Chandrasekaran–Ronen et al. 2023 dove viene ampiamente descritto come uno dei limiti di GPT-4 è che esso non può apprendere in tempo reale, affermazione che comporta la fissità di questi modelli, che restano ancorati a determinati modelli di risposta incapaci di auto aggiornarsi.
33. In argomento Shanahan–Crosby–Beyret–Cheke 2020 dove vengono descritti diversi limiti dell'IA quando comparata all'intelligenza animale come ad esempio l'assenza di comprensione della causalità, la non trasferibilità delle informazioni apprese in altri ambiti e l'assenza di scopi e motivazioni intrinseci.
34. Aspetto rimarcato ulteriormente dalla Commissione europea negli *Orientamenti della Commissione sulla definizione di un sistema di intelligenza artificiale ai sensi del Regolamento (UE) 2024/1689 (AI Act)*, doc. C(2025) 924, nei quali si afferma come "'l'adattabilità' si riferisce alle capacità di autoapprendimento, che consentono al sistema di cambiare comportamento durante l'uso". È opportuno precisare che l'adattività dei sistemi di intelligenza artificiale, nella maggior parte dei casi, non implica un apprendimento continuo. Infatti, il modello apprende durante la fase di addestramento, mentre nella fase di utilizzo non sviluppa nuovi *pattern* di risposta, ma si limita a adattare il proprio comportamento in base agli input ricevuti, senza aggiornare autonomamente i propri parametri. Definizioni troppo ampie, come il riferimento alle capacità di autoapprendimento, potrebbero quindi generare ambiguità interpretative (Almada 2025), in particolare nel trattamento sanzionatorio, alimentando letture che accostano l'adattabilità dell'IA a forme di apprendimento naturale e, in ultima analisi, a ipotesi di rieducabilità attraverso la pena. Per questo motivo, è fondamentale circoscrivere con precisione il concetto di apprendimento automatico, escludendo dal suo perimetro forme di interattività proprie dell'intelligenza naturale e scenari speculativi o fantascientifici.
35. Sul punto si vedano i contributi di Guidotti–Monreale–Ruggieri et al. 2018 e Ribeiro–Singh–Guestrin 2016, che rappresentano vere e proprie pietre miliari sul tema, tentando di dare risposta sul come si sviluppino i modelli di risposta dei sistemi IA.

In letteratura una scuola di pensiero molto diffusa[36] risponde a queste domande affermando che il metodo di ragionamento sul quale sono basate le IA sia la probabilità, in accordo con la sua provenienza dalla computazione. Alcuni esempi del funzionamento probabilistico dell'attuale intelligenza artificiale sarebbero l'*attention mechanism* basato sul transformer e il caso *husky vs wolf*. L'*attention mechanism* è una tecnica di addestramento particolarmente usata per i *large language models* (LLMs) mediante la quale l'algoritmo nella fase di addestramento associa a ciascuna informazione data in input un valore numerico percentuale differente, concentrandosi così sulle informazioni che si presentano più spesso e associando ad esse un peso maggiore[37]; successivamente nella fase di generazione di testo l'algoritmo determinerà quale parola generare sulla base di calcoli probabilistici basati sulla frequenza delle parole usate in contesti simili. Quest'informazione sarà presa sulla base dei dati di addestramento dai quali ha appreso i *pattern* di risposta. Le parole che genererà l'algoritmo saranno dunque quelle alle quali è stato associato lo *score* più alto e baserà su di esse l'output. È facilmente intuibile come la logica sottostante a questo tipo di funzionamento sia la logica della probabilità statistica. Nell'esempio di una traduzione l'algoritmo non si concentrerà sul significato complessivo del testo ma procederà con la traduzione basandosi sulla frequenza delle parole e dando più importanza a quelle che si presentano più spesso.

In modo analogo, nell'esempio *husky vs wolf*, l'algoritmo incaricato di classificare immagini di cani come "husky" o "wolf" ha prodotto una classificazione errata, etichettando alcune foto di husky come lupi. Quando i ricercatori hanno chiesto all'algoritmo sulla base di quale criterio avesse erroneamente classificato un husky come lupo, l'IA ha risposto indicando la presenza della neve: riconosceva una foto come "lupo" basandosi esclusivamente sullo sfondo innevato (*snow pattern*)[38]. Quest'esempio di modello in *over-fitting*[39] rispetto ai dati che riceve sembra dimostrare che l'IA si basi su un'analisi quantitativa, nella quale a fronte del *pattern* ricorrente associa una voce di classificazione. In questo senso il machine learning, differentemente rispetto all'apprendimento naturale[40], individua una maggiore concentrazione di pixel bianchi, la neve, nelle foto rappresentanti i lupi ai quali associa la voce "wolf". Ciò sembrerebbe dimostrare come il machine learning riesca a emulare una parte del pensiero umano basata su un modello di apprendimento *try and error* dal quale però non sempre è possibile risalire a motivazioni fondate. Questa dimensione dell'autonomia dell'IA, che pure è capace di dimostrare livelli d'intelligenza elevati, pone però un problema per le categorie classiche del diritto, creando una notevole incertezza circa la natura giuridica degli output che produce.

4.2. La natura giuridica dell'output dell'intelligenza artificiale

In questo quadro, l'output dell'intelligenza artificiale sembra essere prodotto dalla riproduzione di una legge statistica che, come nel caso *husky vs wolf*, esegue meccanicamente un'analisi quantitativa dei dati senza una reale comprensione causale della situazione che le viene posta[41]. Per tale ragione, molti output dei sistemi di intelligenza artificiale

36. Si vedano i lavori di Bender–Gebru–McMillan-Major–Shmitchell 2021 e Fokas 2023.
37. Vedi Vaswani–Shazeer–Parmar et al. 2017 e per una spiegazione semplificata *Attention mechanism: Overview (Meccanismo di attenzione: Panoramica)*, un video di Google Cloud Tech. Va ricordato che è quest'architettura informatica a costituire il fondamento di tutti i modelli di IA generativa ad oggi disponibili.
38. Nel caso in esempio i ricercatori hanno volutamente fornito all'IA immagini di lupi con associato lo *snow pattern* perseguendo volutamente un modello sbagliato, si veda Ribeiro–Singh–Guestrin 2016.
39. L'overfitting è un problema frequente per i modelli statistici troppo complessi che sviluppano un sovradattamento ai dati di addestramento, il che li rende impraticabili in contesti in cui vi sono dati differenti a causa della loro mancanza di generalità, Ying 2019.
40. Shanahan–Crosby–Beyret–Cheke 2020.
41. Su tutti si veda l'esperimento della stanza cinese di John Searle, nel quale egli, emulando il meccanismo di black box dell'IA, immagina di trovarsi in una stanza chiusa con un dizionario di cinese-inglese, dalla porta gli vengono passati dei biglietti scritti in cinese che lui raccoglie e traduce con il dizionario per comprenderli e

risultano difficilmente spiegabili, in quanto non fondati su catene di ragionamento causale assimilabili a quelle proprie del pensiero umano, ma sulla correlazione statistica tra variabili. In questo senso, affermare che in una fotografia sia rappresentato un lupo piuttosto che un husky in ragione della maggiore presenza di neve, ossia di una più elevata concentrazione di pixel bianchi, equivale a fornire una motivazione meramente descrittiva e statisticamente orientata, insufficiente sul piano logico e priva di reale valore esplicativo.

Questo fenomeno, ovvero quello dell'incidenza della probabilità sull'assunzione di un fatto come motivazione per spiegare una data circostanza, è stato affrontato a suo tempo dalle Sezioni Unite della Corte di Cassazione nella famosa sentenza Franzese, chiamata a pronunciare un principio di diritto su una questione molto simile avente ad oggetto la base motivazionale per giudicare una causa come penalmente rilevante ai fini della condanna. Il tema, lungi dall'essere desueto, è fondamentalmente connesso alla fattispecie dell'intelligenza artificiale e alla natura giuridica dei suoi output, così come al valore giuridico che l'ordinamento è chiamato a dare di essi. Uno dei passaggi chiave della sentenza Franzese ha infatti ad oggetto "i criteri di determinazione e di apprezzamento del valore probabilistico della spiegazione causale", e dunque la legittimità dell'uso di tecniche valutative probabilistiche in tema di pronunce giudiziali.

Già infatti nel caso COMPAS[42], algoritmo di predizione della recidiva, è venuto in essere il limite che la predizione probabilistica può avere nella spiegazione causale degli eventi. Nel caso di specie, usando lo stesso criterio probabilistico di analisi quantitativa dei dati del caso *husky vs wolf*, l'algoritmo COMPAS dava in output maggiori livelli di rischio di recidiva per le persone di colore solo sulla base di un dato statistico, ovvero che la maggior parte delle persone di colore registrate nel database erano poi risultate recidive[43]. In questa situazione ci si trova di fronte a un dato statisticamente corretto ma privo di efficacia giustificativa sotto il piano causale, questo perché l'essere una persona di colore non è causa sufficiente per la giustificazione di una previsione di recidiva. Sotto un profilo causale una spiegazione efficace potrebbe alternativamente essere la presenza di indici, come potenziali motivazioni del reo, che lo possano spingere a commettere un illecito nuovamente, ma a ben vedere quest'indice causale può essere confermato con certezza solo *ex post*, a fatto compiuto.

Su questo punto di intersezione tra XAI e giurisprudenza di legittimità si innesta la questione sulla natura giuridica dell'output dell'intelligenza artificiale. Mentre il settore della XAI ad oggi approfondisce il tema della distinzione tra causazione e spiegazione[44], pure fondamentale per lo sviluppo di sistemi IA più intelligenti, questa stessa questione è già stata parzialmente risolta dalle Sezioni Unite affermando come il calcolo statistico e la probabilità non siano sufficienti per dare una spiegazione sulla causazione degli eventi e che, nel giudizio, essi non possono soppiantare il necessario accertamento processuale del decorso causale nel caso concreto[45]. La conseguenza per il giudice

formulare risposte che poi ritradurrà e ripasserà da sotto la porta, corrispondendo così output corretti e coerenti senza l'effettiva comprensione della lingua cinese. Questo esperimento mentale ha dimostrato come l'IA abbia a disposizione una forma di comprensione solamente sintattica alla quale manca la comprensione semantica della realtà, che è necessaria per poter attribuire a qualunque entità forme di soggettività, Searle 1980.

42. *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS), noto software americano usato per predire la possibilità di recidiva di un imputato, al centro del caso Eric Loomis (*State v Loomis*, 881 NW2d 749, 754 – Supreme Court of Wisconsin), nel quale è stato dimostrato come il software sistematicamente desse in output percentuali maggiori di potenziali recidivi per gli imputati di colore solo sulla base della rilevanza statistica del dato sulla loro carnagione. Vedi Carnat 2024.

43. Lo studio condotto da ProPublica è stato presentato in Larson–Mattu–Kirchner–Angwin, *How We Analyzed the COMPAS Recidivism Algorithm,* 2016 e Angwin–Larson–Mattu–Kirchner, *Machine Bias*, 2016 in propublica.org.

44. Si veda Carloni–Berti–Colantonio 2023.

45. Cass., Sez. Unite Penali, n. 30328 del 10 luglio 2002, *Franzese*, da leggere secondo l'interpretazione data da Viganò 2013: "la mancanza di una legge di copertura con coefficiente statistico pari o prossimo al 100%

è espressa dal principio di diritto che recita "non è consentito dedurre automaticamente dal coefficiente di probabilità espresso dalla legge statistica la conferma, o meno, dell'ipotesi accusatoria sull'esistenza del nesso causale"[46].

In questo senso, tale principio di diritto pare applicabile per estensione analogica anche alla fattispecie del valore giuridico dell'output dell'intelligenza artificiale. Essendo lo stesso solo frutto di un mero calcolo statistico, come nei casi portati ad esempio, esso non costituirà mai il prodotto di un'analisi causale basata su una valutazione *ex post* dei dati rilevanti, ma costituirà tutt'al più una forma di applicazione *ex ante* della legge statistica generale al caso concreto e pertanto suscettibile di errori e inesattezze logiche[47]. Pertanto, alla luce della chiave interpretativa offerta dalla sentenza Franzese, l'output dell'intelligenza artificiale potrebbe al più essere considerato uno strumento di supporto, utile a rappresentare leggi di tipo statistico o scientifico-universale. In tale prospettiva, esso può fungere da elemento ausiliario nel processo decisionale, ma non potrà mai assurgere a criterio autonomo e risolutivo, giacché la decisione finale dovrà sempre fondarsi sul libero apprezzamento del giudice in relazione alle specifiche circostanze del caso concreto.

Tutto ciò, traslato in materia di valore giuridico dell'output dell'intelligenza artificiale, sembra suggerire che non solo l'output dell'IA non potrà avere sufficiente valore spiegazionale per essere posto a fondamento del giudizio, come già anticipato dalla sentenza Franzese a suo tempo e confermato dal caso Loomis, ma allo stesso tempo, ad esso non dovrà neanche essere attribuita una sufficiente facoltà giuridica per poter costituire autonomi rapporti di diritto se sciolto dal rapporto di dipendenza che esso ha rispetto all'agire umano, perché inidoneo a costituire un rapporto di imputazione tra esso e l'ente che lo produce. In questo quadro risulta essere pienamente condivisibile la pronuncia del Consiglio di Stato in tema di inadeguatezza della decisione algoritmica a costituire autonomi rapporti giuridici quando sciolta da un vincolo di attribuzione rispetto ad una decisione umana.

5. Il ruolo dei sistemi d'intelligenza artificiale nell'ordinamento giuridico

Quest'ultimo punto della riflessione apre un inedito scenario sul ruolo che i sistemi d'intelligenza artificiale andranno a ricoprire nei nostri ordinamenti nel contesto recentemente inaugurato dall'Artificial Intelligence Act. Essi, alcuni già ora, verranno senz'altro integrati in molte delle professioni attualmente svolte nella società. Già sono noti i casi sia cinesi[48] che americani[49] di integrazione di IA nel settore giudiziario, ai quali possono essere accompagnati i casi dell'introduzione di questi sistemi in campo medico, artistico e amministrativo. La naturale conseguenza di quest'introduzione dell'IA nei nostri ordinamenti sarà la rilevanza giuridica dei suoi output, che andranno a influire sui rapporti tra consociati.

Per questa motivazione e in continuità con le argomentazioni dei paragrafi precedenti, la soluzione proposta in dottrina sulla possibilità di attribuire forme di personalità giuridica a questi sistemi[50] non pare essere percorribile. Infatti, a forme di personalità giuridica si accompagnano anche conseguenti facoltà come la possibilità di costituire, modificare o far terminare rapporti giuridici, e con esse anche la possibilità di esserne titolari e di essere potenziali soggetti imputabili per azioni legali connesse a questi mutamenti. Non solo gli attuali sistemi di IA non sono suscettibili di essere considerati soggetti

non può impedire l'ascrizione causale di un evento a una condotta, quando l'evento – a una considerazione ex post – non è ragionevolmente spiegabile se non come conseguenza di quella condotta".

46. *Ibidem*.

47. Fenomeno questo alla base delle "allucinazioni" dell'intelligenza artificiale, ovvero di risposte prive di senso logico. Per un'ulteriore riflessione si veda il già citato caso degli *stochastic parrots* Bender–Gebru–McMillan-Major–Shmitchell 2021.

48. Si veda Liu-Liu-Lin 2024.

49. Si veda il già citato caso COMPAS, *State vs Loomis,* 881 NW2d 749, 754 (Win 2016) (Supreme Court of Wisconsin).

50. Si vedano Hallevy 2010; Solum 2020.

agenti[51], sia per la loro mancanza di raziocinio[52] che per la loro natura sistemica, ma come si è visto i loro output sono ad oggi privi di spiegabilità e dunque di base motivazionale che possa fungere contestualmente da base giuridica per rendere tali sistemi imputabili. La conseguenza di questo ragionamento è che un sistema d'intelligenza artificiale sganciato dal rapporto di intrinseca dipendenza che esso ha con l'umano è privo di qualunque valore giuridico, il quale dunque deriva solamente dal quantum che l'umano decide di trasferire in questo sistema.

In questo quadro, l'attribuzione di forme di personalità giuridica a sistemi di intelligenza artificiale potrebbe legittimare forme di responsabilità oggettiva, fondate sull'accertamento causale degli eventi sulla base di considerazioni statistiche *ex ante* piuttosto che considerazioni causali *ex post*, così di fatto abbassando il livello di tutela costituzionale garantito alla tutela dei diritti della persona umana[53] e andando contro quanto statuito sul tema nel nostro ordinamento dalle sentenze gemelle della Corte Costituzionale (n. 348 e 349/2007) e dalla Corte di Cassazione con la sentenza Franzese[54]. Pertanto, la necessità che si pone oggi sotto il profilo giuridico è quella di legare l'output dell'intelligenza artificiale all'azione umana che l'ha causato, ancorando così i profili di responsabilità per gli output prodotti da questi sistemi al principio di personalità della responsabilità giuridica. Questa soluzione, sebbene attualmente adottabile, richiede un ulteriore approfondimento attraverso l'analisi di alcune questioni giuridiche già emerse nella prassi.

5.1. Alcuni esempi

Per dare maggiore concretezza a questa argomentazione, risulta utile esaminare alcuni esempi tratti dalla letteratura che illustrano l'impiego di sistemi IA già esistenti e attualmente utilizzati in diverse professioni. L'analisi di questi esempi potrà facilitare la comprensione del ruolo che questi sistemi potrebbero assumere nel quadro regolamentare delineato dall'AI Act.

I casi selezionati sono tre: proprietà intellettuale, uso d'IA nel sistema giuridico (giudiziario e pubblica amministrazione) e in campo medico. Gli ultimi due sono stati selezionati per affrontare contestualmente il tema della responsabilità, fattispecie di rilievo il cui futuro ancora oggi è incerto in seguito al ritiro dell'AI Liability Directive da parte della Commissione europea[55], che segna

51. Sul punto si vedano i contributi di Floridi 2023 e Floridi 2025, laddove la stessa definizione proposta di IA come *Agere sine Intelligere* risulta problematica sotto molteplici profili giuridici, tra i quali la responsabilità, la soggettività e la causalità. In tal senso va considerata la difficoltà nell'imputare gli output dell'IA ad una sua stessa facoltà agente, dovuta all'assenza di una loro spiegabilità intrinseca. Per alcune voci critiche si vedano Roli–Jaeger–Kauffman 2022; Nyholm 2018.

52. Si veda ancora una volta sul punto il noto argomento della stanza cinese proposto in Searle 1980 per dimostrare come il nodo principale per la riproduzione di una forma d'intelligenza sia la comprensione di quelle funzioni soggettive, come la comprensione semantica del linguaggio, che danno origine all'identità. Al quale può essere accompagnata la già richiamata rappresentazione di IA come *stochastic parrot* presente in Bender–Gebru–McMillan-Major–Shmitchell 2021.

53. Si veda sul punto Simoncini 2019, che lucidamente fa notare come "Vi è un effetto sulla sfera della libertà giuridica ben più rilevante e profondo delle possibili applicazioni tecnologiche; esso, in realtà, va ad intaccare uno dei fattori essenziali del fenomeno giuridico in quanto tale, ovverosia l'idea stessa dell'esistenza di un nesso di causalità tra gli eventi e, conseguentemente, la distinzione 'mezzo-fine' ovvero 'agente-strumento' nella definizione dei comportamenti e delle relazioni" con una possibile conseguente riduzione delle tutela costituzionalmente garantita derivante dalla scelta di far prendere agli algoritmi di IA decisioni giuridicamente rilevanti.

54. Sentenze che sebbene distinte sul piano dell'oggetto della pronuncia, le prime in tema di principio di colpevolezza e la seconda in tema di accertamento del nesso causale, possono essere interpretate alla luce di un unico criterio, ovvero quello della chiara attribuzione di responsabilità individuale per gli illeciti penali, rigettando formule vaghe idonee ad annacquare la responsabilità penale dell'individuo.

55. Commissione europea, *Proposta di direttiva del Parlamento europeo e del Consiglio sull'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale* (AI Liability Directive),

un rilevante vuoto legislativo europeo in tema di responsabilità per gli output prodotti dall'IA. Ad ogni modo si partirà trattando il tema della proprietà intellettuale, forse il migliore per affrontare per prima la questione della natura giuridica degli output dell'intelligenza artificiale.

5.1.1. Proprietà intellettuale

Con l'introduzione dell'intelligenza artificiale generativa, capace di creare contenuti innovativi partendo dai dati in input combinati con quelli del proprio dataset, è venuta in essere anche la relativa questione sulla protezione del diritto d'autore e sulla possibilità di riconoscerlo ai sistemi di IA. Sotto quest'ultimo profilo ha fatto molto discutere a livello internazionale il caso DABUS, discusso dal Tribunale federale australiano[56], il quale, in una decisione sui generis, ha deciso di riconoscere al sistema IA denominato Dabus (*Device for the Autonomous Bootstrapping of Unified Sentience*) il diritto di possedere dei brevetti su due invenzioni da esso generate, riconoscendogli così di fatto anche la qualifica di inventore. Tale decisione è stata poi ribaltata dalla Full Federal Court che ha stabilito come il diritto d'autore e la brevettabilità di un'invenzione siano riconoscibili solo alle persone fisiche, escludendo così di fatto la possibilità di accordare a sistemi IA qualunque diritto di paternità per l'opera computazionale[57]. La questione ha fatto molto discutere[58], anche in relazione al ruolo che oggi hanno i sistemi di intelligenza artificiale generativa nella creazione di contenuti. Infatti è ancora poco chiaro quale sia lo statuto giuridico dell'opera realizzata con lo strumento informatico, proprio per via del *quid pluris* che l'IA vi aggiunge rispetto al contributo umano.

In questo quadro sicuramente sembra essere corretta la critica di forme di attribuzione di diritto d'autore alle macchine posto che, come già fatto notare[59], il lavoro svolto dal sistema IA sulla generazione di un'opera computazionale è sempre eterodiretto dall'essere umano che programma il sistema e definisce in input i suoi obiettivi. Pertanto, all'output dell'IA non può essere riconosciuto il carattere della *creatività*, necessario al fine del riconoscimento del diritto d'autore, perché l'opera prodotta dall'IA non sarà altro che una forma di riproduzione meccanica di *pattern* ricorrenti che l'IA ha imparato a riconoscere nei dati sui quali essa è stata addestrata[60]. In questo senso, il valore giuridico del suo output può invece venire in essere in relazione al riconoscimento del diritto d'autore alla persona fisica che l'ha prodotto avvalendosi dell'IA come strumento. Confermando il giudizio espresso precedentemente sul valore giuridico da attribuire all'output dell'intelligenza artificiale: ovvero come intrinsecamente legato ad un rapporto di dipendenza con l'agire umano.

5.1.2. L'uso dell'IA nel sistema giudiziario

Nel caso di utilizzo dell'intelligenza artificiale nell'ambito dell'ordinamento giuridico il rapporto tra l'output dell'IA e l'umano si capovolge. Mentre nel caso del diritto d'autore l'IA è uno strumento nelle mani umane, mediante il quale l'umano produce un output, nel caso del sistema giudiziario l'output dell'IA esercita una notevole influenza sul giudice umano o sul destinatario del provvedimento amministrativo automatizzato nel caso di uso di IA nella pubblica amministrazione.

Come già visto in relazione al caso COMPAS, la questione attiene al tipo di valore giuridico che

doc. COM(2022) 496, che successivamente alla pubblicazione del report di Hacker 2024, era entrata in una fase di stallo a causa dell'incertezza se propendere per un sistema di responsabilità basato sulla colpa o sulla responsabilità oggettiva, infine definitivamente ritirata dalla Commissione europea. Fattispecie che invece è regolamentata nell'attuale versione della proposta legislativa brasiliana (PL 2338-2023) agli articoli 35-39.

56. Federal Court of Australia, *Thaler v Commissioner of Patents* [2021] FCA 879.

57. Matulionyte 2022.

58. *Ibidem*, per una puntuale critica della decisione della corte federale australiana.

59. *Ibidem*.

60. Per un orientamento opposto si veda Deshpande–Kamath 2020 che, analizzando il caso DABUS nel contesto della legislazione indiana, afferma come la sua opera possa essere brevettabile, essendo assente nel Patents Act indiano del 1970 ogni riferimento alla persona naturale, con invece ampi margini di interpretabilità del termine "*person*", arrivando ad includere anche il Governo ed i suoi organi funzionali.

bisogna riconoscere all'output dell'IA nell'eventualità in cui esso concorra a determinare una decisione umana. Punto di partenza per questa riflessione è che di fatto, la sola presenza di un output dell'IA come elemento di valutazione per il giudice rende lo stesso output idoneo ad avere un valore giuridico autoritativo sull'interpretazione che il giudice darà del caso[61]. Esso contribuirà dunque a definire la prospettiva attraverso la quale il giudice interpreterà i fatti, concorrendo insieme agli altri elementi di prova a definire il caso stesso. Per questa motivazione, come fatto notare da Simoncini, non è sufficiente che la legge richieda, come nel caso di COMPAS, che la valutazione giudiziale si basi anche su ulteriori elementi valutativi e non integralmente sull'output dell'intelligenza artificiale. Questo perché "chi ha preso la decisione, infatti, avrà sempre la possibilità di sostenere che nella decisione si è solo avvalso dell'algoritmo, ma non è stato 'sostituito' da esso"[62]. Ci si trova qui di fronte ad un contesto nel quale, una volta introdotto l'output dell'IA all'interno dell'ordinamento, nel caso di specie all'interno della valutazione giudiziale, esso avrà a prescindere un valore giuridico di fatto, indipendentemente dall'uso che se ne farà. In questo quadro è fondamentale l'applicazione delle regole di design *ex ante* che l'AI Act impone per la progettazione di sistemi d'intelligenza artificiale ad alto rischio per il loro possibile impatto sociale, al fine di prevenire possibili errori di questi sistemi[63]. Ad esse in ogni caso appare opportuno affiancare oneri di motivazione rafforzati per il giudice che si avvalga di output IA nella valutazione della prova o nel giudizio[64]. Quest'ultima esigenza appare necessaria per circoscrivere razionalmente l'ambito d'applicazione dell'IA di fronte alla contestuale possibilità che i suoi output vadano ad influenzare in maniera sproporzionata la valutazione giudiziale.

Accanto a questa questione, in rapporto di dipendenza si pone la questione già affrontata dalla giurisprudenza nostrana della legittimità della decisione automatizzata, in buona parte risolta dalla sentenza del Consiglio di Stato n. 2270/2019, con la quale è stata affermata l'inidoneità di una decisione puramente automatizzata a definire rapporti giuridici tra i consociati, fattispecie già normata dall'art. 22. co.1 GDPR. Senza dilungarsi troppo sul punto[65], parrebbe opportuna l'introduzione formale di un obbligo di decisione umana per tutte quelle questioni sostanziali, o *semantiche*, idonee a influire sui diritti fondamentali dell'individuo, lasciando così aperta la possibilità di utilizzare algoritmi di IA per la risoluzione di questioni procedurali, o *sintattiche*[66], puramente computazionali, le quali sono anch'esse parte del diritto,

61. Argomento già discusso in Simoncini 2019 e Carnat 2024. In quest'ultimo contributo va segnalata la riflessione sull'*automation bias*, oggetto dell'art. 14.4(b) dell'AI Act, ovvero la possibile tendenza ad affidarsi automaticamente o a fare eccessivo affidamento sui risultati prodotti da un sistema di IA, fenomeno che contribuisce a ridurre il margine di indipendenza della decisione umana quando coadiuvata da un sistema IA.
62. Simoncini 2019 p. 81.
63. Si veda il Capitolo 3, sezioni 2 e 3 dell'AI Act (Regolamento (UE) 2024/1689) che stabilisce requisiti e obbligazioni per i provider di sistemi d'intelligenza artificiale ad alto rischio. Accanto ad esso si pone il considerando 61 che stabilisce esplicitamente come "L'utilizzo di strumenti di IA può fornire sostegno al potere decisionale dei giudici o all'indipendenza del potere giudiziario, ma non dovrebbe sostituirlo: il processo decisionale finale deve rimanere un'attività a guida umana", ripreso anche dall'art. 15 della legge 132/2025 che recita "nei casi di impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti".
64. Ampliando un poco l'orizzonte, si veda sul tema dell'incidenza degli strumenti informatici sulla discrezionalità giudiziale Borruso 2002.
65. Già oggetto di cospicua dottrina, si vedano ancora una volta Simoncini 2019 e la giurisprudenza nazionale del TAR Lazio sezione III bis nn. 9224-9230 del 2018 e 12026 del 2016.
66. Distinzione che riprende quanto statuito da Searle 1980 che usa i due termini per definire due differenti tipologie di operazioni: le operazioni sintattiche, come il calcolo, le quali non richiedono necessariamente una effettiva comprensione dell'operazione purché essa venga portata a compimento con successo e le operazioni

soprattutto procedurale ed amministrativo, e per le quali non è necessaria la partecipazione decisionale umana[67].

La *ratio* di questa norma andrebbe rinvenuta nell'impossibilità di risalire a logiche coerenti nei sistemi d'intelligenza artificiale, i quali, fondandosi solamente su procedimenti computazionali sintattici, difficilmente potranno pervenire a logiche decisionali effettivamente spiegabili se impiegati per effettuare decisioni giuridicamente rilevanti. Pertanto, andando oltre il già noto divieto di decisione automatizzata per la determinazione di questioni attinenti ai diritti fondamentali[68], la non spiegabilità delle decisioni automatizzate potrebbe rendere insufficienti le forme di tutela come l'intervento umano[69] o il diritto alla spiegazione delle decisioni algoritmiche[70], ricadendo così in fattispecie di incostituzionalità *ab origine* dell'algoritmo di IA impiegato per processi decisionali per la sua mancanza di trasparenza, già richiamate da Simoncini[71]. Per questa ragione l'intervento legislativo sul punto è auspicabile, al fine di definire un'area di operatività per i sistemi di IA limitata alle sole funzioni sintattiche.

5.1.3. L'uso dell'IA in campo medico

Infine viene in rilievo l'uso di sistemi IA in campo medico, dove numerosi profili dubbi di responsabilità, già in parte sollevati e analizzati dalla giurisprudenza Franzese, sorgono in relazione all'impatto che l'uso di questi sistemi può avere per la responsabilità professionale del medico. Anche qui, come nel caso dell'uso dell'IA nel giudiziario, ci si trova di fronte ad una fattispecie nella quale, una volta introdotto l'output dell'IA come elemento di valutazione del caso, esso inevitabilmente andrà a concorrere con gli altri strumenti d'analisi a disposizione del medico a formare la diagnosi finale. Però, differentemente rispetto al caso dell'uso dell'IA nel giudiziario, verrà qui in rilievo non soltanto l'output dell'IA come elemento valutativo per la diagnosi, ma anche l'uso dell'IA stessa come strumento diagnostico per l'analisi del caso. Risultando così in una sfera di responsabilità ampliata per il medico moderno che dovrà rispondere della valutazione fatta dell'output dell'IA e dell'uso fatto del sistema per la produzione di quel determinato output.

In questo senso, va preliminarmente menzionata la precedenza dell'onere di ripartizione delle responsabilità lungo la catena causale che porta alla generazione dell'output dell'IA, al fine di alleggerire il carico del medico operante. L'AI Act provvede a questo con obbligazioni da porre a carico dello sviluppatore del sistema d'IA (*provider*) e del gestore del sistema (*deployer*)[72], al fine di garantire che il sistema IA, una volta a disposizione del medico, produca output affidabili e con pochi margini di errore e che nel caso in cui invece accada l'opposto, a risponderne non sia il medico ma lo sviluppatore o il gestore del sistema. In questo quadro già si possono identificare tre differenti livelli di causazione, in parte applicabili anche al caso

semantiche, le quali di contro richiedono necessariamente una parte motivazionale di comprensione affinché il risultato sia valido, proprio come lo richiedono le decisioni giudiziali.

67. Si veda sul punto quanto già detto in nota 23, dove viene richiamato il contributo di Borruso 1995, il quale con lungimiranza già intravedeva il possibile campo d'applicazione degli algoritmi nel diritto.
68. Si segnalano inoltre le sentenze della Corte Costituzionale n. 1/1971, n. 139/1982 e n. 367/2004, in controtendenza rispetto all'ordinamento cinese nel quale invece si impiegano sperimentalmente algoritmi di IA per verificare la possibilità di razionalizzare e rendere più efficiente il procedimento decisionale giudiziario, si veda Liu–Liu–Lin 2024.
69. Si veda sul punto Almada 2019.
70. Codificato dall'art. 86 dell'AI Act recante norme sul "diritto alla spiegazione dei singoli processi decisionali".
71. Simoncini 2019.
72. Entrambe queste figure sono definite dall'art. 3(3-4) dell'AI Act e si riferiscono a due differenti momenti della catena di sviluppo dell'IA. Infatti il *provider* sarà il responsabile dello sviluppo del modello d'IA, ovvero del modello di risposta, mentre il *deployer* sarà colui che svilupperà e gestirà l'applicativo dell'IA, il quale conterrà al suo interno il modello adattato alle esigenze particolari per il quale viene successivamente ottimizzato. Le due figure potranno ovviamente coincidere. Per una differenza tra *AI model* e *AI system* nel quadro dell'AI Act si veda Fernández-Llorca–Gómez–Sánchez–Mazzini 2024.

dell'uso dell'IA nel giudiziario, idonei a costituire una prima teorica catena di responsabilità:

1) Responsabilità dello sviluppatore del modello IA (*provider*) per il training dei dati e l'aggiornamento del modello.
2) Responsabilità del gestore del sistema IA (*deployer*) per la gestione dell'applicativo.
3) Responsabilità del medico decisore per la decisione finale.

Una giustificazione per questa preliminare classificazione dei momenti causali rilevanti nella catena di responsabilità per la decisione medica presa con l'ausilio dell'IA la si rinviene in quanto affermato sul punto da Naik *et al.* "it is the underlying data than the algorithm itself that is to be held responsible. Models can be trained on data which contains human decisions or on data that reflects the second-order effects of social or historical inequities"[73]. Mentre l'AI Act copre le fattispecie di responsabilità per i primi due livelli della catena causale, l'ultimo, quello della responsabilità del decisore umano, rimane fuori dal paradigma europeo e viene delegato alla legislazione nazionale[74].

Su quest'ultimo caso, la giurisprudenza Franzese fornisce un'importante chiave interpretativa. Lo scenario tipico è quello di una decisione medica influenzata dall'output dell'IA o addirittura messa in discussione da esso: ad esempio, nel caso in cui una diagnosi errata porti a conseguenze infauste per il paziente e, in retrospettiva, l'IA abbia suggerito un decorso alternativo più favorevole. In questo quadro, messe da parte le questioni sulla correttezza dell'output concernenti i primi due livelli di responsabilità, e quindi assumendo l'esistenza di un output corretto, la questione che viene in rilievo è se l'output dell'IA rappresentante un decorso alternativo favorevole possa legittimare una pronuncia di condanna del medico. Ebbene, la risposta che l'interpretazione della sentenza Franzese ci offre pare essere di segno negativo. Questo perché l'output dell'IA, essendo il prodotto di un calcolo probabilistico, rappresenterà tutt'al più "il coefficiente di probabilità espresso dalla legge statistica" e giammai un reale decorso alternativo degli eventi. Questo aspetto duplice della causalità è riconducibile alla distinzione tra causalità scientifica e causalità giuridica per la quale, se la prima considera la validità di una legge in astratto, la seconda valuta la relazione causale nel caso concreto[75]. Per comprendere la rilevanza di questa distinzione, si può richiamare l'esempio formulato da Viganò[76]: se Tizio avvelena Caio con un veleno che causa la morte nel 100% dei casi, ma prima che il veleno faccia effetto Sempronio gli spara uccidendolo, sarà Sempronio ad aver cagionato la morte, non Tizio, nonostante il veleno fosse letale. Questo perché, ai fini giuridici, la *condicio sine qua non* è rappresentata dall'azione che ha effettivamente determinato l'evento nel

73. Naik–Hameed–Shetty et al. 2022.

74. Per un quadro completo del funzionamento dell'impianto obbligatorio dell'AI Act in campo medico si veda van Kolfschooten–van Oirschot 2024, che inquadra con chiarezza il rapporto tra i tre livelli di responsabilità comprendenti *provider*, *deployer* e decisore finale. Inoltre ad oggi è ancora dubbio se le norme previste dall'AI Liability Directive, comprendenti l'obbligo di divulgazione delle prove (art. 3) e la presunzione di causalità (art. 4), entreranno effettivamente a far parte degli ordinamenti nazionali al fine di coprire anche il terzo livello di responsabilità.

75. Sul tema si veda Blaiotta 2010 che affronta in maniera sistematica il tema della causalità giuridica, comparandola alla causalità scientifica e segnalando che, mentre nella causalità scientifica la nozione di legge scientifica faccia riferimento alla regolarità con la quale un fenomeno si presenta, nel diritto la causalità che rileva ai fini del giudizio è quella che nel caso concreto ha dato origine alla fattispecie oggetto di censura.

76. Viganò 2013: "Riprendendo uno dei tanti esempi di scuola discussi in letteratura, pensiamo a una ipotetica legge scientifica che asserisca che il morso del serpente mocassino è sempre letale per la vittima, la quale invariabilmente decede qualche ora dopo il morso, senza alcuna possibilità di essere salvata mediante la somministrazione di un siero o di altro presidio terapeutico. Se – prima che il veleno del serpente, introdotto furtivamente dà A nel letto di B, abbia sortito il proprio effetto letale – B viene attinto da un colpo di pistola sparatogli da C, è evidente che la morte di B dovrà essere attribuita, in base a una considerazione ex post, a un decorso causale diverso rispetto al morso del serpente (e dunque alla condotta di A), malgrado la natura universale della legge scientifica che assicura, in una prospettiva generalizzante ed ex ante, che tutti coloro che vengono morsi dal serpente sono destinati a sicura morte".

caso concreto, e non da quella che in astratto ha il maggior coefficiente di probabilità nella causazione dell'evento.

Poiché l'output dell'IA rappresenta un'elaborazione statistica *ex ante*, puramente teorica e ipotetica, senza prendere in considerazione il reale decorso causale degli eventi accertato *ex post*, non può essere utilizzato come base per imputare la responsabilità al medico. Legittimare contestazioni della decisione medica sulla base di valutazioni alternative offerte da sistemi IA comporterebbe un'ingiusta espansione della sfera di responsabilità del medico, attribuendogli un obbligo di conformità a una previsione statistica che, per sua natura, non è equivalente a un'analisi fattuale e causale degli eventi effettivamente accaduti. Certamente la sola presenza di una previsione diagnostica da parte dell'IA contribuirà non poco allo sviluppo causale degli eventi e influirà, anche se tacitamente, sulla decisione medica. Per questo motivo sarebbe opportuna la previsione di obblighi di motivazione a presidio della libertà del medico che, se rispettati, possano scagionarlo nel caso di esito infausto quando non ha seguito la raccomandazione potenzialmente positiva dell'IA. In ogni caso il giudizio di responsabilità va condotto sulla ragionevolezza della motivazione fornita dal medico e sull'influenza della sua decisione nel caso concreto, prendendo in considerazione la raccomandazione dell'IA solo in relazione ad essa e non in relazione al potenziale decorso causale alternativo, puramente teorico e incerto.

Contemporaneamente sul punto è bene anche ribadire quanto già affermato in tema di responsabilità del giudice che si avvalga di output di IA per la valutazione giudiziale, ovvero che la raccomandazione dell'IA non può essere considerata come vincolante né come unica *ratio* per il giudizio di responsabilità[77]. Questo al fine di garantire entrambi i casi, quello positivo di tutela del medico diligente che abbia commesso un errore in buona fede, così come quello negativo del medico negligente che abbia basato la sua diagnosi unicamente sull'output dell'IA.

6. Conclusioni

L'analisi condotta evidenzia come la questione dell'autonomia dell'intelligenza artificiale non possa essere affrontata esclusivamente in termini tecnici o definitori, ma impone una riflessione giuridica più ampia, centrata sulla natura e sul valore degli output generati da tali sistemi. Essi contribuiranno non poco alla determinazione in concreto di nuovi e autonomi rapporti di diritto, a prescindere da un loro riconoscimento formale. Come già visto nei casi dell'applicazione di IA in ambito medico e giuridico, ci si trova di fronte a contesti nei quali, una volta introdotto l'output dell'IA, esso inevitabilmente andrà a contribuire allo sviluppo causale degli eventi, influendo sul processo decisionale umano e sulla costituzione o modificazione dei rapporti socio-giuridici tra i soggetti coinvolti.

Questa è una realtà inevitabile che il progresso tecnologico oggi in atto comporta, dove l'introduzione di algoritmi autonomi senz'altro contribuirà a rendere ancora più complessi e intricati i rapporti sociali nei nostri ordinamenti. Ciononostante, esiste una strada per il mantenimento di forme di indipendenza decisionale senza che essa comporti un maggiore carico per l'essere umano. Tale transizione sembra inevitabilmente implicare una rinuncia, da parte dell'essere umano, al controllo diretto su una serie di operazioni che in passato richiedevano il suo intervento e che oggi sono affidate ad algoritmi dotati di autonomia operativa. A questa perdita di controllo non si può rispondere con maggiori responsabilità per l'essere umano, ma piuttosto, ciò che appare oggi opportuno è la sostituzione di antiquate forme di controllo con nuovi criteri d'imputazione della responsabilità basati sulla fiducia.

Il riconoscimento della natura probabilistica degli output dell'IA, prodotti da mere inferenze probabilistiche, prive di un contenuto esplicativo pienamente causale, conduce a escludere la possibilità di attribuire loro rilevanza giuridica autonoma se sciolta dal rapporto di dipendenza con l'agire umano. Per questa motivazione la responsabilità umana permane, anche se in una forma

77. È utile sul punto citare Rubel–Pham–Castro 2019, che con il termine *agency laundering* hanno ben inquadrato questa tendenza a voler "riciclare" le proprie azioni mediante l'uso di algoritmi di intelligenza artificiale, pratica che il diritto non può legittimare e che le teorie della responsabilità oggettiva e della personalità elettronica finirebbero col consentire.

diversa rispetto al controllo diretto sulla produzione dell'output. L'essere umano rimane responsabile per l'affidabilità dell'output prodotto dal sistema di IA. Questa forma di responsabilità sorge al fine di poter garantire tutti coloro che poi, nelle fasi conclusive della catena di responsabilità, potranno interagire con gli output prodotti dal sistema.

Questo nuovo paradigma di responsabilità, non basato su prevedibilità e controllo, ma piuttosto su un onere di affidabilità che, rivolgendosi alla società nel suo insieme, è idoneo a rendere imputabile l'individuo che ha contribuito alla creazione dell'output dell'IA, può aprire uno scenario inedito dove la sfera di responsabilità individuale risulta ampliata e al contempo alleggerita. L'ampliamento deriva dallo svincolarsi di queste forme di responsabilità dai criteri d'imputazione tradizionali di prevedibilità e controllo, per il quale un individuo può essere chiamato a rispondere di qualcosa anche dopo molto tempo e indipendentemente da un rapporto di causazione di tipo diretto. Contemporaneamente si verifica un alleggerimento del peso della responsabilità, laddove all'individuo essa non potrà essere addossata per l'output dell'IA con la formula della colpa per la performance negligente, oggettivamente compiuta da un altro attore, il sistema IA, ma piuttosto all'individuo che si avvale dell'output dell'IA potranno essere richiesti oneri giustificativi e di trasparenza idonei a spiegare il suo rapporto con esso e la sussistenza della sua buona fede.

Questa descrizione sintetica di un nuovo paradigma di responsabilità, capace di superare i confini della tradizionale imputazione colposa, rappresenta oggi solamente un accenno di soluzione. Ulteriori ricerche su questo fronte si rendono necessarie. Tuttavia, quanto emerso delinea una direzione inedita per l'evoluzione del diritto, sollecitata dall'avvento dell'intelligenza artificiale, dove, a fronte di un ampliamento della sfera di responsabilità, può accompagnarsi la realizzazione dell'antico sogno di un suo alleggerimento, grazie alla delega a sistemi d'intelligenza artificiale di numerose attività una volta compiute dall'essere umano.

Riferimenti bibliografici

M. Almada (2025), *What Is an Artificial Intelligence System, Really? Notes on the European Commission's Guidelines on Article 3 (1) of the AI Act*, in "Journal of AI Law and Regulation", vol. 2, n. 1, 2025

M. Almada (2019), *Human intervention in automated decision-making: Toward the construction of contestable systems*, in "Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law", 2019

Y. Bathaee (2018), *The artificial intelligence black box and the failure of intent and causation*, in "Harvard Journal of Law & Technology", vol. 31, 2018

L. Belli, Y. Curzi, W.B. Gaspar (2023), *AI regulation in Brazil: Advancements, flows, and need to learn from the data protection experience*, in "Computer Law and Security Review", vol. 48, 2023

E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell (2021), *On the dangers of stochastic parrots: Can language models be too big?*, in "Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency", 2021

R. Blaiotta (2010), *Causalità giuridica*, Giappichelli, 2010

E. Börger, E. Grädel, Y. Gurevich (2001), *The classical decision problem*, Springer, 2001

R. Borruso (2002), *Discrezionalità e autonomia del giudice: il contributo dell'informatica giuridica*, in "Il Diritto dell'informazione e dell'informatica", 2002, n. 2

R. Borruso (1995), *Il computer, il diritto romano e il giudice*, in "Informatica e diritto", 1995, n. 2

L. Brennan (2023), *AI ethical compliance is undecidable*, in "Hastings Science & Technology Law Journal", vol. 14, 2023, n. 2

S. Bubeck, V. Chandrasekaran, R. Eldan et al. (2023), *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, arXiv preprint, arXiv:2303.12712

G. Carloni, A. Berti, S. Colantonio (2023), *The role of causality in explainable artificial intelligence*, arXiv preprint, arXiv:2309.09901

I. Carnat (2024), *Addressing the risks of generative AI for the judiciary: The accountability framework(s) under the EU AI Act*, in "Computer Law & Security Review", vol. 55, 2024, 106067

R. Deshpande, K. Kamath (2020), *Patentability of inventions created by AI – the DABUS claims from an Indian perspective*, in "Journal of Intellectual Property Law & Practice", vol. 15, 2020, n. 11

D. Fernández-Llorca, E. Gómez, I. Sánchez, G. Mazzini (2024), *An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act*, in "Artificial Intelligence and Law", 2024

L. Floridi (2025), *AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis*, in "Philosphy and Technology" vol. 38, 2025

L. Floridi (2023), *AI as agency without intelligence: on ChatGPT, large language models, and other generative models*, in "Philosophy & Technology", vol. 36, 2023

A.S. Fokas (2023), *Can artificial intelligence reach human thought?*, in "PNAS Nexus", vol. 2, 2023, n. 12

R. Guidotti, A. Monreale, S. Ruggieri et al. (2018), *A survey of methods for explaining black box models*, in "ACM Computing Surveys (CSUR)", vol. 51, 2018, n. 5

P. Hacker (2024), *Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence: Complementary impact assessment*, Study EPRS – European Parliamentary Research Service, 2024

G. Hallevy (2010), *The criminal liability of artificial intelligence entities – from science fiction to legal social control*, in "Akron Intellectual Property Journal", vol. 4, 2010

High-Level Expert Group on Artificial Intelligence (2020), *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, European Commission, 2020

High-Level Expert Group on Artificial Intelligence (2019-a), *Ethics Guidelines for Trustworthy AI*, European Commission, 2019

High-Level Expert Group on Artificial Intelligence (2019-b), *A Definition of AI: Main Capabilities and Disciplines*, European Commission, 2019

Y. LeCun (2022), *A Path Towards Autonomous Machine Intelligence*, Version 0.9.2, in openreview.net, 2022

Z. Liu, S. Liu, Y. Lin (2024), *Building Smart Courts Through Large Legal Language Models? Experience from China*, in "AI from the Global Majority", 2024

R. Matulionyte (2022), *AI as an Inventor: Has the Federal Court of Australia Erred in DABUS?*, in "Journal of Intellectual Property, Information Technology and Electronic Commerce Law", vol. 13, 2022

N. Naik, B.M.Z. Hameed, D.K. Shetty et al. (2022), *Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?*, in "Frontiers in Surgery", vol. 9, 2022, 862322

S. Nyholm (2018), *Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci*, in "Science and Engineering Ethics", vol. 24, 2018, n. 4

OpenAI (2023), *GPT-4 Technical Report*, arXiv preprint, arXiv:2303.08774

M. Pietrangelo, L. Nannipieri (2023), *Intelligenza artificiale e pubbliche amministrazioni, tra incertezze definitorie, usi disomogenei e giudici supplenti*, in "Ital-IA 2023: 3rd National Conference on Artificial Intelligence", Pisa, 29-31 maggio 2023

M.T. Ribeiro, S. Singh, C. Guestrin (2016), *Why should I trust you?: Explaining the predictions of any classifier*, in "Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", ACM, 2016

A. Roli, J. Jaeger, S.A. Kauffman (2022), *How organisms come to know the world: Fundamental limits on artificial general intelligence*, in "Frontiers in Ecology and Evolution", vol. 9, 2022, 806283

A. Rubel, A. Pham, C. Castro (2019), *Agency Laundering and Algorithmic Decision Systems*, in "Information in Contemporary Society: Proceedings of the 14th International Conference, iConference 2019", Springer, 2019

S.J. Russell, P. Norvig (2010), *Artificial Intelligence: A Modern Approach*, Pearson Education, 2010

G. Sartor, A. Omicini (2016), *The autonomy of technological systems and responsibilities for their use*, in "Autonomous Weapon Systems. Law, Ethics, Policy", Cambridge University Press, 2016

J.R. Searle (1980), *Minds, brains, and programs*, in "Behavioral and Brain Sciences", vol. 3, 1980

M. Shanahan, M. Crosby, B. Beyret, L. Cheke (2020), *Artificial intelligence and the common sense of animals*, in "Trends in Cognitive Sciences", vol. 24, 2020, n. 11

A. Simoncini (2019), *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in "BioLaw Journal", 2019

L.B. Solum (2020), *Legal personhood for artificial intelligences*, in W. Wallach, P. Asaro (eds.), "Machine Ethics and Robot Ethics", Routledge, 2020

G. Taddei Elmi, S. Marchiafava (2024), *Sviluppi recenti in tema di Intelligenza Artificiale e diritto (gennaio-aprile 2024)*, in "Rivista italiana di informatica e diritto", 2024, n. 1

G. Taddei Elmi, S. Marchiafava (2023-a), *Sviluppi recenti in tema di Intelligenza Artificiale e diritto (novembre-dicembre 2023)*, in "Rivista italiana di informatica e diritto", 2023, n. 2

G. Taddei Elmi, S. Marchiafava (2023-b), *Sviluppi recenti in tema di Intelligenza Artificiale e diritto (giugno-agosto 2023)*, in "Rivista italiana di informatica e diritto", 2023, n. 2

G. Taddei Elmi, S. Marchiafava (2023-c), *Sviluppi recenti in tema di Intelligenza Artificiale e diritto (gennaio-aprile 2023)*, in "Rivista italiana di informatica e diritto", 2023, n. 1

G. Taddei Elmi, S. Marchiafava (2022), *Sviluppi recenti in tema di Intelligenza Artificiale e diritto: Una rassegna di legislazione, giurisprudenza e dottrina*, in "Rivista italiana di informatica e diritto", 2022, n. 2

A.M. Turing (1950), *Computing machinery and intelligence*, in "Mind", vol. 59, 1950, n. 236

A.M. Turing (1936), *On computable numbers, with an application to the entscheidungsproblem*, in "Journal of Math", vol. 58, 1936

H. van Kolfschooten, J. van Oirschot (2024), *The EU Artificial Intelligence Act (2024): implications for healthcare*, in "Health Policy", vol. 149, 2024, 105152

A. Vaswani, N. Shazeer, N. Parmar et al. (2017), *Attention is all you need*, in "Advances in Neural Information Processing Systems", vol. 30, 2017

F. Viganò (2013), *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, in "Diritto penale contemporaneo", vol. 2, 2013

X. Ying (2019), *An overview of overfitting and its solutions*, in "Journal of Physics: Conference Series", vol. 1168, 2019, 022022