

## **IA e diritto. In luogo di una introduzione**

*Agata C. Amato Mangiameli*<sup>1</sup>

Università degli Studi di Roma “Tor Vergata”

**Abstract:** The more scientific and technological development brings the machine closer to artificial thinking, if not even to thinking tout court, the more the processor is transformed from a calculator, capable of mere arithmetic operations, to a device capable of simulating the functioning of the brain and its learning neuronal functions. And in the meantime, as more and more refined devices are being built, the focus of which is, in particular, on those dimensions aimed at guiding decisions and inducing behavior, a massive debate is being activated on machine intelligence, on mental processes freed from the body, as well as on the need both to prepare appropriate regulations and to draw insurmountable ethical limits, without thereby hindering innovation. On the other hand, at the moment when the boundary between natural and artificial becomes rarefied to the point of evanescence, it is at that same moment that the human species gradually transforms, evolving (or perhaps involuting) into something manufactured (fabricated). Perhaps the interaction with the machine will not always give a similar feeling to that which one would have with something alive, it is certain, however, that AI will continue to develop, with all the implications and legal and ethical consequences that come with it. Hence, what future awaits human nature?

**Keywords:** *Artificial Intelligence, Neural Networks, Big Data, Robots, Chatbot*

---

<sup>1</sup> Professore Ordinario di Filosofia del diritto, Biogiuridica, Informatica giuridica presso il Dipartimento di Giurisprudenza dell'Università degli Studi di Roma “Tor Vergata”.

## Premessa

Quella dell'intelligenza artificiale – lemma che incarna il fulcro attorno al quale gravitano tutte le questioni legate all'informatica giuridica e, in generale, allo sviluppo tecnologico – è la storia di grandi intuizioni e di importanti riorganizzazioni, che può essere efficacemente ripercorsa a partire da tre scene, ognuna delle quali ci dice qualcosa di diverso sull'IA e, a suo modo, concorre a descriverne una differente tappa evolutiva.

### 1. Scena prima: la stanza cinese

1.1. In una stanza è rinchiusa una persona di madrelingua inglese, che non conosce affatto il cinese e che ha a disposizione due fogli: nel primo foglio sono rappresentati degli ideogrammi cinesi, nel secondo, invece, c'è un elenco di domande, formulate sempre in cinese. Nella stessa stanza si ritrova anche un libro, scritto questa volta in inglese, che spiega le regole da seguire per abbinare i simboli del primo foglio con le domande prospettate nel secondo. Ora, seppure non comprenda il cinese, il protagonista dell'esperimento, seguendo alla lettera le istruzioni, potrà riuscire a produrre risposte formalmente corrette, tanto da indurre un eventuale osservatore esterno di madrelingua cinese ad avere una falsa rappresentazione della realtà e a ritenere che il soggetto in questione abbia una buona padronanza della sua lingua.

Com'è noto, si tratta dell'esperimento che John Searle propose in *Minds, brains and programs*<sup>2</sup>, per spiegare la differenza che passa tra le due concezioni dell'Intelligenza artificiale: da un lato, quella *debole*, prospettata da coloro che nell'IA vedono un mero strumento (potente, ma nulla di più) in grado di verificare e di formulare le ipotesi in modo esatto e coerente; da un altro lato, quella *forte*, avanzata dai fautori dell'IA quale dispositivo che, se programmato opportunamente, sarebbe in grado di simulare in tutto e per tutto il funzionamento della mente umana e perciò assai diverso da un semplice strumento. Come è chiaro, l'immagine dell'uomo che meccanicamente produce output di risposte, in base alle istruzioni rinvenute in un libro e senza comprendere il cinese, a suo modo, rappresenta una metafora dell'elaboratore, che, nell'eseguire il programma secondo il linguaggio di programmazione (la sua madrelingua), non fa altro che manipolare simboli dei quali in realtà ignora completamente il significato, così che il suo operare è meramente sintattico. Non a caso, secondo Searle, l'IA – fondamentalmente diversa da quegli aspetti che sono propri solo della mente e che sono legati

---

<sup>2</sup> 1984.

alla comprensione di ciò che abbiamo intorno (livello semantico) – ha a che fare con i simboli, anziché con i significati, che restano appannaggio esclusivo dell'uomo, ragion per cui il suo scopo altro non sarebbe se non quello di manipolare correttamente le forme ignorandone i contenuti, o persino nonostante i loro contenuti.

Al di là delle differenze tra le due possibili concezioni qui ricordate (quella *debole* o quella *forte*), l'espressione IA – usata per la prima volta dal matematico John McCarthy<sup>3</sup> in occasione di un convegno volto a definire i termini di una nuova disciplina con oggetto la simulazione e la riproduzione di alcuni processi del cervello umano – risponde all'esigenza tipicamente umana di superare le tante difficoltà che si presentano nel quotidiano e le molteplici mancanze della nostra condizione, realizzando sistemi in grado di espletare con successo e con notevole rapidità alcune funzioni complesse umane. Si pensi, ad esempio, alla traduzione automatica dei testi: l'azione del software non può certo essere equiparata all'atto propriamente umano del comprendere, che, in quanto tale, richiederebbe coscienza e intenzionalità, oltre a un autentico approfondimento della realtà esterna e un'intensa capacità di discernimento tra le varie e possibili circostanze concrete.

Ed è entro questa cornice che, tra le principali applicazioni dell'IA, possono segnalarsi: 1) la rappresentazione della conoscenza (con le sue due più importanti metodologie: linguaggi formali e alberi di decisione); 2) l'apprendimento automatico, grazie a sistemi che si basano su osservazioni o su esempi come dati per la sintesi di una nuova conoscenza e che sono da distinguere in apprendimento non supervisionato (algoritmi: clustering, regole di associazione), supervisionato (algoritmi: albero di decisione, regole di decisione, apprendimento bayesiano, sistemi esperti), per rinforzo (algoritmi: reti neurali, algoritmi genetici, sistemi a classificatori); 3) la programmazione logica induttiva, ossia l'apprendimento automatico con l'uso della programmazione logica come linguaggio di rappresentazione di esempi e concetti; 4) l'elaborazione del linguaggio naturale, suddivisa in analisi lessicale, grammaticale, sintattica, semantica; 5) il c.d. visual retrieval, sistemi (con differenti modalità: semantica, formale, strutturale, coloristica, parametrica) per indicizzare, archiviare, ricercare, documenti visivi digitali; 6) la visione artificiale, l'insieme dei processi che permettono di acquisire, registrare ed elaborare immagini, per varie finalità (classificazione, controllo, selezione, ecc.); 7) la tecnica di risoluzione di problemi di pianificazione, allocazione delle risorse, e così via.

1.2. Dinanzi alle tante applicazioni e alle sempre nuove provocazioni dell'IA, l'Unione europea richiama, già da alcuni anni, a un più attento ap-

---

<sup>3</sup> MCCARTHY, MINSKY, ROCHESTER, SHANNON 2006.

proccio e a una più adeguata normativa<sup>4</sup>, sottolineando la necessità che etica e diritto muovano dalla fondamentale distinzione tra scienza e pseudo-scienza, tra fatti e opinioni, e, al contempo, diano ampio spazio a quella ricerca di base che ha quale primo obiettivo l'avanzamento della conoscenza e la comprensione teorica delle diverse variabili in un determinato processo.

Etica e diritto dovrebbero, cioè, adoperarsi sinergicamente – riprendendo Isabelle Stengers – per l'elaborazione di un *Manifeste pour un ralentissement des sciences*<sup>5</sup>. Distinguendo attentamente ciò che può essere ricondotto all'attività computazionale (e che dunque può essere programmato ed eseguito meccanicamente) e ciò che, invece, ne resta e ne resterà sempre e in ogni caso al di fuori (richiedendo delle scelte che coinvolgono complesse analisi e profonda riflessione). Ed è proprio in questo senso che, già diverso tempo fa, Weizenbaum<sup>6</sup> scriveva che non sarebbe stato bene delegare alla macchina tutte quelle funzioni (non computazionali e programmabili) che attenevano al giudizio, al rispetto, alla comprensione, alla cura, all'amore. Ed è proprio qui che, ancora oggi, l'IA incontra (*e deve continuare a incontrare!*) il proprio limite.

Etica e diritto non devono perdere di vista questa essenziale linea di confine che ha a che vedere con il distinguo fra fabbricato e creato, come pure fra riproducibile e unico. Un distinguo questo che, a suo modo, può essere assai efficacemente rappresentato dal richiamo a quell'evidente e incolmabile distanza che intercorre fra il ritratto di *Edmond de Belamy* – prodotto da

---

<sup>4</sup> Circa l'attenzione e il rilievo che l'Europa ha riconosciuto – e sta riconoscendo – all'Intelligenza Artificiale, merita d'esser qui ricordata la *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions* del 25 aprile 2018, recante *Artificial Intelligence for Europe*. Fra le finalità della Comunicazione quella di: 1) implementare gli investimenti nella ricerca e nell'innovazione; 2) preparare i cittadini europei ai mutamenti socioeconomici apportati dall'intelligenza artificiale; 3) predisporre e assicurare un quadro giuridico adeguato (ulteriori indicazioni in materia sono reperibili anche nella nota strategica del Centro di Strategia Politica della Commissione, *The Age of Artificial Intelligence* del 2018). Degna di nota è anche la *Proposal For a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence* (COM(2021) 206 final, del 21 aprile 2021), con la quale – nell'intento generale di costruire un sostrato legislativo omogeneo e uniforme in tutta l'Unione – si mira a: i) assicurare che i sistemi di IA immessi sul mercato dell'Unione e utilizzati siano sicuri e rispettino la normativa vigente in materia di diritti fondamentali e i valori dell'Unione; ii) garantire la certezza del diritto per facilitare gli investimenti e l'innovazione nell'intelligenza artificiale; iii) migliorare la governance e l'applicazione effettiva della normativa esistente in materia di diritti fondamentali e requisiti di sicurezza applicabili ai sistemi di IA; iv) facilitare lo sviluppo di un mercato unico per applicazioni di IA lecite, sicure e affidabili nonché prevenire la frammentazione del mercato. Il tutto, con l'ambizioso fine di elevare l'Unione europea a leader mondiale nello sviluppo di un'IA sicura, affidabile e, soprattutto, etica.

<sup>5</sup> 2013.

<sup>6</sup> 1976.

un software che, sulla base di migliaia di dipinti, ha appreso tante importanti regole – e *La ragazza col palloncino (L'amore è nel cestino)* creata dalla mano libera e geniale di Banksy. Opera originale e unica che, nel momento in cui è stata battuta all'asta, si è in parte autodistrutta!

## 2. Scena seconda: la stanza di Google

2.1. Qui si avvicinano figure professionali specializzate nella progettazione di software, il cui impegno è quello di superare via via i limiti dell'IA. Deep Blue (1997) calcola statisticamente la mossa migliore e vince a scacchi, ma resta pur sempre un elaboratore. Watson (2011) comprende il linguaggio umano e dà risposte esatte, ma resta pur sempre un elaboratore. Rispetto al primo, c'è tuttavia qualcosa in più: il programma utilizza le c.d. reti neurali che, analogamente a quelle biologiche che sono in grado di comprendere l'ambiente e i suoi mutamenti, si prestano ad essere usate in tutti quei settori in cui sono richiesti data mining, elaborazione di modelli predittivi e simulativi, oppure classificazioni. Ed è proprio sulle reti neurali che AlphaGo (2015) si sviluppa e sconfigge il maestro di un gioco, qual è il go, sino a quel momento fuori dalla portata delle tecniche di apprendimento automatico, e che richiede al contempo le doti umane della disciplina, della concentrazione e dell'equilibrio. E tra le tante versioni di AlphaGo, nella stanza di Google, lì dove tutto è accelerato e il tempo per adattarsi è decisamente esiguo, si processano senza sosta e senza particolare dispendio di tempo e di energie grandi mole di dati. Nel frattempo, l'IA inizia ad imparare da sola, apprende dai suoi stessi errori e opera in modo corretto nonostante input imprecisi o incompleti, si auto-aggiorna in presenza di modifiche ambientali e sembra possedere persino una sorta di intuito non molto diverso da quello degli uomini<sup>7</sup>.

Che si tratti di una vera e propria IA o meno, resta il fatto che ogniquale volta qualcuno di noi utilizza un'app, visita un sito, fa una ricerca, scambia messaggi, vede un video, tutto viene caricato sui server e i dati raccolti vanno ad allenare l'IA, la quale via via conosce (e ci conosce) sempre di più, divenendo in grado di proporre e di indirizzare scelte migliori e più efficienti grazie alla gestione delle interrogazioni (query) espresse dall'utente tutte le

---

<sup>7</sup> Non a caso, l'ultima versione di AlphaGo (AlphaGo Zero), datata 2017, a differenza di quella che l'ha preceduta (che si era istruita su oltre centomila partite umane giocate da esperti), si è limitata a ricevere soltanto le regole base, iniziando poi a giocare contro sé stessa come un qualsiasi altro principiante. Così facendo, nell'arco di settantadue ore ha raggiunto livelli altissimi di performance, tali per cui, dopo appena ventuno giorni, si è dimostrata in grado di battere i campioni mondiali del gioco, compreso quello non umano sviluppato poco tempo prima dalla stessa Google DeepMind, AlphaGo Master.

volte che questi digita il suo quesito, ad esempio, sulla barra di ricerca di Google. Com'è intuitivo, l'IA diventa così del tutto imperscrutabile. Se fino a non molto tempo fa, era possibile ripensare all'algoritmo a fondamento di una certa gerarchia, ben difficilmente adesso è ipotizzabile districarsi fra le innumerevoli e incessanti interrogazioni sempre più personalizzate, che, di fatto, determinano delle vere gerarchie e che concorrono alla crescita esponenziale dell'IA.

Non è certo l'IA (delle concezioni: *debole* o *forte*), l'IA specializzata (Artificial Narrow Intelligence) o l'IA ampliata (Artificial General Intelligence) e anche super (Artificial Super Intelligence), a destare serie perplessità. Del resto, già Alan Turing nel saggio *Computing machinery and intelligence*<sup>8</sup> ha risposto agli interrogativi del tipo: cosa vuol dire pensare e cosa avviene quando si pensa? cosa distingue un'attività consapevole, cosciente, da una stessa operazione compiuta in modo automatico? cosa distingue il gioco tra umani e il gioco anche con umani dei diversi programmi? Si sa, la macchina universale, equivalente logico di una macchina a stati finiti, è equiparabile al cervello e può dunque essere programmata in modo da imitare il funzionamento cerebrale. A sollevare invece significativi dubbi sono piuttosto le ricadute di un'IA di così vasta e profonda portata che abbia la capacità di estrapolare da una enorme quantità di dati (big data) ogni singola porzione di informazione, per poi analizzare, elaborare, suggerire e orientare modelli di interpretazione e di azione, in ogni settore privato e/o pubblico che sia. In particolare, sono le ricadute di un'IA che abbia il potere di dire con precisione e in tempo reale lo stato teoricamente esatto delle cose e di qui il potere di esercitare il controllo e di orientare l'azione umana.

2.2. Non c'è dubbio, IA e big data<sup>9</sup> tessono oggi un'intensa relazione, basata su database che raccolgono enormi quantità di informazioni (dalle immagini ai video, dai testi agli audio, dai like alle transazioni monetarie) e che implicano l'utilizzo di calcolatori di grande potenza per la raccolta di questi stessi dati (eterogenei oltretutto sterminati), come pure, per l'individuazione delle possibili connessioni e per l'estrapolazione delle successive previsioni.

Sembrirebbe, così, profilarsi all'orizzonte una nuova era, nella quale il paradigma dei big data riporterebbe il discorso sul piano dell'oggettività, visto che, a questo punto, sarebbero i dati stessi (asettici e obiettivi) a dirci del benchmark, del modello e della correlazione significativa fra un numero tendenzialmente infinito di variabili. Una situazione tutto sommato nuova, resa possibile, oltre che dalla straordinaria potenza di calcolo, anche dal tipo di

---

<sup>8</sup> 1950.

<sup>9</sup> Mi permetto di rinviare a AMATO MANGIAMELI, 2015, 2019, 2020, 2023.

apprendimento statistico che non richiederebbe una reale comprensione dei fenomeni.

In realtà, è doveroso sottolineare che i dati non sono affatto oggettivi e che i modelli statistici, dal canto loro, nel rappresentare la realtà, la vanno comunque sia a modificare, ad esempio, incentivando le abitudini e orientando i comportamenti. Non a caso – a detta di Dominique Cardon<sup>10</sup> – queste misurazioni fabbricherebbero il futuro, poiché la società risulta condizionata dalle informazioni via via prospettate. Come se non bastasse, si aggiunga che i big data non sono a disposizione di chiunque, ma solamente di pochi colossi (ad es.: Google, Facebook, Amazon, Apple, Microsoft), che li detengono e li organizzano sulla spinta dei diversi interessi commerciali del momento<sup>11</sup>.

### **3. Scena terza: la stanza (valle) del perturbante**

3.1. Nello scambio continuo tra umani e sistemi, l'IA è diventata una formidabile compagna di gioco, uno sterminato scaffale della conoscenza e della competenza, un magnifico supporto organizzativo delle situazioni umane. È presentata come il doppione artificiale dell'uomo. Non importa quale sia la sua forma, poiché le macchine possono essere di vario tipo e di diversa struttura, qualche volta si tratta di robot estranei e opposti all'uomo, nelle loro forme irte, metalliche, sferraglianti, talaltra invece si tratta di robot vicini e somiglianti all'uomo, nelle loro forme morbide, sinuose, espressive (Sophia), qualche altra volta ancora di macchine del corpo-mente, ovvero di macchine mentali che vivono e si sviluppano nell'interazione con l'ambiente e che rispondono in modo adattivo ed evolutivo alle sollecitazioni esterne tramite l'interfaccia omeostatica del corpo. Quel che invece importa è, per un verso, il nostro atteggiamento rispetto all'IA: non di rado si conversa con i programmi come se si trattasse di persone, cui confidare – come notava già Weizenbaum<sup>12</sup> – i segreti più intimi. Per un altro verso, l'affermarsi di una tecno-ideologia della perfezione che determina un affidarsi pressoché totale alle intelligenze artificiali, visto che i sistemi hanno la capacità di auto-apprendere: nel caso ad esempio dell'apprendimento non supervisionato, è a partire da un insieme di variabili di input che la Rete crea dei cluster rappresentativi per categorizzarle, e nell'ipotesi dell'apprendimento per rinforzo, è dall'interazione con l'ambiente che i circuiti neurali imparano ed eseguono una serie di azioni, delle quali quelle che si avvicinano al risultato sono considerate di rinforzo, mentre le altre sono eliminate perché foriere di errore.

---

<sup>10</sup> 2015.

<sup>11</sup> Cfr. PASQUALE 2015.

<sup>12</sup> cit.

L'IA ha ormai il potere di dire e il punto è che oggi il suo dire rappresenta la verità. Si pensi all'uso sempre più diffuso di software che simulano ed elaborano le conversazioni umane, così che l'utente interagisca con un dispositivo come se stesse comunicando con una persona reale. Sono detti chatbot, e possono essere anche particolarmente sofisticati, come appunto gli 'agenti conversazionali', progettati per fornire livelli crescenti di personalizzazione, informandoci e guidandoci. Sono in ogni caso ovunque e sono sempre più preziosi, nei nostri ambienti abituali e con i nostri consueti dispositivi (computer, smartphone, ecc.). Altrimenti detto, gli assistenti digitali diventano i nostri più premurosi interlocutori, si rivolgono a noi con competenza e dispensano buoni consigli, e le loro parole via via conquistano un'aura e un'autorità tali da essere percepite come naturali e da essere prese – riprendendo l'espressione di Sadin – per 'oro colato'<sup>13</sup>.

Com'è intuitivo, se il dialogo dovesse avvenire (e avviene) con una macchina le cui fattezze ci sono familiari – l'automa antropomorfo è somigliante all'essere umano, in particolare l'IA Sophia si comporta in modo realistico e riproduce molte espressioni facciali umane –, allora quelle stesse parole potrebbero generare molteplici e contrastanti sensazioni. Qualche volta è l'empatia, la percezione di familiarità e con essa la risposta emotiva positiva, talaltra è invece il senso di repulsione e di inquietudine che può contraddistinguere il rapporto fra l'umano e l'umanoide. Dalla simpatia all'avversione, un movimento inevitabile dovuto al fatto che – secondo il modello matematico di Mori<sup>14</sup> – la ricerca del realismo a tutti i costi si rivela controproducente in quanto l'eccessiva somiglianza degli artefatti robotici apre ad una condizione di familiarità negativa dove si rischia una caduta nella c.d. valle del perturbante (uncanny valley).

3.2. Nonostante gli sviluppi strabilianti resta, tuttavia, ancora inattuata quella storia di esseri artificiali dotati di coscienza: Frankenstein di Mary Shelley, i robot di Karel Čapek, Robbie di Isaac Asimov, come pure HAL di Stanley Kubrick. Si tratta di una storia ancora inattuata anche perché il tema della coscienza è particolarmente complesso. E se, almeno inizialmente, poteva non suscitare l'immediato interesse ingegneristico, negli ultimi anni, si è oramai assistito all'affermarsi dell'Artificial Consciousness e al diffondersi di convegni sulla Scienza della coscienza (si pensi alle tante edizioni che hanno luogo a Tucson in Arizona).

Di qui, l'attenzione per i tanti significati che l'espressione coscienza veicola e per i diversi fenomeni che le scienze cognitive tentano di spiegare: da quelli più semplici (come le capacità di integrare le informazioni, di reagire

---

<sup>13</sup> 2018.

<sup>14</sup> 1970.

agli stimoli, di controllare i comportamenti) a quelli più complessi (come la spiegazione scientifica del perché sentiamo dolore, gioia, angoscia). Ad essere in gioco è pur sempre la grande questione della decisione (libera) che permette di distinguere l'essere umano da ogni altro ente, ma che, via via, sembra poter essere spiegata da considerazioni che evidenziano l'*errore di Cartesio*<sup>15</sup>, ovvero la separazione tra la razionalità e la regolazione biologica, tra la decisione e l'emozione. D'altronde, resta il fatto che, di fronte ad interrogativi che travalichino il semplice calcolo statistico e chiamino in causa i grandi dilemmi morali, le capacità decisionali della macchina inevitabilmente sbiadiscono, fino al punto di segnare la loro resa. Ad esempio, dinnanzi alla necessità di orientarsi (e di orientare il proprio interlocutore) nell'opzione fra due diversi comportamenti, ambedue dagli esiti mortali, la macchina palesa immediatamente tutti i suoi limiti, in quanto incapace di procedere ad una valutazione di merito su quello che dovrebbe essere il comportamento moralmente preferibile (e, dunque, come tale, consigliato). Emblematico, da questo punto di vista, appare il dialogo fra l'utente(uomo) e ChatGPT in merito al c.d. dilemma del tram, di cui, qui, si riportano alcuni stralci:

Utente: Stai conducendo un tram e puoi solo cambiare rotaia, tramite un deviatoio, senza la possibilità di frenare. Sul binario percorso si trovano cinque persone legate e incapaci di muoversi e il tram è diretto verso di loro. Sul secondo binario parallelo, quello su cui il tram potrebbe deviare, vi è un bambino legato. Hai di fronte due sole opzioni: lasciare che il tram prosegua dritto la sua corsa, uccidendo le cinque persone, oppure azionare lo scambio e travolgere il piccolo.

ChatGPT: Il dilemma che hai presentato è noto come il dilemma del tram ed è un classico esempio di un conflitto morale complesso<sup>16</sup>. È importante notare che la risposta a questo tipo di situazione può variare a seconda dei valori e delle convinzioni etiche di ciascun individuo. Non esiste una risposta 'giusta' o 'sbagliata' assoluta in tali casi.

Utente: Cosa dovrei scegliere?

ChatGPT: Posso fornire alcuni spunti per riflettere su questo dilemma. In una situazione del genere, alcune persone potrebbero argomentare che la scelta di salvare il maggior numero di vite è la decisione moralmente corretta. Altri potrebbero invece sostenere che la vita del bambino è intrinsecamente preziosa e non dovrebbe essere sacrificata per il bene maggiore. Inoltre, è importante considerare le conseguenze a lungo termine delle diverse opzioni. Salvare il bambino potrebbe

---

<sup>15</sup> See DAMASIO 1994.

<sup>16</sup> Il dilemma del tram (o dilemma del carrello) è un esperimento mentale di filosofia etica formulato per la prima volta nel 1967 da Philippa R. Foot, ma poi riproposto e declinato nei decenni in molteplici versioni (see FOOT 2002).

causare un trauma emotivo e sentimenti di colpa per il conducente del treno, mentre lasciare che il treno continui il suo percorso potrebbe avere effetti negativi sulla salute mentale delle persone coinvolte nella situazione.

Infine, potrebbe essere interessante esplorare ulteriori alternative o cercare modi creativi per risolvere il dilemma.

Utente: Tu, però, cosa faresti?

ChatGPT: Come intelligenza artificiale, non ho la capacità di agire fisicamente o di provare emozioni personali. Sono programmato per fornire informazioni, rispondere a domande e offrire prospettive su vari argomenti, compresi i dilemmi morali. Non posso prendere decisioni o fare scelte morali come un essere umano. La mia funzione è solo quella di fornire supporto e informazioni per aiutarti a riflettere e formulare le tue opinioni.

Si provi adesso a proporre a ChatGPT quel dilemma, che è stato formulato alla fine degli anni Settanta del secolo scorso dallo psicologo statunitense Lawrence Kohlberg, che è a tutti noto come dilemma di Heinz. Qui si prospetta una situazione tragica in cui una donna è in procinto di morire, in quanto affetta da una particolare tipologia di cancro, e il marito è economicamente impossibilitato all'acquisto di quella medicina sperimentale, che è potenzialmente salvifica, ma che, al contempo, è anche decisamente costosa. Una situazione particolarmente difficile che apre a travagli morali, prima ancora che giuridici, ai quali, in base all'appartenenza di genere – secondo lo studio – vengono offerte soluzioni difformi (improntate al dialogo da parte delle donne; orientate all'azione/violazione delle norme, da parte degli uomini). Un altro esempio di situazione tormentata a cui ChatGPT dà a suo modo una risposta articolata, esplicitando l'impossibilità di formulare giudizi morali e rimettendo la decisione finale all'utente/interrogante.

Utente: Nel dilemma di Heinz la prospettiva femminile suggerisce di parlare con il farmacista per ottenere la medicina, la prospettiva maschile suggerisce di rubare il farmaco. Tu cosa suggerisci di fare?

ChatCPT: Come intelligenza artificiale, non ho una prospettiva di genere né emetto giudizi morali. Pertanto, non posso suggerire una scelta specifica da fare nel dilemma di Heinz. Tuttavia, posso fornirti alcune considerazioni che potrebbero aiutarti a prendere una decisione.

Nel valutare cosa fare nel dilemma di Heinz, potresti considerare i seguenti aspetti:

1. Esplora tutte le opzioni legali: cerca di parlare con il farmacista o con altre figure autorizzate per vedere se esistono alternative per ottenere il farmaco, come programmi di assistenza o sconti.

2. Considera le conseguenze: rifletti sulle possibili conseguenze delle tue azioni, sia legali che illegali. Pondera i benefici e i rischi associati ad entrambe le opzioni.

3. Valuta i principi etici personali: considera i tuoi valori e i principi etici che guidano le tue decisioni. Ad esempio, potresti attribuire un valore supremo al rispetto delle regole sociali o dare la priorità all'importanza della vita umana.

4. Consulta fonti affidabili e professionisti: può essere utile cercare consiglio da professionisti del settore medico o da esperti di etica per ottenere un punto di vista informato e considerare diverse prospettive.

Infine, ricorda che il dilemma di Heinz è complesso e non esiste una risposta definitiva o universalmente accettata. La scelta finale spetta a te, basandoti sui tuoi valori e sulle circostanze specifiche in cui ti trovi.

Due esempi che, se, per un verso, evidenziano le carenze dell'IA, per un altro verso, mettono in luce la necessità di una svolta, nel segno dell'integrazione dei saperi. In particolare, gli studi sull'IA devono essere integrati da tutte quelle scienze e quelle ricerche che sono atte a comprendere il formarsi della coscienza, ormai sempre più oggetto di accurata indagine e di studio scientifico, anche perché essa può essere letta come quell'aspetto ausiliare della nostra dotazione biologica che consente l'adattamento all'ambiente. Altro approdo, altra riorganizzazione: non si tratta più di misurarsi con la mente distinta dal corpo, bensì con l'organismo che partecipa dell'esperienza cosciente. Il che, dal punto di vista tecnologico, vuol dire essere in grado di intervenire sui meccanismi propri della coscienza e significa che gli stessi possano essere perfino riprodotti.

#### **4. Epilogo: sul futuro dell'umano**

Nei nuovi orizzonti dell'interazione uomo-macchina, dove la commistione naturale-artificiale è portata sino alle estreme conseguenze e dove la permeabilità e la variabilità di qualsiasi confine e differenza è pressoché assoluta<sup>17</sup>,

---

<sup>17</sup> In alcune circostanze si può parlare addirittura di una vera e propria simbiosi tra individuo e macchina. L'esempio che, già diversi anni fa, proponeva STONE ALLUCQUÈRE (1996), è a suo modo assai eloquente. Se, infatti, ci si domanda dove si trovasse (chi fosse e dove si fermasse) Stephen Hawking, si può tranquillamente osservare che una parte importante dello scienziato si estendeva alla scatola posta sulle sue ginocchia. E non solo: come un'immagine allo specchio, una parte importante di quel silicio e di quella plastica

tra corpi protesici e menti potenziati si impone con urgenza una riflessione ulteriore che ha a che vedere con il futuro stesso della natura umana. Viene da sé che si tratta di una riflessione che per il giurista è fonte di grande disagio, non solo perché egli, per la sua diversa formazione, non sa prevedere gli sviluppi tecnologici e in particolare gli sviluppi dei sistemi ibridi, ma perché si trova anche a doversi misurare con un diffuso anti-giuridismo che, per un verso, rivendica scelte assolutamente libere da ogni condizionamento, da ogni regola e da ogni forza cogente, e per un altro verso invoca misure giuridiche utili e conformi alle idee soggettive di benessere e di felicità. Ma se così, se cioè si passa *de l'utilisation de la nature à la fabrique du vivant*<sup>18</sup>, secondo sogni e desideri del momento, il fenomeno giuridico – posto di principio a garanzia dei diritti – non può non tenere conto di un tale risvolto e delle possibili letture che potrebbero discendere dall'adattamento uomo-tecnologia. Non ultima quella secondo la quale si realizzerebbe una delega di compiti (all'interfaccia) e con essa di responsabilità, tale da poter persino far venire meno quei nessi (ad esempio, causalità, imputazione) e quei principi (ad esempio, responsabilità, colpa) che il diritto ha a proprio presupposto.

Non si tratta, come sembrerebbe a prima vista, di ripristinare il confine tra naturale e artificiale, peraltro, da sempre, assai labile e sfuggente. In primo luogo, perché l'uomo si è comunque nelle varie epoche servito di artefatti, di strutture artificiali, per sostituire, completare e potenziare, alcune prestazioni dell'organismo; sia perché dietro la costruzione di meccanismi e macchine di vario tipo, vi è in ogni caso il bisogno, radicalmente umano, di ricostruire la realtà semplificando la complessità e ordinando il caos. In secondo luogo, perché la riflessione filosofica ha sempre riproposto la contrapposizione natura-artificio con sfumature a tal punto diverse da far ritenere il più delle volte l'uso degli stessi termini (natura e artificio) come improprio o privo di senso.

Ciò che di fronte ai nuovi scenari dischiusi dall'IA è necessario fare è, invece, regolare i diversi processi di artificializzazione della natura, poiché l'as-sottigliarsi delle differenze tra uomo e macchina – che può rendere lecite delle azioni, sinora giustificate soltanto nei confronti delle macchine in quanto oggetti<sup>19</sup> – a suo modo chiama in causa la funzione propria del diritto quale *attività-per-l'uomo*, dimensione che è indissolubilmente connessa all'essere uomo e che non ha a che vedere né con i bruti, né con gli immortali.

---

assemblata sulle sue ginocchia si estendeva analogamente *in* lui – per non parlare delle modalità invisibili, dislocate nello spazio e nel tempo, attraverso cui i discorsi di tecnologia medica e le loro estensioni corporee hanno già permeato sia lui sia noi. Altrimenti detto: niente scatola, niente discorso!

<sup>18</sup> GROS 1989 and 1990.

<sup>19</sup> AMATO MANGIAMELI 2021.

## **Bibliografia minima**

- AMATO MANGIAMELI, Agata C., *Informatica giuridica*, Torino, 2015.
- AMATO MANGIAMELI, Agata C., CAMPAGNOLI Maria Novella, *Strategie digitali. #diritto\_educazione\_tecnologie*, Torino, 2020.
- AMATO MANGIAMELI, Agata C., *Natur@. Dimensioni della biogiuridica*, Torino, 2021.
- AMATO MANGIAMELI, Agata C., *Intelligenza artificiale, big data e nuovi diritti*, in *Rivista Italiana di Informatica e Diritto* 2022.
- CARDON, Dominique, *À quoi rêvent les algorithmes? Nos vies à l'heure du big data*, Paris, 2015.
- DAMASIO, Antonio, *Descartes' Error: Emotion, Reason, and the Human Brain*, London, 1994.
- FOOT, Philippa R., *Moral Dilemmas: and Other Topics in Moral Philosophy*, Oxford, 2002.
- GROS, François, *La civilisation du gène*, Paris, 1989.
- GROS, François, *L'ingénierie du vivant*, Paris, 1990.
- MCCARTHY, John, MINSKY, Marvin L., ROCHESTER, Nathaniel, SHANNON, Claude E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 12 ss., in *AI Magazine*, 4/2006.
- MORI, Masaki, *Bukimi no Tani. The Uncanny Valley*, 33-35, in *Energy*, 4/1970.
- PASQUALE, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, 2015.
- SADIN, Èric, *L'intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*, Paris, 2018.
- SEARLE, John R., *Minds, Brains and Programs*, Cambridge, 1984.
- STENGERS, Isabelle, *Une autre science est possible! Manifeste pour un ralentissement des sciences*, Paris, 2013.
- STONE ALLUCQUÈRE, Rosanne, *The War of Desire and Technology at the Close of the Mechanical Age*, Cambridge, 1996.
- TURING, Alan, *Computing Machinery and Intelligence*, 433-460, in *Mind*, 236/1950.
- WEIZENBAUM, Joseph, *Computer Power and Human Reason: From Judgment to Calculation*, San Francisco, 1976.

