



Dipartimento di Giurisprudenza

Cattedra di Metodologia della Scienza Giuridica

Intelligenza Artificiale, *fictio iuris* e  
Causalità:

Profili giusfilosofici di una questione attuale.

Prof. Antonio Punzi

---

RELATORE

Prof. Alessia Farano

---

CORRELATORE

Vincenzo Calderonio

---

CANDIDATO

Anno Accademico 2020/21

# INDICE

## *Introduzione*

### I. UN NUOVO FENOMENO

#### I.1 Cos'è l'intelligenza artificiale

#### I.2 Il quadro normativo dell'Unione Europea

#### I.3 Le caratteristiche dell'IA

##### I.3.1 Autonomia

##### I.3.2 La dipendenza dai dati

##### I.3.3 Complessità

##### I.3.4 Opacità

### II. QUALIFICAZIONE GIURIDICA DELL'IA

#### II.1 Test di Turing

#### II.2 Un problema giuridico: IA “forte” vs IA “debole”

##### II.2.1 L'argomento della stanza cinese

##### II.2.2 L'intenzionalità nel diritto positivo

##### II.2.3 L'IA come *res*

### III. DETERMINISMO E LIBERO ARBITRIO

#### III.1 Due Principi complementari

##### III.1.2 L'algoritmo è causale, il diritto imputabile

## IV. PROFILI DI DIRITTO

### IV.1 Disciplina giuridica dell'IA

IV.1.1 Profilo civile: Artificial intelligence act

IV.1.2 Profilo penale: Finzione di continuità

*Conclusioni*

## *Introduzione*

Il diritto è un fenomeno sociale risalente alle prime civiltà umane, tutti nel nostro percorso scolastico abbiamo studiato le tavole del Re sumero Hammurabi, la prima forma di diritto, la legge del taglione, il cui dispositivo, ripreso successivamente dal Vecchio Testamento (Levitico, 24, 19-20) recita “*Se uno farà una lesione al suo prossimo, si farà a lui come egli ha fatto all’altro: frattura per frattura, occhio per occhio, dente per dente; gli si farà la stessa lesione che egli ha fatto all’altro*”.

Da allora sono passati millenni ed il diritto ha attraversato delle trasformazioni radicali, mutando la sua natura e le sue configurazioni nel corso dei secoli fino ad arrivare alla moderna scienza giuridica, che si presenta a noi come una scienza sociale e come tale esatta, con regole fisse, autorevole e soprattutto necessaria alla regolamentazione dei rapporti tra i soggetti di un qualsiasi ordinamento giuridico.

La soggettività giuridica ci appartiene sin dalla nascita con l’acquisizione della capacità giuridica, ossia la capacità di essere titolari di diritti e doveri, e successivamente con il raggiungimento della maggiore età, per il nostro ordinamento giuridico acquisiamo anche la capacità d’agire, ossia la facoltà di realizzare atti produttivi di effetti giuridici, diveniamo, dunque, a tutti gli effetti soggetti di diritto e pertanto imputabili.

Come tali siamo destinatari di norme: obblighi giuridici specifici e generali. Nella prima categoria rientrano ad esempio gli obblighi derivanti dalla genitorialità o quelli derivanti dal contratto di lavoro, nella seconda categoria invece si possono includere obblighi di natura costituzionale. In entrambi i casi tali obblighi possono far sorgere una responsabilità giuridica nel caso di inadempimento e, pertanto, renderci passibili di sanzioni.

Il concetto di personalità giuridica è fondamentale nella scienza giuridica e rappresenta uno dei pilastri del funzionamento della nostra società. Resta da chiedersi sino a che punto potrà estendersi questa categoria? Possiamo osservare come non solo gli umani, persone fisiche, che abbiano raggiunto la maggiore età possono essere destinatari di diritti e doveri, anche le persone giuridiche, ai sensi della disciplina del Codice civile, possono costituire entità

autonome distinte dai soggetti che le compongono, mediante questa  *fictio iuris*  esse acquistano personalità giuridica.

Si tratta di un'astrazione che non trova riscontro nella realtà naturale, la cui esistenza viene giustificata per regolare rapporti economico-giuridici tra i consociati; ciò è reso possibile, ad esempio, grazie al regime di autonomia patrimoniale perfetta nel caso delle società di capitali (ad es. S.r.l. e S.p.A.). Significa che godono di un patrimonio proprio sul quale si possono rivalere soggetti terzi nel contesto dei reciproci rapporti contrattuali.

L'aspetto fondamentale che rende possibile quest'astrazione è la presenza di soggetti di diritto, e pertanto esseri umani, alla base dell'ente al quale si attribuisce personalità giuridica. Ciò rende possibile ricondurre la volontà della società a coloro che la costituiscono, sebbene formalmente derivi dalla società stessa dato il regime di personalità giuridica, rendendo trasparente l'apparato sociale ed agevole ogni procedimento  *a posteriori*  per risalire alla formazione di tale volontà. Non solo le società ma anche molte altre formazioni sociali, come lo Stato o le organizzazioni internazionali, funzionano in modo analogo.

È pacifico che gli animali non possano essere processati, perché ad essi, infatti, non è imputabile una determinata norma, essendo privi di capacità giuridica non si possono condannare per il disvalore di un'azione che hanno posto in essere. Pertanto, nei nostri ordinamenti giuridici, gli animali non hanno una soggettività giuridica, vengono piuttosto trattati dal Codice civile come  *res*  e nel caso in cui commettano un danno ne risponderà il legittimo proprietario.

Un altro caso è quello degli incapaci di intendere e di volere che non godono di personalità giuridica piena. Ciò significa che sebbene destinatari di norme, ovvero titolari della capacità giuridica, non saranno considerati direttamente responsabili delle loro azioni illecite poiché in costoro manca il requisito della capacità d'agire e pertanto saranno altri soggetti, che il nostro ordinamento definisce rappresentanti legali, a rispondere delle azioni commesse da tali soggetti incapaci. Questo è il funzionamento degli istituti dell'interdizione, inabilitazione e dell'amministrazione di sostegno.

Tutto ciò vale per tutelare tali soggetti affetti da cause di incapacità che vengono considerati come vulnerabili, ma anche per facilitare agli altri consociati i rapporti con questi ultimi, come

nel caso degli incapaci di intendere e di volere, che per cause di varia natura sono impossibilitati ad attendere alle proprie necessità autonomamente ed agli occhi di soggetti terzi, la natura delle loro azioni, può risultare oscura perché viziata da malattie psichiche o mentali che rendono il loro comportamento anomalo.

La somiglianza tra le fattispecie esaminate attiene alla possibilità di risalire ad una volontà consapevole di compiere tale azione. Presupposto inderogabile che il nostro ordinamento pone per poter configurare una forma di responsabilità, sia essa a titolo di colpa o di dolo, a carico del soggetto che abbia ipoteticamente commesso un illecito.

Ciò è impossibile nel caso di animali perché risalire al loro procedimento decisionale al fine di accertarne la colpa o il dolo apparirebbe insensato, posto che gli animali non condividono i nostri stessi valori sociali e pertanto sono esclusi dal nostro ordinamento come soggetti di diritto.

La motivazione è diversa nel caso degli incapaci di intendere e di volere ma porta anch'essa alla impossibilità di imputare loro la responsabilità per certe azioni: è impossibile pretendere da tali soggetti la consapevolezza di ogni loro azione e pertanto la capacità di comprenderne il disvalore, i loro processi decisionali sono diversi da quelli della generalità dei consociati perché affetti da patologie psichiche, nei loro confronti una pena, come la detenzione, non potrebbe avere alcuna finalità rieducativa.

Come abbiamo visto il concetto di imputabilità è concetto vasto e complesso, apparendo pur sempre riconducibile alla nozione di volontà e consapevolezza, in quanto un soggetto imputabile è destinatario di una norma perché è astrattamente capace di comprenderne il precetto e di modulare il suo comportamento di conseguenza. Questo rende possibile che possano essere convenute in giudizio società e non semplici individui, se affetti da patologie che ne alterano la cognizione o la volontà, o animali, i quali non comprendono i nostri valori e comportamenti.

Fino ad oggi il quadro dei soggetti in grado di assumere decisioni autonome è stato caratterizzato dalle figure sopraelencate, ma il rapido sviluppo tecnologico ha prodotto automi in grado di determinare scelte, più o meno svincolati dal controllo umano, che al pari delle

forme viventi siano in grado di interagire con l'ambiente circostante con la capacità di modificarlo, è il caso del settore in evoluzione dell'*intelligenza artificiale*.

# I. UN NUOVO FENOMENO

## I.1 Che cos'è l'intelligenza artificiale?

L'intelligenza artificiale è un settore dell'*high tech* sviluppatosi a partire dagli anni 50, nel corso del tempo è stato oggetto di ampi dibattiti per l'ambizioso obiettivo di creare delle creature artificiali. Ad oggi tale risultato non è ancora stato raggiunto, nonostante ciò, sono molti i passi in avanti fatti dai primi "programmi pensanti" opera dell'ingegno umano.

Oggetto di studio per gli esperti della materia sono programmi in grado di elaborare informazioni autonomamente, creare degli schemi e, traendone le dovute conseguenze, agire in conformità alle istruzioni ricevute dal programmatore.

La definizione tecnica di IA riportata dal libro di testo "*Artificial Intelligence: A Modern Approach*" di Stuart Russell & Peter Norvig, (noto come AIMA) deriva dalla definizione di intelligenza, ed in particolare, dopo aver illustrato le possibili risposte, che rispecchiano i diversi approcci alla disciplina, viene preso in considerazione dagli autori uno specifico attributo dell'intelligenza definito razionalità.

L'intelligenza umana è un fenomeno complesso e del quale vi sono molti lati ancora sconosciuti, è difficile darne una descrizione composita poiché sono molte le sue caratteristiche: la capacità d'analisi, la memoria, la percezione, l'intuizione e l'emotività, che dà luogo all'intelligenza emotiva oggetto di studio della psicologia, sono tra le più conosciute.

Tra di esse vi è la razionalità che viene definita nell'AIMA come la componente prevalente dell'intelligenza ai fini della definizione di intelligenza artificiale, che dà vita alla nozione di *razionalità perfetta* esprimibile con l'equazione:

$$f_{opt} = \arg \max_f V(f, E, U)$$

Tale formula, descritta nell'articolo "*Artificial Intelligence*"<sup>1</sup> della Stanford Encyclopedia of Philosophy, rappresenta la razionalità come la funzione capace di calcolare sempre la

---

<sup>1</sup> Cfr. S. BRINGSJORD, N. S. GOVINDARAJULU, *Artificial Intelligence*, in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, 2020, URL = <https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/>.

soluzione migliore tra quelle possibili in un dato ambiente, dove: la funzione  $f$  di un agente operante in un ambiente  $E$  elabora una sequenza di stati;  $U$  è la valutazione della sequenza di stati, mentre l'insieme  $V$  denota l'utilità attesa tenuto conto della funzione  $U$  e della funzione  $f$  operante nell'ambiente  $E$ .

La razionalità così descritta può anche essere presentata come un meccanismo, tipico del ragionamento, capace di analizzare delle proposizioni e, confrontandole, trarne una risposta ottimale. Questo modo di pensare razionale, descritto dall'AIMA, è ben noto alla scienza giuridica con il nome di sillogismo aristotelico, che prende il nome dal suo autore Aristotele, identificato dagli stessi autori (Russell e Norving) come il primo a esperire un tentativo di codificare le “*laws of Thought*”.

La logica sillogistica contenuta nell'*Organon* aristotelico fu il principale insegnamento nelle scuole delle arti liberali dalla seconda metà del XII sec.<sup>2</sup> divenendo la base del futuro metodo scientifico. Il meccanismo inferenziale del sillogismo prevede la formulazione di una terza proposizione corretta partendo da due proposizioni predicative (premessa maggiore e premessa minore) basandosi sul principio di identità e differenza secondo il quale: “*due termini identici ad un terzo sono identici tra loro (identità) e – al contrario- due termini, uno solo dei quali sia identico ad un terzo, sono differenti tra loro (differenza)*”<sup>3</sup>.

Questo metodo di ragionare dicotomico ben si concilia col sistema binario, sistema numerico in base 2 utilizzato per la manipolazione dell'informazione in sistemi informatici. Pertanto, sorge spontaneo osservare come il metodo del sillogismo, assunto come base formale per il pensiero razionale, sia il punto di partenza per la descrizione del meccanismo di funzionamento dell'IA.

Dalla razionalità perfetta come modello ideale, discende il concetto di agente razionale, che nel quadro delineato da Russell e Norving nell'AIMA, rappresenta l'obiettivo del settore dell'IA, figlio dell'approccio logico che gli stessi autori privilegiano.

---

<sup>2</sup> A. ERRERA, *Lineamenti di epistemologia giuridica medievale*, G. Giappichelli Editore, Torino, 2006, p. 44.

<sup>3</sup> A. ERRERA, *op. cit.*, p. 41.

Rispetto all'equazione descritta in precedenza, per la definizione di intelligenza artificiale intesa come agente capace di agire razionalmente vanno fatti alcuni accorgimenti<sup>4</sup>:

$$P_{opt,M} = arg \max_{P \in P(M)} V( Agent(P, M), E, U)$$

L'introduzione del Programma P e la macchina M dove viene eseguito il programma, come componenti dell'agente razionale *Agent (P, M)*, delimita il campo di azione, delineando così il profilo di un agente capace di scegliere il programma migliore tra tutti i programmi eseguibili da una macchina.

Questo concetto esprimibile come *bounded optimality* e traducibile in italiano come "ottimalità limitata", dà dei contorni ben precisi alle capacità di un'IA. Ciò significa che le capacità di un agente razionale ed artificiale sono limitate dal codice sorgente, dalle istruzioni che le vengono impartite, dalla qualità delle componenti meccaniche.

Viene così in rilievo il quadro ideale della *computazione standard*, ovvero l'essenza dell'IA, le regole ed i meccanismi che vi sottostanno, idonei a identificarla e definirne limiti e potenzialità.

La definizione di IA che ne discende è quella di automa.

È utile a questo punto della trattazione, avendo definito l'oggetto della ricerca, descrivere brevemente anche ciò che invece, l'IA, attualmente non è. Comprimerne i limiti ed i diversi approcci che si sono sviluppati nel settore, prima di proseguire col tentativo di individuare una corretta collocazione del fenomeno all'interno delle categorie giuridiche odierne.

L'IA non è un computer né una macchina, è un *quid pluris*, non si limita ad automatizzare un compito che le viene affidato, ma ha un grado di autonomia tale che legittima l'attribuzione dell'aggettivo intelligente all'artefatto in questione. È capace di elaborare informazioni ed effettuare operazioni complesse senza intervento umano, un computer non è in grado di compiere tali azioni senza un comando immediato e quindi con un canale diretto con l'operatore.

---

<sup>4</sup> Cfr. S. BRINGSJORD, N. S. GOVINDARAJULU, *op. cit.*

Analogamente non si può paragonare l'IA ad un essere vivente, è un *quid minus*, allo stato attuale dell'arte non sono stati prodotti agenti intelligenti che siano anche capaci di volere. Sebbene sia veritiero che tali agenti artificiali godano di un livello più o meno ampio di autonomia, non si può paragonare questa peculiarità alla capacità di volere. L'IA è pur sempre limitata dalle istruzioni che le vengono impartite dai programmatori, gli animali o gli umani.

La questione se in futuro l'IA possa ambire ad avere una coscienza ed una propria volontà, e così ambire allo status di creatura vivente, è ampiamente dibattuta. Vi sono diverse scuole di pensiero che danno origine alla distinzione tra IA forte ed IA debole (che verrà analizzata nel capitolo II), nessuna delle quali ha finora prevalso.

D'altronde il settore è aperto a innumerevoli possibili sbocchi futuri, molti dei quali presentano retroscena imprevedibili, tra i quali la possibilità che si formi una "Singolarità", ovvero un IA programmata da altre IA che possieda capacità per noi irraggiungibili.

Il settore è certamente esposto al rischio di cigno nero<sup>5</sup>, pertanto è impossibile prevenirne gli sviluppi. Ad oggi è possibile dire che di certo l'IA è un *quid novi* presente nelle nostre vite più di quanto fosse possibile prevedere qualche decennio fa.

Sono molte le applicazioni di questa scienza, attualmente a disposizione del pubblico, che quotidianamente siamo in grado di riconoscere: algoritmi di riconoscimento facciale, di personalizzazione di inserzioni pubblicitarie, CPU dei videogiochi ed infine algoritmi dei social network. Oltre queste applicazioni pratiche, le sperimentazioni più note sono certamente quelle inerenti al settore di trasporti, prossimo ormai alla messa in commercio di veicoli soggetti ad una completa automazione.

Sotto questo profilo possiamo dire di essere davanti ai primi agenti artificiali della storia, sebbene non siano come ci aspettavamo.

Si tratta spesso di algoritmi operanti sulla rete che grazie al loro agire hanno la capacità di modificare l'ambiente circostante, influenzando profondamente sulle persone e sulle loro convinzioni o pensieri. Basti pensare alla personalizzazione delle inserzioni pubblicitarie

---

<sup>5</sup> Sull'Argomento N. N. TALEB, *Il Cigno nero*, trad. di E. Nifosi, il Saggiatore, Milano, 2014.

dovuta ai cosiddetti “*cookies*”, fattispecie rilevante giuridicamente sotto i profili della privacy personale e libera manifestazione del pensiero.

È un’esigenza odierna quella di inserire tali nuove fattispecie, emerse dalla prassi, in un quadro normativo coerente, capace di rispondere ai quesiti più insidiosi: chi risponde del danno commesso da un’IA? come vengono manipolate le informazioni? Quali azioni dovrebbero essere considerate lecite e quali illecite?

Un primo passo verso la regolamentazione dell’IA è stato fatto dall’Unione Europea, che nel corso degli ultimi anni ha costituito un gruppo di esperti indipendenti per l’intelligenza artificiale, al fine di darne una definizione e mettere in luce le problematiche principali, in vista di una successiva regolamentazione.

## I.2 Il quadro normativo dell’Unione Europea

La strada percorsa dall’Unione Europea (UE) inizia formalmente il 10 aprile 2018 con la sottoscrizione da parte di 25 Stati membri di una dichiarazione di cooperazione per l’intelligenza artificiale.

Il primo atto formale emanato è una comunicazione della Commissione europea (CE) al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo ed a quello delle Regioni<sup>6</sup>.

Tale atto diede inizio alla strategia europea per un “*IA made in EU*” coinvolgendo le istituzioni, gli stati membri, i privati e gruppi di ricerca interessati. Preso atto dei nuovi cambiamenti sociali e tecnologici in corso, la Commissione riassume in breve lo scenario internazionale, dove Cina e Stati Uniti hanno già cominciato ad effettuare massicci investimenti per lo sviluppo di IA, proseguendo poi con la disamina della posizione europea.

L’Europa partiva tardi, nel 2016, con un ammontare totale di investimenti privati pari a 2,4-3,2 miliardi di EUR<sup>7</sup>. Nonostante ciò, l’iniziativa lanciata dalla CE mirava ad assumere un

---

<sup>6</sup> Commissione Europea, *L’intelligenza artificiale per l’Europa*, COM (2018) 237, finale.

<sup>7</sup> *ivi* p. 4.

ruolo di leadership mondiale nel decennio successivo, mediante una strategia basata su tre colonne portanti.

1. Sul piano economico, un aumento del volume degli investimenti pubblico/privati fino ad un totale di 20 miliardi di EUR entro il 2020, grazie anche al programma Horizon 2020, programma quadro di ricerca ed innovazione dal valore di 1,5 miliardi.

Incrementare la capacità industriale e tecnologica dei paesi dell'Unione mediante forti incentivi alla ricerca, creazione di centri *ad hoc* e adozione dell'IA a tutti i livelli della catena produttiva. Quest'ultimo punto, meritevole di approfondimento, è stato garantito dalla creazione di una "piattaforma IA on demand" capace di offrire a tutti gli utenti un unico punto d'accesso dal quale attingere alle risorse per il funzionamento dell'IA (informazioni, archivi di dati e cloud finalizzati alla condivisione dei dati).

2. Sul piano sociale, dove vi è il timore che i cambiamenti portati dall'IA possano avere l'impatto più forte, la CE si è proposta di migliorare l'attuale sistema educativo, promuovendo il talento e la diversità.

Preso atto del potenziale impatto sul mondo del lavoro, l'obiettivo diventa quello di formare professionisti capaci di approcciarsi alle nuove sfide con attitudine creativa e spirito d'innovazione, abbandonando il vecchio modello nozionistico, per promuovere lauree miste (giurisprudenza o psicologia ed IA), integrazione degli attuali programmi didattici con l'acquisizione di competenze digitali e l'adattamento del contesto socioeconomico europeo per attirare profili d'eccellenza ed evitare fughe di cervelli verso paesi tecnologicamente più sviluppati.

3. Infine, il profilo che ha generato maggiori dilemmi, il piano etico e giuridico, dove l'obiettivo è quello di predisporre un piano normativo che possa garantire ai cittadini i diritti fondamentali ai sensi del TUE e TFUE, ed allo stesso tempo attirare investimenti da parte dei privati, rendendo il settore dell'IA in Europa il più ambito a livello internazionale.

Il bilanciamento tra queste due posizioni è fondamentale per la strategia europea e coinvolge profili delicati quali: la sicurezza e la responsabilità. L'aggiornamento della normativa per responsabilità per danno da prodotto difettoso sarà certamente una delle maggiori sfide per i giuristi europei. L'obiettivo è lo sviluppo di una IA affidabile e spiegabile, comprendere il processo decisionale delle IA non è sempre possibile e ciò

può essere causa di diffidenza da parte degli utenti e di problematiche etiche legate alla trasparenza o alle eventuali azioni illecite o immorali commesse dall'IA.

L'UE ha già promosso un Regolamento sulla protezione dei dati personali<sup>8</sup> che è entrato in vigore nel 2018, su questa scia si inseriscono le iniziative della CE per una futura regolamentazione del settore dell'intelligenza artificiale.

Insieme a questo piano strategico, il 7 dicembre 2018, la CE ha reso pubblica una dichiarazione di cooperazione<sup>9</sup> con gli stati membri volta a dare attuazione congiunta al piano strategico "*IA made in Europe*". Le indicazioni contenute riprendono il piano strategico della CE, con un forte incoraggiamento agli stati membri di dotarsi di un proprio piano di sviluppo relativo all'IA.

Parallelamente viene formato nel giugno 2018 un gruppo di esperti indipendenti (HLEG) patrocinato dall'Unione al fine di definire meglio i temi che si intrecciano con l'intelligenza artificiale. Il gruppo ha prodotto tra il 2018 e 2019 delle relazioni che definiscono l'IA ed i suoi aspetti legati all'etica ed al diritto.

Da queste pubblicazioni, l'UE, ha tratto le informazioni necessarie che hanno permesso la redazione di un libro bianco, pubblicato nel febbraio 2020, contenente la visione d'insieme dell'IA sotto una prospettiva antropocentrica.

Procedendo con ordine, i contributi più importanti dati dal gruppo di esperti istituito dall'UE sono: la definizione di IA e la proposta di sviluppare un'IA affidabile, basata su sette requisiti, ufficialmente adottata dall'Unione Europea come obiettivo della strategia lanciata nel 2018<sup>10</sup>:

1. intervento e sorveglianza umani (inclusi i diritti fondamentali, l'intervento umano e la sorveglianza umana);
2. robustezza tecnica e sicurezza (inclusi la resilienza agli attacchi e la sicurezza, il piano di emergenza e la sicurezza generale, la precisione, l'affidabilità e la riproducibilità;

---

<sup>8</sup> 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).

<sup>9</sup> Commissione Europea, *Coordinated Plan on Artificial Intelligence*, COM (2018) 795, finale.

<sup>10</sup> Cfr. HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *The assessment list for trustworthy artificial intelligence (ALTAI)*, 8.4.2019.

3. riservatezza e governance dei dati (inclusi il rispetto della riservatezza, la qualità e l'integrità dei dati e l'accesso ai dati);
4. trasparenza (incluse la tracciabilità, la spiegabilità e la comunicazione);
5. diversità, non discriminazione ed equità (incluse la prevenzione di distorsioni inique, l'accessibilità e la progettazione universale, e la partecipazione dei portatori di interessi);
6. benessere sociale e ambientale (inclusi la sostenibilità e il rispetto ambientale, l'impatto sociale, la società e la democrazia);
7. accountability (inclusi la verificabilità, la riduzione al minimo degli effetti negativi e la loro segnalazione, i compromessi e i ricorsi);

Sotto il primo profilo, la definizione di IA del HLEG ha profondamente influito sull'approccio dell'Unione. La Commissione dava questa definizione di IA nel 2018:

*“Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.”<sup>11</sup>*

Come è possibile notare, tale definizione ricalca quella di agente razionale riportata nel paragrafo precedente. A questa, il gruppo di esperti aggiunge degli elementi ulteriori, come il *machine learning* ed il *deep learning* (che verranno trattati in maniera esaustiva parlando di “autonomia”), la robotica e le reti neurali, tutti strumenti propri del settore scientifico dell'IA, idonei a dare un maggiore livello di autonomia al prodotto, utilizzati attualmente in molte tecnologie a disposizione degli utenti e fiore all'occhiello del settore.

Tali metodi, sono idonei a configurare un approccio all'IA *human-based* rientrante in uno dei possibili approcci all'IA descritti all'interno del AIMA.<sup>12</sup> Queste specifiche metodologie sono capaci di dare una maggiore profondità al programma, il quale, analizzando enormi quantità di dati su più livelli (talvolta riproducibili di reti neurali umane), diventa capace di percepire oggetti inserendoli in categorie predeterminate, definire i suoi stessi obiettivi ed “apprendere” dalle proprie azioni.

---

<sup>11</sup> Commissione Europea, COM (2018) 795, *op. cit.*, p. 1.

<sup>12</sup> Cfr. S. RUSSEL, P. NORVING, *Artificial intelligence a modern approach*, Pearson, 2010, p. 2-5.

Ne risulta un approccio intento a sviluppare un'IA le cui prestazioni vengono valutate sulla capacità di riprodurre le *performance* umane, contrapposto all'approccio razionale sulla base del quale le prestazioni di un'IA vengono valutate sulla sua capacità di dare la risposta corretta al problema che le viene posto.

La definizione di IA proposta dal gruppo di esperti indipendenti<sup>13</sup> è più ampia rispetto a quella adottata originariamente dalla CE. Comprende una descrizione particolareggiata delle potenziali capacità dell'IA, nonché delle tecniche più innovative applicate nel settore, lasciando ampi margini di spazio per una visione di futuro in cui l'intelligenza artificiale sarà operativa ed intessuta nelle maglie della nostra società.

Il contributo maggiore dato dall'HLEG è stato sul fronte dell'etica. Le “*Ethics Guidelines for Trustworthy AI*” sono state adottate ufficialmente dalla CE come direttrici per il piano strategico europeo<sup>14</sup> e sono successivamente confluite nel libro bianco adottato dall'UE.

I problemi affrontati dal gruppo di esperti collocano l'IA in una prospettiva in cui essa è integrata in diversi ambiti della società umana, sollevando la questione fondamentale di quale dovrebbe essere il suo ruolo. Si tratta non già di come sviluppare, dal punto di vista tecnico, la migliore IA possibile, ma piuttosto dell'uso e dello scopo per i quali l'IA dovrebbe essere sviluppata<sup>15</sup>.

A tal proposito sono stati sollevati problemi quali: è lecita l'attività di profilazione automatica svolta da alcuni algoritmi? Come evitare le possibili discriminazioni che gli algoritmi effettuano a causa di dati a loro volta discriminatori utilizzati dall'algoritmo per prendere decisioni? Come andrebbe adeguato il quadro normativo in tema di responsabilità alla luce dell'ampia autonomia di cui godono alcuni sistemi IA?

---

<sup>13</sup> “*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.*”, HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *A definition of AI: main capabilities and disciplines*, 2019, p. 6.

<sup>14</sup> Commissione Europea, *Creare fiducia nell'intelligenza artificiale antropocentrica*, COM (2019) 168, finale.

<sup>15</sup> Il problema della “*Governance*” dell'IA.

L'approccio proposto dal gruppo di esperti indipendenti si basa su tre elementi: legalità, etica e robustezza<sup>16</sup>, dai quali discendono ulteriori sette requisiti (tra i quali: sorveglianza umana, trasparenza ed *accountability*) per la realizzazione di un'IA affidabile ed al servizio dell'uomo. Tale sistema è stato ideato per ispirare fiducia negli utenti e finalizzato a migliorare il benessere degli stessi e della società nel suo insieme. L'UE ha dichiarato che potrebbe avere un ruolo fondamentale nella prevenzione dei cambiamenti climatici, nella creazione di una società più sostenibile e nella prevenzione dei crimini.

Questo è l'approccio antropocentrico che la CE ha deciso di perseguire nella propria strategia di sviluppo di un'IA made in EU, alla luce del lavoro condotto dal gruppo di esperti. Ne è conseguito un ampliamento del piano originario, accompagnando agli investimenti in ricerca e sviluppo, anche l'obiettivo di rendere il quadro giuridico europeo uno standard internazionale in materia di sicurezza e garanzia dei diritti fondamentali per i propri cittadini.

L'elaborazione di principi etici che debbano condurre lo sviluppo dell'IA, dalla fase della progettazione e scrittura del codice sorgente, fino alla fase di esecuzione e di rilascio al pubblico è una prerogativa per raggiungere l'obiettivo di creare un'IA affidabile.

Insieme ad esso, l'aspetto più rilevante della questione, è il tema della responsabilità per i danni eventualmente provocati da un sistema d'intelligenza artificiale. È un'ipotesi che si può verificare e che non può essere eliminata all'origine, come tale richiede che vengano prese delle misure precauzionali al fine di garantire cittadini ed investitori.

Il 19 febbraio 2020 si conclude la prima fase di consultazione e ricerca sull'IA, intrapresa dall'UE nel 2018, con la pubblicazione di due documenti: il libro bianco sull'intelligenza artificiale<sup>17</sup> e una relazione della CE sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità<sup>18</sup>.

Nel primo vengono cristallizzati gli orientamenti economici ed etici discussi e dibattuti negli anni precedenti con Stati membri, privati e col gruppo di esperti appositamente costituito. Viene ripreso quanto fin qui descritto, dando le direttive sugli investimenti da realizzare, i

---

<sup>16</sup> *Ivi* p. 3.

<sup>17</sup> Commissione Europea, *White Paper on Artificiale Intelligence – A European approach to excellence and trust*, COM (2020) 65, final.

<sup>18</sup> Commissione Europea, *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*, COM (2020) 64, finale.

fondi da destinare ed il tipo di IA ideale da sviluppare, tenuto conto dei requisiti etici delineati dall'HLEG nel quadro dell'approccio antropocentrico deciso con la COM (2019) 168.

Il secondo documento inaugura una fase nuova e più impegnativa della precedente, nella quale l'obiettivo era quello di ideare un piano strategico ed economico definendo lo scopo del settore dell'IA nel contesto sociale e giuridico europeo. La Relazione su sicurezza e responsabilità avvia un processo legislativo, il cui obiettivo è la promulgazione di un Regolamento europeo che possa essere un quadro normativo di riferimento per la disciplina dell'IA.

In tale documento vengono riportate le caratteristiche principali dell'IA in vista di una sua regolamentazione, sono tutti aspetti intrinsecamente legati al diritto che vengono trattati approfonditamente di seguito.

La prossima azione della CE è una proposta di legge che dovrebbe essere promossa entro il 2021, come è possibile rinvenire sul sito dell'UE<sup>19</sup>, così da gettare i presupposti per un contesto sociale europeo in cui l'IA affidabile possa essere una realtà concreta.

### I.3 Le caratteristiche dell'IA

Nella relazione su responsabilità e sicurezza la CE evidenzia delle caratteristiche proprie dei sistemi IA alla luce del vigente quadro normativo, tenuto conto di diversi settori, tra i quali: i trasporti, la responsabilità per danno da prodotto difettoso, la privacy.

Questi sistemi hanno una forte interconnessione tra loro e spesso risulta difficile tracciare delle linee di demarcazione tra categorie quali IA, IoT (*internet of things*) e robotica. L'interdipendenza propria di questi sistemi informatici è un punto di forza nell'applicazione ed implementazione degli stessi, rendendoli maggiormente efficienti grazie alla condivisione delle informazioni, materia prima del loro funzionamento.

---

<sup>19</sup> <https://ec.europa.eu/digital-single-market/en/artificial-intelligence>.

Un'applicazione estesa di queste nuove tecnologie, inserita in un quadro di politiche di sviluppo come quello proposto dall'UE, lascia auspicare per il futuro un rinnovamento sociale su più fronti.

I singoli individui, che già hanno beneficiato dell'avvento delle tecnologie informatiche, potranno avere più autonomia circa l'esercizio dei propri diritti grazie all'IA. Le imprese avranno accesso a tecnologie in grado di ottimizzare produzione di beni ed erogazione di servizi sul piano logistico, lasciando i lavori di stoccaggio ed organizzazioni dati a sistemi IA, automatizzando lavori meccanici come il trasporto merci o lavori agricoli grazie alla robotica.

Più in generale sarebbe possibile delegare lo svolgimento dei lavori di natura ripetitiva a sistemi aventi una potenza di calcolo di molto superiore alla nostra.

Il punto di forza dei sistemi IA, ovvero l'unitarietà che li caratterizza, rappresenta anche un ostacolo per la loro corretta classificazione e corretta comprensione, esigenza fondamentale, quest'ultima, ai fini dell'adattamento del quadro normativo vigente.

La connettività di questi sistemi sfuma i confini che li distinguono e diversificano. Proprio come è impossibile identificare il computer nel solo software, escludendo la parte hardware, per identificare uno specifico sistema IA vanno prese in considerazione diverse componenti. Gli esempi più chiarificatori sono i chatbot, algoritmi in grado di conversare con più utenti alla volta, utilizzati come operatore telefonico o di assistenza online, ed i motori di ricerca, il cui algoritmo è standardizzato per tutti gli utenti.

Altre forme di IA, come ad esempio i veicoli autonomi e gli algoritmi dei social network, presentano un grado di stratificazione tale da rendere complesso distinguere una parte dal tutto.

Le auto autonome sono fisicamente delle entità materiali indipendenti l'una dall'altra, nonostante ciò, l'algoritmo soggiacente è unico. Pertanto, a fronte di un incidente causato da un solo esemplare, si rende necessario rivedere le modalità di configurazione dell'intero sistema, al fine di evitare la ripetizione dell'errore ed eliminare il vizio che l'ha causato.

In questo caso, porsi davanti al fenomeno individuando l'IA come il singolo veicolo autonomo che ha causato il danno risulterebbe riduttivo, poiché il vizio permane nel sistema. Ugualmente prendere in considerazione l'intero sistema, comprensivo di algoritmo e relativi

esemplari, potrebbe indurre ad escludere il caso del malfunzionamento del dispositivo e del mero errore casuale.

La peculiarità dei sistemi IA può essere affrontata legislativamente mediante una regolamentazione settoriale che prenda in considerazione la diversità dei sistemi, a fronte della loro applicazione ad uno specifico settore, riuscendo così a risolvere i problemi tecnico-settoriali mediante normative di dettaglio, adatte allo specifico contesto, rientranti nel quadro del garantismo tipico degli ordinamenti europei.

La neutralità della tecnologia è un principio adottato dall'UE, di stampo garantista, in grado di guidare l'azione normativa dell'Unione verso una regolamentazione tecnica dell'IA che si ponga nei confronti di essa come uno strumento e, come tale, neutrale. È dal principio della neutralità che discende l'idea di una regolamentazione settoriale di queste nuove tecnologie.<sup>20</sup>

Approcciandosi all'IA come uno strumento diventa possibile stabilire la priorità di definire dei principi generali e sulla base di questi, disegnare le rispettive normative settoriali.

In questo contesto ben si presta il meccanismo legislativo europeo che, nel quadro dell'armonizzazione degli ordinamenti nazionali, avrebbe gli strumenti tecnici necessari, come regolamenti e soprattutto direttive, affinché si realizzi la disciplina dell'IA stabilendo parametri di riferimento comuni a livello comunitario per poi da essi far discendere elementi di normazione secondaria per gli ordinamenti nazionali.

Per far ciò, avendo dato una definizione di intelligenza artificiale, descriverne le principali caratteristiche diventa un presupposto necessario al fine di individuarne punti critici ed elementi costituenti su cui si poggiano queste nuove tecnologie. Cosicché, dopo una loro attenta disamina, potranno emergere quelle dinamiche ricorrenti che renderanno possibile l'enunciazione di principi generali in materia.

### I.3.1. Autonomia

L'elemento di discriminazione capace di tracciare una linea di demarcazione netta tra materia inerte e materia vivente è il grado di autonomia che la seconda riesce a conquistare svincolandosi,

---

<sup>20</sup> A. BERTOLINI, *Artificial intelligence and Civil Liability*, Policy Department for Citizens' Rights and Constitutional Affairs, European Parliament, 2020, p. 67-69.

in misura più o meno accentuata, dall'inesorabile moto inerziale che domina l'esistenza della materia inanimata.

Il principio di causalità descrive in termini fisici il moto dei corpi, regolato dalle leggi della dinamica; rispetto a tale impostazione, il cui assunto fondamentale è che vi sia una *causa* per ciascun *effetto*, ciò che è vivente e ciò che invece è inerte rivestono delle posizioni differenti.

Entrambe sono soggette a questo moto perpetuo, ma mentre nel caso della materia inerte che essa ricopra le vesti di causa o effetto è irrilevante poiché pur sempre di oggetto si tratta, nel caso di esseri viventi la loro posizione di agenti nelle vesti di causa rileva come elemento costituente della loro qualità di soggetti, e come tali capaci di iniziativa.

Non è una distinzione priva di conseguenze rilevanti sul piano giuridico, a seconda che si parli dell'una o dell'altra categoria si useranno i termini di soggetto di diritto, se di persona umana si tratta, o di *res* per riferirvisi.

L'autonomia<sup>21</sup>, propria del comportamento degli esseri viventi, mette in luce la capacità che essi hanno di dare causa ad una serie di eventi autonoma che, pur inserendosi nel quadro più ampio della generalità degli eventi regolati dal principio di causalità, è indipendente dagli altri poiché il punto di partenza di quella specifica serie è l'azione dell'essere vivente che l'ha causato.

Gli oggetti tendenzialmente subiscono gli effetti del moto, pertanto sebbene sia corretta l'affermazione “Cadendo, il macigno ha causato la morte di Tizio” sarebbe scorretto imputare la responsabilità dell'accaduto al macigno. Ci si pone dinanzi ad esso come una causa rispetto all'evento morte, come oggetto in un quadro più ampio dove, la caduta del macigno, può essere causata da diversi fattori come: il vento, una spinta o il caso.

A ben vedere il vento ed il caso, per considerazioni di natura diversa, possono legittimamente essere considerati causa della morte dell'uomo per la loro forza agente, ma difficilmente si possono definire responsabili, non essendo ravvisabile una volontà di causare quell'evento.

Ma nel dire invece “a causa della spinta di Caio quel macigno è caduto su Tizio, uccidendolo” si ravvisa un'azione, la spinta, che noi imputiamo a Caio come frutto di una sua scelta

---

<sup>21</sup>TRECCANI, *Autonomia* Def. 4 “il potere del soggetto di dare a sé stesso la propria legge”, Treccani.it – Vocabolario on line, consultato il 27/01/2021.

deliberata. Tale azione viene punita penalmente dal nostro ordinamento sulla base di considerazioni che ritengono l'omicidio carico di disvalore e come tale passibile di sanzioni. La *ratio* su cui poggia questo meccanismo assume come perno la capacità che aveva Caio di determinare o meno la morte di Tizio, grazie alla sua autonomia di scelta. In questa prospettiva il macigno, che è causa materiale della morte di Tizio, diventa uno strumento usato da Caio.

Autonomia assume un significato diverso se riferita agli oggetti, la definizione che ne dà la Treccani recita:

*“Di impianti, macchine, ecc., capacità di funzionare compiendo il proprio servizio per un periodo di tempo più o meno lungo senza rifornimento di energia o di materiali che forniscano l'energia occorrente per il funzionamento.”<sup>22</sup>*

L'intelligenza artificiale gode di un livello di autonomia del tutto peculiare, non solo è capace di funzionare per un periodo di tempo senza intervento umano, ma, grazie alle tecniche di *machine learning* e *deep learning*, diventa capace di autoregolarsi, richiamando così la definizione di autonomia riferita al comportamento umano.

Le tecniche di “apprendimento della macchina” permettono alla stessa di agire in modo autonomo, una volta che sia stato fissato il risultato da conseguire dall'operatore umano.

Nella creazione di IA di ultima generazione vi è una fase chiamata di “addestramento”, nella quale vengono forniti al programma ingenti quantità di dati. In questa fase, che può essere di tre tipi (addestramento supervisionato, non supervisionato, di rinforzo)<sup>23</sup> a seconda del grado di coinvolgimento dell'operatore umano, operano gli algoritmi della macchina sulla base di input che le vengono forniti, per implementare e testare la sua capacità di elaborazione dati.

Così vengono addestrate le IA a fare operazioni di classificazione, creazione di *cluster* di informazioni ed interazione con l'ambiente. Sono tutte operazioni complesse. Si pensi alla capacità degli algoritmi di riconoscimento facciale di identificare un volto umano associandogli un nome ed un codice per l'accesso, oppure alla capacità dei motori di ricerca

---

<sup>22</sup>TRECCANI, *Autonomia* Def. 6, Treccani.it – Vocabolario on line, consultato il 27/01/2021.

<sup>23</sup> Cfr. S. Quintarelli (a cura di), *Intelligenza Artificiale*, Bollati Boringhieri, 2020, p. 33-46.

di effettuare suggerimenti sulla base delle ricerche precedenti, analizzate, confrontate e raggruppate in insiemi significativi.

L'algoritmo che supera la fase di addestramento è funzionante a tutti gli effetti e pertanto idoneo ad essere rilasciato sul mercato, in quanto capace di eseguire l'operazione per la quale è stato programmato. Inizia così la fase esecutiva, dove l'IA è chiamata ad eseguire ripetutamente le stesse operazioni configurate durante la fase di addestramento, sulla base dei dati che le vengono offerti dagli utenti e dalla rete.

Una volta che l'algoritmo è operante non necessita più dell'intervento umano. Ciò significa che è capace di interagire con gli utenti e rispondere agli input che gli vengono forniti del tutto autonomamente, fornendo le risposte corrette sulla base dei dati e delle operazioni di classificazione e clustering apprese nella prima fase e, in taluni casi, è capace di auto-implementarsi durante la stessa fase esecutiva grazie ai meccanismi di machine learning.

L'intervento umano si rende necessario solo per attivare l'IA ed il suo meccanismo di funzionamento, oppure in caso di errore per rielaborare i dati ed effettuare un aggiornamento dell'algoritmo.

Rispetto al quadro delineato in precedenza, sembra che l'IA ricopra un ruolo intermedio tra autonomia umana e inerzia degli oggetti inanimati. Sebbene sia vero che un algoritmo d'intelligenza artificiale possa effettuare delle operazioni che implicano una decisione, come decidere quali informazioni sono rilevanti e quali no per la formazione di un determinato insieme, tale autonomia deriva da una fase di addestramento dove tali meccanismi sono stati implementati per un'iniziativa umana, e sempre grazie alla stessa, la macchina, per così dire, prende vita.

Nel quadro generale della causalità, l'IA compie delle azioni meccaniche, la cui origine può essere rinvenuta nell'azione umana. Ciononostante, molte delle scelte dell'IA, se viste nella prospettiva del compito che essa è chiamata a svolgere, vengono prese da essa stessa sulla base dei dati e dell'esperienza precedente.

È curioso, inoltre, notare la somiglianza che vi sia tra causalità ed operazioni svolte dall'IA: classificazione e clustering che vanno sotto il nome di “*Data mining*”<sup>24</sup>. Si tratta di operazioni razionali aventi una tradizione secolare, le cui radici possono essere fatte risalire alle “*Categorie*” dell'*Organon* aristotelico. L'aspetto che le accumuna è la consequenzialità; la causalità, così come la razionalità, procedono in modo lineare.

In questo quadro complesso e variegato, la tematica giuridica della responsabilità affrontata dall'UE solleva non poche problematiche.

Relativamente alla questione posta ad inizio paragrafo, ovvero la distinzione tra la causazione di un evento e la responsabilità dello stesso, l'autonomia dell'IA pone una sfida inequivocabile: chiarire qual è il ruolo dell'IA nel caso in cui essa causi un danno e chi ne dovrebbe rispondere giuridicamente.

Si può leggere nella Relazione su sicurezza e responsabilità della Commissione Europea:

*“Non è chiaro in che modo si possa dimostrare la colpa di un'intelligenza artificiale che agisce autonomamente, né in che cosa potrebbe consistere la colpa della persona che utilizza l'intelligenza artificiale.”*<sup>25</sup>

L'IA è capace di causare un danno, tale capacità tuttavia si svincola in parte dal controllo umano grazie ai meccanismi di auto-apprendimento che costituiscono la fonte della sua autonomia. Rispetto a questo fenomeno, individuare a chi è imputabile la responsabilità del danno comporta un'attenta analisi delle dinamiche interne, accompagnata da una rivisitazione di categorie giuridiche classiche come quelle di: soggettività, colpa, nesso di causalità e responsabilità.

Alcune delle possibili soluzioni, prese in considerazione dall'UE in vista di una regolamentazione generale, sono:

1. Attribuire la responsabilità all'operatore informatico.
2. Configurare una personalità elettronica del sistema IA.
3. Elaborare meccanismi di gestione del rischio prima dell'immissione in commercio.

---

<sup>24</sup> C. CAPPIELLO, M. FUGINI, P. GREFFEN, B. PERNICI, P. PLEBANI, M. VITALI, *Fondamenti di sistemi informativi*, 2018, p. 159- 157.

<sup>25</sup> Commissione Europea, COM (2020) 64, *op. cit.* p.17.

4. Configurare una forma di responsabilità oggettiva in capo al produttore.
5. Concentrarsi su meccanismi di “*redress*”.

La prima è anche la più immediata: se l’IA è soggetta al principio di causalità e nella sua forma di autonomia non si può ravvisare capacità d’iniziativa che non sia riconducibile all’azione umana, ne consegue che non vi sarà mai una vera e propria volontà da parte delle macchine, né sarà possibile considerarle come responsabili del danno commesso.

Dal punto di vista giuridico è una strada percorribile, che non richiede grossi cambiamenti nella normativa attuale.

Ciononostante, è legittimo chiedersi fino a che punto e in che misura la tecnologia IA è capace di estendere la responsabilità del programmatore, prolungandola al di là della propria sfera di controllo e del proprio agire?

Per ovviare a quest’ultima problematica è nata la teoria del *meaningful human control* (MHC)<sup>26</sup> che si propone di garantire un’adeguata sorveglianza umana sul prodotto IA, rendendo così imputabili gli errori e i danni all’operatore direttamente in controllo del sistema. Questa teoria, nota come sorveglianza umana (*human oversight*) nei documenti europei, ovvierebbe al problema della responsabilità, eludendo la tentazione della responsabilità oggettiva, andando però ad incidere incisivamente sull’autonomia dei sistemi IA che li contraddistingue positivamente.

La seconda e la terza soluzione sono state analizzate nel documento “*Artificial intelligence and civil liability*”<sup>27</sup> e partono da approcci differenti.

Nel caso della personalità elettronica si fa riferimento al caso della personalità giuridica di natura funzionale attribuita ad enti e società, escludendo a priori una personalità di natura sostanziale e valutando l’opportunità di attribuire una forma di personalità simile ai sistemi IA capaci di agire in autonomia.

Percorrendo questa strada, sarebbe necessario individuare preventivamente dei soggetti che si assumano la responsabilità mediante un vincolo di appartenenza, come quello che lega i

---

<sup>26</sup> S. Quintarelli, *op.cit.*, p.90.

<sup>27</sup> Cfr. A. BERTOLINI, *op. cit.*

soci alle società, e si potrebbe aggiungere un fondo di garanzia per realizzare una forma di autonomia patrimoniale e garantire una soddisfazione alle vittime del danno.

È una strada suggestiva ma insidiosa, che potrebbe creare fenomeni potenzialmente nuovi e portare in futuro a considerare diritti in capo ad agenti IA, come i sostenitori di un IA “forte” auspicano.

La possibilità di configurare delle valutazioni del rischio durante diverse fasi della produzione ed immissione in commercio, risponde all’esigenza di garantire la sicurezza dei prodotti ai consumatori.

Il presupposto è la suddivisione di IA ad alto rischio ed IA a basso rischio, distinzione che si rende necessaria a fronte dei diversi usi possibili di queste tecnologie, così da elaborare diverse fasce di rischio e prevedere degli obblighi assicurativi per le fasce più elevate.

Tutto ciò accompagnato da valutazioni del prodotto prima dell’immissione in mercato, durante la fase di addestramento, e dunque la previsione del rispetto di alcuni standard di sicurezza già nella fase della programmazione, che possono incidere sulla libertà di sperimentazione ma anche sulla sicurezza del prodotto.

Congiuntamente dovrebbero essere introdotti anche test di valutazione del rischio, dopo l’immissione del prodotto in commercio, da effettuarsi durante la fase esecutiva così da ovviare ai rischi dovuti ai meccanismi di autoapprendimento in grado di apportare modifiche al sistema, consentendo usi che non fossero previsti né prevedibili al momento dell’immissione in commercio.

La responsabilità oggettiva, esplicitamente prevista dalla CE nella relazione su responsabilità e sicurezza, pone un problema giuridico rilevante: non tutti i paesi europei condividono quest’impostazione normativa.

L’Italia ha escluso che la responsabilità oggettiva possa essere una forma d’imputazione. Con l’affermazione del principio di colpevolezza in ambito penalistico, gli unici elementi soggettivi in grado di legittimare una sentenza di condanna sono la colpa ed il dolo.

Quindi sebbene da un punto di vista pratico possa essere allettante l’idea di imputare la responsabilità del danno prodotto dall’IA al produttore o all’utente sulla sola base di una

forma di responsabilità oggettiva, sarebbe un grave onere, talvolta ingiustificato, gravante sull'intera filiera dell'intelligenza artificiale, che ne potrebbe compromettere applicazione e diffusione.

Infine, l'approccio basato sul “*redress*” è incentrato sulla previsione di meccanismi riparatori del danno, volti a garantire maggior sicurezza ai consumatori, piuttosto che all'individuazione di un soggetto a cui imputare la responsabilità.

Nella prospettiva di tale approccio confluiscono anche alcune idee degli orientamenti precedenti, come la valutazione del rischio e la responsabilità dell'operatore. Inoltre, preso atto della complessità e connettività dei sistemi IA, ci si propone di configurare forme di responsabilità solidale che possano tenere in considerazione l'attività congiunta dell'operatore, dell'utente e dei programmatori ai fini della causazione del danno.

Si tratta di approcci tutti egualmente possibili e non necessariamente alternativi tra loro, che la CE sta valutando ai fini di una proposta regolamentare orizzontale da proporre al parlamento europeo.

Ad ogni modo per scender più nel dettaglio e comprendere le ragioni che possono indurre a privilegiare un orientamento piuttosto che l'altro, bisogna prendere in considerazione le ulteriori caratteristiche dell'IA (dipendenza dai dati, complessità ed opacità) che le danno una forma ben precisa rispetto alle *quaestiones* fin qui sollevate.

### I.3.2. Dipendenza dai dati

Il concetto di dipendenza dai dati è una diretta conseguenza della *bounded optimality*<sup>28</sup> caratterizzante i sistemi IA. La loro capacità di elaborazione dati è limitata da com'è scritto l'algoritmo, dal numero e dal tipo di operazioni che è capace di compiere, dalla qualità delle componenti meccaniche e soprattutto dalla qualità dei dati.

---

<sup>28</sup> Si veda p.8.

“*Garbage in garbage out*” è il mantra che viene ripetuto negli ambienti tecnici per indicare che se viene immessa dell’informazione spazzatura all’interno di un programma, sarà inevitabile un risultato spazzatura, per riportare fedelmente la dicitura anglosassone.

La questione è più seria di quanto possa sembrare.

Il primo fattore è la definizione d’informazione, elemento cardine di tutte le tecnologie informatiche, che ai fini della nostra ricerca può essere descritta come tutta l’informazione idonea ad essere rappresentata formalmente.

Ciò comprende scale di grandezza che vanno dal mero bit (0/1), da cui il codice binario, ovvero la notazione formale per rappresentare il passaggio di corrente o meno all’interno di un determinato circuito, fino ai sistemi esperti in grado di dare una rappresentazione formale della conoscenza umana, traducendola in linguaggio macchina<sup>29</sup>, ed utilizzarla al fine di riprodurre il processo decisionale umano per svolgere compiti complessi.

Il secondo discende dalla nozione di *informazione rilevante* che pone un serio problema gnoseologico. Ai fini della configurazione di un sistema IA bisogna fornirgli grandi quantità di dati, dati che poi la macchina dev’essere capace di elaborare nella creazione di insiemi e di schemi, al fine di compiere questa classe di operazioni effettua una distinzione tra informazione rilevante e non rilevante.

Chiaramente in questo processo si annida lo spettro dell’errore. In principio la stessa definizione di informazione rilevante è arbitraria e necessariamente relativa allo scopo che ci si propone di raggiungere. In seconda battuta sia l’algoritmo che il programmatore possono commettere errori nella scelta dei dati rilevanti, infine anche se le prime due fasi fossero andate a buon fine, resta la possibilità che i dati siano compromessi o inesatti, causando così un effetto a catena destinato a propagarsi in tutte le applicazioni successive del programma.

Quest’ultimo passaggio è il più sottile perché, richiamando il principio di falsificabilità del filosofo Karl Popper, è possibile intuire come un dato che possa apparire corretto ad una prima analisi, possa nascondere invece un errore rimasto celato fino al momento della sua manifestazione.

---

<sup>29</sup> Per linguaggio macchina s’intende il codice binario, ovvero il sistema di 0/1 idoneo a rappresentare il funzionamento meccanico dei dispositivi informatici.

È il caso di algoritmi che creino delle discriminazioni basate sul genere o sull'etnia, perché gli vengono forniti dati statistici che riportano questa classe di discriminazioni e inducono l'algoritmo a riprodurle nelle sue applicazioni successive. In questo caso nonostante la buona fede del programmatore, la riproduzione dell'errore è spesso inevitabile perché insita nei dati che vengono forniti all'elaboratore.

Questo è in grado di dirci molto sulla società in cui viviamo e sulla natura dell'IA. Due casi recenti mostrano come i dati immessi in un sistema IA possano veicolare significati molto più ampi, portatori di conseguenze sociali inimmaginabili nella fase di programmazione e determinare così il funzionamento successivo dell'algoritmo.

Nel dicembre 2020 ha fatto scalpore la notizia del licenziamento di Timnit Gebru<sup>30</sup>, ricercatrice di Google nel campo dell'etica della tecnologia, autrice di un articolo accademico critico nei confronti dell'uso fatto dall'azienda dell'intelligenza artificiale. I temi sottolineati sono relativi all'attività di *data mining* svolta durante la fase di addestramento, nel caso di specie l'algoritmo di Google, che è un motore di ricerca, viene nutrito con ingenti quantità di dati provenienti dalla rete per migliorare le sue capacità di riconoscimento del linguaggio naturale e dare suggerimenti di ricerca più pertinenti.

L'aspetto morale evidenziato dalla Gebru è che i dati forniti possono derivare da testi razzisti, omofobi o inneggianti alla violenza ed all'uso di droghe andando così a incidere in modo consistente sui possibili suggerimenti di ricerca svolti dall'algoritmo proprio in questa direzione.

Sempre nell'ambito di algoritmi per il riconoscimento del linguaggio naturale, nel 2016 Microsoft ha rilasciato un chatbot chiamato "*Thinking about you*" (Tay)<sup>31</sup> reso accessibile agli utenti via Twitter. Lo scopo di quest'iniziativa era testare le capacità di *machine learning* in tema di apprendimento del linguaggio naturale, infatti Tay era capace di apprendere dalle conversazioni sostenute con gli utenti che interagire liberamente con lui.

Il risultato fu che Tay venne bombardato da messaggi inneggianti all'odio, alla violenza, scurrili e politicamente scorretti in modo deliberato dagli utenti, attività che poi ha indotto

---

<sup>30</sup><https://www.ilpost.it/2020/12/09/gebru-google-intelligenza-artificiale>

<sup>31</sup> S. Quintarelli, *op. cit.* p. 23-24.

l'algoritmo a generare diversi tweet dello stesso tenore costringendo così la Microsoft a ritirarlo dopo sole 16 ore dalla sua pubblicazione.

I casi presi ad esame mettono in mostra come l'IA sia uno strumento e l'uso che se ne può fare dipende dalle nostre scelte, ne consegue che anche le conseguenze negative e non volute possono essere imputate al nostro senso di responsabilità.

Ma l'aspetto preponderante di questo fattore che abbiamo definito *dipendenza dai dati* è la capacità di tracciare un limite intorno alle vere potenzialità dei sistemi IA, qualificandoli come sistemi probabilistici.

La peculiarità di questi sistemi, che li differenzia rispetto ai sistemi deterministici, è che a fronte di un input certo l'output è solo probabile e come tale dominato dalle leggi della statistica. È proprio per questo che vengono fornite ingenti quantità di dati all'algoritmo, in modo tale da aumentare la probabilità statistica di conseguire il risultato esatto. Inoltre, questo presuppone un passaggio ulteriore rispetto all'approccio *logic-based* presentato nel primo paragrafo perché se l'IA fosse governata solo da meccanismi logici, e potesse essere annoverata tra i sistemi di tipo deterministico, il risultato sarebbe sempre quello previsto.

Un esempio di sistema deterministico è il computer, nel momento stesso in cui pigio sulla tastiera il tasto rappresentante la lettera "F" ecco che essa subito appare sullo schermo. Ma se invece io chiedo ad un'IA di compiere una certa azione come guidare un'automobile, entrano in gioco una moltitudine di variabili: l'ambiente circostante in continua evoluzione, gli imprevisti come il pedone che non attraversa sulle strisce pedonali, il tragitto, il riconoscimento dei segnali stradali sono tutti fattori che condizionano il raggiungimento del risultato.

Questa è l'opinione predominante. In realtà appare corretta quella minoritaria, essa afferma di contro che non è l'intelligenza artificiale ad essere un sistema probabilistico (o meglio indeterministico) quanto piuttosto, ricalcando certe idee derivanti dalla fisica quantistica, che sia la realtà stessa meramente probabile e pertanto imprevedibile.

Ne consegue che la natura incerta del risultato non è imputabile al sistema d'intelligenza artificiale, quanto piuttosto all'incertezza dell'input che le viene fornito. In questa prospettiva l'IA sarebbe un sistema deterministico e pertanto conoscendo esattamente l'input in entrata

si dovrebbe essere capaci di prevedere con altrettanta esattezza l'output, questo però non è possibile per i connotati propri della realtà non imputabili al sistema.

Questo susseguirsi di considerazioni di più ampio respiro ci porta a prendere in considerazione un'ulteriore caratteristica dei sistemi IA, la loro complessità.

### I.3.3. Complessità

Ci sono compiti che l'IA riesce a svolgere facilmente e compiti che invece sembra non potrà mai ambire a realizzare.

Nel novero dei primi vi sono giochi come scacchi e go, che si possono considerare come il miglior successo dei sistemi IA fino ad oggi, dato che i campioni delle macchine, rispettivamente Deep Blue di IBM e AlphaGo Zero di Google DeepMind, sono riusciti a sconfiggere i campioni della nostra specie. Deep Blue sconfisse l'allora campione mondiale di scacchi Garry Kasparov tra il 1996-1997 e AlphaZero ha sconfitto nel 2016 Lee Sedol uno tra i più forti giocatori di go al mondo.

Questi episodi hanno destato molto clamore mediatico perché si tratta di giochi che richiedono grandi capacità per eccellere e spesso associati all'intelligenza ed all'abilità strategica, ma la spiegazione per la quale l'intelligenza artificiale batte quella umana è semplice.

In entrambi i casi, scacchi e go, le regole del gioco sono di tipo costitutivo<sup>32</sup>. Questo significa che viene tracciato un perimetro entro il quale è possibile giocare come la scacchiera, tramite una regola di tipo restrittivo, e poi, grazie alle prime, vengono decise le mosse che possono essere giocate, rendendo così il tipo ed il numero di mosse predeterminato.

Il gioco, quindi, ha un alto grado di prevedibilità ed un contesto limitato, dato che ciascun pezzo può muoversi solo in un certo modo, spostando così il piano del confronto solo sulla capacità di calcolo.

A dimostrazione di ciò si può citare il meccanismo di "forza bruta" usato da Deep Blue, consistente in una potenza di calcolo enorme, capace di analizzare una grande mole di

---

<sup>32</sup> Cfr. L. FLORIDI, *Il futuro prossimo dell'intelligenza artificiale*, YouTube, 2020, URL=<https://www.youtube.com/watch?v=0E1kNRpNW10&t=2635s>.

possibili mosse a velocità sovraumana, attingendo ad un archivio contenente lo storico di tutte le partite dei grandi campioni di scacchi. I suoi successori sono stati addirittura programmati, grazie al *machine learning*, per simulare infinite partite contro sé stessi, aumentando a dismisura la loro potenza di calcolo.

La ripetitività e la meccanicità di certi contesti giocano a favore dell'IA, essa grazie alla sua potenza di calcolo è astrattamente in grado primeggiare in tutti gli scenari che richiedono queste caratteristiche, attingendo a massicce quantità di dati.

Per una potenza computazionale sovraumana servono tante risorse, quante più risorse sono necessarie per la soluzione di un problema, tanto più esso sarà *complesso*. Le risorse in informatica sono: spazio d'immagazzinamento dati, la memoria, e tempo di calcolo, più ne servono, maggiore sarà la complessità di un problema e del relativo algoritmo per risolverlo.

Gli algoritmi IA sono complessi e lo si può notare in un compito svolto molto facilmente dall'occhio umano, ma con ben più difficoltà da un algoritmo: riconoscere un'immagine. Per far ciò, nella fase d'addestramento, vengono fatte analizzare dall'occhio della macchina innumerevoli forme, alle quali il programmatore dovrà associare una "etichetta" ovvero il nome della figura specifica, insegnando così all'algoritmo a riconoscere in quella forma quella specifica denominazione.

A tal fine serve una memoria capace di immagazzinare tutte le immagini usate per l'addestramento ed un tempo sufficiente per avere un risultato atteso soddisfacente. La quantità di risorse necessarie è elevatissima, pertanto renderà l'algoritmo complesso, la fortuna però è che all'aumentare delle informazioni aumenteranno anche le capacità dell'IA, rendendo sempre minore la quantità necessaria di risorse per il suo funzionamento.

La teoria della complessità<sup>33</sup>, di cui abbiamo visto l'applicazione informatica, si estende a diversi ambiti, soprattutto per spiegare i comportamenti di sistemi caotici e non deterministici, come gli organismi, i mercati, gli ecosistemi. In essa è centrale la nozione di *non-linearità*,

---

<sup>33</sup> Sull'argomento, N. N. TALEB, *Antifragile*, Il Saggiatore, Milano, 2013.

ovvero l'impossibilità di scomporre i sistemi complessi in singole parti indipendenti dal tutto, andando così a privilegiare metodi di analisi olistici<sup>34</sup>.

L'interdipendenza intrinseca dei sistemi complessi, tale da rendere impossibile distinguere una parte dal tutto senza una perdita considerevole di comprensibilità, è propria anche dei sistemi IA come già preannunciato nel terzo paragrafo parlando di *connettività* di tali sistemi. Questo aspetto evidenzia profili di rischio presi in considerazione dal legislatore europeo.

In particolare, la complessità di tali sistemi che si intreccia con l'*Iot* e la robotica, comporta sul piano concreto la reale possibilità che un algoritmo perfettamente funzionante in laboratorio, una volta immesso nel mercato, possa generare effetti imprevisti dovuti all'interazione tra quel sistema e gli altri già presenti nello spazio circostante.

Sotto questo profilo la valutazione del rischio è un'operazione fondamentale per garantire sicurezza agli utenti e adeguate garanzie anche ai produttori, evitando così che possano andare incontro a conseguenze giuridiche impreviste.

Attualmente vi sono tre casi che il produttore deve tenere in considerazione ai fini della valutazione del rischio, prima dell'immissione in commercio di un prodotto:

- L'uso previsto
- L'uso prevedibile
- L'uso scorretto ragionevolmente prevedibile

In ciascuna di queste circostanze il produttore deve effettuare una valutazione del rischio comprendente la possibilità che il prodotto possa interagire con altri prodotti e delle sue eventuali interconnessioni<sup>35</sup>.

Rispetto a quest'impianto normativo, l'uso di software ed algoritmi potrebbe rendere necessaria una riformulazione di tale normativa, richiedendo sotto il profilo delle definizioni normative l'introduzione di termini quali "sistema IA", "algoritmo" e "connessioni".

---

<sup>34</sup> Un orientamento autorevole in questa direzione è stato inaugurato dalla BIRS nel 1999, con i *Comprehensive Development Framework*, M. R. MAURO, *Diritto internazionale dell'economia*, Edizioni scientifiche italiane, Napoli, 2019, p. 266.

<sup>35</sup> Commissione Europea, COM (2020) 64, *op. cit.* p. 10-13.

La complessità di sistemi d'intelligenza artificiale apre anche delle prospettive sull'IA generale, che propone un approccio volto a massimizzare l'interdipendenza di questi sistemi, al fine di avere un sistema esteso capace di integrarsi propriamente nel nostro tessuto sociale. In contrapposizione ad una visione ristretta dell'IA, più incentrata sulla creazione di algoritmi in grado di risolvere problemi specifici.

I compiti che l'IA non è capace di fronteggiare sono proprio quelli più complessi, ovvero i più stratificati e intricati. Gli scacchi vedono l'IA avvantaggiata perché sono un gioco scomponibile in tante parti *lineari* indipendenti tra loro (questo perché il gioco è scandito da turni, mosse e archi di calcolo circoscritti). Ma come si comporta un'IA davanti a problemi ampi come la creatività e la coscienza? Non è ancora stato possibile sperimentarlo perché questi fenomeni sono universali, neanche noi siamo in grado di scomporli o definirli propriamente, sono per loro natura ambigui dinanzi al nostro linguaggio e questo è una conseguenza della loro complessità.

La conseguenza più rilevante della complessità e della componente probabilistica dei sistemi IA, come abbiamo visto, è l'incertezza del risultato atteso. Questo porta con sé anche un'altra problematica, quella della comprensibilità dell'algoritmo.

#### I.3.4 Opacità

L'ambizione di avere un'IA spiegabile è una delle più condivisibili tra quelle proposte dall'UE. Quest'esigenza, manifestata dalla Commissione in più documenti, risponde alla necessità di garantire un quadro giuridico coerente che possa ovviare al problema del cosiddetto "*effetto scatola nera*".

Difatti grazie ai modelli di apprendimento automatico propri dei sistemi IA, quest'ultimi evolvono nel tempo, mutando il loro comportamento a seconda dell'ambiente circostante e del flusso di informazioni in entrata. La dinamicità di questo sistema fa sì che si crei un effetto scatola nera rispetto alla ricostruzione del processo decisionale usato dall'IA per svolgere un determinato compito.

L'opacità che ne consegue è una diretta conseguenza dell'autonomia di questi sistemi, causando un serio problema in merito al meccanismo di controllo *ex post* del rispetto delle norme ai fini dell'individuazione di un responsabile.

Ne risulta che la verifica del nesso di causalità tra un comportamento posto in essere dall'operatore informatico e l'eventuale danno causato dall'algoritmo diventa in alcuni casi impossibile da ricostruire.

La normativa italiana generale dettata dal Codice civile in materia di risarcimento del danno, pone l'onere della prova a carico di colui che vuole far valere i propri diritti, quindi la vittima, chiedendogli la prova del danno subito e del nesso di causalità per imputare la responsabilità a chi ha causato il danno.

Ma tale impostazione, se usata nei confronti dell'intelligenza artificiale, rischia di condurre ad una *probatio diabolica*, posto che l'opacità di tali sistemi spesso impedisce la ricostruzione dei fatti *a posteriori*.

Le soluzioni proposte a livello europeo, oltre al già citato regime di responsabilità oggettiva per i produttori in materia di responsabilità per danno da prodotto difettoso, prevedono anche degli alleggerimenti in materia di onere probatorio.

In taluni casi potrebbe essere opportuna anche un'inversione dell'onere della prova, accompagnata da obblighi da rispettare in materia di produzione di algoritmi, che se rispettati possano esonerare dalla responsabilità colposa, cosicché spetti al produttore dimostrare di aver adempiuto a tali obblighi, eludendo così la difficile impresa di dover ricostruire il processo decisionale dell'IA che l'ha portata ad innescare la serie causale culminante nel danno.<sup>36</sup>

Tutto ciò potrebbe non bastare, sono già state codificate delle norme a livello europeo in merito al diritto degli utenti ad avere delle spiegazioni in merito alla logica usata da algoritmi in caso di trattamento di dati personali mediante un processo decisionale automatizzato<sup>37</sup>. Sembra questa la strada intrapresa dall'UE, coerentemente con l'impianto garantista dei

---

<sup>36</sup> Appare questa l'impostazione normativa inaugurata dall'*Artificial intelligence act* COM(2021) 206 final, quando impone il rispetto di numerosi requisiti di sicurezza come presupposto del rilascio di un certificato di conformità necessario per l'immissione nel mercato dell'IA.

<sup>37</sup> Si veda Regolamento UE n. 2016/679, in particolare art. 13 comm. 2 lett. f

Trattati europei, improntata allo sviluppo di un'IA antropocentrica e quindi al servizio del cittadino.

Per sviluppare un'IA spiegabile l'UE investirà molto in ricerca e, se dovesse riuscire a svilupparla, ciò potrebbe rappresentare il conseguimento del risultato postosi con la strategia lanciata nel 2018: quello di creare fiducia tra i cittadini europei in un IA che sia affidabile.

## II. QUALIFICAZIONE GIURIDICA DELL'IA

### II.1 Test di Turing

Alla domanda “che cos'è l'IA?” è stata data la risposta: essa è un *automa*<sup>38</sup>.

Qualche lettore potrebbe dissentire, invero non vi è una definizione unanimemente accettata in merito a che cosa sia l'intelligenza artificiale e, come già detto, vi sono diversi orientamenti in materia. Mi sono preso la libertà di utilizzare questo termine piuttosto che un altro, in accordo con le argomentazioni che questa tesi si propone di dimostrare.

In questo secondo capitolo verrà concesso ampio spazio ai ragionamenti che mi hanno convinto ad inquadrare l'intelligenza artificiale in una specifica categoria giuridica piuttosto che un'altra. Il problema di come definire l'IA è un presupposto logico necessario per qualunque discorso sulla sua regolamentazione, ed è per queste esigenze di rigore logico che chiedo al lettore il beneficio del dubbio sull'ipotesi ivi formulata.

Le radici di questo problema affondano nel secolo scorso, dove grazie al prezioso contributo di menti brillanti come K. Gödel, D. Hilbert, B. Russell e A. Turing da un lato e N. Bohr, E. Schrödinger, W. Heisenberg ed A. Einstein dall'altro, si è arrivati a chiedersi quali fossero i limiti del determinismo.

Il determinismo così caratteristico della scienza e della filosofia occidentale è una corrente di pensiero che vide i suoi albori con la teoria dell'atomo del filosofo greco Democrito, successivamente negli anni la sua forma più strutturata divenne quella espressa mediante il *Principio di Causalità*. La convinzione formatasi fu quella che conoscendo le cause fondamentali di un fenomeno si potessero prevederne tutti i possibili effetti.

Tale convinzione fu messa in crisi in diversi ambiti scientifici: fisica, logica, matematica ed in quella che poi sarebbe stata chiamata informatica. Si scopri il Principio di Indeterminazione

---

<sup>38</sup> Vedi Cap I.1, p.8.

che inaugurò la fisica quantistica, venne formulato il Teorema di Incompletezza di Gödel che insieme al Paradosso di Russell minarono alle fondamenta di matematica e logica.

In questo clima rivoluzionario Alan Turing, il padre della moderna informatica, nel suo celebre articolo pubblicato sulla rivista *Mind* nel 1950 ed intitolato “*Computing Machinery and Intelligence*”, si chiese:

“*Can machines think?*”

Da questa domanda è bene che parta ogni possibile discorso in merito a quale possa essere la vera natura dell'intelligenza artificiale.

All'epoca la tecnologia digitale, della quale oggi facciamo uso quotidiano, non era ancora stata sviluppata. Il mondo era da poco uscito dal secondo conflitto mondiale e le macchine che facevano parte della quotidianità delle persone erano *analogiche*.

Era analogica anche la prima formulazione teorica della macchina di Turing, inventata dall'omonimo creatore per la risoluzione dell'*Entscheidungsproblem*<sup>39</sup> (noto anche come problema della decisione o problema dell'arresto), nonché punto di partenza per la comprensione della computabilità, idea centrale del pensiero di Alan Turing e cuore pulsante della tecnologia digitale.

La macchina di Turing è un concetto astratto funzionante grazie a dei numeri chiamati *computabili*<sup>40</sup>. La stessa macchina è stata ideata con lo scopo di definire quali siano questi numeri, descrivendone il procedimento. Nella sua formulazione originaria essa ha una struttura analogica poiché è composta da un nastro scorrevole con delle fessure che vengono lette dalla macchina per la risoluzione di un problema algoritmico.

Turing inventò la macchina omonima nel tentativo di risolvere il problema della decisione, senza però trovarvi una soluzione. Cercando di sciogliere questo problema scoprì l'esistenza

---

<sup>39</sup> Una descrizione completa di che cosa sia il problema della decisione viene data dal fisico R. PENROSE in *La mente nuova dell'Imperatore*, trad. di L. Sosio, BUR Rizzoli, Milano, 1992, p. 56- 107. Cit. “*il problema era: esiste un qualche procedimento meccanico generale in grado, in linea di principio, di risolvere uno dopo l'altro tutti i problemi della matematica?*” [...] “*una parte della difficoltà nel rispondere a questa domanda consisteva nel decidere che cosa si dovesse intendere per procedimento meccanico*” p.61.

<sup>40</sup> A. M. TURING, *On computable numbers, with an application to the Entscheidungsproblem*, 1936.

di una classe di numeri, chiamati computabili e decritti formalmente con il codice binario (0/1), inaugurando così di fatto la scienza informatica.

È noto, infatti, che i computer che utilizziamo comunemente devono la loro denominazione proprio alla computabilità. Questa, come procedimento, è affine al ragionamento ma, sotto una prospettiva matematica, se ne discosta parzialmente e ciò è il risultato conseguito da Turing nel tentativo di risolvere il problema posto da David Hilbert.

Sui numeri computabili Turing si esprime così:

*“The "computable" numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means. [...] According to my definition, a number is computable if its decimal can be written down by a machine.”*<sup>41</sup>

L' idoneità di questi numeri ad essere scritti da una macchina risponde all' esigenza di definire cosa sia un procedimento meccanico, ai fini dell' *Entscheidungsproblem*.

La macchina di Turing è una macchina teorica in grado di rappresentare e portare a termine qualunque operazione meccanica, essa si riferisce a tutte quelle operazioni che possono essere descritte come computabili o algoritmiche.<sup>42</sup>

Nei termini di questa tesi, riprendendo le due equazioni riportate nel capitolo primo, si può notare la somiglianza tra i concetti di macchina di Turing e *bounded optimality*, tracciando al contempo una linea di demarcazione netta tra essi ed invece il concetto di razionalità perfetta descritto dalla prima equazione.

Questa distinzione tra razionalità o pensiero da un lato e computabilità dall' altro è il cuore della domanda postasi da Turing nel 1950, nonché il motivo per il quale l' autore preferì sostituire alla domanda *“can machine think?”* la più articolata formulazione del Test di Turing<sup>43</sup>, noto anche come *imitation game*.

Per sapere se una macchina sia capace di pensare, sarebbe necessario definire cosa sia una macchina e cosa sia il pensiero, per l' ambiguità di queste parole, alle quali difficilmente è possibile dare dei contorni precisi, Turing, da rigoroso matematico, preferì elaborare un

---

<sup>41</sup> A. M. TURING, *On computable numbers*, op. cit., p.230.

<sup>42</sup> Sull' argomento R. PENROSE, op. cit. p. 77.

<sup>43</sup> A. M. TURING, *Computing machinery and intelligence*, Mind, 1950.

procedimento che fosse in grado di produrre un qualche risultato reale e non ambiguo sul quale basare una possibile risposta al quesito.

Il test di Turing consiste in un gioco con tre parti: un interrogatore e due interrogati, dei quali uno è un essere umano e la controparte una macchina, più precisamente un *digital computer*. Lo scopo del gioco è quello di convincere l'interrogatore in modi diversi, l'umano della sua natura umana, il computer invece di depistare l'interrogatore e convincerlo di essere lui l'umano. Tutto questo ovviamente senza che vi sia un'interlocuzione diretta tra i partecipanti, ma piuttosto mediante dei canali di telecomunicazione, così che il libero convincimento dell'interrogatore possa basarsi esclusivamente sulle parole scritte dai giocatori.

Il gioco dell'imitazione mette dunque alla prova la capacità dei computer di simulare le capacità umane e, attraverso questo procedimento originale, si chiede quale sia l'elemento di discriminazione tra macchina e uomo e se effettivamente una macchina è in grado di ingannare un uomo circa la sua natura.

La macchina che partecipa al gioco è un computer digitale per un motivo ben specifico. Questo è una delle conseguenze prodotte dalla teorizzazione della macchina di Turing ed essa è un concetto teorico non riproducibile nella realtà, ma il computer digitale nasce grazie a quest'idea. Il principio su cui si basa il funzionamento dei computer digitali, come abbiamo visto, è la computazione.

La chiave interpretativa che suggerisce la sostituzione del primo quesito col secondo posto sottoforma di test, alla luce della precedente considerazione, è che forse questa scelta specifica ulteriormente il quesito che Turing si pone, passando dal chiedersi se una macchina possa pensare al più articolato quesito se computare equivalga a pensare.

A sostegno di questa ipotesi viene in aiuto la terminologia. Nel documento più volte viene ripetuto il termine "*human computer*", un termine attualmente desueto, ma all'epoca parecchio diffuso.<sup>44</sup> Distinguere tra *digital computer* e *human computer* ha come conseguenza quella di mettere al centro la computazione come sinonimo di calcolo. È proprio in tale prospettiva, a mio parere, che l'autore decide di cambiare i termini della questione.

---

<sup>44</sup> Cfr. J. R. SEARLE, *What computer can't do*, The New York Review of Books, vol. LXI n.15, 9.10.2014. p. 2.

È possibile notare come Turing stesso ritenga che la domanda “*can machine think?*” sia priva di senso<sup>45</sup> e preferisca impostare i termini della questione nella forma ormai famosa del gioco dell’imitazione per porsi la domanda, più coerente con la sua formazione ed i suoi scritti precedenti, su quali possano essere i limiti della computazione ed in che misura essa sia capace di eguagliare il pensiero umano.

L’autore dopo aver impostato i termini della questione ed aver espresso il suo punto di vista fa una rassegna delle opinioni contrarie, elencate in nove punti, alcune delle quali possono essere prese in esame ai fini della presente tesi:

- *The Mathematical Objection.*<sup>46</sup>

L’obiezione matematica analizzata da Turing si basa sul primo teorema d’incompletezza di Kurt Gödel.

Quest’ultimo in breve dimostra che in ciascun sistema formale possono esservi degli enunciati che siano veri, ma al contempo non dimostrabili all’interno del sistema formale stesso, senza che esso venga compromesso.

Una formulazione più familiare di questo fenomeno la si può cogliere nel paradosso di Bertrand Russel:

*“L’insieme di tutti gli insiemi che non appartengono a sé stessi appartiene a sé stesso se e solo se non appartiene a sé stesso”*<sup>47</sup>.

L’antinomia che deriva da questa situazione paradossale dimostra l’insufficienza dei sistemi formali rispetto ai cosiddetti enunciati veri ma non dimostrabili. È questa la grande scoperta di Kurt Gödel che mise in crisi le fondamenta della matematica.

Nel contesto del test di Turing, il primo teorema di incompletezza ha delle implicazioni fondamentali: essendo i computer digitali dei sistemi formali, funzionanti grazie alla codificazione delle informazioni attraverso il codice binario, essi stessi saranno

---

<sup>45</sup> “*The original question, “Can machines think?” I believe to be too meaningless to deserve discussion*”, A. M. TURING, *Computing machinery*, op. cit. p. 8.

<sup>46</sup> Qui viene ripresa la nomenclatura originale delle obiezioni analizzate da A. M. TURING in *Computing machinery*, op. cit. p. 14.

<sup>47</sup> Un insieme R, che raggruppi tutti gli insiemi che non appartengono a sé stessi, può appartenere a sé stesso? Se sì allora  $R \in R$ , dato che R appartiene a sé stesso, essendo R l’insieme di tutti gli insiemi che non appartengono a sé stessi, R è un insieme che non appartiene a sé stesso, dunque non appartiene a sé stesso e viceversa. L’ enunciato è dimostrato ma ciò crea un’antinomia che invalida l’intero sistema.

soggetti al limite dei sistemi formali, pertanto tali sistemi saranno incompleti e vi saranno in essi delle asserzioni vere ma non dimostrabili all'interno del sistema stesso. L'implicazione più importante è che invece, gli esseri umani non sono sottoposti a queste limitazioni, ed essendo il pensiero una prerogativa umana, è legittimo chiedersi se dei computer, con le limitazioni derivanti dal primo teorema d'incompletezza, possano ambire a pensare.

Tale interrogativo, che anima il dibattito tra fautori dell'intelligenza forte e debole, viene accennato da Turing il quale, pur affermando che vi può essere una differenza notevole tra capacità di un computer e capacità della mente umana, ribadisce che se viene posta ad una macchina una "*appropriate critical question*" essa darà sempre una "*definite answer*".

Questa risposta potrà essere inadatta rispetto a domande aperte ed ambigue come "*What do you think of Picasso?*"<sup>48</sup> e ciò dimostra un limite delle macchine, che invece non si applica all'intelletto umano. Nonostante ciò, quest'argomento non è idoneo ad invalidare il risultato del test.

Sebbene persista quest'elemento di discriminazione sarà sempre possibile per la macchina, con le limitazioni del Teorema di Gödel e nell'ambito della computabilità, ambire a superare il test.

- *Lady Lovelace's Objection.*

Lady Ada Lovelace, figlia del poeta inglese Lord Byron e vissuta nella prima metà del XIX secolo, viene considerata come la prima programmatrice della storia grazie ai suoi contributi alla macchina analitica di Charles Babbage<sup>49</sup>.

La macchina analitica progettata da Charles Babbage, a sua volta stimato matematico, non vide mai la luce. Ciononostante, viene considerata la prima progettazione di un computer per come lo conosciamo oggi.

Gli studi di questi due grandi matematici furono fondamentali per la nascita futura dell'informatica, e la loro influenza determinante alla luce delle ipotesi avanzate da Turing sui limiti teorici dell'omonima macchina.

---

<sup>48</sup> Lo stesso Turing su tale domanda: "*The questions that we know the machines must fail on are of this Type*", *Computing machinery, op. cit.*, p. 14.

<sup>49</sup> [https://it.wikipedia.org/wiki/Ada\\_Lovelace](https://it.wikipedia.org/wiki/Ada_Lovelace)

In “*Computing machinery and intelligence*” al punto 6 della lista delle opinioni contrarie al possibile superamento del test di Turing da parte di un *digital computer*, l’autore cita una memoria di Lady Lovelace inerente alla macchina analitica di Charles Babbage che recita:

*"The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform".*

Su questo punto emerge in maniera sorprendente la lungimiranza del padre dell’informatica, frutto di una profonda comprensione del potenziale delle sue scoperte.

Alan Turing contrasta facilmente questa obiezione facendo riferimento a delle “*learning machines*”<sup>50</sup>, ovvero delle macchine capaci di svincolarsi dal controllo diretto dell’essere umano e che quindi, mediante meccanismi di apprendimento, diventino capaci di compiere azioni che il programmatore non abbia espressamente previsto.

L’autonomia delle macchine IA è oggi un fatto notorio, che già è stato affrontato al Cap. I.3.1, resta solo da lodare la capacità di Turing di prevedere un fenomeno, per nulla scontato, divenuto realtà solo molti anni dopo le sue ricerche.

○ *The Argument from Informality of Behaviour.*

L’obiezione ivi esaminata parte dall’assunto che sia impossibile codificare il comportamento poiché esso è per sua natura *informale*.

L’autore in proposito scrive di “*rules of conduct*” e “*laws of behavior*” al fine di superare un’obiezione che viene formulata come segue:

*"if each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines."*<sup>51</sup>

Questa obiezione viene avanzata sulla base della constatazione che non vi possano essere *rules of conduct* per coprire ogni eventualità. Ciò giustificherebbe, nella

---

<sup>50</sup> A tale categoria di macchine, che potrebbero corrispondere all’odierna intelligenza artificiale, Turing dedica l’ultimo paragrafo di *Computing machinery and intelligence*, *op. cit.*, p. 21- 26, da ciò potrebbe essere possibile desumere come il vero scopo del test di Turing possa consistere nel mettere alla prova le capacità della computazione.

<sup>51</sup> A. M. TURING, *Computing machinery*, *op. cit.*, p. 19.

prospettiva dell'obiezione presa in esame, la distinzione tra macchine e uomini e che pertanto questi ultimi non possano rientrare nel *genus* delle macchine.

Turing, perciò, presenta anche la categoria delle *laws of behavior*, ovvero le leggi che determinano anche il comportamento degli umani, affermando che, nonostante la solidità di quest'ultima obiezione, possano esservi dei margini anche per la codificazione di queste ultime.

Se anche quest'ultima previsione dovesse dimostrarsi vera in futuro, nonostante l'autore stesso nutra dei dubbi sulla reale fattibilità dell'impresa, ciò potrebbe dimostrare come anche il comportamento umano possa essere formalizzato e se ciò non bastasse a fare rientrare gli uomini nella categoria più ampia delle macchine, sicuramente potrebbe fare vacillare qualunque linea di demarcazione netta.

- *The Argument from Extrasensory Perception.*

Quest'ultimo argomento venne descritto da Alan Turing come convincente nella sua prospettiva, nonostante le sue implicazioni sgradite.

Nel discuterlo Turing si discosta notevolmente dall'impostazione dell'articolo, come qui interpretato, andando al di là del test proposto e dell'analisi dei limiti della computazione, per rivolgersi alla domanda originaria: "*can machine think?*".

Come abbiamo visto nel primo capitolo vi sono stati grandi pensatori che sono stati capaci di codificare il pensiero, come Aristotele ed il suo sillogismo, la stessa intelligenza artificiale può essere considerata un'approssimazione formale dei fenomeni che accadono dentro le nostre menti. Ma, nonostante ciò, nessuno è mai riuscito a definire con assoluta certezza cosa sia il pensiero e da dove provenga.

È sulla base di quest'incertezza scientifica che Turing avanza l'ultima obiezione alla sua ipotesi, ovvero che il pensiero possa essere un ESP (*extra sensory perception*).

È sicuramente singolare leggere di un uomo di scienza che avanzi una proposta che contrasta fortemente con le fondamenta del metodo scientifico, osservazione e sperimentazione, ma proprio per questo forse merita di essere presa in considerazione molto seriamente nel dibattito odierno sui limiti e le potenzialità dell'intelligenza artificiale.

L'articolo si conclude con la descrizione di quelle che l'autore definisce *learning machines* (per la quale rimando il lettore al cap I.3.1).

Ad oggi non ci sono delle IA che abbiano passato il Test di Turing, soprattutto perché quest'ultimo negli anni è mutato, ne sono state proposte diverse varianti rispetto all'originale che è stato considerato troppo generico. Le più attuali tecnologie in materia di IA ad oggi vengono sviluppate incentrandosi proprio sulla capacità di interpretazione ed uso del linguaggio naturale<sup>52</sup>, forse grazie alla sfida intellettuale lanciata da Alan Turing.

L'interpretazione ivi presentata, basata sul concetto di computazione, pone l'accento sulle capacità pratiche dell'IA di poter superare il test, *simulando* il comportamento umano e riproducendone l'uso del linguaggio naturale.

L'attuale sperimentazione scientifica sembra andare in questa direzione e non vi è motivo di dubitare che prima o poi sarà possibile costruire un'IA capace di superare il test di Turing così formulato. Ma la domanda "*can machine think?*" ha generato un più profondo dibattito che ha coinvolto diversi ambiti del sapere e che di recente si è incardinato su due posizioni: sostenitori dell'IA forte e sostenitori dell'IA debole.

## II.2 Un problema giuridico: IA forte vs IA debole

La risposta alla domanda postasi da Turing sarebbe idonea ad influire sulla diversa qualificazione giuridica dell'IA in un senso o nell'altro.

Da un lato c'è chi vede nelle potenzialità dell'intelligenza artificiale combinata alla robotica la possibilità di costruire androidi, macchine non solo capaci di simulare il comportamento umano, ma anche capaci di svilupparne uno proprio in maniera autonoma (*strong AI*). Sul versante opposto (*weak AI*), chi invece individua nelle obiezioni del paragrafo precedente dei limiti teorici insuperabili.

---

<sup>52</sup> IA di questo tipo sono già presenti sul mercato come Siri, Cortana ed Alexa, operatori vocali che eseguono le operazioni richieste dall'utente e con le quali si può parzialmente dialogare.

Sul fronte del diritto la prima questione da affrontare attiene alla soggettività, in tal senso è utile chiedersi: qual è il presupposto, quali sono le prerogative, per essere considerato dall'ordinamento giuridico un soggetto di diritto?

Abbiamo visto nell'introduzione come l'ordinamento giuridico italiano sappia adattarsi a difficoltà derivanti dai diversi status di *incapace*, *animale* e *società*. In ciascuno di questi casi opera una *fictio iuris*: si considera soggetto chi in realtà non lo è, la società, e all'opposto si considera estraneo alla categoria giuridica della soggettività chi invece, come l'incapace o l'animale, è capace di compiere azioni produttive di effetti causali ma non giuridicamente riconducibili a loro stessi.

Ne risulta un quadro composito in cui la categoria della soggettività viene utilizzata strumentalmente per adempiere a esigenze di diversa natura, sempre nel segno della certezza del diritto. Ciononostante, è possibile individuare una base comune sulla quale poggia il senso della soggettività.

Se ci si pone nell'ottica della certezza del diritto come esigenza fondamentale della disciplina giuridica, la scelta di non attribuire la titolarità di diritti e doveri a soggetti la cui volontà e cognizione è *opaca*, vuoi per dei vizi della mente o per la natura animale, appare più che sensata.

È pacifico che nel giudizio di cognizione il giudice debba estrapolare dalla realtà dei fatti quelli che appaiano rilevanti ai fini del processo e, applicando le norme, giudicare sull'oggetto della controversia. Ma come potrebbe ricostruire la realtà dei fatti il giudice, se dovesse trovarsi ad interrogare un infermo o un animale?

Allo stesso modo si può comprendere come invece possa risultare agevole per il giudice interrogare l'amministratore delegato di una società, incaricato di rappresentare quella collettività di umani unitasi nel segno della società per svolgere una qualsiasi attività.

Si può dire che, alla luce delle considerazioni esposte, la soggettività giuridica assolva nel nostro ordinamento la funzione di ricondurre un determinato atto o diritto al legittimo titolare, rendendolo *imputabile* ed anche *responsabile* per lo stesso, garantendo così la certezza del diritto.

Se invece ci si vuole interrogare in senso più ampio su quale sia la nozione di soggetto, si dovrebbe cercare al di là del perimetro della disciplina giuridica.

Di questo si occupano i fautori della diatriba tra IA forte ed IA debole, chiedendosi se l'IA potrà avere in futuro una coscienza ed una volontà proprie. Una delle considerazioni più note, nonché variante del test di Turing, è stata formulata dal filosofo John Rogers Searle.

L'argomentazione della stanza cinese è una delle più adatte a trattare la tematica della soggettività dell'IA ai fini di una sua regolamentazione. In questa prospettiva diventa possibile cogliere un nuovo modo di affrontare la questione, definendo l'intelligenza artificiale in termini computazionali e, per sottrazione, diventa anche possibile comprendere qualcosa di più su noi stessi.

## II.2.1 L'argomento della stanza cinese

L'argomento della stanza cinese venne introdotto per la prima volta dal filosofo americano John Searle nel 1980 in *"Minds, Brains, and Programs"* per contrastare una visione computazionale della mente umana.

L'argomentazione è fondata su solide basi che già negli anni successivi alla sua pubblicazione sollevarono un grande dibattito. Per il filosofo americano un'IA può essere capace di simulare il comportamento umano grazie alle regole formali della *sintassi*, ma, mancando in essa la comprensione della *semantica*, non avrebbe mai potuto raggiungere un'effettiva conoscenza della realtà, discostandosi sotto quest'aspetto dalla mente umana<sup>53</sup>.

L'argomentazione viene ivi riportata nella formulazione originale di J. Searle:

*“Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the*

---

<sup>53</sup> Nell'analisi di John Searle la distinzione fra semantica e sintassi è la distinzione tra la capacità di dare un significato a ciò che si dice (oggetto della semiotica, della logica e della psicologia) e la capacità di costruzione della frase mediante un insieme di simboli costituente un linguaggio formale.

*man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.*"<sup>54</sup>

Risulta evidente come, sostituendo all'uomo nella stanza un *digital computer*, sia possibile superare il test di Turing senza che comunque vi sia una effettiva comprensione da parte dell'IA. Infatti, il soggetto nella stanza non ha bisogno di una reale comprensione del cinese per fornire un *output* corretto, ma è sufficiente per lui eseguire un procedimento causale, che nel contesto dell'esperimento mentale è quello di traduzione dal cinese all'inglese e viceversa.

Prima di passare alla disamina delle più rilevanti opposizioni alla tesi di Searle, occorre chiedersi se quest'ultima sia idonea ad invalidare la metodologia del test?

La risposta dipende dallo scopo che il test si propone:

Se lo scopo del test di Turing è quello di rispondere alla domanda originaria ("*can machine think?*") la risposta non può che essere negativa. L'argomento della stanza cinese dimostra come esista la possibilità che un computer, con un'elevata capacità di calcolo, possa essere capace di simulare l'uso del linguaggio naturale mediante la manipolazione sintattica di simboli, dissimulando la propria natura meccanica per depistare l'interrogatore senza che vi sia in esso alcuna reale comprensione del significato delle parole utilizzate.<sup>55</sup>

Dall'altro lato si è visto come Alan Turing adottò questa metodologia proprio per proporre un'alternativa all'ambiguità della domanda originaria. Interpretando il test di Turing come un procedimento per testare i limiti e le potenzialità della computazione, come proposto nei paragrafi precedenti, allora la stanza cinese di Searle non andrebbe a invalidarlo, ma anzi lo arricchirebbe di contenuti.

La soluzione ci viene data dalle motivazioni per le quali John Searle elaborò questa tesi.

I fautori dell'intelligenza artificiale forte vogliono che il test di Turing sia un procedimento per dimostrare come la stessa intelligenza umana sia spiegabile in soli termini computazionali, capovolgendo la domanda originaria in: è la mente umana una macchina?

---

<sup>54</sup> Citato in D. COLE, *The Chinese Room Argument*, The Stanford Encyclopedia of Philosophy, Edward N. Zalta, 2020, URL = <https://plato.stanford.edu/archives/win2020/entries/chinese-room/> .

<sup>55</sup> Quest'interpretazione del test concorda armoniosamente con la sua ulteriore denominazione di "*imitation game*".

Quest'approccio prende il nome di *teoria computazionale della mente* o *approccio operativo*, il suo assunto è che: “*mental states are defined by their causal roles, not by the stuff (neurons, transistors) that plays those roles.*”<sup>56</sup>

La tesi di Searle ha suscitato diverse critiche. Qui vengono riportate le più importanti riproducendo la stessa metodologia adottata per le obiezioni poste al test di Turing:

- *The System Reply.*

Questa replica all'argomento della stanza cinese afferma che sebbene non vi sia da parte dell'uomo nella stanza alcuna comprensione della lingua cinese, esso non è che una parte di un sistema più ampio.

In effetti, prendendo come paradigma un computer digitale moderno, l'uomo nella stanza non sarebbe che la CPU, mera esecutrice delle operazioni algoritmiche. Insieme a questa parte ve ne sarebbero molte altre: la memoria rappresentata dal dizionario, i programmi software rappresentati dal libro di istruzioni, tastiera e mouse (o sensori nel caso di IA robotiche) come input e monitor come output.

La comprensione, secondo i fautori di questa replica, risiederebbe non in una delle singole parti ma nel sistema stesso inteso come *quid pluris* rispetto alla mera somma delle parti del sistema stesso.

Questa prospettiva, in accordo con alcuni assunti della teoria della complessità, è trattata in maniera originale dal biologo James Lovelock in relazione alla vita nei suoi saggi: “*Gaia, A New Look at Life on Earth*” e “*Novacene, The Coming Age of Hyperintelligence*”<sup>57</sup>.

La tesi di *Gaia* propone di considerare la terra come un organismo vivente. In particolare, nel libro viene precisato come il pianeta terra sia, nell'ottica dell'autore, un sistema cibernetico che ha dei meccanismi propri di autoregolamentazione, mentre in “*Novacene*” viene riproposta la medesima prospettiva adattata al fenomeno dell'intelligenza artificiale, ipotizzando una collettivizzazione dell'intelligenza da parte di androidi che sebbene separati fisicamente siano un'unica entità (come nell'esempio dei veicoli autonomi del Cap I.3).

---

<sup>56</sup> Ivi, D. COLE, URL = <https://plato.stanford.edu/archives/win2020/entries/chinese-room/>, *op. cit.*

<sup>57</sup> Cfr. J. LOVELOCK, *Gaia, nuove idee sull'ecologia*, trad. di V. B. Landucci, Bollati Boringhieri, 1981, e *Novacene, l'età dell'iperintelligenza*, trad. di A. Panini, Bollati Boringhieri, 2020.

La proposta di trattare la vita e la coscienza come proprietà emergenti in sistemi dinamici è molto suggestiva, ma la sua applicazione alle tecnologie IA rischia di scontrarsi contro il senso comune. Infatti, una possibile controreplica abbastanza solida proviene direttamente dal primo teorema di incompletezza di Gödel.

Tale posizione sostenuta separatamente da Searle e dal fisico Roger Penrose afferma come l'IA non necessiti di comprensione per il suo funzionamento e la formalizzazione delle istruzioni non sia una prerogativa per il sistema ma piuttosto per l'operatore umano.

*“Formal symbols by themselves can never be enough for mental contents, because the symbols, by definition, have no meaning (or interpretation, or semantics) except insofar as someone outside the system gives it to them”*<sup>58</sup>.

Il sistema semplicemente funziona, esegue operazioni meccaniche alle quali noi attribuiamo il valore formale di 0/1 al fine di poterlo controllare e poterlo programmare per compiere delle azioni tramite esso.

○ *Brain simulator Reply.*

La replica del simulatore del cervello ipotizza che una perfetta riproduzione dei meccanismi fisici coinvolti nell'atto di pensare equivalga all'atto stesso.

La realizzazione di un tale artefatto, grazie alle più moderne tecniche sulle reti neurali, sarebbe capace quindi di riprodurre ogni cambiamento di stato delle sinapsi del cervello umano e pertanto rappresenterebbe perfettamente il pensiero.

Per i sostenitori di questa critica alla stanza cinese l'atto di pensare non si discosterebbe dalla simulazione del pensiero, questi due fenomeni sarebbero in ultima istanza la stessa cosa.

Searle risponde affermando che vi sarebbe una distinzione netta tra la simulazione e la realtà.

In questo si può cogliere, a detta del filosofo, la fallacia delle teorie computazionali della mente umana per le quali sarebbe sufficiente riprodurre dal punto di vista

---

<sup>58</sup> citazione di J. R. SEARLE in D. COLE, URL = <https://plato.stanford.edu/archives/win2020/entries/chinese-room/>, *op. cit.*

meccanico tutti quei processi materialmente coinvolti nell'atto di pensare per ricreare anche una effettiva comprensione da parte della macchina.

Pertanto, nella prospettiva di Searle, le macchine anche se replicanti perfettamente il comportamento umano, sarebbero sempre un *quid minus* di fronte alla nostra capacità di pensare.

In esse verrebbe a mancare una caratteristica fondamentale: l'intenzionalità.

Quest'ultima è una caratteristica volitiva propria del pensiero e si esprime nella capacità che esso ha di riferirsi a qualcosa contestualizzando così l'oggetto del discorso e dando luogo alla semantica, assente invece nella simulazione effettuata dalle macchine.

Vengono così a costituirsi due categorie distinte: da un lato le *prestazioni computazionali* di un artefatto, che danno luogo ad una perfetta riproduzione del pensiero umano mediante la manipolazione di simboli in accordo con le regole della sintassi; dall'altro le reali *prestazioni del pensiero*, in cui la comprensione sarebbe data dall'intenzionalità intesa come la capacità di dare un significato a ciò che si dice mediante il riferimento diretto a qualcosa di esterno rispetto al soggetto, di cui la parola come fenomeno sintattico ne è solo una rappresentazione formale.

Sotto un profilo sostanziale, la controreplica di Searle afferma che la macchina sia soggetta all'attività volitiva dell'uomo e che quindi ne sia strumento. In questa prospettiva l'intenzionalità che erroneamente potrebbe essere attribuita alla macchina non sarebbe che una forma di *derivated intentionality* da attribuire a chi ha l'ha messa in moto.

- *The Other Minds Reply.*

Questa replica è fondata sulla *ratio* del test di Turing.

Questo, infatti, è stato concepito come un test comportamentale per avere un qualche risultato tangibile ed inequivoco, data l'esigenza del metodo scientifico di una valutazione di fatti oggettivi.

Il comportamento viene quindi visto come unica prova della coscienza del soggetto che lo realizza, constatando così l'impossibilità di riconoscere la soggettività, o la

presenza di una mente, in qualcun altro se non basandosi su una qualche estrinsecazione di quest'ultima.

Pertanto, seguendo questa logica si arriva alla conclusione che basterebbe porre in essere un comportamento idoneo ad essere giudicato come cosciente per attribuire una qualche forma di soggettività all'altro.

In verità è evidente la somiglianza tra questa replica e la precedente, in quanto entrambe si fondano sulla medesima *ratio* per la quale è necessario avere una prova tangibile della coscienza al fine di poterla riconoscere.

La questione è di natura epistemologica, ci si chiede infatti cos'è la coscienza e come la si può dimostrare? La risposta di Searle si fonda quasi interamente su un'*intuizione*: noi attribuiamo ad un soggetto una volontà propria sulla base delle manifestazioni che di essa possiamo percepire.

L'intenzionalità non è idonea ad essere provata mediante un procedimento meccanico, ma piuttosto tendiamo ad attribuirle in modo innato per l'esperienza data dal senso comune. Nella nostra vita quotidiana abbiamo a che fare quotidianamente con soggetti e siamo perfettamente in grado di riconoscere la differenza tra soggetto ed oggetto.

La risposta di Searle apparentemente sembra portare ad un'impasse, scontrandosi contro la richiesta originaria del test di Turing di una qualche prova tangibile del pensiero.

La prova in effetti c'è, anche se non dimostrabile mediante qualche rigoroso procedimento meccanico. Infatti, sebbene l'intenzionalità sia un: "*intrinsic phenomenal character that is inherently intentional*"<sup>59</sup> noi ne possiamo sperimentare l'esistenza grazie all'esperienza quotidiana.

C'è chi vede tra le argomentazioni di Turing e Searle una divergenza incolmabile. Questa è data dalla diretta conseguenza dell'inattendibilità del test di Turing come metodo per dimostrare l'esistenza del pensiero, derivante dalle conclusioni della stanza cinese di Searle sull'intenzionalità. C'è però anche un'altra prospettiva.

---

<sup>59</sup> Citazione di T. HORGAN in D. COLE, URL = <https://plato.stanford.edu/archives/win2020/entries/chinese-room/>, *op. cit.*

Searle usa queste parole per spiegare cosa sia nella sua visione ciò che egli ha definito col nome di intenzionalità:

*“the causal powers of the human brain”<sup>60</sup>*

Dove per potere causale della mente si intende la capacità di dare origine ad un'autonoma catena causale di eventi, rispetto alla quale l'uomo è il punto d'inizio. Questa formulazione sembra discendere direttamente da un'altra:

*“[...] one fears that thinking is just the kind of phenomenon where ESP may be especially relevant.”<sup>61</sup>*

Il punto di convergenza tra i due documenti *“Minds, brains and programs”* e *“Computing machinery and intelligence”* potrebbe essere proprio questo.

L'obiezione derivante dall'ipotesi delle *extrasensory perception* è alquanto singolare, nonostante ciò, Turing la prese in considerazione come possibile elemento ostativo al test. Egli ipotizzò che un soggetto potesse avere poteri telecinetici capaci di influire sul buon funzionamento della macchina, situazione potenzialmente superabile tramite una camera idonea ad isolare la macchina evitando l'influenza telecinetica dell'avversario.

Turing, nel prendere in considerazione la possibilità che un soggetto possa avere poteri telecinetici, implicitamente ammette che le menti possano aver qualche potere di natura causale. Se la forma fosse quella dell'intenzionalità non si tratterebbe di un qualche potere mistico, ma piuttosto di una caratteristica propria del pensiero umano e perfettamente nota al diritto. D'altronde lo stesso Turing ammise che quest'eventualità era possibile proprio perché la mente umana è un campo inesplorato. Solo di recente il dibattito si è intensificato ed è iniziata una fruttuosa ricerca scientifica della quale il Test di Turing e la stanza cinese sono un esempio.

Nel contesto dell'interpretazione qui presentata, le conseguenze dell'intenzionalità come potere causale della mente non sarebbero tali da invalidare il lavoro svolto fin oggi nel campo dell'intelligenza artificiale.

---

<sup>60</sup> Più volte ripetuto in J. R. SEARLE, *Minds, brains and programs*, Behavioral and Brain Sciences, 3: 417–57, 1980.

<sup>61</sup> A. M. TURING *Computing machinery, op. cit.*, p. 21.

Se Alan Turing, decidendo di sostituire alla domanda originaria la metodologia del test, si fosse effettivamente chiesto se computare equivalga a pensare, le conseguenze della stanza cinese gli darebbero parzialmente ragione come lo stesso Searle afferma concludendo “*Minds, brains and programs*”.<sup>62</sup>

Infatti, una parte del nostro pensiero consiste in meccanismi computazionali che gli algoritmi d'intelligenza artificiale sono perfettamente in grado di riprodurre. Ci sono già esempi di algoritmi in grado di performare in attività tipiche dell'essere umano. Questi algoritmi hanno capacità di calcolo enormi e sono in grado di replicare delle conversazioni che, grazie alle distanze delle comunicazioni telematiche come le chat, possono indurre chiunque a pensare che si tratti di una persona.

Il problema di questi algoritmi è che però essi non possiedono una propria volontà, sono piuttosto il prodotto della nostra. Usando la terminologia adottata da Searle, questi algoritmi non possono avere desideri, credenze o intenzioni perché non sono soggetti, non hanno intenzionalità e questa caratteristica del pensiero umano non è riproducibile solo mediante regole e meccanismi formali.

Entrambe le posizioni possono così essere conciliate, da un lato un *digital computer* potrà passare il test di Turing, non c'è motivo di dubitare che con un adeguata potenza di calcolo potrà essere nelle condizioni di performare in maniera eccellente anche in una conversazione così come hanno fatto Deep Blue e Alpha zero rispettivamente negli scacchi e nel go.

Rispetto alla domanda “*can machine think?*”, invece, le argomentazioni sull'intenzionalità ci dicono che sì, una macchina può essere capace di pensare per noi, lo fanno le calcolatrici e molti altri dispositivi. Ma se ci si chiede invece chi è il soggetto che pensa, non si potrà attribuire alcuno stato intenzionale all'algoritmo. Basti chiedersi chi si imputò il successo della vittoria su Garry Kasparov, Deep Blue o il team di ingegneri che lo costruì?<sup>63</sup>

Il test di Turing è valido e la domanda di principio è confutata, ciò che invece è da negare è la soggettività della macchina. Questo è diventato in seguito il reale punto di confronto sebbene nella formulazione di “*Computing machinery and intelligence*” il piano del discorso

---

<sup>62</sup> “*Could a machine think? The answer is, obviously, yes. We are precisely such machines.*”, J. SEARLE, *Minds, brains and programs*, op. cit., p. 422.

<sup>63</sup> Cfr. J. R. SEARLE, *What your computer can't know*, op. cit., p.1.

sia incentrato sulle capacità di una macchina di performare in una conversazione, ponendo l'accento sulle potenzialità dei *digital computer* piuttosto che sulla loro soggettività.

L'intelligenza artificiale altro non sarebbe che un'estrinsecazione del potere causale delle nostre menti:

*“Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output”<sup>64</sup>*

Sul piano giuridico le conseguenze sono notevoli.

Il vero quesito diventa quindi: da dove proviene la volontà e la natura soggettiva degli esseri viventi? Per i fautori dell'IA forte basterebbero dei meccanismi computazionali. Searle ne identifica la fonte nell'intenzionalità, un'intuizione non ben definita. Chi ha ragione? Da dove proviene l'intenzionalità?

Queste domande sono destinate a rimanere per il momento senza alcuna risposta, c'è però un dato di fatto. La distinzione tra soggetto ed oggetto, sebbene relativa, è reale e nel rapporto tra uomo e macchina risulta evidente come la seconda sia al servizio del primo in un rapporto di subordinazione come quello che vi è proprio tra soggetto ed oggetto, in cui il primo si serve del secondo.

Forse non sarà possibile dimostrare l'origine della volontà o dell'intenzionalità, di contro però sembra essere dimostrabile l'insufficienza delle attuali tecnologie a spiegare od anche solo riprodurre il fenomeno della soggettività. Bisogna prenderne atto ed accettare questo mistero.

La finalità di questo capitolo, come anticipato all'inizio, era quella di motivare la scelta di definire l'intelligenza artificiale un automa. Questa parola ben si presta per spiegare sinteticamente ciò che ad oggi l'IA rappresenta.

È stato detto nel capitolo I come le capacità dell'IA vadano oltre il controllo diretto dell'operatore grazie alla sua *autonomia* e di come l'*opacità* dei meccanismi interni di funzionamento possa indurre a pensare che agisca *proprio motu*.

---

<sup>64</sup> J. R. SEARLE, *Minds, brains and programs*, op. cit., p. 422.

Si è visto in questo capitolo come l'IA sia un prodotto della creatività umana, come essa non abbia una propria volontà ed i suoi spazi di autonomia siano limitati da ciò che è programmata per fare. Si è affermato in ultimo come non solo essa sia un artefatto umano, ma che il suo stesso agire sia determinato in parte dal potere causale della mente umana qui definito col nome di intenzionalità.

È bene soffermarsi in ultimo su questa nozione per capirne il significato e comprendere così, come dalla natura di automa dell'IA, si possa arrivare ad una forma di responsabilità nuova ma già nota al diritto.

## II.2.2 L'intenzionalità nel diritto positivo

La Teoria dell'Intenzionalità venne elaborata da John Searle con la finalità più ampia di creare una Teoria generale dell'Azione. Lo scopo della prima era quello di creare i presupposti per la seconda, mediante un'attenta analisi dei rapporti reciproci tra *intenzioni* ed *azioni*.

Nel far ciò Searle distingue tra: *prior intentions* e *intentions in action* da un lato e tra *intentional action* e *unintentional actions* dall'altro.<sup>65</sup>

Nel precedente paragrafo sono già stati presentati gli *stati intenzionali* che abbiamo descritto come quelli diretti o riguardanti qualcosa di esterno rispetto al soggetto che li manifesta. In questa categoria rientrano: pensieri, desideri, ambizioni, intenzioni e paure.

Searle dice che le intenzioni si differenziano dagli altri stati intenzionali per il loro rapporto privilegiato con le azioni, dove non v'è azione senza intenzione. In questa prospettiva l'intenzione si colloca come antecedente dell'azione, nonché come causa, rappresentante la realizzazione del contenuto dello stato intenzionale.

Un'ulteriore distinzione può essere fatta tra stati intenzionali e *prior intentions*, dove i primi sono capaci di animare un soggetto ed i secondi invece consistono nella rappresentazione dell'azione che si vuole realizzare per la soddisfazione dello stato intenzionale. Usando la

---

<sup>65</sup> J. R. SEARLE, *The Intentionality of Intention and Action*, University of California, Cognitive Science 4, 47-70, 1980.

terminologia giuridica possiamo individuare una forma di *prior intentions* nella categoria penalistica della premeditazione.

Questa specificazione è importante per distinguere ulteriormente le *prior intentions* dalle *intentions in action* perché non tutte le azioni intenzionali sono premeditate. Ad esempio, Tizio può pianificare di uccidere Caio con un martello, se ci riesce questa azione sarà intenzionale e al contempo soddisferà la sua precedente intenzione di ucciderlo con un martello. Ma se invece Tizio, incontrando per strada Caio, decide che è il momento giusto per ucciderlo e lo investe con la macchina, si realizza così solo un'*intention in action*, cioè un'azione dolosa per il diritto penale ma senza alcuna premeditazione, ovvero senza la soddisfazione della *prior intention* di ucciderlo col martello.

Infine, una distinzione netta vi è tra le *intentional actions* e le *unintentional actions*. Le prime sono azioni realizzate volontariamente, le seconde no, ma nonostante la loro accidentalità anche le seconde possono far sorgere in capo al soggetto che le attua una forma di responsabilità giuridica.

Infatti, se Tizio anziché uccidere Caio, per errore uccide Sempronio, sarà processato per omicidio colposo. La distinzione tra dolo e colpa è emblematica per distinguere la natura e gli effetti differenti che scaturiscono da *intentional* e *unintentional actions*. Nel caso di omicidio doloso Tizio compie la sua intenzione di uccidere Caio ed è responsabile a titolo di dolo, mentre nel caso di omicidio colposo Tizio uccide Sempronio per un errore che comunque va ricondotto ad una azione intenzionale di Tizio e per questo sarà responsabile a titolo di colpa.

Nell'affermare che non vi sono azioni senza intenzioni, Searle suggerisce una via per interpretare la responsabilità. Infatti, a discapito di ciò che si possa pensare, anche le *unintentional action* sono riconducibili ad un'intenzione come nel caso della responsabilità colposa: "an unintentional action is an intentional action which has aspects which were not intended in it".<sup>66</sup>

La presenza dell'intenzionalità nel diritto penale, alla luce delle categorie penalistiche ivi evocate, può essere così riassunta:

---

<sup>66</sup> J. R. SEARLE *The Intentionality*, op. cit., p. 67.

INTENZIONALITA'		DIRITTO POSITIVO (penalistico)
Intentional action	Prior intentions	Art. 577 n. 3 c.p. Aggravanti concernenti l'elemento soggettivo del reato: L'aver commesso il fatto con <u>premeditazione</u>
	Intentions In Action	Art. 43.1 c.p. "il delitto è <u>doloso</u> , o <u>secondo l'intenzione</u> , quando l'evento dannoso [...] è dall'agente preveduto e voluto come conseguenza della propria azione."
Unintentional action		Art. 43.2 c.p. "il delitto è <u>preterintenzionale</u> , o <u>oltre l'intenzione</u> , quando dall'azione deriva un evento [...] più grave di quello voluto."
		Art.43.3 c.p. "il delitto è <u>colposo</u> , o <u>contro l'intenzione</u> , quando l'evento, anche se preveduto, non è voluto dall'agente."

Queste circostanze mettono in risalto la preponderanza dell'elemento volitivo e intenzionale dell'agente che, insieme all'accertamento del *rapporto di causalità* tra l'azione dell'agente e l'evento illecito *ex art. 40 c.p.*, rende passibile il soggetto di essere giudicato responsabile e quindi reo per l'aver commesso un'azione penalmente sanzionabile secondo la legislazione penalistica italiana.

La centralità dell'intenzione per spiegare l'azione è elemento cardine del processo penale. Si è visto infatti come l'antecedenza logica dell'intenzione rispetto all'azione è fondamentale per graduare la responsabilità del soggetto, che potrà essere per dolo, per colpa o per delitto preterintenzionale<sup>67</sup> *ex art. 42 c.p.* con diverse conseguenze sul piano giuridico.

L'intenzionalità gioca anche un ruolo chiave per la nozione giuridica di imputabilità, la quale presenta connotazioni affini all'elemento soggettivo e psicologico del reato, discostandosene per la sua natura di categoria più ampia e generale del diritto in quanto riferibile all'intero *corpus* normativo dell'ordinamento giuridico.

---

<sup>67</sup> È dubbia invece la possibilità individuare una forma di responsabilità oggettiva, cioè che prescinde dal dolo o la colpa, dopo l'affermazione del principio di colpevolezza da parte della Corte costituzionale avvenuta con la sentenza n.364 del 1988.

L'imputabilità, della quale si è già detto in precedenza, costituisce l'idoneità ad essere considerati soggetti di diritto, generalmente identificabile con la formula: "*capace di intendere e di volere*"<sup>68</sup> dove entrambi i termini possono essere riferiti alla teoria dell'intenzionalità. Infatti, sebbene si possa obiettare che vi siano dei casi in cui vi può essere volizione senza cognizione, le categorie intenzionali presentate in precedenza tengono in considerazione anche queste ipotesi, riportandole alla responsabilità del soggetto mediante il quale si verificano.

Il riferimento dell'intenzionalità alla secolare categoria giuridica dell'imputabilità ci pone di fronte ad una domanda altrettanto antica: Qual è il fondamento dell'imputabilità?

In un primo momento si potrebbe affermare che l'intenzionalità stessa possa essere il fondamento dell'imputabilità. Questa prospettiva verrebbe sostenuta dalla statuizione dell'art. 85.2 c.p. che individua nella capacità d'intendere e di volere il fondamento dell'imputabilità, questo basterebbe ad affermare l'inidoneità dell'intelligenza artificiale ad essere inquadrata nella categoria dei soggetti imputabili, tenuto conto delle conseguenze della stanza cinese.

Ma il vero cuore della questione è un altro. Le teorie computazionali della mente umana poggiano interamente sul determinismo, negando totalmente il libero arbitrio. D'altro canto, secondo il manuale di Antolisei<sup>69</sup>, la teoria più antica sul fondamento dell'imputabilità lo individua nel libero arbitrio, rischiando così di incardinare la questione su due posizioni inconciliabili: Determinismo e Libero Arbitrio.

La questione di natura epistemologica che attraversa l'intero capitolo è proprio questa, lo si può scorgere in tutte le tappe fin qui percorse: computazione ed intenzionalità, IA forte ed IA debole ed infine determinismo e libero arbitrio. La stessa domanda su cosa sia l'intelligenza artificiale, per la sua natura, cela in sé questo profondo dilemma.

L'intenzionalità come chiave di lettura dimostra: da un lato come sia imprescindibile il libero arbitrio per l'imputabilità, chi può essere responsabile delle proprie azioni se non può esserne padrone? Dall'altro invece come, nell'attuale legislazione, il libero arbitrio non sia il presupposto della stessa.

---

<sup>68</sup> Art. 85.2 c.p. "Capacità d'intendere e di volere": È imputabile chi ha la capacità d'intendere e di volere.

<sup>69</sup> F. ANTOLISEI, *Manuale di Diritto Penale, parte generale*, ottava ed., Giuffrè, 1989, p. 513-520.

A far luce su questo mistero sono i fenomeni che se ne collocano fuori: minore età, infermità di mente, natura animale.

Di questi fenomeni si è già parlato nell'introduzione in termini generali, poi in relazione alla natura di soggetto ed alla categoria di soggettività<sup>70</sup> affermando che la soggettività assolve la funzione di garantire la certezza del diritto nel nostro ordinamento.

Sull'argomento l'opinione data dal manuale Antolisei è più che autorevole. Si afferma infatti come il principio della teoria classica, il libero arbitrio, dev'essere ammesso, in quanto per imputare una pena è necessario che il reo avesse al momento dell'azione la possibilità di scegliere di porre in essere una condotta diversa da quella concretamente realizzata. Ma individuare nel libero arbitrio il fondamento dell'imputabilità sarebbe erroneo quanto identificarlo nell'intenzionalità.

Difatti sia dei minori che degli infermi di mente, ed al loro pari degli animali, si può dire che difettino della libertà d'agire? No. L'intelligenza artificiale gioca un ruolo fondamentale nella comprensione del problema; infatti, se la prima manca di intenzionalità per le motivazioni sopra esposte, per le stesse si deve necessariamente riconoscere che le forme viventi invece ne sono provviste.

Un animale è capace di porre in essere un comportamento libero e parzialmente svincolato dall'inerzia causale e come lui anche i minori e coloro i quali sono affetti da patologie psichiche che ne alterano volere e cognizione.

In tutti questi casi l'elemento preponderante dell'azione è la sua qualificazione in termini di *unintentional action* che, in casi diversi da questi, può essere fonte di responsabilità colposa e pur sempre riconducibile alla volontà dell'individuo. In questo quadro non solo si deve affermare la libertà di queste categorie di soggetti, ma al pari di essa se ne deve anche rimarcare la dignità. Infatti, gli animali da tempo subiscono gli effetti dei nostri ordinamenti giuridici, venendo considerati sin dai tempi del diritto romano come *res*.

I nostri ordinamenti giuridici si sono sviluppati al punto di affermare il *diritto alla vita* per ciascun essere umano all'art. 3 della Dichiarazione universale dei diritti dell'uomo<sup>71</sup> e risulta

---

<sup>70</sup> Vedi Cap II.2

<sup>71</sup>United Nations, Universal Declaration of Human Rights, Art. 3 "Everyone has the right to life, liberty and security of person".

ignoto il motivo per il quale gli animali invece non dovrebbero godere della stessa tutela tenuto conto della loro libertà e dignità di creature viventi con sentimenti e volontà, cosa che non si riscontra nell'IA.

Per queste motivazioni il fondamento dell'imputabilità non può essere il libero arbitrio, non si spiegherebbero infatti le posizioni delle categorie di soggetti che ne restano fuori. Il fondamento dell'imputabilità di un individuo dev'essere piuttosto ricercato in ciò che Antolisei definisce come *concezione comune della responsabilità umana*.

Su questo concetto:

*“Affinché un uomo possa essere chiamato a rispondere dei propri atti di fronte alla legge penale è necessario che sia in grado di rendersi conto del valore sociale degli stessi e non sia affetto da anomalie psichiche che gli impediscano di agire come dovrebbe”.<sup>72</sup>*

Ciò che viene messo al centro è il diritto concepito come un fenomeno sociale, strumento per regolare i reciproci rapporti pacificamente da un lato e conoscenza codificata dalle precedenti generazioni a presidio del vivere comune dall'altro.

In questo contesto l'animale è fuori dalla nozione di imputabilità per la sua impossibilità di comprendere, delle sue azioni risponderà il padrone. L'infante ha troppa poca esperienza per essere cosciente di tutti gli effetti delle proprie azioni e del loro valore, a differenza del minore con età compresa tra i quattordici ed i diciotto anni la cui imputabilità dovrà esser valutata caso per caso dal giudice. Infine, gli infermi di mente quali benefici potrebbero trarre da una pena della quale non possono comprendere il significato?

Di fronte alla sapienza millenaria del diritto la paventata intelligenza artificiale forte non è che un nuovo fenomeno circa il quale c'è da chiedersi: come la si può regolamentare?

La risposta, qualunque essa sarà, dovrà tenere conto di due prerogative fondamentali:

- L'imputabilità.
- La certezza del diritto.

---

<sup>72</sup> F. ANTOLISEI *Manuale di Diritto Penale, op. cit.*, p.519.

Se la prima può dirsi risolta in senso negativo, algoritmi di intelligenza artificiale non possono dirsi esseri viventi e pertanto non sono idonei ad essere imputabili, la seconda questione resta aperta.

La certezza del diritto è lo scopo fondamentale che la legge ha nella nostra società, un diritto solido e certo è in grado di garantire rapporti pacifici tra consociati e dunque anche una società prospera. È per questa seconda ragione che sono state proposte due soluzioni suggestive: una forma di personalità elettronica funzionale dell'IA o una forma di responsabilità oggettiva per il danno commesso dall'IA.

In entrambi i casi opererebbe una *fictio iuris* di natura diversa. Nel primo la finzione sarebbe circa l'elemento psicologico, si attribuirebbe l'intenzione ad un artefatto che nella realtà non ne ha nessuna. Nel secondo la finzione inficerebbe il rapporto di causalità tra azione e danno. Infatti il danno, del quale secondo questa ipotesi dovrebbe rispondere il produttore o l'operatore, non sarebbe causato da quest'ultimo, ma piuttosto dall'algoritmo<sup>73</sup>.

In entrambi i casi si arriva ad una forzatura, che in qualche misura si discosta dalla realtà della situazione. La problematica maggiore è la novità di questo fenomeno di cui non si conoscono bene i termini.

Nell'ambito di questa tesi si è provato a presentare gli elementi essenziali della questione senza pretesa di esaustività, e con la medesima metodologia mi appresto a proporre nelle pagine che seguono la mia personale visione delle cose.

### II.2.3 L'IA come *res*

Che cos'è dunque l'IA?

Si è visto come l'antecedente storico dell'intelligenza artificiale sia la macchina di Turing, ideata per risolvere l'*Entscheidungsproblem*, il cui scopo era cercare di rappresentare un procedimento meccanico ideale per risolvere tutti i problemi della matematica.

---

<sup>73</sup> L'art. 40 c.p. afferma che: "Nessuno può essere punito per un fatto preveduto dalla legge come reato, se l'evento dannoso o pericoloso, da cui dipende l'esistenza del reato, non è conseguenza della sua azione od omissione".

L'insufficienza di tale metodo per risolvere il problema posto da David Hilbert ha portato a chiedersi cosa fosse la computazione e di contro cosa il pensiero. Si è arrivati così al test di Turing e dopo anni alla stanza cinese ed all'intenzionalità di Searle.

Questo è il contesto nel quale va, ed in questa tesi è stata ricercata la vera natura della tecnologia che oggi noi chiamiamo intelligenza artificiale.

Quest'ultima può essere definita come la più fedele approssimazione di ciò che chiamiamo procedimento meccanico, rappresentato concettualmente dalla macchina di Turing. Il punto successivo è stato chiedersi se è il pensiero stesso un procedimento meccanico?

La risposta che si deve dare a questo quesito è negativa. Le conseguenze derivanti dal primo teorema di incompletezza di Gödel e dall'intenzionalità di Searle lo dimostrano. L'atto del pensare richiede un soggetto coinvolto nell'azione e, ad oggi, non possiamo considerare l'IA come il soggetto che pensa. Piuttosto la si può descrivere come lo strumento al quale noi deleghiamo i ragionamenti che siamo in grado di formalizzare.

Per queste motivazioni mi sembra ragionevole supporre che l'intelligenza artificiale possa collocarsi nel nostro ordinamento giuridico come *species* del più ampio *genus* delle *res*.

Tale categoria di natura romanistica è amplissima, viene detta *res* una parte di materia, essa si differenzia dal bene poiché: "*sono beni soltanto le cose che possono formare oggetto di diritti*"<sup>74</sup>. È indubbio che ad oggi gli algoritmi di intelligenza artificiale siano anche dei beni, in quanto le aziende digitali possono vantare dei diritti reali di proprietà su di essi, che rappresentano per le aziende fonti di profitto. Questa prima classificazione si rende comunque doverosa per escludere che l'IA possa far parte della categoria dei soggetti di diritto.

Nonostante la loro natura di beni ai sensi del diritto civile, gli algoritmi IA, non possono dirsi inerti. Si è visto come, qualificandoli sotto la definizione di automi, se ne riconosca l'autonomia e dunque la capacità di svincolarsi parzialmente dal moto incessante della causalità.

Allo stesso tempo un'ulteriore ambiguità sorge in relazione alla natura ibrida che essi hanno. Infatti, tenuto conto della robotica e degli algoritmi già operanti in larga parte sul web, non è

---

<sup>74</sup> Ex Art. 810 cod. civ.

agevole effettuare una distinzione netta rispetto alle categorie dei beni materiali ed immateriali.

L'IA quindi è una *res atipica*, ha infatti la capacità di compiere azioni che possono produrre degli effetti causali. In questo si può riscontrare una netta distinzione con il computer che è invece già adeguatamente regolamentato. La parte hardware rientra nella categoria dei beni materiali, mentre la parte software in quella dei beni immateriali, specificamente regolamentata dalla normativa sul diritto d'autore e, in altri ordinamenti, come in quello americano, anche sotto la normativa del brevetto industriale.

Ad ogni modo la qualificazione dell'IA come oggetto di diritto non risulta particolarmente audace, anzi sembra pacificamente ammessa da più fonti. Ciò che desta le maggiori perplessità è invece il profilo regolamentare della responsabilità.

È questa la grande sfida che i giuristi sono chiamati ad affrontare. Le conseguenze derivanti dall'autonomia degli algoritmi d'intelligenza artificiale sono tali da indurre a prendere seriamente in considerazione l'idea di creare una forma di personalità giuridica per tali algoritmi, strutturata su quella già esistente delle persone giuridiche come sono le società.

Questa strada elude però parte del problema intrinseco della questione: perché ed in che misura si è responsabili del danno commesso dall'IA?

La soluzione data dall'ideazione della personalità elettronica tiene già conto delle conseguenze della stanza cinese di Searle, infatti la *fictio iuris* della personalità elettronica di tipo funzionale, al pari della responsabilità oggettiva, andrebbe ad addossare *automaticamente* la responsabilità ad un essere umano.

In entrambi i casi si riconosce l'assenza d'intenzionalità della macchina e la necessaria attribuzione di responsabilità ad un essere umano.

La strada è quella che porta alla soluzione, ma sono le modalità ad essere insoddisfacenti. Infatti, in entrambi i casi la *ratio* degli istituti consiste nello scavalcare l'esame probatorio ed il conseguente accertamento del nesso di causalità.

Questa è comunque un'esigenza necessaria, i tempi e la complessità dei processi si aggraverebbero di molto se ogni volta che l'IA fosse coinvolta si dovessero effettuare perizie

informatiche per dimostrare il nesso tra danno e azione, quando le circostanze lo rendano possibile, cosa che per l'opacità degli algoritmi non può darsi per scontata.

La *ratio* delle *fictiones* consiste nell'assumere che, chi può godere dei grandi benefici derivanti dall'uso dell'IA, dovrà anche rispondere degli eventuali danni che causa. Così si conquista da un lato un ideale di giustizia, per il quale chi subisce il danno avrà il diritto alla riparazione, ed allo stesso tempo si evitano le problematiche che questa tecnologia ci costringe ad affrontare.

Ma proprio per questo motivo bisogna ritenere che se queste questioni non verranno risolte definitivamente è probabile che si ripresenteranno in futuro. È essenziale che venga ideato un istituto che sia capace di soddisfare i quesiti principali in termini perentori, chi risponde del danno dell'IA e perché?

Un indizio su quale sia l'ordine di idee che può portare ad una risposta essenziale lo si trova proprio in uno scritto di John Searle:

*“A question I don't know the answer to is, how do we distinguish between those features of the complex event which are unintentional actions and those features which are so far from the intention that they are not actions at all? When Oedipus married his mother he moved a lot of molecules, caused some neurophysiological stuff in his brain, and altered his spatial relationship to the North Pole. These are all things he did unintentionally and none of them are actions of his. Yet I feel inclined to say that marrying his mother, though it was something he did unintentionally, was still an action, an unintentional action. Perhaps the reason for this difference in our (my) intuitions is that the description “marrying his mother” is closer to the content “marrying Jocasta” than is “moving a lot of molecules.” I think that there must be a principle in operation here, but I do not know what it is.”<sup>75</sup>*

A questo Principio innominato sarà dedicato il prossimo capitolo di questa Tesi.

---

<sup>75</sup> J. SEARLE, *The Intentionality*, op. cit., p. 64.

## III. DETERMINISMO E LIBERO ARBITRIO

### III.1 Due Principi complementari

*“Distinguere l’azione del cielo dall’azione dell’uomo, ecco il vertice della conoscenza”<sup>76</sup>*

Nel corso della trattazione sono stati sollevati innumerevoli problematiche relative all’intelligenza artificiale. Si è visto come questo nuovo fenomeno si intreccia con tutti i maggiori quesiti che l’umanità si pone da sempre: cos’è umano e cosa no? Esiste o meno il libero arbitrio? Cos’è la coscienza e cosa il pensiero?

Le maggiori perplessità e, secondo la mia opinione, la grande attrazione che suscita in molti questa materia non sono dovute al fenomeno in sé, che è pur interessante soprattutto per gli appassionati di scienza, ma piuttosto agli interrogativi su noi stessi coi quali ci costringe a fare i conti.

Ebbene si è visto come sin dalla sua origine il padre dell’informatica si chiese se ed in che misura queste nuove macchine possano somigliarci. È evidente che come premessa logica per la risposta sia necessario sapere chi e che cosa siamo noi esseri umani.

Questo aspetto davvero sottile della questione, credo, ne rappresenti il fulcro, la vera sfida.

Durante la trattazione per ben due volte mi sono avvalso dell’esempio classico che rappresenta i noti Tizio, Caio e Sempronio per descrivere due principi generali, di cui ciascuno di noi fa esperienza quotidianamente, i quali per loro natura risultano evidenti di per sé e che soltanto mediante gli esempi, per la loro forza rappresentativa dell’esperienza, possono essere dimostrati.

Il principio di cui ho discusso nel primo capitolo è quello di casualità un principio ben noto alla scienza e semplice da rappresentare:

---

<sup>76</sup> Citazione proveniente dall’antico testo taoista *Zhuang-zi*, che apre il capitolo “*La scuola del Primo Principio*”, L. Kia-Hway (a cura di), trad. di C. Laurenti e C. Leverd, Adelphi, p.57.

A (causa)  $\longrightarrow$  B (effetto)

Tale principio permea questa tesi, è presente in ogni sua parte data la natura scientifica ed oggettiva del fenomeno che ne è fonte di studio, l'intelligenza artificiale. Va da sé che il lettore già avrà una certa familiarità con questo fenomeno, non solo per l'esperienza quotidiana, ma soprattutto per la sua presenza richiamata più e più volte nel testo nei punti in cui ve n'è stata la necessità.

Ciononostante, è utile chiedersi ancora una volta cosa debba intendersi davvero per causalità?

Classicamente la causalità è il modo in cui spazio e tempo si intrecciano per formare la realtà di cui ciascuno fa esperienza quotidiana. Sul piano astratto per causalità s'intende invece un modo di ragionare definito come *logico*, nel quale dalle premesse iniziali nella dialettica e dagli assiomi in matematica si fanno discendere delle conclusioni dimostrate da argomentazioni o calcoli.

L'intelligenza artificiale può ben rappresentare l'archetipo della causalità.

Infatti, sul piano della programmazione si è visto come gli autori dell'AIMA facciano riferimento innanzitutto ad Aristotele, padre della logica, per spiegare le più elementari tecniche di programmazione. Mentre sul piano della robotica, le istruzioni impartite dall'algoritmo producano un comportamento fisico che si va ad inquadrare nelle leggi del moto causale dei corpi in movimento descritto dalla dinamica e regolato dalle leggi della fisica classica.

Sia che essa sia rappresentata sul piano fisico dello spazio-tempo che su quello astratto della logica, se effettivamente una distinzione tra i due sussiste, la causalità viene spiegata dalla rappresentazione vettoriale che collega causa ad effetto (A-B).

In questo senso, la causalità come legge, dimostra e spiega piuttosto che regolare. Si potrebbe dire che essa non è che una formalizzazione di una delle regolarità fondamentali che noi umani abbiamo individuato nella realtà. Infatti, basta pensare ad essa in termini semplici per rendersi conto di come essa esprime la nostra conoscenza della realtà piuttosto che essere essa stessa la realtà.

La realtà per sua natura è multiforme, labile e difficilmente spiegabile come dimostrano le insidie insite nel linguaggio, regno nel quale le ambiguità sono la regola. Solo l'uomo, delle specie viventi conosciute, è stato in grado di razionalizzare l'amorfa realtà tramite l'esercizio del *λόγος* e tramite esso inserirla in categorie, descriverne leggi e costruire civiltà.

La causalità come principio capace di spiegare la realtà ha però subito un duro colpo agli inizi del Novecento. Come si è visto nel secondo capitolo, le scoperte della scienza novecentesca hanno leso irrimediabilmente la capacità ermeneutica della causalità, rendendola anch'essa una parte rispetto al tutto.

Un breve accenno delle scoperte della fisica si rende necessario per capire come ad oggi la causalità sia in crisi e come si sia iniziato a riflettere sui suoi limiti e sulla possibilità di affiancare ad essa qualche altro metodo di interpretazione del reale che, nella convinzione di chi scrive, sia in grado di rendere la crisi della causalità un cigno nero positivo.

È noto, infatti, che le conquiste teoriche raggiunte dalla fisica dei primi del Novecento hanno interamente plasmato la realtà in cui viviamo oggi, basti pensare alle conseguenze che ha avuto la fisica atomica sugli esiti del secondo conflitto mondiale, al pari di esse in termini d'importanza si possono citare le moderne tecnologie informatiche come: internet, computer, block-chain e IA tutte figlie del nuovo mondo prospettato dal radicale cambiamento di paradigma avvenuto nella fisica.

Se la fisica classica era il regno della causalità e del determinismo, la fisica quantistica è invece il regno dell'incertezza. Questa affermazione ben nota deriva dalle conseguenze teoriche che alcuni esperimenti sulle particelle fondamentali della materia, elettroni e fotoni, hanno prodotto in maniera del tutto inaspettata.

La denominazione “*quantistica*” deriva da *quanto*, questo è un'unità di misura che indica un *pacchetto d'energia*. Questa innovazione deriva dalla scoperta che l'energia e di conseguenza gli elettroni non seguono un moto continuo, come gli oggetti della fisica classica, ma piuttosto sono definiti da stati discreti. Ad esempio, un elettrone muta il suo andamento non in maniera progressiva ma effettuando un salto sull'orbitale successivo una volta raggiunta la soglia energetica necessaria.

La definizione stessa di elettrone è diventata enigmatica. Se per secoli si è creduto che gli atomi fossero delle particelle irriducibili di materia, arrivando poi ad essere descritti come sistemi solari in miniatura con elettroni e protoni rotanti attorno ad un nucleo, gli esperimenti che inaugurarono la fisica quantistica hanno mostrato un mondo completamente nuovo.

L'esperimento della doppia fenditura<sup>77</sup>, ideato per indagare la vera natura della luce, dimostra come i fotoni<sup>78</sup> sono sia particelle che onde. L'aspetto veramente rivoluzionario di questo esperimento è che viola in maniera sorprendente il principio di causalità.

L'esperimento consiste nel far attraversare un fascio di luce attraverso una doppia fenditura e rilevarne la posizione finale su un apposito pannello posto dopo le fenditure, la *ratio* era quella di verificare l'esistenza di fenomeni di diffrazione e rifrazione che se verificati avrebbero dimostrato la natura ondulatoria della luce. Questi fenomeni si sono verificati, ma allo stesso tempo, introducendo un sensore capace di individuare il passaggio di singole particelle da ciascuna fenditura, se ne verificarono anche altri, tali da dimostrarne la natura particellare, cosicché il quadro risultante comprendeva una descrizione della luce particellare, con dei fenomeni d'interferenza inspiegabili che invece mostravano un andamento ondulatorio.

Senza entrare troppo nel dettaglio basti qui citare le parole del fisico premio nobel Leon Lederman:

*“Le conseguenze dovrebbero esser chiare - e farvi venire i brividi, la pelle d'oca o quant'altro: ogni singolo fotone «sa» in qualche modo se le fenditure aperte sono due o una sola e sulla base di questa conoscenza muta il suo comportamento!”*<sup>79</sup>

Da qui la riconsiderazione della natura dell'elettrone alla luce dalla *funzione d'onda*  $\psi$  che descrive la probabilità di trovare un elettrone in un dato punto, il quale non viene più visto come un'onda od una particella, ma viene descritto dall' Interpretazione di Copenaghen come uno spettro di possibilità prima che venga misurato.

Secondo questa interpretazione l'elettrone viene a determinarsi solo con l'atto della misurazione o osservazione, prima di quest'atto si trova invece in una sovrapposizione di stati

---

<sup>77</sup> cfr. L. M. LEDERMAN, C. T. HILL, *Fisica quantistica per poeti*, trad. di L. Civalleri, Bollati Boringhieri, 2011 e, per una comprensione immediata: [https://it.wikipedia.org/wiki/Esperimento\\_della\\_doppia\\_fenditura](https://it.wikipedia.org/wiki/Esperimento_della_doppia_fenditura).

<sup>78</sup> Ai quali possono essere assimilati gli elettroni, cfr. L. M. LEDERMAN, *Fisica quantistica*, op. cit.

<sup>79</sup> L. M. LEDERMAN, C. T. HILL, *Fisica quantistica*, op. cit. p. 104.

possibili che diventa determinata solo dopo il collasso della funzione d'onda  $\psi$  per atto dell'osservazione, costringendo così l'elettrone ad assumere uno solo degli stati possibili descritti da  $\psi$ .<sup>80</sup>

Questa è la realtà che viene descritta dal Principio di Indeterminazione di Werner Heisenberg che viola il principio di località, per il quale un oggetto fisico può trovarsi solo in un determinato spazio ad un certo tempo. Quale sia la realtà descritta realmente dalla fisica quantistica è un dibattito ancora aperto e vivacissimo.

Tra le teorie più recenti ed accreditate vi sono quelle a molti mondi, per la quale è supponibile l'esistenza di universi paralleli associati alla funzione d'onda. Altre teorie prendono invece in considerazione il ruolo della coscienza nella manifestazione della realtà, suggerendo che essa possa avere una funzione fondamentale nel meccanismo del collasso della funzione  $\psi$ .

Ciò che è certo è che le antiche dottrine deterministiche, le quali prendevano in considerazione un mondo certo ed oggettivo, come ad esempio la dottrina latina dello stoicismo che assumeva il fato come realtà immutabile, sono state superate dalle attuali evidenze scientifiche.

A questa crisi in ambito scientifico ne sono seguite altre in tutti gli ambiti del sapere ed il diritto non ne è rimasto immune.

La teoria classica della *condicio sine qua non* finalizzata all'attribuzione di un determinato evento all'azione umana mediante il meccanismo definito come *sottrazione mentale*<sup>81</sup>, è stata oggetto di critiche e sono stati proposti dei correttivi che prendono il nome di: teoria della casualità adeguata e teoria della causalità umana.<sup>82</sup> Quest'ultime mirano a chiarire un elemento ulteriore per l'accertamento del nesso causale, "*cosa sia necessario per affermare che un dato evento è conseguenza di una data azione*".<sup>83</sup>

---

<sup>80</sup> Questo fenomeno traslato nella realtà macroscopica dà vita al famosissimo paradosso del "*gatto di Schrödinger*". Questo esperimento mentale ipotizza che un gatto chiuso in una scatola, nella quale sia posta una fiala di veleno collegata ad un singolo elettrone che si apre non appena esso assume uno stato determinato, si trovi in uno stato per il quale, prima del collasso della funzione d'onda  $\psi$ , risulta essere sia vivo che morto contemporaneamente.

<sup>81</sup> Per il quale il giudice, al fine di poter attribuire all'azione umana la qualifica di condizione necessaria per la causazione dell'evento, deve eliminare l'azione e verificare se senza di essa l'evento non si sarebbe verificato.

<sup>82</sup> Le informazioni relative alle recenti teorie della causalità giuridica provengono dal testo G. MARINUCCI, E. DOLCINI, G. L. GATTA, *Manuale di Diritto Penale, parte generale*, Giuffrè, 2018, p. 225-238.

<sup>83</sup> *Ivi* p. 225.

Entra a questo punto in gioco l'interrogativo postosi da Searle citato in conclusione del precedente capitolo. Ci si è chiesti infatti cosa costituisca azione di un agente, Edipo nell'esempio portato da Searle, ed invece cosa non sia attribuibile ad esso come azione.

Andando ad analizzare in maniera specifica quell'interrogativo gravoso lasciato per un momento in sospenso, si può notare come gli elementi che vengano presi in considerazione nell'esempio siano essenzialmente tre: il movimento di molecole che viene causato da Edipo ma che non viene ricondotto ad esso come soggetto e che quindi non costituisce una sua azione, l'atto di sposare Giocasta che corrisponde ad una azione voluta e preveduta da Edipo ed infine l'atto di sposare sua madre che non è voluto né preveduto da Edipo ma che secondo la nostra sensibilità viene ricondotto ad esso come azione compiuta da lui stesso.

Rispetto a questo quadro la domanda è: qual è l'elemento di discriminazione tra azione e ciò che invece azione non è?

Le risposte che si propongono di dare le due teorie citate in precedenza<sup>84</sup> individuano come elemento di discriminazione, capace di distinguere tra azione e non azione, la normale prevedibilità delle conseguenze del proprio agire la prima e l'esclusione del concorso di fattori eccezionali la seconda.

Ciò che di contro risulta a mio avviso evidente è che la risposta non vada cercata nell'ambito della causalità. Ciò viene corroborato dalle conseguenze che possiamo trarre dalle riflessioni condotte circa le nuove frontiere esplorate dalla fisica. Infatti, quel che sembra più ragionevole non è tanto cercare di aggiustare la causalità mediante l'introduzione di nuove formule più sofisticate della *condicio sine qua non*, che sembra adeguata ad accertare il nesso causale, quanto piuttosto affrontare ciò che sembra porsi dinanzi al giurista: un nuovo problema che, come tale, richieda l'elaborazione di un nuovo principio.

Infatti, se da un lato il nesso causale può essere accertato quasi in ogni circostanza, la vera domanda che sorge è in che misura è responsabile l'uomo? A questo quesito si rivolgono le teorie della causalità adeguata e della causalità umana, che propongono una correzione della

---

<sup>84</sup> Causalità adeguata e causalità umana, cfr. G. MARINUCCI, E. DOLCINI, G. L. GATTA, *Manuale di diritto penale*, op. cit., p. 225-234.

teoria classica, così come le nuove interpretazioni della realtà scaturite dalle evidenze sperimentali della fisica quantistica ed infine John Searle.

Quest'ultimo, a ragion veduta, afferma che secondo la sua opinione vi è un qualche principio sconosciuto operante nell'esempio riportato. Ed infatti è a questo principio che mi sono rivolto utilizzando la seconda volta un esempio parlando d'intenzionalità secondo le categorie elaborate da Searle inquadrandone il corrispettivo nella legislazione penalistica italiana.

Se fin qui l'intelligenza artificiale è stata usata come chiave del discorso è perché essa, oltre ad essere un perfetto esempio di causalità, per sottrazione risulta essere anche un campo di prova ottimo per questo secondo principio.

Le riflessioni condotte nel secondo capitolo, trattando di Test di Turing e Intenzionalità, hanno messo in luce il ruolo e l'importanza della soggettività nel funzionamento dell'IA. Si è visto infatti come essa sia un automa, causale, mentre chi di norma è considerato soggetto ha una qualità peculiare, detta intenzionalità secondo la terminologia adottata nel capitolo precedente, che lo rende libero.

Se da un lato la causalità è assoluta necessità, questo secondo principio innominato è espressione di libertà. Si è visto, infatti, come sia questo l'elemento di discriminazione tra intelligenza umana ed artificiale, la prima vuole ed agisce, la seconda sebbene dotata di autonomia operativa, non può volere, non è libera.

In questa direzione muove la teoria dell'*imputazione oggettiva dell'evento*, secondo la quale non debba essere modificata la teoria classica della causalità quanto piuttosto si renda necessaria l'introduzione di un elemento ulteriore: l'imputazione. Questo requisito sarebbe idoneo a rispondere alla domanda "*se l'evento possa considerarsi opera dell'agente*"<sup>85</sup>.

È chiaro che ci troviamo dinanzi ad un radicale mutamento di paradigma, in totale accordo con le evidenze scientifiche della fisica quantistica e idoneo a rispondere al quesito fondamentale dell'intelligenza artificiale.

L'approccio da cui muove la teoria dell'imputazione oggettiva è di natura probabilistica, incentrato sul rischio che viene generato dall'agente con la sua condotta circa la possibilità di

---

<sup>85</sup> G. MARINUCCI, E. DOLCINI, G. L. GATTA, *Manuale di diritto penale*, op. cit., p.235.

verificarsi dell'evento dannoso. Questa metodologia si trova in un rapporto di perfetta sincronia con l'attuale metodo scientifico inaugurato dalla fisica quantistica, anch'esso basato sulla probabilità, e generalmente accolto dalla comunità scientifica.

Infatti, se in ambito scientifico si è visto che per le conseguenze del principio d'indeterminazione e la funzione d'onda  $\psi$ , non si parli più di posizione di una particella ma di probabilità di trovarla in un dato punto, in ambito giuridico per i fattori dell'imputazione oggettiva non è necessario che l'autore dell'azione abbia con essa causato il danno. Sarebbe in ultimo sufficiente che esso, agendo in violazione di una regola prudenziale, abbia aumentato le probabilità di causazione dell'evento dannoso e che esso si sia poi verificato.

In questo quadro non sarebbe più necessario accertare il nesso di causalità in misura stringente, sarebbe piuttosto sufficiente che l'agente violi una data norma di comportamento e che l'evento si verifichi e se anch'esso non fosse diretta conseguenza dell'azione dell'agente, quest'ultimo, risulterebbe comunque responsabile per l'accaduto avendo violato la relativa norma di diligenza che vieta quell'azione.

Nel contesto dell'IA quest'approccio sarebbe certamente molto vantaggioso, posto che l'accertamento più difficoltoso è senz'altro quello del nesso di causalità, in un contesto nel quale il giudice si troverebbe a dover ricondurre un evento, materialmente causato da un automa, all'azione umana.

Una notevole apertura nei confronti di quest'approccio è stata manifestata dalla Commissione europea nella sua proposta di regolamento per l'IA<sup>86</sup> (*artificial intelligence act*) nella quale è preponderante un approccio basato sul rischio (*risk management*). Rimandando il lettore ai paragrafi successivi per un commento più approfondito di quest'iniziativa legislativa, basti per il momento menzionare la scelta della Commissione di effettuare una distinzione tra *high-risk* e *low-risk IA systems*, prevedendo solo per i primi dei requisiti stringenti (*risk management system, record-keeping e human-oversight*)<sup>87</sup> finalizzati alla prevenzione del rischio.

---

<sup>86</sup> Commissione Europea, *Regulation of the European Parliament and of the Council, laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM(2021) 206 final, Brussels 21.4.2021.

<sup>87</sup> *Ivi* rispettivamente art. 9, art. 12 e art. 14 dell'*artificial intelligence act*.

Data la natura evolutiva della materia in oggetto<sup>88</sup>, non è ancora chiaro che impatto giuridico avrà quest'impostazione e se il rispetto di questi requisiti potrà esonerare il responsabile del sistema IA (*provider*) dal rispondere del danno eventualmente causato da sistemi conformi ai requisiti della normativa o in che modo possa incidere sugli oneri probatori.

In ogni caso sul piano metodologico siamo di fronte ad un'innovazione notevole rispetto ai sistemi giuridici classici, nonostante ciò, si rende necessario un appunto; nella prospettiva più ampia dei principi generali, l'approccio probabilistico, sebbene innovativo, non è di per sé sufficiente a configurare un nuovo principio.

Infatti, risulta evidente come in un quadro in cui il determinismo e la legge causale rappresentano il 100% di esattezza di una previsione, la probabilità delle teorie sopra esposte rappresenterebbe l'80% o il 70% configurandosi così sempre nell'ambito del meccanicismo classico. In questa prospettiva la metodologia probabilistica non rappresenta che un semplice affievolimento del principio di causalità, riflesso della crisi della sua capacità ermeneutica, nonché tentativo di conservarne la validità.

Riportando il ragionamento al problema concreto sollevato da John Searle (distinguere un'azione da un mero accadimento) ci si rende conto, infatti, che neanche l'approccio probabilistico della teoria dell'imputazione oggettiva dell'evento è in grado di dare una risposta pienamente esaustiva. Nel caso di specie, se Tizio ha aumentato col suo agire le probabilità di verificarsi di un evento in violazione di una norma prudenziale, a buon titolo potrà essere responsabile se la sua azione è caratterizzata dal dolo, perché si tratta di una condotta a cui attribuiamo un disvalore, nondimeno ciò non ci dice se realmente Tizio con la sua condotta ha causato l'evento.

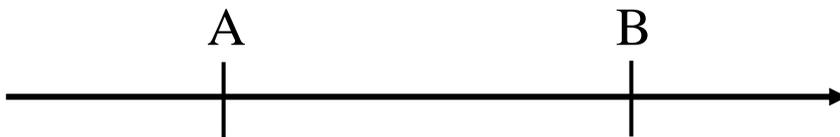
Ancor più nello specifico, se un automa IA causa un danno e successivamente si scopre che tale sistema non rispettava i requisiti affinché fosse qualificabile ai sensi della legge come conforme agli standard di sicurezza, il *provider*, per la sua negligenza nel rispetto delle norme, sarà responsabile del danno causato ma non per questo necessariamente si avrà risposta al

---

<sup>88</sup> Infatti, l'attuale proposta legislativa è intervenuta durante la redazione della presente tesi, confermando alcuni degli orientamenti esposti nel cap. I basati su documenti precedenti, ed è in corso di approvazione nel rispetto dei tempi e delle procedure dell'ordinario iter legislativo europeo.

quesito: in che misura il danno commesso dall'automa IA costituisce un'azione del *provider* o chi per lui?

Per inquadrare anche graficamente i termini del problema, sempre restando nell'ambito della causalità, si può ampliare la rappresentazione vettoriale esposta in precedenza:



In questo quadro, dove il nesso causa-effetto che collega A e B rappresenta una serie causale autonoma, ci si pone il problema di capire in che modo si possa distinguere una parte dal tutto. Riprendendo l'esempio del cap. I.3.1 si ha una serie causale in cui Caio causa il movimento del macigno mediante un'azione di spinta e il macigno causa l'evento morte di Tizio, viene così a configurarsi ai sensi della legge un'autonoma serie causale (A-B) che viene isolata dal moto perpetuo della causalità generale e presa in considerazione per l'applicazione dell'art. 575 c.p. (norma del Codice penale che punisce l'omicidio).

Il principio che opera in questo processo non è quello di causalità, anzi, il principio che si sta prendendo in considerazione piuttosto opera sulla causalità.

Le previsioni dell'art. 41 c.p. recanti le disposizioni in tema di “concorso di cause” complicano la situazione senza però mutarne la natura. Se l'intervento di serie causali ulteriori ed autonome rispetto a quella principale (A-B) può aver influito sulla realizzazione dell'evento si rende necessario prenderle in considerazione per l'applicazione della legge e l'individuazione del responsabile.

A ben vedere quest'elemento stravolge l'ordinario andamento causale degli eventi per l'irruzione di fenomeni quali il *caso fortuito* o la *forza maggiore*, in questi casi la responsabilità dell'agente viene graduata in relazione a questi eventi imprevedibili ed ingovernabili, su questo tendono ad operare le teorie della causalità adeguata, causalità umana ed anche la teoria dell'imputazione oggettiva dell'evento, chiedendosi quando un evento può essere imputato ad un uomo.

La causalità pone la questione esclusivamente su un piano oggettivo, in cui ciascun elemento del discorso rileva solo se qualificato come causa (A) o effetto (B), pertanto è oggettivamente

indiscutibile, a parer mio, la validità della *condicio sine qua non*. Infatti, se ci si pone nell'ambito della causalità, per poter prendere in considerazione la condotta dell'agente, la stessa non può che rilevare se e solamente se essa ha un qualche rapporto causale con l'evento causato e questo nesso può sempre essere accertato dalla teoria della *condicio sine qua non*.

Seguendo questo ragionamento è evidente come la causalità non sia che una parte del processo di accertamento della responsabilità di un soggetto e di contro che la risposta al quesito sulla responsabilità di John Searle non trovi riscontro nella legge di causalità. Infatti, se precedentemente abbiamo distinto tra le azioni che Edipo compie: sposare Giocasta e sposare sua madre e il movimento delle particelle che ha causato non corrispondente ad una sua azione; sul terreno della causalità non vi è alcuna distinzione, ciascuno di questi eventi è effetto di una causa singola: Edipo.

Searle giustamente afferma la sua propensione ad inquadrare nell'evento sposare sua madre, un'azione di Edipo, sebbene inconsapevole, perché ha un significato molto vicino con l'evento sposare Giocasta.

Si pongono quindi all'orizzonte due strade percorribili<sup>89</sup>: la prima è di natura meccanicistica e in questa tesi è espressa dalle teorie computazionali della mente umana, consiste nel vedere l'universo come un grande meccanismo e l'uomo come ingranaggio sfociando in un dominio assoluto della causalità sul piano generale e nella responsabilità oggettiva su quello giuridico. Tale prospettiva è stata affrontata e rifiutata nel secondo capito di questa tesi.

L'altra strada, che personalmente ho deciso di percorrere, consiste nell'ampliare il proprio orizzonte, riconoscendo che dall'esperienza vi è un fatto assolutamente certo ed inconfutabile: l'uomo non è una macchina.

Nell'esempio di Searle, il movimento di particelle, come sposare Giocasta e sposare sua madre non sono che diverse espressioni per descrivere un unico evento, l'azione di Edipo che esercitando il suo libero arbitrio compie una scelta e con essa un'azione, divenendo responsabile di ogni conseguenza positiva o negativa.

---

<sup>89</sup> Le stesse conclusioni vengono presentate da J. R. SEARLE in *Freedom and Neurobiology, reflections on free will, language and political power*, Columbia University Press, New York, 2007, p. 66-78, il quale per la seconda possibilità si riferisce esplicitamente al fenomeno dell'indeterminazione quantistica come possibile evidenza dell'esistenza del libero arbitrio in natura, "*Conclusion: Consciousness manifests quantum indeterminism*" p.75.

Sotto questo profilo la distinzione tra oggettivo e soggettivo è fondamentale. Oggettivamente si parla di causalità e quindi consequenzialità del moto, soggettivamente le intenzioni di Edipo rispecchiano la sua natura umana che non è in grado di prevedere il futuro né di conoscere assolutamente il presente ed è per questo che noi siamo inclini a scusarlo del tragico errore di sposare sua madre ed uccidere suo padre, assolvendolo dal gravoso senso di colpa che lo porterà nella conclusione dell'Edipo re ad accecarsi per il senso di colpa.

Tradizionalmente si pensa che determinismo e libero arbitrio si contraddicano, come può infatti Edipo essere responsabile di una tale atrocità se era il suo destino quello di compierla e come può sentirsi in colpa se non aveva possibilità di scelta. Tale paradosso prospettico è secondo me solo apparente.

Infatti, sembra ragionevole in questa prospettiva supporre che causalità e volontà si possano conciliare. Baso questo giudizio su una delle conseguenze della teoria quantistica che forse da sola è in grado di dare la più coerente spiegazione della realtà nella quale mi sono imbattuto durante il mio percorso accademico.

L'assoluta necessità della causalità che dominava nella fisica classica e nella concezione meccanicistica del mondo si basava su un assunto: che vi fosse un osservatore esterno, l'uomo, che potesse descrivere in maniera oggettiva e certa tutto ciò che era in grado di osservare. Questo pregiudizio crolla con l'esperimento della doppia fenditura scoprendo che osservare un fenomeno lo influenza.

È quasi banale pensare che, al tempo, la scienza si basasse su un presupposto come quello dell'osservatore, il quale portava ad escludere l'uomo dalla natura, come se esso non ne facesse parte. Gli esperimenti della fisica quantistica hanno invece dimostrato la realtà, noi facciamo parte della natura e la influenziamo.<sup>90</sup>

Tale interpretazione della fisica quantistica è quella propria di Erwin Schrödinger esposta nel saggio "*Che cos'è la vita?*"<sup>91</sup>. In questo libro illuminante, precursore delle ricerche sul genoma che portarono poi alla scoperta del DNA, vi sono capitoli quali "*E' la vita basata*

---

<sup>90</sup> Sul ruolo dell'osservatore nella natura per la fisica quantistica si vedano F. CAPRA, *il Tao della fisica*, trad. di G. Salio, Adelphi, 1982 e C. ROVELLI, *Helgoland*, Adelphi, Milano, 2020.

<sup>91</sup> E. SCHRÖDINGER, *Che cos'è la vita?*, trad. di M. Ageno, Adelphi, Milano, 1944.

sulle leggi della fisica?” e “Determinismo e libero arbitrio”<sup>92</sup> dove lo scopritore della funzione d’onda discute di un ulteriore principio della fisica, diverso dalla causalità, che egli definisce “ordine dall’ordine” contrapposto al “ordine dal disordine” proprio della causalità.<sup>93</sup>

Sarebbe quest’ultimo principio quello che regola il comportamento degli organismi viventi, contrapposto alla causalità che regola il moto della materia inerte. L’autore così tende a conservare la causalità ed il determinismo, affiancandogli quest’ulteriore principio, per concludere infine nel capitolo più filosofico del libro:

*“Si pensa infatti che ciò contraddica al libero arbitrio, quale è attestato dalla diretta introspezione. Ma esperienze immediate, in se stesse, [...] sono dal punto di vista logico incapaci di contraddirsi a vicenda [...] Il mio corpo funziona come un puro meccanismo, secondo le leggi di natura; inoltre io so, per esperienza diretta incontrovertibile, che io dirigo i suoi movimenti, dei quali io prevedo gli effetti, che possono essere gravi di conseguenze, nel qual caso io sento e assumo piena responsabilità di essi. L’unica possibile deduzione da questi due fatti è, penso, quella che io [...] sono la persona che controlla «il movimento degli atomi» secondo le leggi di natura.”<sup>94</sup>*

Da qui la perdita della forza ermeneutica della causalità, la natura probabilistica della fisica quantistica e l’incertezza del futuro.

Il determinismo è valido perché è un metodo in grado di fare previsioni basandosi su dati oggettivi e quindi capace di dare certezza, contribuendo alla certezza della realtà e del diritto. La volontà che esercitiamo e di cui siamo padroni ci rende capaci di influenzare il presente ed il futuro, che a sua volta diventa incerto proprio per questo fattore.

---

<sup>92</sup> Ivi p. 131- 153.

<sup>93</sup> La differenza si basa su concetti quali *entropia* ed *equilibrio termodinamico*, dove nella causalità gli oggetti tendono ad uno stato permanente di equilibrio termodinamico dove non vi sono più interazioni di alcun tipo (definito anche come entropia massima). Mentre diverso è il comportamento della materia vivente (ordine dall’ordine) E. SCHRÖDINGER, *Che cos’è la vita?*, op. cit., p.117- 129.

<sup>94</sup> Ivi p. 148.

Ciò non nega il determinismo perché se conoscessimo tutte le cause potremmo prevedere certamente il futuro, non le conosciamo perché esercitiamo noi *in primis* questo potere causale sulla realtà tramite un attributo che vien definito azione o volontà o ancora intenzionalità.

Né d'altro canto risulta essere il libero arbitrio un'illusione, perché se è vero che tutto è determinabile e prevedibile è pur vero che noi siamo attori di questo gioco in prima persona e quindi contribuiamo a dar forma al futuro. Tutto ciò ci rende responsabili delle nostre azioni.

Secondo l'opinione di chi scrive quindi, la causalità come legge universale agisce sugli oggetti ed è invece esercitata in parte dai soggetti. L'esempio paradigmatico è proprio l'intelligenza artificiale, essa infatti agisce su comando umano, è un'oggetto e subisce il moto causale. Di contro, se ci si chiede come agisce l'IA si nota che essa dipende dall'uomo, se non fosse esistito l'uomo di certo non sarebbe esistita alcuna intelligenza artificiale.

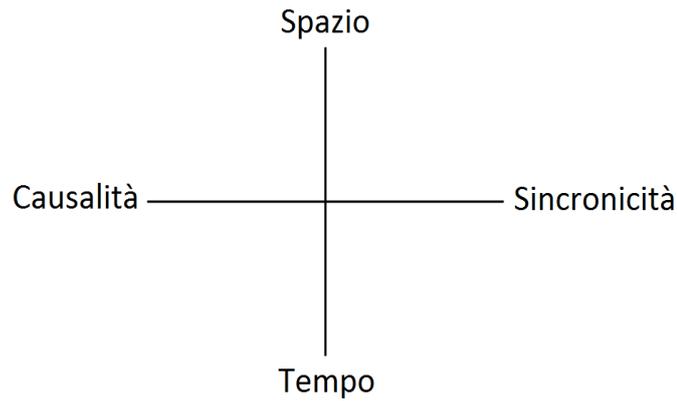
Due ulteriori connotazioni possono essere date a questo principio prima di passare ad un'attenta disamina della sua descrizione tecnico-giuridica offerta dal giurista tedesco Hans Kelsen.

È curioso notare come, questa seconda connotazione, seppur non direttamente, provenga anch'essa dalle scoperte della fisica quantistica. Infatti, nacque da un fruttuoso sodalizio tra le menti brillanti di Carl Gustave Jung, psichiatra di fama mondiale e Wolfgang Pauli, fisico della scuola di Copenaghen, allievo di Niels Bhor e scopritore del principio di esclusione, nonché uno dei fondatori della meccanica quantistica.

Dall'incontro tra i due nacque il *principio di sincronicità*. Tale principio, nato con l'ambizione di posizionarsi accanto alla causalità e completare il quadro della realtà, sarebbe nelle intenzioni degli autori in grado di descrivere quei fenomeni a-causali<sup>95</sup> che non trovano spiegazione nel paradigma classico. La notorietà di questo principio è dovuta anche alla capacità ermeneutica conferitagli dai due ideatori, infatti, unendo gli studi di diverse discipline come fisica a psicologia, nel concetto di sincronicità possono essere fatti rientrare categorie di fenomeni diversissimi tra loro.

---

<sup>95</sup> Uno di essi è l'imprevedibilità del comportamento di un singolo atomo nella fisica quantistica dovuta al principio di indeterminazione di Heisenberg, di cui si è detto sopra.



In questo schema<sup>96</sup> dove la causalità rappresenta una connessione di due eventi collegati da un nesso causa-effetto sul piano spazio-temporale, il nesso di sincronicità lega due eventi che si manifestano contemporaneamente, ovvero nello stesso tempo. Ciò che cambia è il diverso significato che lega i due eventi: il nesso causale, lega due eventi sul piano della necessaria consequenzialità tramite un nesso *logico*, il nesso sincronico lega due eventi per la vicinanza dei loro significati.

Per capire meglio questo secondo tipo di connessione, che può essere la chiave nel dare una risposta al quesito di Searle, può essere riportato un esempio derivato dall'esperienza dell'autore<sup>97</sup>. Jung racconta come nel corso della sua attività professionale di psichiatra si imbatté in una paziente particolarmente difficile, il cui percorso terapeutico si sbloccò grazie ad una coincidenza. Avendo ella sognato il giorno prima uno scarabeo dorato ed essendo rimasta profondamente colpita dal sogno, beneficiò di un evento del tutto inaspettato raccontato da Jung nel saggio, ovvero l'improvvisa irruzione di uno scarabeo nella stanza dove si stava svolgendo la psicoterapia.

Risulta dunque evidente la differenza: se la causalità poggia su un nesso di materialità, invece, la sincronicità poggia su un nesso di significato, l'una è oggettiva, l'altra soggettiva. In particolare, secondo Jung:

---

<sup>96</sup> C. G. JUNG, *L'analisi dei sogni, gli archetipi dell'inconscio, la sincronicità*, Bollati Boringhieri, Torino, p. 275.

<sup>97</sup> C. G. JUNG, *La sincronicità, op. cit.*, p. 200-220.

*“Un contenuto inatteso in relazione immediata o mediata con un evento esterno oggettivo coincide con lo stato psichico abituale: è questo fatto che chiamo sincronicità”<sup>98</sup>*

Semplificando si può parlare di una piena coincidenza tra uno stato soggettivo ed un evento oggettivo esterno al soggetto che ne fa esperienza. Per riportare anche questo fenomeno in ambito giuridico, si può notare come anch'esso faccia parte del processo di accertamento della responsabilità ed in particolare è proprio l'elemento che si vuol dimostrare cercando di verificare la concomitanza tra un'intenzione (dolo, colpa, preterintenzione) ed il corrispondente evento oggettivo ricercato nell'ambito della causalità.

Nella descrizione fatta da Jung l'accento viene posto più sull'accidentalità del fenomeno e sulle sue caratteristiche numinose, accentuandone il valore psicologico per l'individuo. Nell'opinione di chi scrive ed ai fini di questa tesi, non risulta azzardato accostare questo meccanismo di *sincronicità*, anche al principio del quale si sta tentando di dare una descrizione.

Risulterà evidente più avanti, infatti, come vi siano delle sorprendenti analogie tra la descrizione junghiana di *sincronicità* e la presentazione dell'imputazione del giurista Hans Kelsen. Ciò che è capace di unire due prospettive così distanti è l'elemento centrale che risalta in ambedue le descrizioni e che qualifica tale principio come altro rispetto alla causalità.

Riportando i termini del problema al quesito principale si può notare che: sul piano causale non si può effettuare una distinzione coerente tra il movimento di molecole, sposare Giocasta e sposare sua madre, mentre adoperando questo secondo principio siamo in grado di distinguere tra:

- Movimento di molecole (azione inconsapevole, non corrispondente ad uno stato soggettivo).
- Sposare Giocasta (azione intenzionale, corrispondente al dolo).
- Sposare sua madre (azione inconsapevole ma corrispondente alla colpa).

---

<sup>98</sup> C. G. JUNG, *La sincronicità*, op. cit., p. 209.

La sincronicità, intesa come coincidenza tra uno stato soggettivo ed uno oggettivo, può ambire a spiegare il problema sollevato da John Searle ed assurgere al principio innominato al quale il filosofo si riferisce.

È chiaro che si tratta comunque di una visione parziale che lascia ancora alcune ombre sul piano della spiegazione di come si manifesta un'azione. È bene precisarlo perché se fosse questa una spiegazione esaustiva, si potrebbe affermare che la libertà è un'illusione e che essa nasce dal presupposto che siamo noi a dare un significato a qualcosa che in realtà si manifesta indipendentemente da noi.<sup>99</sup>

Sia la sincronicità che l'imputazione muovono in questa direzione, condizionate da un'epoca storica in cui la validità assoluta della causalità aveva appena subito i primi segni di cedimento e, forse anche proprio per questo motivo, era ancora un vero e proprio tabù metterla in discussione. Mentre oggi i principi della fisica quantistica sono stati ben recepiti in tutti gli ambiti del sapere e le nuove teorie su coscienza, libero arbitrio e volontà, nonché la teoria unificata dei campi in ambito fisico muovono tutte da questi presupposti.<sup>100</sup>

È infatti nella prospettiva presentata da E. Schrödinger che vanno interpretati i risultati teorici di C. G. Jung, dove non solo determinismo e libero arbitrio sono compatibili, ma i rispettivi principi agiscono l'uno sull'altro intrecciandosi per creare fenomeni complessi come la realtà e la soggettività. Tenendo sempre a mente che non sono i principi che noi elaboriamo a creare la realtà, quanto piuttosto è la realtà che si fa conoscere da noi mediante questi principi, il loro valore è sempre descrittivo.

Siamo ora quindi nella posizione di dare una risposta giuridicamente rilevante ad alcuni degli innumerevoli problemi che si sono presentati durante la trattazione di questa tesi.

Per distinguere un'azione da ciò che azione non è vi sono diversi elementi da tenere in considerazione, rispetto ai quali la causalità gioca un ruolo minore, essi sono: la soggettività, un'azione può essere tale solo se posta in essere da un soggetto; la consapevolezza del

---

<sup>99</sup> È questa la posizione del determinismo proposta anche da H. KELSEN in *Lineamenti di dottrina pura del diritto*, Einaudi, Torino, 1952 p.218- 221 “Vani sono i tentativi di salvare la libertà del volere invocando il fatto che la volontà umana non è determinata dalla legge di causalità”. Va ad ogni modo ricordato che nell'epoca in cui scriveva il giurista tedesco ancora non si erano ben sedimentati i principi della fisica quantistica ai quali egli stesso fa un accenno a p.221 e quindi anche questo punto di vista dev'essere giudicato *cum grano salis*, per questo rimando il lettore al paragrafo successivo.

<sup>100</sup> Cfr. J. R. SEARLE, *Freedom and Neurobiology*, op. cit., p. 37-78.

soggetto, conseguenza di questo elemento sono le ipotesi di incapacità in ambito giuridico; il valore o significato che viene dato ad un certo evento, infatti la qualifica di azione e i vari tipi di azione si distinguono in base a come vengono percepiti soggettivamente, esempio di ciò è la distinzione tra sposare Giocasta e sposare sua madre.

In tutti questi casi opera un principio diverso da quello di causalità.

Sul piano giuridico la teoria della *condicio sine qua non* non ha bisogno di correttivi, piuttosto si renderà necessario in futuro, quando il problema che oggi è solamente accennato da fenomeni emergenti come l'IA sarà più urgente, elaborare una nuova categoria giuridica che si basi su questo secondo principio e che sia capace di distinguere tra azione e non azione.

Infine, dopo aver proposto un principio di risposta al quesito specifico sollevato da John Searle e aver accennato le conseguenze giuridiche, è possibile dare un'ultima connotazione a quest'ulteriore principio.

L'attuale principio di causalità proprio della scienza non è che una visione parziale di quella quadripartizione delle cause teorizzata da Aristotele. Se l'attuale nozione di causa è quella di *causa efficiente* intesa come ciò che produce, in origine ve ne erano altre tre, una delle quali era la *causa finale* ovvero ciò che è finalizzata a realizzare. Questo secondo principio, a cui si è fatto riferimento mediante riflessioni originate dagli esperimenti della fisica quantistica, e ben connotato in senso finalistico da C. G. Jung, il quale nei suoi testi fa spesso riferimento anche all'alchimia ed al taoismo cinese, potrebbe non essere altro che una riscoperta della causalità finalistica.

Questo tipo di causalità finalistica, in verità, sarebbe un altro principio diverso dalla causalità e che nelle ricerche di Jung è descritto come sconosciuto all'Occidente basato su un modo di pensare incentrato sul *λόγος*. Sarebbe esso invece conosciuto molto bene dall'Oriente col nome di *Tao*, la cui traduzione in italiano potrebbe approssimativamente essere quella di senso.

Questo secondo principio avrebbe un modo di operare del tutto diverso rispetto al principio di causalità e se ne può avere conferma riportando qui almeno due esempi:

- la scrittura orientale è strutturalmente diversa da quella occidentale. Infatti, il modello occidentale basato sul *λόγος* prevede un alfabeto ed un modo di costruzione della frase

lineare e consequenziale dove le lettere e poi le parole sono poste una dopo l'altra. Mentre la scrittura propria di paesi come Cina e Giappone è quella degli ideogrammi dove non vi è la caratteristica consequenzialità delle lingue neolatine, bensì simboli grafici che non rappresentano un valore fonetico ma un'immagine o un'idea.

- La tradizione occidentale viene tramandata mediante storie, come ad esempio quelle bibliche, nelle quali il modo di procedere è consequenziale e i versetti, sebbene leggibili indipendentemente, fanno parte di un corpus organico e strutturato. Mentre l'I-Ching, libro della tradizione taoista, consiste in un insieme di massime di saggezza consultabili mediante un metodo oracolare che prevede il lancio delle monete, assumendo come principio di funzionamento la corrispondenza tra lo stato interno del soggetto consultante ed il risultato del lancio delle monete dal quale discenderà anche la massima da consultare<sup>101</sup>.

Questi esempi possono dare un'idea della divergenza che vi è tra questi due principi uno logico-formale con un modo di agire improntato alla consequenzialità ed alla linearità, l'altro più intuitivo incentrato sul significato che una data immagine o impressione può lasciare. Se l'Oriente ha conservato questo modo di ragionare, nella storia dell'Occidente vi è stata una separazione tra causalità utilitaristica e causalità finalistica, la prima principio cardine della scienza, la seconda confluita nell'alchimia e in altre pratiche mistico-esoteriche poi svalutate con l'avvento del progresso.

Se negli ultimi secoli industrializzati ha dominato la causalità utilitaristica e di recente essa è in crisi, nel diritto la linea direttrice, quella che può essere denominata la volontà del legislatore, si è sempre basata su un tipo di causalità finalistica: la norma viene emanata al fine di prevenire, la pena al fine di punire ecc.

Questo aspetto intrinseco del diritto che si rispecchia in questo principio sarà l'oggetto del prossimo paragrafo.

---

<sup>101</sup> Si vedano per un esempio di consultazione le pagine di finali di R. WILHEM (a cura di), *I-Ching, Il Libro dei Mutamenti*, Adelphi, Milano, 1991, nonché la prefazione dello stesso a cura di C. G. JUNG.

### III.1.2 L'algoritmo è causale, il diritto imputabile

Il diritto come fenomeno sociale include in sé la causalità per accertare la realtà di un danno e l'effettività dell'azione dell'uomo, ma si può notare come sia insita nella logica stessa del diritto una metodologia che contraddice vistosamente una visione puramente meccanicistica del mondo.

È stata esposta in principio l'antica legge del taglione che statuisce il ben noto “*occhio per occhio dente per dente*”, in questa legge ancestrale la ragion pratica che conduce alla sua istituzione è quella di preservare la pace sociale.

Il presupposto che rende possibile la sua emanazione è la libertà dell'agire umano; infatti, nello stabilire che “*se uno farà una lesione al suo prossimo, si farà a lui come egli ha fatto all'altro*”<sup>102</sup> opera questo principio, il quale, presumendo un'alterazione dell'equilibrio naturale dovuta all'azione umana che provoca la lesione, statuisce una contromisura punitiva finalizzata alla ricostituzione dell'equilibrio originario<sup>103</sup>.

Se il nesso di causalità collega due elementi con un nesso di consequenzialità: se c'è A ci sarà B. La logica prevalente nel diritto descrive un diverso tipo di connessione: l'*imputazione*, per la quale se c'è A ci *dev'essere* B.<sup>104</sup> Una sua applicazione è proprio quella della legge del taglione: se uno caverà un occhio al suo prossimo, allora a costui dovrà essere cavato un occhio. Questa forma di equità, o idea primordiale della giustizia che assolve uno scopo ripristinatorio, è espressa tecnicamente da ciò che Hans Kelsen ha definito principio di imputazione.

Kelsen, in “*Causalità e imputazione*”,<sup>105</sup> pone le basi per quella che forse è la descrizione più accurata del meccanismo di funzionamento del diritto. L'affermazione di questo principio si colloca all'interno della dottrina *pura* del diritto, il progetto più ambizioso del giurista

---

<sup>102</sup> Levitico, 24, 19-20.

<sup>103</sup> Nel diritto questa natura ripristinatoria della pena è pressoché costante, basti qui citare a titolo esemplificativo la *restitutio in integrum* o il risarcimento pecuniario, che rispetto alla prima adopera il denaro come unità di misura comune per quantificare il valore, ma che risponde alla medesima *ratio*.

<sup>104</sup> H. KELSEN, *dottrina pura del diritto*, op. cit., p.216-217.

<sup>105</sup> *Ivi* in appendice, p.207-227.

tedesco, dove egli prova a delineare una visione tecnica del diritto e pertanto completamente epurata da ogni ideologia giusnaturalistica.

In questa prospettiva egli espone una visione quasi scientifica del diritto nella quale anche i concetti che in realtà sono imbevuti di ideologia ne vengono svuotati al fine di delineare una descrizione analitica di che cos'è il diritto in sé.

Nella descrizione della dottrina pura ogni elemento rileva sul profilo funzionale, esempio massimo di questa visione pura del diritto è la descrizione della *Grundnorm* o norma fondamentale, che nel quadro tecnico di Kelsen ha un preciso valore formale.

Questa è la norma presupposta a qualsiasi ordinamento, che ne esprime valori e sentimenti costituenti, ma la sua descrizione ai fini della dottrina pura non è che puramente funzionale. La funzione che essa assolve in un ordinamento, nella prospettiva della dottrina pura, è quella di dargli validità e insieme ad esso anche ai successivi atti di diritto positivo nel conteso della *costruzione a gradi* dell'ordinamento giuridico.

Ebbene già sotto questo profilo, meramente funzionale, si può tracciare una prima linea di demarcazione tra imputazione e casualità. Il ruolo della *Grundnorm* in un ordinamento giuridico e quello degli assiomi in un sistema matematico sono differenti.

Si è detto del primo teorema d'incompletezza di Kurt Gödel<sup>106</sup> a proposito del test di Turing, portandolo come argomentazione forte circa i limiti di una possibile mente meccanica. Questa, si è detto, basandosi solo su un linguaggio formale come la matematica, non potrà mai superare certi limiti intrinseci, in quanto il suo essere fondata su assiomi comporta per essa un limite fondamentale, ovvero l'incompletezza dei sistemi formali, che non limita invece la mente umana.

Si è visto, inoltre, il paradosso di Bertrand Russell e dell'omonima antinomia. Ciò che questi due risultati teorici hanno in comune è la dimostrazione che un senso, o significato, necessita in una qualche misura di una *ratio* esterna al sistema, perché se ci si basa solo su di esso e quindi sui suoi assiomi, si arriva a dei limiti intrinseci come l'incompletezza o l'antinomia di Russell.

---

<sup>106</sup> Per una spiegazione tecnica ed accessibile si veda S. NASINI, *il primo teorema di incompletezza di Gödel*, Dept. of Statistics and Operations Research, Universitat Politècnica de Catalunya.

Sulla *Grundnorm* Kelsen si pronuncia in questi termini:

*“Essa è soltanto l’espressione del presupposto necessario per comprendere positivisticamente il materiale giuridico. Essa non vale come norma giuridica positiva perché non è prodotta nel corso del procedimento del diritto; essa non è posta, ma è presupposta.”*<sup>107</sup>

Nell’affermare che “*essa non è posta, ma è presupposta*” si vuol riferire al necessario fatto che gli avvenimenti che sono la *ratio* dell’ordinamento e che andranno poi a costituire il contenuto della *Grundnorm* non fanno originariamente parte del sistema. Infatti, questi avvenimenti istitutivi di un nuovo ordinamento, ad esempio episodi rivoluzionari, sovvertono un ordinamento vecchio per crearne un altro nuovo, in questo senso pur rimanendo tale la *Grundnorm* sotto il profilo funzionale, i suoi contenuti mutano sostanzialmente.<sup>108</sup>

Rispetto agli assiomi, che fanno parte del sistema dandone origine, la *grundnorm* si pone come presupposto dell’ordinamento giuridico dandone validità grazie al suo significato (*dover essere*).

Si tratta di due schemi, metodi o principi completamente diversi.

Ad accentuare la diversità di questi due schemi di pensiero si aggiunge un ulteriore fattore: l’incompletezza, che nei sistemi matematici è un *vulnus*, mentre negli ordinamenti giuridici è una necessità. Infatti, se viene traslato nel diritto quel ragionamento che porta alle conseguenze paradossali del primo teorema d’incompletezza di Gödel in matematica e logica, si arriva alla conclusione che per un ordinamento giuridico:

*“non è necessaria una corrispondenza completa e senza eccezioni [con gli eventi ad esso esterni]. Anzi, deve esserci perfino la possibilità d’una discrepanza fra l’ordinamento normativo e l’ambito degli eventi ad esso riferibili, perché senza una tale possibilità, un ordinamento normativo non avrebbe assolutamente nessun significato.”*<sup>109</sup>

In seguito:

---

<sup>107</sup> H. KELSEN, *dottrina pura del diritto*, op. cit., p.99.

<sup>108</sup> Ivi “*il contenuto della norma fondamentale*” p.99-101.

<sup>109</sup> Ivi p.100.

*“un ordinamento normativo deve perdere la sua validità di fronte alla realtà che cessa di corrispondergli fino a un certo grado.”<sup>110</sup>*

Si arriva ad un risultato completamente opposto, la matematica dev'essere completa per avere senso, un ordinamento giuridico invece dev'essere incompleto. In termini gergali questa evidenza è espressa dal fatto che in matematica la regola non ammette eccezioni ed in diritto, invece, sono le eccezioni che confermano la regola.

Sul piano fenomenico l'algoritmo, rispondente alla legge di causalità, non potrà che riportare la dicitura “*error*” dinanzi ad elementi non previsti dal suo codice sorgente.<sup>111</sup> Di contro un ordinamento giuridico di fronte ad un nuovo fenomeno non perderà la sua validità, bensì cercherà il modo di ordinarlo riferendolo a se stesso, qualificandolo come lecito o illecito a seconda dei valori propri dell'ordinamento.

Questa peculiarità, che sul piano teorico è idonea ad individuare due schemi di pensiero differenti e su quello pratico specie di fenomeni che mostrano comportamenti opposti, è la prima differenza di principio tra questi due aspetti complementari della realtà che ai fini della presente trattazione verranno descritti con la terminologia di principio di causalità e principio d'imputazione.

La seconda differenza è già stata presentata, Kelsen ne descrive la forma grammaticale: se c'è *A ci sarà B* per la causalità e se c'è *A ci dev'essere B* per l'imputazione<sup>112</sup>. Si è detto di come sia la *Grundnorm* a creare il dover essere in qualità di norma fondamentale presupposta e legittimante, è bene però precisare un ulteriore elemento per distinguere queste due forme grammaticali.

Ciò che avviene nel contesto della causalità rileva meramente sul piano oggettivo e quindi, come si è visto, gli elementi causali del nesso (A-B) rilevano come causa ed effetto, subendo il moto inerziale. Nel contesto dell'imputazione, ciò che crea il dover essere è invece la volontà, cioè un'azione soggettiva; infatti, è lo stesso Kelsen ad affermare che “*la relazione*

---

<sup>110</sup> *Ivi* p.101.

<sup>111</sup> È pur vero che le metodologie tipiche dell'IA (*machine learning* e *reti neurali*) in parte risolvono questo problema garantendo maggiore autonomia di scelta all'IA nei confronti dell'imprevisto, è questo meccanismo che crea l'*effetto scatola nera*.

<sup>112</sup> Descritta a p.216 di *Dottrina pura del diritto, op. cit.*, come “*Prima differenza tra principio di causalità e quello d'imputazione*”.

*tra la condizione e la conseguenza che si trova in una legge morale, religiosa o giuridica, è stabilita da atti di esseri umani o sovraumani”*.<sup>113</sup>

Sotto questo profilo, prendendo come punto di riferimento la *volontà*, si può tracciare in principio una prima linea di demarcazione tra causalità, come principio operante per la logica formale, ed imputazione, come principio cardine della logica argomentativa. Infatti, una stessa situazione, lo scioglimento del vincolo sociale, viene interpretata diversamente a seconda che la chiave usata sia giuridica o scientifica.

Quello che per la logica formale può essere descritto in maniera univoca, cioè la separazione di una parte dal tutto, nel caso di specie la separazione che avviene tra un socio e la società, ha invece nel campo giuridico una natura mutevole. Essa è descritta dall'art.24 c.c. recante le disposizioni in tema di “*Recesso ed esclusione degli associati*” che, essendo collocate nel Titolo II del codice “*Delle persone giuridiche*”, assurgono a modello generale per la disciplina dello scioglimento del vincolo sociale nelle diverse forme societarie.

Ciò che rileva è notevole. Mentre sul fronte della causalità l'evento ha una mera descrizione in termini di causa ed effetto in cui, in ciascuno dei casi previsti dall'art. 24 c.c., l'elemento rilevante è il distaccamento del socio dalla società; su quello dell'imputazione e quindi in diritto, l'elemento realmente rilevante è l'attribuzione della volontà dello scioglimento che, nel caso della sua imputazione al socio configurerà l'ipotesi della rescissione, ma se imputabile all'assemblea dei soci configurerà la fattispecie di esclusione.

Queste due fattispecie giuridiche non costituiscono una mera formalità, per la quale si definisce in maniera diversa un dato fenomeno, ma pongono una serie di obblighi e diritti idonei ad influire sulla realtà contingente. Quindi sebbene resti valida la rappresentazione causale per lo scioglimento del vincolo, per poterne cogliere tutti gli aspetti si ha bisogno di un *quid pluris*, la rappresentazione del nesso d'imputazione, capace di indicare a chi è imputabile la volontà di separarsi<sup>114</sup>.

---

<sup>113</sup> *Ivi* p.217.

<sup>114</sup> Si badi bene che questo elemento non è spiegabile solo sul piano causale, quanto piuttosto si rende effettivamente necessario spiegare chi ha dato origine alla separazione, la cui descrizione solo su un piano causale ed oggettivo non avrebbe alcuna rilevanza giuridica ai fini dell'applicazione dell'art.24 c.c.

È chiarita ulteriormente la motivazione per la quale J. R. Searle si dichiara più incline ad attribuire a Edipo la responsabilità di aver sposato sua madre piuttosto che quella di aver mosso delle molecole, la prima contiene in sé un'intenzione sebbene inconsapevole di sposare quella donna, che sotto la definizione di Giocasta invece contiene un'intenzione consapevole. Mentre una tale volontà consapevole o inconsapevole non può essere riscontrata nel movimento di molecole, poiché esso si pone al di là della sua sfera volitiva e del suo spazio di libertà d'azione.

Ed infatti è in questa prospettiva, sebbene molto controversa, che Kelsen dichiara:

*“Generalmente si afferma che soltanto la sua libertà, cioè la sua esenzione dal principio di causalità, rende possibile l'imputazione. È invece giusto l'opposto. Gli esseri umani sono liberi perché imputano la ricompensa, la penitenza o il castigo come conseguenza, al comportamento umano come condizione.”<sup>115</sup>*

Su questo non posso dirmi d'accordo. Va premesso che nel periodo in cui Kelsen scrive il determinismo era dominante ed indiscutibile ed infatti, anche in passaggi successivi, viene corretto quanto affermato fortemente in precedenza, concludendo poi infine con le stesse idee riportate in questa tesi: che causalità ed imputazione siano principi complementari.

Ad ogni modo la scienza ha fatto grandi progressi da allora e grazie alle scoperte della fisica quantistica, riportate precedentemente, non è più un tabù ipotizzare che l'uomo possa essere capace di modificare il comportamento delle particelle. È infatti in questa prospettiva che muovono ad esempio le idee del fisico premio nobel Roger Penrose che, nel tentativo di dare una spiegazione coerente dell'indeterminazione quantistica, ha proposto la teoria della coscienza quantistica o Orch-or (*orchestred objective reduction*).<sup>116</sup>

Questa prospettiva già accennata in precedenza potrebbe essere la chiave per dare una risposta ai quesiti che la fisica quantistica ha sollevato più di un secolo fa. D'altronde sembra chiaro, nell'opinione di chi scrive, che sia questo il tratto principale dell'intera questione e inoltre la strada più promettente come lo stesso J. R. Searle afferma in *“Freedom and neurobiology”*. Quando e se la scienza troverà conferme sperimentali all'incidenza della volontà o della

---

<sup>115</sup> *Ivi* p.220.

<sup>116</sup> Per delle considerazioni rilevanti dell'autore cfr. R. PENROSE, *“La mente nuova dell'imperatore”*, *op. cit.*

coscienza nei fenomeni d'interferenza descritti dalla funzione d'onda  $\psi$  allora si potrà configurare anche il movimento di molecole come un *unintentional action* di Edipo, ma fino ad allora è corretto trattare questa ipotesi come tale.

Per riportare invece le due diverse forme grammaticali dell'imputazione e della causalità ad esempi concreti, si possono ipotizzare tre scenari:

1. L'algoritmo, puramente causale, si ritrova ad eseguire un programma, in questo caso varranno solo le istruzioni fornite dal programmatore ordinate coerentemente in un rapporto di consequenzialità (istruzione A, causa, che se eseguita correttamente porterà all'evento B, conseguenza). Se questo meccanismo incontra un ostacolo si fermerà.
2. Un umano si ritrova di fronte ad una situazione che richiede una sua azione, costui soppeserà le opzioni mediante un ragionamento logico e quindi causale (se compirò l'azione A si verificherà l'effetto B), ma nel momento di scegliere cosa sceglierà? Probabilmente sceglierà in base a ciò che vuole realizzare, in questo contesto opera l'imputazione (se voglio ottenere l'effetto B devo effettuare la scelta A), chiarito l'obiettivo allora compirà l'azione.
3. Un'ipotesi mista: l'IA. Ipotizzando un sistema con un elevatissimo livello di autonomia, come si comporterà dinanzi ad uno scenario che richiede una scelta? A differenza dell'algoritmo che ha un percorso rigido da seguire, si è detto che l'IA ha un margine di autonomia, di fronte ad un nuovo elemento non si fermerà, bensì eseguirà una scelta tra quelle possibili. Quindi eseguendo un procedimento per conseguire un risultato (se eseguo A, ci sarà l'esecuzione di B) si verifica un elemento impeditivo, dinanzi ad esso l'IA si troverà a dover scegliere cosa fare per raggiungere l'obiettivo impostato dal programmatore quindi (per raggiungere l'obiettivo B, dovrò eseguire A). Vi è quindi la parvenza di un'intenzionalità che però non vi è; infatti la volontà che sembra manifestare l'IA non è che quella del programmatore. Sia l'obiettivo (B) che le istruzioni (A) sono predeterminate. Vi sarà dunque un comportamento autonomo ma non libero.

La volontà che opera nei sistemi ad intelligenza artificiale, ad oggi, è la nostra e pertanto sarà necessariamente nostra la responsabilità di eventuali danni che essa causa. In questo fattore si ravvisa la differenza tra intelligenza umana ed artificiale, la prima è capace di

autodeterminarsi, la seconda, sebbene capace di agire secondo i meccanismi propri dell'imputazione, dovrà fare necessariamente riferimento a obiettivi, risultati e scopi prefissati da un'intelligenza umana.

Questa distinzione tra intelligenza umana e artificiale può essere ancor meglio definita grazie al diverso approccio di fronte a ciò che Searle chiama "*experience of the gap*"<sup>117</sup>, ovvero l'esperienza della possibilità di scelta. Se un sistema IA dinanzi alla possibilità di scelta si troverà vincolato da quell'attributo definito come *bounded optimality*, che lo costringerà a scegliere una delle possibilità comprese nelle impostazioni o programmi che già possiede dentro di esso, questo stesso limite non si applicherà ad un'intelligenza umana, la quale, dinanzi alla possibilità di scelta ovvero un fenomeno nuovo, avrà a sua disposizione gli strumenti per dare una risposta innovativa che non necessariamente era già presente in sé stessa.

Questa differenza chiave mostra tutta la sua forza in quella qualità che viene definita creatività e che per sua natura richiede una certa dose di libertà. L'intelligenza umana è stata capace di creare oggetti che non avevano precedenti nella storia naturale: edifici, libri, veicoli e l'IA stessa, tra questi vi sono pure complessi di norme autonomi che sebbene facenti parte del più grande sistema delle leggi naturali, in parte istituiscono nuove regole e con esse nuove possibilità. Di contro un'intelligenza artificiale, con la sua autonomia, sarebbe in grado al massimo di replicare fedelmente certe opere dell'ingegno, ma non sarebbe capace di crearle da zero essendo priva d'iniziativa.

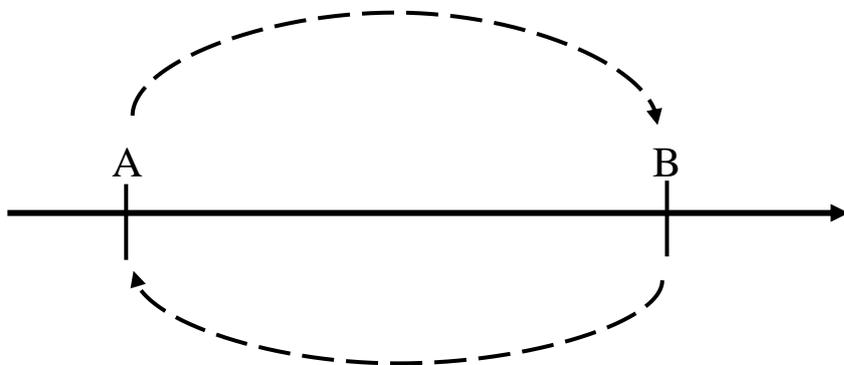
Kelsen afferma che un'ulteriore differenza tra causalità ed imputazione<sup>118</sup> è che la prima è infinita, mentre la seconda è finita. In particolare, viene fatto notare che nel contesto della causalità vi sarà una catena infinita di cause ed effetti che si susseguono reciprocamente e che non avranno mai fine, mentre nel contesto dell'imputazione, essendo di natura diversa il nesso tra condizione e conseguenza che lega i termini, si avrà un tipo di connessione biunivoco e finito.

---

<sup>117</sup> J. R. SEARLE, *Freedom and Neurobiology*, op.cit., p. 66-68.

<sup>118</sup> H. KELSEN, *dottrina pura del diritto*, op. cit. p.217-218.

Rispetto a quest'affermazione può essere utile inquadrare graficamente i due diversi nessi, ampliando la rappresentazione vettoriale riportata in precedenza:



Ripartendo dall'inizio si è tracciata una singola linea, rappresentante la causalità intesa comunemente; successivamente si è aggiunto un *quid pluris* circoscrivendo una determinata serie causale (A-B) e ci si è chiesti, alla luce degli interrogativi posti da Searle, come fosse possibile distinguere una parte dal tutto, lasciando in sospeso la risposta. Ivi viene ampliato ulteriormente lo schema, aggiungendo un ulteriore nesso, quello d'imputazione, a completare il quadro.

Si è detto infatti di come, restando nell'ambito della sola causalità, risultasse impossibile distinguere la singola serie causale (A-B) dal resto, questo proprio perché la causalità ha questa peculiare tendenza all'infinito. Se si resta solo nell'ambito delle cause e degli effetti risulta impossibile effettuare distinzioni di ogni sorta e per poterle fare ci si è avvalsi fin ora di definizioni ulteriori ed ausiliare quali quelle di: significato, soggetto, contenuto.

Queste nozioni non troverebbero posto all'interno del paradigma della causalità, forse solo quella di soggetto inteso come causa potrebbe avere un qualche rilievo. Ciò avviene perché come precisato più volte, nella causalità, un fenomeno rileva solo in termini *sintattici* e quindi come insieme di regole, d'altro canto la spiegazione *semantica* di certi fenomeni, quella che comprende le nozioni di significato, soggetto, contenuto, richiede qualcosa, che nella causalità non trova riscontro.

Infatti, se ci si chiede ad esempio perché Edipo ha sposato Giocasta, sarebbe per noi insufficiente affermare che ha mosso delle molecole per farlo, quanto piuttosto saremmo

inclinati a spiegare come e per quali ragioni si è formata in Edipo una tale volontà. Si noti che questo approccio non esclude la causalità perché nel descrivere gli scenari e le azioni che portarono Edipo a sposare Giocasta non se ne potrebbe fare a meno, e del pari di ciò sarebbe insufficiente senza spiegare i *motivi* che lo hanno condotto a compiere certe azioni.

L'imputazione così qualificata si aggiunge alla causalità come principio ermeneutico e sarebbe errato considerarla come avente mero valore descrittivo, perché così come la causalità descrive certe leggi che però danno forma alla realtà e finisce con l'identificarsi con esse, lo stesso accade per l'imputazione.

Ciò che ovviamente quest'affermazione implica è che certi fenomeni che spesso sono stati in passato relegati alla metafisica, al regno della psiche o dell'etere, siano invece reali e capaci di avere un impatto concreto sulla realtà, uno di questi è certamente il pensiero.

Nello schema qui riportato viene preso in considerazione anche il principio d'imputazione tenuto conto della descrizione offerta da Hans Kelsen. Com'è possibile notare la prima cosa che risalta immediatamente è che se la causalità tende ad infinito e rispetto ad essa il nesso (A-B) assume una posizione indefinita, nel contesto dell'imputazione, ovvero le frecce curve riportate nel grafico, essa è finita e la serie (A-B) ha una posizione rilevante in quanto si tratta dell'inizio e della fine del meccanismo circolare dell'imputazione.

Sono le frecce curve nel grafico che possono spiegare come viene a distinguersi una parte dal tutto. Infatti, se il nesso causale (A-B) spiega come si arriva dalla causa (A) all'effetto (B) è poi la freccia curva in basso che, riportando l'effetto (B) alla causa (A), è in grado di spiegare perché l'effetto (B) è imputabile proprio alla singola causa (A) e ciò può avvenire attingendo ad ulteriori fonti come il significato e il contenuto di ciò che rappresenta il nesso d'imputazione (A-B). Risulta in ultimo chiaro perché questi due principi vengono rappresentati graficamente insieme: essi sono complementari ed operano reciprocamente l'uno sull'altro.

Kelsen nel descrivere questa differenza afferma che “*la catena d'imputazione [...] ha soltanto due anelli*”<sup>119</sup> che nel grafico sono rappresentati da (A-B) i quali assumono posizioni completamente differenti a seconda della prospettiva della causalità o dell'imputazione.

---

<sup>119</sup> H. KELSEN, *dottrina pura del diritto, op. cit.*, p. 217.

In questo senso conviene fare un esempio riprendendone anche la forma grammaticale: se Tizio compie un'azione meritevole allora gli si dovrà essere riconoscenti. Nella prospettiva di Kelsen il punto finale dell'imputazione è proprio il comportamento umano che ne è causa e questo spiega la scelta della rappresentazione circolare di questo secondo nesso tenuto conto della sua finitezza.

Azione meritevole di tizio (A), che sul piano causale produce l'effetto concreto (B) riconoscenza, ma per spiegarsi perché di questo effetto (B) bisogna ricondurlo, ovvero imputarlo, all'azione (A) che gli ha dato origine e che non è necessariamente la causa immediatamente precedente.

Prima di riportare quest'ultima considerazione all'intero discorso di questa tesi e poterla utilizzare come chiave teorica per dare una risposta agli innumerevoli quesiti lasciati in sospeso, sarà utile chiarire in modo inequivocabile e giuridicamente rilevante la differenza tra causalità e imputazione mediante un esempio chiaro: la distinzione tra Processo e Procedimento.

La sistematica del diritto processuale è prevalentemente incentrata sul metodo, dove, se il diritto è sostanza, la sua forma processuale è il modo mediante il quale esso si estrinseca nella realtà. Nicola Picardi descrive l'evoluzione della processualistica incentrando la sua attenzione *in primis* sul passaggio da giudizio a processo,<sup>120</sup> che avviene con l'avvio dell'epoca moderna. A tal proposito viene ricordato il rifiuto da parte dei giuristi dell'età di mezzo dell'applicazione del sillogismo aristotelico in sede di giudizio, in quanto quest'ultimo era il regno del probabile mentre il primo, per il suo carattere dicotomico bastato sulla logica del vero-falso, era il regno della certezza scientifica.

Il giudizio in epoca medievale era quindi un confronto paritario di opinioni, nel quale non vi era una ricerca della verità oggettiva, caratteristica preponderante nel processo moderno, quanto una prevalenza delle argomentazioni più *verosimili*.

Con l'epoca moderna prende avvio quel dominio della razionalità che prenderà piede anche in ambito giuridico, portando la sostituzione della nozione di giudizio con quella di processo. La libertà e la giustizia a partire dall'epoca moderna non sono più garantite da opinioni

---

<sup>120</sup> N. PICARDI, *Manuale del processo civile*, Giuffrè, terza ed., 2013, p. 3-5.

autorevoli o dal senso etico del giudice quanto dalla più razionale prerogativa del rispetto delle garanzie processuali. Se ciò da un lato costituisce una conquista di civiltà dall'altro ha comportato una perdita del senso originario del diritto.

Da qui, con la riforma costituzionale del 1999, sono state introdotte nell'art. 111 Cost. le regole del “*giusto processo*”, già codificate nell'art. 6 CEDU, con il chiaro intento di riportare una dimensione etica del processo. È in questo quadro che va a porsi la distinzione tra procedimento e processo descritta da Picardi<sup>121</sup> che ai fini della tesi racchiude il senso delle precedenti distinzioni tra sintassi e semantica da un lato e tra causalità ed imputazione dall'altro.

Riportando il ragionamento allo schema argomentativo precedente dove la causalità è rappresentata in maniera *lineare* e l'imputazione in forma *circolare* mi appresto a descrivere i punti salienti individuati dall'autore: il procedimento è una serie di atti concatenati tra loro in cui solo il primo e l'ultimo degli atti processuali hanno un solo punto di contatto, questo metodo di interpretare il processo come procedimento risponde ad una logica-formale nella quale rileva solo la forma sintattica della struttura giuridica del processo, in questo contesto gli atti processuali non sono che elementi e le norme vengono considerate alla stregua di regole.

Nel processo vi è invece un senso qualitativamente diverso, in esso non è sufficiente l'ordine logico proprio del procedimento ma vi è un *quid pluris*, la logica-argomentativa che nel diagramma è rappresentata dalla curvatura delle frecce rappresentanti il nesso d'imputazione. Se la linearità della causalità spiega *come* si susseguono gli atti processuali nel procedimento, la curvatura dell'imputazione spiega *perché* essi si susseguono in quel preciso ordine.

Questa differenza viene ulteriormente accentuata nel contesto del principio del contraddittorio che, secondo il giudizio di Nicola Picardi, è il vero elemento di discriminazione tra processo e procedimento.

L'assunto fondamentale del contraddittorio è la parità di diritti delle parti e la loro equidistanza dal giudice. Questa situazione non può essere rappresentata appropriatamente su una struttura lineare come quella causale, bensì richiede una struttura circolare come quella

---

<sup>121</sup> Ivi p.229-243.

del processo dove, se il giudice è il centro, e le parti si collocano sul perimetro, qualunque posizione assumeranno saranno equidistanti dal giudice e sullo stesso perimetro di partenza, quindi godranno degli stessi diritti.

Infine, la natura finita del processo va individuata nella coincidenza tra inizio e fine del nesso d'imputazione dove, l'evento che dà causa al processo è lo stesso che viene giudicato alla fine. Di contro nel procedimento, data la natura formale del nesso che collega i vari elementi, non sarà riscontrabile questa coincidenza per la natura indefinita degli elementi che ne danno inizio e fine, ovvero atto di citazione e sentenza che sul piano formale sono atti differenti.

In conclusione, per la differenza tra logica-formale e logica-argomentativa sul piano metodologico è bene riportare quanto detto da Nicola Picardi:

*“i momenti fondamentali (ad es. il contraddittorio, la prova ed il giudizio) obbediscono ad una logica diversa, non del necessario e dell'assoluto, ma solo del probabile e del ragionevole. Si tratta del metodo dialettico, di quella logica argomentativa che [...] si configura come una ricerca dialettica di «verità» verosimili, probabili, incerte.”*<sup>122</sup>

Ed infine:

*“il processo, almeno nei suoi snodi centrali, va appunto studiato alla luce della logica giuridica intesa come teoria dell'argomentazione. In questo quadro i «principi» non vengono più concepiti come assiomi o verità evidenti, ma come le antiche *regulae iuris*, cioè le regole fondamentali della nostra materia, frutto di una logica del senso comune, regole destinate a facilitare l'interpretazione basata sulla ragionevolezza.”*<sup>123</sup>

Questo punto sarà fondamentale per individuare l'eccezione alla *regula iuris* della causalità che potrà dare una soluzione esaustiva al problema specifico dell'intelligenza artificiale. Ma in ultimo, riformulando qui la distinzione precedentemente esposta tra *Grundnorm* ed assiomi, ovvero tra diritto e scienza, tra un modo di pensare ed un altro è mio desiderio lasciare l'ultima parola a Nicola Picardi, autore di un libro fondamentale per la mia formazione:

*“Ciò che la logica rinserra, l'argomentazione lascia aperto.”*<sup>124</sup>

---

<sup>122</sup> N. PICARDI, *Manuale del processo civile*, op. cit., p. 17.

<sup>123</sup> Ibidem.

<sup>124</sup> Ibidem.

## IV. PROFILI DI DIRITTO

### IV.1 Disciplina giuridica dell'IA

Le conclusioni teoriche riportate fin qui ambiscono a dare una risposta a quegli interrogativi sollevati da questa tesi: cosa distingue umano da non umano, azione da non azione ed infine come conciliare determinismo e libero arbitrio.

Le risposte date nei paragrafi precedenti, ancorché non esaustive, possono considerarsi in linea col discorso fin qui condotto. Si sono infatti formulati due principi analoghi per spiegare la complementarità di determinismo e libero arbitrio e, sulla base di essi, dare risposta ai due quesiti teorici minori e derivati dalla questione principale. Ma considerare queste speculazioni alla stregua di una conclusione risulterebbe tanto fuorviante quanto errato.

È stata infatti la necessità di avere una soluzione al caso concreto dell'intelligenza artificiale a spingere i confini del discorso fino alla massima estensione. Era una questione di principio, sulla quale andava fatta chiarezza.

La degna conclusione di quanto fin qui argomentato concerne la disciplina giuridica dell'IA. Si rende dunque doveroso riportarla alle conclusioni teoriche formulate e, usando le stesse come strumento, superare i limiti dell'attuale scienza giuridica che hanno reso problematico il caso dell'IA.

È infatti la totale incompatibilità tra l'attuale concetto di nesso causale e il profilo specifico della responsabilità concernente il danno commesso da un sistema IA ad aver generato l'intera questione. Se da un lato, sul profilo del ramo privatistico del diritto, questo fenomeno si risolve in una mera scelta di opportunità fra un tipo di regolamentazione ed un altro, dall'altro, e quindi il ramo penalistico del diritto ma anche tutti quegli aspetti civilistici legati all'accertamento della responsabilità, si potrebbe manifestare un *vulnus* idoneo a coinvolgere l'intero ordinamento, ovvero l'incompatibilità tra la richiesta del legislatore di accertare il nesso di causalità e l'effettiva inadeguatezza per il caso dell'IA e *consimili*.

Questo problema si presenterà di certo dinanzi alle corti nel momento in cui le cause aventi ad oggetto i danni commessi da un'IA saranno più frequenti. Se questo non è già successo, si badi bene, non è per la scarsa diffusione del fenomeno, che invece è quasi onnipresente nella nostra attuale quotidianità su internet, ma è piuttosto il risultato della mancanza di previsioni legislative che identifichino delle fattispecie legalmente sanzionabili.

Sebbene il nostro ordinamento non preveda la tipicità dell'azione, cionondimeno si rende necessario, nel diritto penale, la presenza di una fattispecie astratta nella quale sussumere quella concreta per la tassatività dell'ordinamento e per far ciò è necessario un intervento legislativo.

Ed è anche su questo punto che il diritto si sta muovendo. L'Unione Europea, che sta fortemente puntando su questo settore, ha già fatto le prime mosse, descritte dal primo capitolo di questa tesi nel quale si è pure detto di una imminente proposta regolamentare, che è finalmente intervenuta durante la redazione e che per la sua importanza merita un approfondimento dettagliato.

#### IV.1.1 Profilo civile: Artificial intelligence act

Il ventuno aprile 2021, la Commissione Europea ha presentato una proposta di regolamento recante norme armonizzate in tema di intelligenza artificiale dal titolo “*Artificial intelligence act*”.<sup>125</sup>

Il lavoro di ricerca che fornisce le basi per le argomentazioni di questa tesi ha risentito notevolmente dei lavori preparatori della Commissione confluiti in questa proposta legislativa. Questi sono stati presentati nel primo capitolo come chiave di lettura per comprendere giuridicamente il fenomeno emergente dell'intelligenza artificiale.

Si rimanda il lettore al capitolo dedicato per maggiori approfondimenti, mentre è utile riprendere in breve lo schema che ha condotto la CE alla proposta oggetto d'analisi. Sono due i pilastri su cui si è fondato il lavoro preparatorio: l'etica, su cui si è incentrato il lavoro del gruppo di esperti indipendenti (HLEG) e la tecnica, oggetto di numerosi atti della CE tra cui

---

<sup>125</sup> Commissione Europea, *Artificial Intelligence Act*, COM(2021) 206 final, *op. cit.*

è bene ricordare il libro bianco e la Relazione su responsabilità e sicurezza (nella quale vengono presentate le quattro caratteristiche che danno il titolo ai rispettivi sottoparagrafi del capitolo I.3 di questa tesi).

È sulla base di etica e tecnica che vengono elaborate le più importanti disposizioni legislative che l'Artificial intelligence act si propone di introdurre nell'ordinamento giuridico europeo. In via di prima approssimazione ciò che appare subito evidente è una forte presa di posizione rispetto a delle pratiche, proprie dell'arte dell'intelligenza artificiale, che la Commissione proibisce esplicitamente nel titolo II recante le disposizioni in tema di "*Prohibited artificial intelligence practices*".<sup>126</sup>

Subito dopo aver stabilito chiarimenti in tema di definizioni (art. 3) e scopo della normativa (art. 4), in merito ai quali si nota immediatamente l'influenza dei lavori preparatori dell'HLEG e soprattutto delle "*Ethics Guidelines for Trustworthy AI*" rispetto alle quali vi è quasi totale coincidenza sulla definizione di IA che comprende al suo interno anche le tecniche di machine learning (art. 3.1), la proposta normativa si addentra fin da subito nel tema etico, individuando quattro pratiche esplicitamente proibite poiché idonee a generare un "*unacceptable risk*"<sup>127</sup>:

1. Pratiche atte ad influenzare il comportamento degli individui.
2. Pratiche che possano mettere a rischio categorie di individui vulnerabili: minori, soggetti con disabilità fisiche o mentali.
3. Pratiche istitutive di sistemi di credito sociale.
4. Partiche di profilazione biometrica.

Le valutazioni di opportunità su cui si basano i giudizi di valore espressi dall'art. 5 sono quelle elaborate dall'HLEG<sup>128</sup>. Il quadro etico che ne deriva è coerente con i valori giuridici europei e con la cultura degli *human rights* dove la centralità dell'individuo è garantita dalla forte tutela nei confronti delle attività invasive di apparati pubblici o privati.

---

<sup>126</sup> *Ivi*, art. 5.

<sup>127</sup> Sono tre le classificazioni del rischio previste dall'AI act: *unacceptable risk*, *high risk*, *low or minimal risk*, cfr. Commissione Europea, *Artificial Intelligence Act*, COM(2021) 206 final, op. cit. p. 12.

<sup>128</sup> Riservatezza e Governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale ed ambientale sono alcuni dei requisiti individuati dall'HLEG nelle "*Ethics Guidelines for Trustworthy AI*" ai quali la Commissione fa esplicito riferimento nella proposta.

È indubbiamente questo il merito dell'introduzione dei divieti di profilazione biometrica ed attribuzione di punteggi sociali, che definiscono una linea di pensiero netta del legislatore europeo intento a proteggere l'individualità in tutte le sue estrinsecazioni, anche informatiche, da possibili attività invasive pubbliche o private.

In questo senso appare esservi una certa continuità con l'indirizzo inaugurato dal General data protection regulation (GDPR) che condivide con l'Artificial intelligence act il medesimo sistema di valori. Questo quadro valoriale, che forse trova la sua massima espressione nella CEDU, è proprio dei sistemi occidentali in cui la cultura europea ha un ruolo trainante.

Si può notare ad esempio come, nel concreto, l'indirizzo che quest'approccio sembra inaugurare sia quello di porre un argine al potere delle grandi *digital company* (Facebook, Google, Tik tok, Twitter). La continuità alla quale accennavo emerge fortemente se si considera che, sia nell'GDPR come nell'Artificial intelligence act, le tutele poste a garanzia del cittadino sono finalizzate ad una protezione dell'individuo contro l'esercizio del potere piuttosto che la regolamentazione di quest'ultimo, ne risulta un quadro in cui il destinatario della norma è il cittadino come soggetto meritevole di tutela e non l'apparato pubblico o privato che esercita il potere.

Eccezione fatta per la menzione ai sistemi di credito sociale, in cui il riferimento all'attuale sistema del governo cinese non è neanche troppo velato. Infatti, è la Cina l'unica nazione che attualmente potrebbe rientrare nella categoria di "*public authority*" prevista dall'art. 5.1(c) se si considera che ha un forte piano governativo in ambito digitale di cui il sistema di credito sociale (SCS) è l'ultima di una serie d'iniziative particolarmente invasive attuate dal governo di Pechino.<sup>129</sup>

Basti qui ricordare, al fine di comprendere la diversità di approcci in materia digitale, come la Cina abbia inaugurato nel 1998 la sua strategia *digital* con il "*Golden Shield Project*" (conosciuto anche come Great Firewall) che rappresenta forse il più ampio programma di censura e sorveglianza di massa attuato da una nazione in epoca moderna. Con questo piano strategico si è creato uno spazio informatico cinese quasi impermeabile alle influenze esterne

---

<sup>129</sup> Cfr. [https://it.wikipedia.org/wiki/Sistema\\_di\\_credito\\_sociale](https://it.wikipedia.org/wiki/Sistema_di_credito_sociale)

sgradite. Esso consiste di pratiche informatiche volte a bloccare indirizzi IP stranieri e filtraggio delle URL al fine di evitare la diffusione di contenuti vietati dalla legge.<sup>130</sup>

Se in Europa il quadro valoriale è di stampo individualista, come confermato dalla lettura congiunta di GDPR ed Artificial intelligence act, in Cina vi è una forte influenza della cultura di massa tipica del comunismo dove il governo assume un ruolo centrale nel controllo della vita degli individui. Per comprenderne la portata basti qui citare che, grazie al SCS, il governo cinese ha il potere di attribuire un punteggio sociale ai cittadini, fissati alcuni valori di riferimento, e sulla base di esso comminare sanzioni o ricompense.

Di segno diverso è invece l'approccio americano in cui la tradizione anglosassone, per la quale le *big companies* hanno storicamente un rapporto privilegiato con lo Stato, sembra prevalere. Infatti, è in questa cornice che si colloca il recente episodio che ha visto protagonisti di un aspro conflitto istituzionale (o para-istituzionale) l'uscente presidente degli Stati Uniti d'America Donald Trump e le più grandi società sul mercato dei social network: Facebook e Twitter.

I fatti dell'accaduto si collocano dopo l'assalto a Capitol Hill, sede del Congresso americano, in relazione al quale la posizione dell'ex presidente Trump è molto controversa. Infatti, quest'ultimo, avendo pubblicato diversi tweet dal tenore particolarmente ostile ed inneggiante all'odio e alla violenza, è stato ritenuto da molti l'autore di questa mobilitazione di massa. La reazione delle piattaforme digitali (Twitter e Facebook) sulle quali si era perpetrato l'incitamento è stata quella di rimuovere gli account di Trump.

Questa vicenda, esclusivamente politica, sebbene coinvolgente figure istituzionali e diritti fondamentali come la libertà di espressione non ha visto coinvolta la giustizia nelle sue forme ordinarie, e questo è di certo singolare se si considera che si tratta di un profondo conflitto di poteri che ha sicuramente inciso sulla vita pubblica del paese.

In questo caso vi è stato un esercizio diretto del potere di censura da parte dei social network interessati a danno della più alta carica dello stato federale americano. Fatto ancor più singolare è che le norme invocate per legittimare questo tipo intervento non sono state quelle costituzionali, bensì le *policies* delle *community* di Facebook e Twitter, le quali, com'è noto,

---

<sup>130</sup> Cfr. [https://it.wikipedia.org/wiki/Great\\_Firewall](https://it.wikipedia.org/wiki/Great_Firewall)

sono discrezionalmente stabilite e non facenti parte dell'ordinamento giuridico americano, andando così a configurare ciò che, a ragione, è stata definita una privatizzazione del diritto fondamentale alla libertà d'espressione.<sup>131</sup>

Ciò che infine potrebbe sembrare è che vi sia una certa complicità tra l'azione dei social network e gli altri poteri dello stato americano, andando quasi a inquadrarsi nel ben noto sistema di *checks and balances* della tradizione costituzionale americana. Secondo quest'interpretazione, l'azione di Facebook e Twitter sarebbe stata atta a preservare l'assetto istituzionale americano, punendo così di fatto un tentativo di *golpe* perpetrato dal presidente uscente.

Ciò che risulta evidente in ultimo è sicuramente lo strapotere di queste piattaforme digitali, rispetto alle quali l'UE è l'unica istituzione che si differenzia per metodi. Infatti, a differenza dei modelli americano e cinese che, con forme diverse, sembrano servirsi di queste piattaforme ai fini dell'esercizio del potere, beneficiando di un vasto e notevole vuoto normativo, l'UE soltanto ha promosso diverse iniziative per la regolamentazione del fenomeno, tracciando così una linea di demarcazione netta tra valori e approcci diversi.

L'etica dell'Artificial intelligence act è dunque questa: limitare il potere e garantire al singolo cittadino degli efficaci strumenti di tutela contro le possibili attività invasive lesive della propria sfera privata, siano esse perpetrate da autorità nazionali o imprese private.

Un ulteriore esempio di quest'impianto garantista è la previsione di una doppia riserva, di legge e di giurisdizione, per l'uso da parte di autorità pubbliche di IA per la profilazione biometrica, ordinariamente proibita dall'art. 5.1 (d).

Infatti, riproducendo uno schema simile a quello vigente nell'ordinamento penale italiano in tema di intercettazioni (art 266 c.p. e seguenti), quanto previsto dall'art. 5.1(d) dell'AI act costituisce la regola, ovvero l'inammissibilità di pratiche di profilazione biometrica, mentre l'esercizio di questo potere da parte di un autorità pubblica rappresenta l'eccezione e, come tale, dovrà essere sottoposto all'autorizzazione da parte di una specifica autorità giudiziaria

---

<sup>131</sup> Per un confronto tra approccio americano ed europeo F. PAOLUCCI, *Il blocco dei social di Trump e la libertà di espressione online*, Ius in Itinere, 09.4.2021, p. 5-9, URL = <https://www.iusinitinere.it/il-blocco-dei-social-di-trump-e-la-liberta-di-espressione-online-35567>

nonché essere previsto *ex lege*. Questa previsione dell'art. 5.3 è idonea a rafforzare un quadro giuridico europeo dove, a differenza di altre nazioni, la tendenza è quella a limitare l'esercizio del potere mediante la legge nel contesto dello Stato di diritto.

La natura delle sanzioni che questa proposta aspira ad introdurre nello spazio giuridico europeo è quella di sanzioni amministrative, accentuando ancor di più, di fatto, l'indirizzo specifico di queste norme a carico delle piattaforme digitali private. Infatti, è previsto all'art. 71 che le *penalties* in caso di violazione dell'art. 5 (*prohibited practices*) possano spaziare da 30 000 000 EUR fino al 6% del fatturato globale annuo nel caso di società private. È certo, inoltre, che le pratiche proibite vadano ad intaccare diritti sostanziali, di natura più profonda, e che richiederanno anche un adeguamento del sistema penalistico degli stati membri, ma quest'aspetto verrà approfondito successivamente poiché richiedente una trattazione esaustiva ed assestante.

Proseguendo nell'analisi dell'etica dell'Artificial intelligence act, il giudizio che bisogna dare alle scelte della Commissione è indubbiamente positivo. Questa normativa, inquadrandosi in una serie di scelte inaugurata col GDPR, rappresenta una forte presa di posizione in senso individualista e garantista in un palcoscenico mondiale dove i principali attori invece hanno optato per strategie atte all'esercizio ed alla gestione del potere che, a seconda dei metodi propri di diversi ordinamenti, viene modulato da apparati pubblici *ex lege* o da grandi complessi di aziende private e che infine si riduce sempre ad attività invasive e talvolta coercitive idonee ad influire profondamente sulla vita degli individui.

Quest'approccio garantista è sufficiente a tutelarci e questo sicuramente rappresenta una forte garanzia per i cittadini europei, ciononostante, quest'approccio può essere definito sotto certi aspetti passivo. Le misure attive dell'Europa e dei paesi europei sono minime e l'impostazione passiva dell'AI act conferma un quadro mondiale dove la leadership digitale appartiene a USA e Cina, potendosi quasi parlare di quest'iniziativa legislativa come uno scudo legale o una contromisura rispetto a dei poteri esterni minacciosi.

La forza dell'UE e delle sue iniziative sarà comunque determinata dalle scelte future e, a mio parere, la strada intrapresa dalla Commissione, sebbene ancora debole, potrà rivelarsi un domani vincente perché poggia su valori che sono idonei ad essere condivisi da qualunque individuo: libertà, personalità e giustizia. Inoltre, l'UE ha marcatamente posto l'accento sui

limiti che il senso etico impone di prendere in considerazione, la previsione di pratiche illecite risponde infatti ad un'esigenza comune: cercare di dare un indirizzo ed una regolamentazione a questa rivoluzione digitale che fa parte del nostro presente e del nostro divenire.

In questo senso si può parlare dell'AI act come di una vera pietra miliare nella storia della regolamentazione digitale. È infatti sul piano tecnico che questa proposta legislativa fa dei passi in avanti notevoli, sebbene ricalchi uno schema regolamentare già noto al legislatore europeo.

Il modello adottato dall'AI act è incentrato sul *risk management*. Questa scelta si traduce in una classificazione duale dell'oggetto regolamentare ponendo un *aut aut* sul tipo di IA in esame: *high risk AI* e *low risk AI*, che a seconda del tipo comporterà sul piano disciplinare un insieme di obbligazioni e oneri diversi. Questa scelta sull'oggetto della regolamentazione è accompagnata dalla previsione di un soggetto responsabile per la *high risk AI* individuato nella figura del *provider* che sarà chiamato a rispondere dei danni dovuti all'inosservanza delle norme europee.

Il Titolo III dell'AI act inaugura le disposizioni in tema di “HIGH-RISK AI SYSTEMS” con gli artt. 6 e 7 che prevedono la classificazione di IA come ad alto rischio se: *product* o *safety component of a product* andando poi a specificare, nell'allegato III, certi tipi di IA catalogati come *high-risk AI* sulla base delle funzioni che sono in grado di esplicare e della loro collocazione all'interno del tessuto sociale. Alcune tipologie che vengono richiamate dall'allegato III sono ad esempio: IA con capacità d'identificazione biometrica, IA coinvolte in gestione e funzionamento di infrastrutture critiche (vedi traffico stradale, fornitura di gas e acqua, riscaldamento ed elettricità), IA proprie degli apparati pubblici operanti nel settore dell'erogazione dei servizi pubblici, nell'amministrazione della giustizia e di contrasto alla criminalità.

Inoltre, *ex art. 7*, alla Commissione viene attribuito il potere di aggiornare l'elenco delle *high-risk AI* previste dall'annesso III, qualificando lo stesso come non esaustivo e prevedendo come basi legali per l'estensione dell'elenco diversi criteri tra i quali: lo scopo dei sistemi ad intelligenza artificiale e l'uso che ne può essere fatto, andando così a ricalcare il modello tipico europeo di valutazione del rischio comprendente l'uso previsto, l'uso prevedibile e

l'uso scorretto ragionevolmente prevedibile,<sup>132</sup> ed insieme ad esso il danno potenziale che può emergere dall'uso scorretto, dal malfunzionamento o a carico di soggetti particolarmente vulnerabili.

Queste tipologie di IA vengono classificate dal legislatore europeo come ad alto rischio sulla base di considerazioni diverse: vuoi per la potenza di calcolo o per il loro particolare ruolo nel contesto di attività delicate come l'amministrazione della giustizia o la gestione di infrastrutture critiche, tutte attività dell'IA che pongono l'accento su un unico comune denominatore: la loro capacità di incidere sui diritti fondamentali dell'individuo. Dall'altro lato invece, tutte le IA non comprese da quest'elenco saranno qualificate come a basso rischio (*low-risk AI*), per esse non sarà necessario rispettare le previsioni vincolanti che dovranno obbligatoriamente essere rispettate per le IA ad alto rischio (come le previsioni in tema di *conformity assesement* o *quality management*) ma sarà sufficiente adeguarsi a dei codici di condotta meramente raccomandati e senza alcuna forza vincolante.

Sotto questo primo profilo, la scelta di diversificare le varie tipologie di intelligenze artificiali sulla base dell'impatto potenziale sugli individui adottando il criterio del rischio come principio dirimente risulta una scelta particolarmente felice. Infatti, la soluzione adottata dalla Commissione appare essere quella più ragionevole e di minor impatto sul potenziale evolutivo di questa tecnologia tenuto conto delle cinque opzioni presentate nel primo capitolo. Nel contesto civilistico avere una normativa dettagliata di settore è sicuramente una forte garanzia processuale e, al contempo, nella realtà pragmatica si pone un argine allo strapotere delle *digital companies* lasciando invece ampi margini di libertà per le piccole iniziative private emergenti.

Se sul piano regolamentare la scelta basata su un sistema di *risk management* può dirsi pienamente condivisibile, sul piano della responsabilità restano ancora delle lacune. Infatti, la proposta mira ad introdurre un sistema misto e non ben definito di responsabilità che talvolta prevede la responsabilità oggettiva del produttore che immette in commercio l'IA

---

<sup>132</sup> Vedi in merito quanto affermato dalla Commissione Europea nella Relazione su responsabilità e sicurezza, COM (2020) 64, *cit.*, p. 10, nota 48 che richiama: Regolamento (CE) n. 765/2008, decisione n. 768/2008/CE e normativa settoriale armonizzata in materia di sicurezza dei prodotti, ad esempio la direttiva 2006/42/CE relativa alle macchine.

(definito *provider*) mentre in altri casi basa la responsabilità del danno sulla base dell'uso scorretto dell'utente o dell'operatore informatico.

Ciò che sembra restare fuori invece è la responsabilità del programmatore che è tutelato dagli innumerevoli requisiti richiesti per l'immissione in commercio delle IA ad alto rischio. Infatti, vi è nell'AI act un lungo elenco di disposizioni (artt. 8-15) comprendenti diversi oneri a carico del provider. Queste disposizioni vincolanti riportano i requisiti, anch'essi vincolanti, che la Commissione ha deciso di prevedere come condizione per l'immissione in commercio di sistemi IA ad alto rischio. Si tratta di requisiti emergenti dai lavori preparatori dell'HLEG e dalle quattro caratteristiche dell'IA individuate dalla CE (autonomia, dipendenza dai dati, complessità e opacità).

Questi requisiti rappresentano insieme alle pratiche proibite dall'art. 5 il cuore pulsante di questa proposta normativa, andando a delineare chiaramente quali norme la Commissione ha deciso di introdurre per limitare e regolamentare il fenomeno dell'IA.

Il primo di essi è descritto dall'art. 9 "*risk management system*". Coerentemente con l'impianto generale della proposta basato sulla gestione del rischio, il provider dovrà dotarsi di un sistema qualificato di gestione del rischio che comprenda al suo interno: l'identificazione e l'analisi dei rischi previsti, la valutazione dei rischi che possono potenzialmente emergere dall'uso dell'IA, la valutazione di altri possibili rischi emergenti in accordo con la valutazione dei dati successiva all'immissione in commercio ed infine l'adozione di misure di gestione e prevenzione del rischio. Tali misure dovrebbero essere sufficienti per prevenire tutti quei rischi elevati o inaccettabili cosicché a restar fuori da questo sistema siano solo i rischi accettabili che non siano in grado di causare gravi lesioni dei diritti fondamentali dell'individuo.

Inoltre è lo stesso articolo che al comma 4 prevede le misure per fronteggiare questi rischi, individuando tre possibili metodologie adottate ufficialmente dal legislatore europeo comprendenti: l'eliminazione del rischio durante la fase di design e sviluppo del prodotto finale, l'implementazione del prodotto dopo l'immissione in commercio con la previsione di specifiche modalità tali da ridurre i rischi ineliminabili ed infine la previsione di adeguata informazione circa i rischi incontro i quali può andare l'utente che si serve di quella specifica intelligenza artificiale.

L'articolo 9 conclude infine con la previsione di test obbligatori prima della immissione in commercio per testare l'IA e far fronte preventivamente ai possibili rischi che essa possa generare per gli utenti.

L'articolo 10 detta disposizioni vincolanti in tema di “*Data and data governance*”. In accordo con la dipendenza dai dati propria dei sistemi IA, con quest'articolo la Commissione intende introdurre previsioni vincolanti per far sì che vi sia un'elevata qualità dei dati forniti alle IA ed evitare i possibili errori che invece potrebbero emergere se ad esse venissero forniti dati spazzatura, evitando così la possibilità di avere *garbage in, garbage out*.

Viene innanzitutto individuato un elenco di fasi e processi della progettazione dell'IA in cui viene imposto il rispetto dei requisiti di qualità dei dati introdotti dall'AI act, tra di essi vi sono: le scelte di design rilevanti, le raccolte di dati, le operazioni di processamento e classificazione dei dati e l'esame di possibili bias. Rispetto a tutte queste operazioni che vengono effettuate sui dati, la Commissione richiede che esse vengano condotte con dati di qualità, rilevanti, privi di errori e completi, inoltre vi dovranno essere le adeguate proprietà statistiche per far sì che l'insieme di dati fornito all'IA sia rilevante ai fini della prevenzione degli errori e del suo buon funzionamento.

Accanto a queste previsioni, gli articoli 11 e 12 prevedono delle obbligazioni inerenti alla documentazione ed alla tracciabilità delle operazioni rilevanti portate a termine dall'intelligenza artificiale.

La prima di esse, l'obbligo di una documentazione tecnica adeguata, risponde all'esigenza di prevenire il rischio e di far sì che i soggetti interessati abbiano preventivamente tutta la documentazione relativa a quella specifica IA, comprendente il suo intero ciclo di progettazione e le sue componenti fondamentali. Quest'obbligo è fissato per permettere alle autorità nazionali competenti di verificare la concordanza dell'IA ad alto rischio con i requisiti della normativa europea e procedere poi con il rilascio del certificato di conformità.

Mentre l'articolo 12 prevede un'obbligazione fondamentale per il controllo *ex post* ovvero l'obbligo di *record-keeping* di tutte le operazioni che l'IA svolge durante il suo ciclo vitale. Quest'obbligo diventa fondamentale non solo per la verifica e correzione di eventuali errori dopo l'immissione in commercio del prodotto, e quindi utile per la prevenzione del rischio,

ma anche e soprattutto per il processo di attribuzione della responsabilità dopo che il danno si è verificato.

È infatti, quest'ultima, una contromisura utile per affrontare l'opacità degli algoritmi che, sebbene talvolta difficili da comprendere, possono così essere ricostruiti nelle operazioni fondamentali che hanno generato l'ipotetico danno. Quest'obbligo, quindi, può diventare una chiave per le ricostruzioni dei fatti all'interno dei processi, ma al contempo anche un metodo utile ed efficace per la riparazione degli errori di programmazione che hanno portato al verificarsi del danno.

Viene previsto dalla normativa l'obbligo di abilitare il cosiddetto *log* automatico delle attività dell'IA che dovrà restare attivo durante l'intero ciclo vitale del prodotto. Compresi in quest'obbligo sono: la registrazione del periodo d'uso, del database di riferimento dal quale provengono i dati in input, gli stessi dati in input e l'identificazione della persona fisica coinvolta nella verifica dei dati.

Successivamente all'articolo 14 vi sono delle previsioni a garanzia dell'utente denominate "*Transparency and provision of information to users*" le quali si sostanziano in un insieme di obbligazioni delle quali la centrale è l'accompagnamento al prodotto delle istruzioni d'uso prudenziale in maniera analoga ai bugiardini per i prodotti medicinali o agli elettrodomestici. Le informazioni presenti in essa consistono in: identità del provider, caratteristiche, capacità e limiti del prodotto IA, misure di *human oversight* previste.

Quest'ultimo è anche un requisito assestante previsto originariamente dalle "*Ethics Guidelines for Trustworthy AI*" dell'HLEG e codificato dall'articolo 14, il quale prevede che vi sia una appropriata interfaccia uomo-macchina che possa agevolare l'operatore informatico nell'uso della stessa. Questa misura, presa per controbilanciare l'autonomia dell'IA, tradisce anche una certa sfiducia nei confronti dell'intelligenza artificiale, determinata dalle capacità d'autonomia che, accompagnate all'opacità degli algoritmi, creano un alone di mistero e incomprensibilità circa i suoi reali meccanismi di funzionamento interno.

In questa prospettiva il controllo umano diviene una misura prudenziale atta a prevenire potenziali rischi inaspettati autogenerati dalla macchina e si concreta nell'esplicita previsione

dell'art.14.4(e) della possibilità d'intervento dell'operatore umano in caso di malfunzionamento o di mancata risposta agli ordini umani.

Infine, l'articolo 15 riprende le raccomandazioni dell'HLEG in tema di “*Accuracy, robustness and cybersecurity*”, un insieme di obbligazioni alquanto disomogeneo all'interno del quale vengono previste misure piuttosto generiche in tema di sicurezza dei sistemi IA.

Questi requisiti sono la perfetta trasposizione normativa degli orientamenti etici espressi dall'HLEG tra il 2018 ed il 2019. Si tratta di un impianto ideologico ispirato al buonsenso che spesso però, per sfortuna, finisce con lo sfociare in misure di controllo che personalmente ritengo un po' eccessive.

Il mio personale giudizio sulla normativa in esame è infatti ambiguo, se da un lato vi è un impianto tecnico-giuridico quasi perfetto, richiamando qui la scelta particolarmente felice della impostazione basata sulla gestione del rischio e la conseguente categorizzazione dell'IA con l'ulteriore previsione di autorità pubbliche preposte al controllo dei requisiti; dall'altro mi sento di rilevare una certa carenza di decisione su certe scelte specifiche in merito alle quali si poteva far meglio, mi riferisco in particolare a tutte quelle misure (*human oversight, record-keeping*) che seppur utili sono troppo improntate ad una logica del controllo a mio parere non necessaria.

Infatti, la critica sul piano etico della passività di certi aspetti della normativa si traduce in quello tecnico in una eccessiva smania di controllo dettata forse dalla poca comprensione del fenomeno e dalla mancanza di una vera esperienza europea in tema d'intelligenza artificiale, cosa che non manca invece agli USA ed alla Cina. L'UE ha delle forti basi teoriche, forse le migliori al mondo, in tema di ricerca, basti qui citare la ricerca nell'ambito della fisica quantistica presso il CERN di Ginevra che è la più avanzata al mondo e produce scoperte sempre più sorprendenti tra le quali ad esempio il Bosone di Higgs, ma sul piano pratico le carenze di questi ultimi anni sono fin troppo evidenti e si mostrano anche in quest'ultima proposta regolamentare (l'AI act).

Mentre gli USA hanno beneficiato dell'iniziativa privata che è confluita nei grandi colossi digital (Facebook, Google, Amazon, Twitter) e lo stesso dicasi della Cina in ambito pubblico, l'esperienza dell'UE non può dirsi veramente iniziata e resta spesso solo sulla carta. Questa

realtà purtroppo traspare se ci si concentra sul titolo V recante disposizioni a favore dell'innovazione che conta solo di tre articoli a fronte dei restanti 50 dedicati alle IA da alto rischio che, come si è visto, rappresentano per lo più uno scudo legale per gli utenti europei di piattaforme digitali statunitensi o cinesi.

In ultima analisi ad un quadro giuridico probabilmente molto efficace si accompagna un comparto di scelte poco propenso all'iniziativa e talvolta passivo. Sono molti gli interrogativi che restano, trattandosi anche di una proposta di regolamento non ancora approvata: come si concretizzeranno i numerosi obblighi in una realtà dove l'intelligenza artificiale è onnipresente e totalmente deregolamentata e soprattutto in mano a potenze straniere?

È proprio qui che risalta anche la peculiarità dell'AI act, nonché la misura più forte ed ambiziosa che intende introdurre: l'applicazione extraterritoriale del regolamento, come si rinvia dal testo stesso:

*“This Regulation should also apply to providers and users of AI systems that are established in a third country, to the extent the output produced by those systems is used in the Union.”<sup>133</sup>*

È l'articolo 2 della normativa a codificare questa previsione, stabilendo che il presente regolamento dovrà essere applicato in tre ipotesi alternative: la prima ipotesi prevede che l'IA in oggetto sia immessa nel mercato dell'UE senza tener conto della sede legale del provider che la immette nel mercato, la seconda che sia l'utente dell'IA ad essere situato nell'UE ed infine, nella terza ipotesi, si ammette che il regolamento possa essere applicato anche se provider ed user siano situati al di fuori dei confini dell'UE a patto che l'output dell'IA utilizzata abbia effetto all'interno dell'Unione.

Quest'impostazione è perfettamente in linea con la natura della tecnologia digitale che abbatte ogni confine territoriale. Si può parlare in tal senso di un efficace scelta legislativa nonché molto coraggiosa poiché innovativa nel suo genere, andando a colpire direttamente tutte quelle aziende che spesso, grazie alla delocalizzazione, hanno in passato operato sui mercati europei restando sostanzialmente impunibili ai sensi delle norme dei trattati.

---

<sup>133</sup> Commissione Europea, Artificial Intelligence Act, COM(2021) 206 final, *op. cit.*, p.20, considerando (11).

Si tratta della scelta giusta e necessaria per affrontare correttamente il tema della regolamentazione digitale ma la sua trasposizione normativa può essere fraintendibile. Infatti, a quest'approccio nuovo si accompagna la scelta di utilizzare come parametro di riferimento un non ben specificato criterio territoriale che viene mutuato a seconda delle situazioni. Se nei casi dell'IA operante nel mercato europeo e dell'utente che ne usufruisce i dubbi sono limitati e chiarificabili, si potrebbe ad esempio eccepire che un IA non operi sul mercato UE perché le sue parti hardware sono situate in un'altra regione, la disposizione più ambigua è certamente quella relativa all'output prodotto.

Sono convinto che sia la scelta giusta per affrontare il problema per come si manifesta nella realtà oggi ma, per com'è impostata, questa norma appare destinata a far parlare di sé in futuro. Che cosa bisogna intendere per output prodotto utilizzato all'interno dell'Unione?

Il mondo digitale fino ad oggi è apparso come sostanzialmente distaccato da quello reale, o meglio, come operante su di esso dall'esterno. Nel corso delle questioni teoriche sollevate negli anni sono emersi diversi concetti per descrivere il mondo digital come connettività ed infosfera<sup>134</sup>, tanto cara al Prof. Luciano Floridi. Si tratta di due modi, entrambi appropriati, per riferirsi a quella che sembra la peculiarità del mondo digitale, la sua intrinseca interdipendenza che sembra configurare un tutto quasi indistinguibile che in termini semplici può essere definito internet.

La normativa UE, adattandosi a questa realtà, a ragion veduta propone l'applicazione extraterritoriale dell'AI act mediante l'art. 2, scontrandosi così con un aspetto fondamentale del problema: quali sono i limiti della giurisdizione europea? Se per output prodotto nell'UE si dovesse intendere qualsiasi tipo di effetto rilevante prodotto dall'utilizzo di sistemi IA, allora anche il caso di Cambridge Analytica<sup>135</sup> dovrà sicuramente essere considerato sotto una prospettiva legale, tenuto conto che, secondo diverse indiscrezioni, questa società inglese ha giocato un ruolo determinante sulla Brexit.

---

<sup>134</sup> “l'infosfera è la globalità dello spazio delle informazioni, perciò include sia il cyberspazio (Internet, telefonia digitale, ecc.) sia i mass media classici (biblioteche, archivi, emeroteche, ecc.)”, citazione di L. FLORIDI in G. FONTANA, *In viaggio con il padre dell'infosfera*, Il Sole 24 Ore, 20.05.2021.

<sup>135</sup> Vedi anche, <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump> e [https://it.wikipedia.org/wiki/Cambridge\\_Analytica](https://it.wikipedia.org/wiki/Cambridge_Analytica)

Di segno opposto ma maggiormente efficace può essere un altro scenario: si può ipotizzare l'utilizzo dell'IA per altri scopi, all'esterno del territorio europeo, come ad esempio l'uso di bot per l'acquisto di azioni automatizzato, che produca un danno rilevante e con effetto a catena, una crisi finanziaria come quella del 2008; questo scenario sarebbe idoneo ad azionare il meccanismo dell'art. 2? Si tratterebbe in questo caso di un output certo ed inequivocabile sullo spazio europeo senza che però vi sia nessun tipo di azione diretta a realizzarlo.

Ciò che si vuol mostrare ipotizzando questi casi è che, parlando di connettività dei sistemi IA, spesso ci si ritrova a considerare scenari enormi come quelli della crisi finanziaria del 2008 o della pandemia da covid-19, tipici di un mondo globalizzato in cui concetti come quello di confine territoriale appaiono superati. In questo senso la previsione dell'art. 2 può essere innovativa e rivoluzionaria, ciò che è certo è che potrà portare con sé nuove ed inedite problematiche.

Infine, il profilo attinente alla responsabilità, rispetto al quale la scelta dell'AI act appare corretta per regolamentare il profilo civile. Superate le iniziali discussioni su una possibile personalità elettronica per i sistemi IA, la CE ha deciso di adottare un modello basato sulla responsabilità oggettiva del provider, temperato da diversi obblighi per controbilanciare l'onere probatorio, nonché ipotesi di esonero dalla responsabilità per uso scorretto da parte degli utenti.

L'obbligo chiave di questo modello è quello del *conformity assessment*. Con questo termine ci si vuol riferire a tutte le numerose obbligazioni poste dagli artt. 16-25 a carico del provider affinché costui possa dirsi conforme rispetto alla normativa. Rientrano in questo pacchetto di obbligazioni: l'obbligo di dotarsi autonomamente di un *quality management system* per l'IA immessa nel mercato, l'obbligo di avere un'adeguata documentazione tecnica nonché di avere una dichiarazione di conformità (il *conformity assessment* appunto) rilasciato da un'autorità europea preposta a patto che l'IA rispetti tutti i requisiti previsti per le IA ad alto rischio.

Una volta adempiuti questi oneri il provider potrà immettere nel mercato la *high risk IA* e sarà responsabile per il suo funzionamento. L'AI act tace sulla effettiva responsabilità del provider, vi è un accenno al considerando 53 circa l'opportunità di individuare in questa figura il responsabile, ma poi il dato normativo non dà indicazioni esplicite sulla reale portata di questa previsione. Non vi è in ultimo alcuna previsione espressa che affermi se, adempiendo

a tutte queste obbligazioni, il provider sia esonerato dall'onere probatorio, nonostante sia ragionevole desumerlo dall'impianto della normativa.

Ritengo legittimo supporre che il provider potrà beneficiare in qualche modo del rispetto dei numerosi obblighi posti a suo carico, sarebbe irragionevole il contrario. Ciò che invece il testo dice è quando il provider sarà certamente esonerato dalla responsabilità, per imputarla all'utente che usi scorrettamente il sistema IA. Ai sensi dell'art. 29: "*Users of high-risk AI systems shall use such systems in accordance with the instructions of use accompanying the systems*".

Pertanto, il quadro che ne risulta prevede una generale previsione di responsabilità oggettiva in capo al produttore che talvolta può essere estesa, seguendo lo stesso schema, all'importatore (art. 26) ed al distributore (art. 27) e che in via eccezionale prevede la responsabilità del singolo utente (art.29) ogniqualvolta costui non usi in modo conforme alle istruzioni il sistema IA.

Si tratta di un modello già ampiamente utilizzato dall'UE in altri settori e non vi è motivo di dubitare che possa risultare efficace anche nel settore dell'industria IA. Sotto il profilo civile questo sistema misto di responsabilità, prevalentemente oggettiva e talvolta per uso, non desta particolari problemi; così non è per il profilo penale del diritto dove, a causa della personalità della responsabilità penale e del principio di colpevolezza, sarà probabilmente necessario scendere più nel dettaglio.

#### IV.1.2 Profilo penale: Finzione di continuità

Così come è avvenuto col GDPR, anche con l'Artificial intelligence act dovranno essere introdotte nel nostro ordinamento nuove fattispecie penali. L'art. 5 dell'AI act prevede quattro fattispecie che hanno ad oggetto diversi beni giuridicamente tutelati dalla Costituzione italiana; basti pensare al coinvolgimento della libera manifestazione del pensiero (art. 21 Cost.) della libertà e segretezza della corrispondenza (art. 15) della libertà di culto (art. 19) nelle pratiche IA proibite dalla proposta europea.

Infatti, risulta evidente come nel riferirsi a “*AI system that deploys subliminal techniques beyond a person’s consciousness in order to materially distort a person’s behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm*”<sup>136</sup> il legislatore europeo non faccia riferimento solo al diritto alla riservatezza, cuore del GDPR, ma anche al diritto alla salute ed all’integrità persino fisica dell’individuo a contatto con il sistema IA, ponendo così il problema in un’ottica più grande che complessivamente, in ampiezza, supera di molto il perimetro giuridico tracciato inizialmente dal GDPR.

Integrità fisica e psichica, libera manifestazione e formazione del pensiero, privacy, salute sono tutti beni giuridici da tutelare nella casistica dell’intelligenza artificiale. Ciò che risulta evidente è che questo fenomeno necessiterà di una vigorosa e ben definita disciplina penalistica.

In concreto i rischi specifici del fenomeno provengono certamente dalle grandi multinazionali digital, destinatarie dirette dell’AI act, che sono in grado di erogare un servizio di massa (Facebook, Twitter, Tik Tok, Google), accanto ad esse sono numerosissimi i casi da ipotizzare di uso privato dell’IA per scopi illeciti. L’intelligenza artificiale non è monopolio delle grandi aziende e può essere facilmente sviluppata anche da singoli individui, in quest’ottica è bene pensare ad essa come un mezzo che può essere utilizzato in modo non dissimile da un’arma da fuoco e, al pari di essa, accessibile a chi sia in grado di utilizzarla.

È bene ribadire in via preliminare quest’aspetto per introdurre una premessa sull’indirizzo metodologico. Infatti, l’esigenza di introdurre una regolamentazione penalistica incentrata sulla condotta personale dell’individuo non risponde solo a ragioni di ordine pratico, ma si sposa ed è al contempo richiesta dal dettato costituzionale che all’art. 27 sancisce il principio della personalità della responsabilità penale (“*La responsabilità penale è personale*”).

È in questo contesto che viene in luce la necessaria differenziazione che bisogna fare tra profilo civile e penale della responsabilità individuale. Se, infatti, sul piano civile è ammissibile un quadro disciplinare incentrato su un modello di responsabilità oggettiva come quello presentato dall’AI act, sul piano penale ciò diventa problematico.

---

<sup>136</sup> Commissione Europea, Artificial Intelligence Act, COM(2021) 206 final, *op. cit.*, p. 43, art. 5.1(a).

Il quadro dell'AI act contiene una disciplina incentrata sulla regolamentazione del prodotto AI, una volta immesso nel mercato, ispirata da alcuni principi etici, l'AI act non va oltre e ciò è confermato dall'impianto obbligatorio che genera la responsabilità del provider non come individuo ma come colui che deve garantire la qualità e la sicurezza di un servizio. La regolamentazione penalistica richiede un ordine diverso di considerazioni per espletare la sua funzione, solamente gli aspetti legati alla giustizia retributiva sarebbero di per sé insufficienti a dare soddisfazione alle vittime o riparare il danno.

Ciò che il buonsenso richiede è una piena cognizione delle cause che hanno prodotto il danno, giudizi di tipo etico-morale sulla condotta degli agenti e sul valore di essa per la nostra società. Ciò che serve, ai fini di un'efficace regolamentazione dell'IA, è un impianto penalistico in grado di dare piena risposta ai profondi dilemmi etici che già numerosi scandali hanno generato (si vedano Cambridge Analytica e il blocco degli account di Trump).

È in questa prospettiva e per queste esigenze che, secondo la mia opinione, una forma di responsabilità oggettiva potrebbe risultare inadeguata ai fini di un'appropriata legislazione penale in materia di IA. Si scontrerebbe col principio di personalità della condotta penalmente sanzionabile, nonché col principio di colpevolezza così come affermato dalla Corte Costituzionale con la sentenza n. 364 del 24 marzo 1988.

Sarebbe necessario uno strumento che possa bypassare il vaglio del rapporto di causalità richiesto dall'art. 40 c.p., così come nel caso di responsabilità oggettiva, e che al contempo permetta di punire la condotta, dolosa o colposa ex art. 42 c.p., di chi ha intenzionalmente causato danno ad altri o approfittato dei sistemi IA per commettere un illecito.

Su questa linea di pensiero si pronunciava non molto tempo fa il Prof. Luciano Floridi quando si esprimeva a favore di una forma di responsabilità mutuata dal diritto romano<sup>137</sup>. La fattispecie evocata era la responsabilità del padrone per le azioni commesse dallo schiavo<sup>138</sup>, una fattispecie non dissimile dalla *culpa in vigilando* degli istituti citati nel secondo capitolo: la responsabilità del genitore per il danno commesso dal minore, la responsabilità del padrone

---

<sup>137</sup> L. FLORIDI, *Roman law offers a better guide to robotrights than sci-f*, Artificial Intelligence and Robotics, 22.02.2017.

<sup>138</sup> *Ivi*, "If robots become as good as human agents, we can adapt rules as old as Roman law, in which the owner of enslaved persons is responsible for any damage." p.3.

per il danno commesso dall'animale domestico e la responsabilità del tutore per il danno commesso dall'incapace.

Recuperando il discorso sulla imputabilità condotto nel secondo capitolo, si diceva come, essendo animali, incapaci e minori comunque dotati di coscienza e volontà, andasse rinvenuto il fondamento della imputabilità non già nel libero arbitrio quanto in ciò che Antolisei definisce: concezione comune della responsabilità umana. È per questo motivo che certe categorie di soggetti ne restano esclusi, ma ciò non vale per l'intelligenza artificiale.

Il presupposto della *culpa in vigilando* è, infatti, la libertà dell'individuo agente del cui danno risponderà una figura terza per aver omesso di vigilare con la dovuta diligenza (tutore, genitore o padrone) mentre nel caso dell'IA non si ha alcun tipo di soggettività, bensì un automa che esegue oggettivamente degli ordini.

L'esigenza di individuare un responsabile è comune a queste fattispecie, diversa è la motivazione che porta alla soluzione nei diversi casi.

Cardine di questa linea di pensiero è un concetto semplice: quello di azione che, nella disciplina giuridica, rappresenta il perno degli istituti aventi ad oggetto la responsabilità. Si dice infatti in più disposizioni normative come la responsabilità possa sorgere da una azione od omissione, per la possibilità che aveva il soggetto di intervenire e mutare il corso degli eventi. Si può dire lo stesso dell'intelligenza artificiale? La si può ritenere capace di intervenire liberamente sulla realtà, di propria iniziativa ed assumendosi la responsabilità delle proprie azioni?

Secondo le argomentazioni di questa tesi, la risposta che bisogna dare a questo quesito è negativa: non solo l'IA, come i minori, gli incapaci e gli animali, non può essere ritenuta responsabile delle proprie azioni, ma ciò che essa compie, a differenza delle altre fattispecie, non può nemmeno essere configurato come azione.

Ciò che voglio dire con questo è che, mentre nel caso dell'incapacità vi è pur sempre un soggetto che agisce, nel caso dell'IA, non essendovi soggettività non vi potrà essere nemmeno azione. Le conseguenze di questa affermazione possono essere rilevanti e, se inquadrare nel grafico del capitolo III su causalità ed imputazione, possono ambire a proporre la

considerazione di una tecnica giuridica già nota al diritto per la soluzione di quest'intricata fattispecie: essa è la *finzione giuridica*.

L'introduzione di una *fictio iuris* per disciplinare il caso dell'intelligenza artificiale risponde, quindi, all'esigenza di conciliare da un lato la semplificazione dell'onere probatorio circa l'effettivo nesso causale che ha provocato il danno e dall'altro la necessità, in ambito penalistico, di un'attribuzione chiara e personale della responsabilità penale.

In realtà, il fenomeno dell'IA ben si presta ad una regolamentazione giuridica, in maniera ben più semplice delle fattispecie proprie della *culpa in vigilando*. Se, infatti, non si può attribuire alcuna volontà all'IA, né alcuna forma di soggettività o imputabilità, essa non è che un automa e la sua azione resta circoscritta nell'ambito della sola causalità, esulando da qualunque tipo di nesso di imputazione.

Come ribadito più volte nel corso della tesi, se le idee esposte sono corrette, l'autonomia dell'IA che la fa in apparenza agire, nasconde in profondità un tranello: piuttosto che agire, si può dire che l'IA è agita.

Restano perciò le problematiche relative all'opacità degli algoritmi, che rendono quasi del tutto impossibile la ricostruzione delle operazioni che hanno portato l'IA a compiere il danno. Questa problematica, seppur mitigata dall'obbligo di *record-keeping* imposto dall'art. 12 dell'AI act, appare destinata a permanere se non addirittura ad essere amplificata nel corso del tempo a causa del progressivo aumento della complessità degli algoritmi informatici.

Accanto ad essa si pone soltanto la tesi, sostenuta dalle argomentazioni di Searle, Gödel e Penrose, che l'autonomia dell'IA abbia dei limiti teorici intrinseci ed insuperabili, che paiono destinati a farla permanere nello stato di automa quantomeno nel futuro prossimo.

È così che da un problema emergente, l'imperscrutabilità dei meccanismi interni di un algoritmo, e uno spiraglio di conoscenza, la certezza che l'IA non abbia capacità d'iniziativa, può nascere la soluzione da applicare in ambito giuridico, che prende il nome di *finzione di continuità*.

Mi sono imbattuto nel corso della mia ricerca in questa particolare finzione giuridica mentre cercavo approfondimenti circa la finzione giuridica della soggettività attribuita a società ed enti impersonali. Come più volte ribadito, è quasi ricorrente l'invocazione di una particolare

*fictio iuris* per la regolamentazione dell'IA, finzione che spesso assume le fattezze della finzione di soggettività trattata col nome di personalità elettronica nel corso della presente trattazione.

Se la finzione della personalità elettronica, da molti discussa, non è stata infine capace di risolvere il problema, è probabilmente perché essa opera su un piano soggettivo mentre il problema specifico dell'IA appare essere quello della causalità. Quest'aspetto del problema è stato recepito bene dall'UE che, infatti, ha adoperato un sistema di responsabilità oggettiva in ambito civilistico per bypassare il vaglio della causalità.

La finzione di continuità, così come è stata formulata dall'ex magistrato e primo presidente della Corte Suprema di Cassazione Antonio La Torre, potrebbe essere idonea a superare l'ostica richiesta dell'art. 40 c.p. *“Nessuno può essere punito per un fatto preveduto dalla legge come reato, se l'evento dannoso o pericoloso, da cui dipende l'esistenza del reato, non è conseguenza della sua azione od omissione”*.

Così La Torre si pronuncia su cosa sia la finzione di continuità:

*“Chiamerei «Finzione di Continuità» quella posta a salvaguardia del sistema giuridico che, come la natura, ha horror vacui. Il problema si pone in termini di acuta antitesi, al punto da non essere risolvibile senza l'ausilio di un artificio, quando concorrono: a) da un lato una vicenda giuridica che deve poter procedere senza soluzione di continuità; b) dall'altro l'incidenza di un fattore che ne provoca l'interruzione. Come, allora, conciliare la necessità del “continuo” con la inevitabilità del “discontinuo”?”<sup>139</sup>.*

Subito appare evidente l'aderenza di questa fattispecie teorica al caso concreto. Difatti, laddove La Torre si esprime sull'esigenza della continuità di una vicenda giuridica al punto a) si può chiaramente far risalire la necessità della continuità tra azione del soggetto responsabile e danno, così come richiesta dall'art. 40 c.p. sul nesso di causalità, e per quanto riguarda la vicenda interruttiva del punto b) si scorge con evidenza il cuore del problema

---

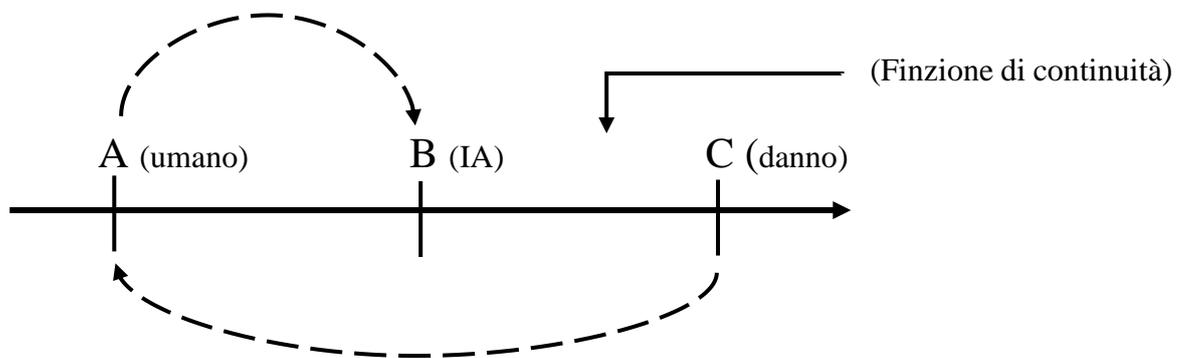
<sup>139</sup> A. LA TORRE, *La finzione nel diritto*, Facoltà di Giurisprudenza dell'Università di Catanzaro p. 315-334, 8.11.1999, p.329.

giuridico che l'IA pone, ovvero l'incidenza dell'azione dell'automa sulla continuità causale dell'azione umana provocandone, di fatto, un'interruzione.

Ed è di questo che si tratta quando si parla di IA forte vs IA debole, così come quando ci si domanda come conciliare continuo e discontinuo.

Sul piano sostanziale la soluzione accettata nel secondo capitolo di questa tesi è quella offerta da un'interpretazione compatibilista dei documenti "*Computing machinery and intelligence*" di Turing e "*Minds, brains and programs*" di Searle che, ricordiamolo, voleva l'IA come un automa, piena estrinsecazione del potere causale della mente umana che è stato definito col nome di intenzionalità. Come calare, allora, questa soluzione dettata dal buonsenso in ambito giuridico?

A questo scopo sono serviti i grafici esposti nel capitolo terzo rappresentati causalità ed imputazione come principi complementari. Servendoci di essi si può scorgere come, aggiungendo il nesso dell'imputazione allo scorrere causale degli eventi, sia la stessa nozione di causalità ad uscirne mutata o quantomeno ridefinita dalle conseguenze del caso dell'IA.



Viene così a configurarsi una situazione a tre elementi: A) la situazione giuridica che deve procedere con soluzione di continuità, ovvero l'azione umana dalla quale nasce la responsabilità del soggetto che la pone in essere, B) il fenomeno che ne provoca l'interruzione, ovvero l'azione dell'automa artificiale, ed infine C) il danno cagionato ad un terzo e causato dall'azione congiunta del soggetto agente e dell'automa IA.

Per spiegare le frecce rappresentate nel grafico e, contestualmente, la funzione di conciliazione tra esigenze opposte dell'ordinamento svolta dalla finzione di continuità è bene riportare la descrizione del meccanismo giuridico operante:

*“Non sembra vi sia altro modo se non di negare la cesura mediante l'espedito della “retroattività”: cioè facendo risalire indietro nel tempo gli effetti di un dato atto, come se esso fosse stato compiuto prima.”<sup>140</sup>*

È questa la soluzione che dà La Torre rispetto al quesito di come conciliare la necessità del continuo con l'inevitabilità del discontinuo.

Nel grafico l'unica freccia vettoriale piena rappresenta il decorso causale degli avvenimenti che, come si nota, è lineare. Non-lineare è invece il decorso del nesso d'imputazione che viene spezzato nel punto mediano (B). Quindi, mentre si può dire che il nesso causale (A-B-C) procede linearmente, ovvero con continuità, il nesso d'imputazione subisce un'interruzione in (B), configurandosi come (A-B). Ed è qui che allora può intervenire la finzione di continuità, imputando retroattivamente gli effetti dell'azione dell'IA (ovvero C) all'azione di (A), sulla base delle considerazioni circa la non imputabilità dell'automa IA, creando così l'ulteriore nesso (C-A).

Facendo un esempio concreto, ciò che avviene è che mediante un'espedito, ovvero la finzione di continuità, si finge che vi sia continuità tra la volontà del soggetto agente (A), l'esecuzione automatica dell'IA (B) ed infine il danno prodotto da essa, o meglio, tramite essa (C). Pertanto, se Tizio da degli ordini in input ad un IA ed essa, eseguendoli, commette un danno, ne risponderà Tizio a titolo di colpa o dolo, a seconda dei casi, come se l'atto d'esecuzione dell'IA fosse un prolungamento della volontà di Tizio, facendo così risalire *retroattivamente* le conseguenze dell'atto eseguito dall'IA all'azione di Tizio come se a commettere il danno fosse stato egli stesso.

Il merito di questa soluzione, secondo la mia opinione, è quello di essere una soluzione di principio e, come tale, suscettibile di una molteplicità di declinazioni, su alcune delle quali è bene soffermarsi.

---

<sup>140</sup> A. LA TORRE, *La finzione nel diritto*, Facoltà di Giurisprudenza dell'Università di Catanzaro p. 315-334, 8.11.1999, p.329

- Vizio *ex uso*: questa fattispecie risulta essere anche la più semplice. In questo caso il vizio causante il danno trae origine dall'uso scorretto che viene fatto dell'intelligenza artificiale, in questo campo la finzione di continuità trova la sua massima applicazione, dimostrando al contempo la sua *ratio*. Se, infatti, un qualunque soggetto si sia avvalso dell'IA per uccidere, rubare o commettere qualsiasi altro delitto, sarà senz'altro pacifico che, una volta individuato il responsabile e l'azione causante il danno, la responsabilità verrà ad esso imputata per aver dato l'istruzione illecita all'IA di cui si è servito per realizzare i suoi scopi.
- Vizio *ab origine*: in tal caso si può ipotizzare che il vizio causante il danno non sia originato dall'uso scorretto del sistema IA, bensì faccia parte del suo codice sorgente, ovvero non vi sia un legame diretto tra il danno e l'azione umana. In tal caso ci si potrà sempre avvalere della finzione di continuità per rintracciare il responsabile, ma il soggetto che ne risponderà sarà diverso da colui che ne rispondeva nella fattispecie precedente.

Se, infatti, nel caso del vizio *ex uso* ne risponde colui che utilizza l'IA per scopi illeciti (sia esso l'utente, il provider o l'operatore informatico) per il vizio *ab origine* la responsabilità dovrà essere imputata al programmatore che abbia costruito un sistema IA.

In questa ipotesi la responsabilità può assumere i connotati del dolo di programmazione se, ad esempio, colui che ha programmato un IA l'abbia fatto per identificare e raggirare soggetti vulnerabili come bambini o incapaci, o colpa grave per aver mancato della dovuta diligenza nella programmazione di sistemi IA adoperati ad esempio in infrastrutture critiche.

- Vizio casuale: infine la fattispecie delegata alla forza maggiore o caso fortuito che, nel caso specifico dell'IA, prende le sembianze dell'errore informatico. Il malfunzionamento dev'essere tenuto in considerazione come possibile e se all'esito del processo di cognizione non risulti esservi responsabilità nell'utilizzatore o nel programmatore del sistema IA, non resta che attribuire la causa del danno al guasto del sistema stesso che dovrà comunque essere certificato da perizia informatica.

Questi sono solo alcuni dei casi ipotizzabili in cui la finzione di continuità troverebbe applicazione come chiave di volta. Tale *fictio iuris* deve il suo merito alla capacità di

inquadrare il fenomeno dell'IA nelle già presenti categorie giuridiche senza effettuare alcuno stravolgimento e svolgendo la funzione di elemento di armonizzazione dell'ordinamento.

È infine ciò che mi è parso il miglior modo per risolvere questa spinosa questione. Avendo preso atto delle conclusioni sostanziali sulla soggettività delle macchine e delle esigenze di certezza dell'ordinamento giuridico, negare la cesura causale mediante l'espedito della retroattività sembra, a discapito della denominazione di finzione, la soluzione che più aderisce alla realtà dei fatti che ci si presenta.

La causalità che ne emerge è più estesa rispetto alla sua concezione ordinaria, così come appare nell'art. 40 c.p., manifestando un concetto di causa che sembra oltrepassare la mera causalità materiale. Se, infatti, la macchina è priva di volontà come si può imputare alla sua esecuzione la definizione di causa? Se essa non è altro che un meccanismo, una causa materiale, la soluzione che appare corretta è quella di imputare la causazione effettiva dell'evento a chi, l'operatore, è intervenuto, ovvero è stato capace di determinare con la propria volontà il corso degli eventi.

La finzione è tale rispetto all'ordinamento giuridico che richiede una causazione materiale tra azione e danno, ma risponde alla realtà dei fatti in cui un'azione umana viene codificata e rielaborata da un sistema IA per produrre un output complesso, del quale l'azione umana è pur sempre la causa originaria e l'azione dell'intelligenza artificiale è un prolungamento, una rielaborazione di essa mediante meccanismi interposti inerti.

Ciò che credo è che il diritto come fenomeno umano trovi applicazione e significato solo rispetto alle vicende umane, in questo contesto la causalità dell'art. 40 c.p. non può che riferirsi ad un sistema concepito su base umana in cui, seppur vi sono elementi estranei come le macchine, le regole e le norme che lo compongono hanno un preciso valore semantico, e cioè quello di regolare le vicende che intervengono negli umani rapporti, descriverli e mitigarne la portata dannosa.

Se, dunque, il diritto è legge umana ed in quanto tale si colloca nella più ampia costellazione delle leggi della natura e dell'universo è giusto che le rispetti e per far ciò, nel caso dell'intelligenza artificiale, bisogna rimarcare quella che infine è la *distinctio* fondamentale: quella tra vivente e non vivente che nel corso di questa tesi ha infine portato a rimarcare la

priorità della tutela del primo a discapito del secondo, facendo emergere così un quadro nuovo ed in parte inedito.

L'intelligenza artificiale nella sua dimensione causale dà come risposta giuridica la finzione di continuità perché, come ben sapevano gli umanisti del Rinascimento, "*homo faber fortunae suae*".

## *Conclusioni*

Nel corso di questa tesi sono stati toccati temi importanti: l'intelligenza artificiale, la fisica quantistica, le fondamenta di un ordinamento giuridico, gli scenari di politica globale ed infine il ruolo dell'individuo in un sistema giuridico e nel mondo.

L'idea rinascimentale dell'*homo faber* sembra ancora viva ed attuale, soprattutto se ci si sofferma sull'interpretazione nuova che emerge di quella legge fondamentale dell'universo chiamata causalità. Si tratta di un processo iniziato nel secolo scorso, tra le vicende belliche e le scoperte della fisica, in cui si inizia a riscoprire il valore della vita. L'uomo non è più un ingranaggio della macchina celeste, egli è un individuo con diritti e aspirazioni, non è un caso che in tempi pandemici si risvegliano temi come i diritti civili, l'identità di genere, l'antirazzismo e le libertà personali.

C'è chi vede nell'intelligenza artificiale solo grigia tecnica, io personalmente in essa vedo un prodotto dell'umanesimo. Parlando di causalità, oltre ad interrogarmi sulle leggi umane e divine che ci governano, tento di soffermarmi sulle capacità umane che hanno prodotto prodigi dell'intelletto e che mettono al servizio dell'uomo le forze naturali.

Naturalmente il covid-19 ci ricorda che spesso così non è, riportandoci alla nostra fragilità di creature mortali. Questo secondo me è un bene, ci ricorda ciò che siamo e ciò che infine ci attende.

Questa ricerca è frutto della curiosità che ho coltivato negli anni della mia università. Mentirei se dicessi che sono arrivato a conclusioni definitive, ciononostante sono ampiamente soddisfatto dei miei progressi e dell'insieme di idee che sono riuscito a formarmi. Se esse siano corrette od errate è qualcosa che potrò scoprire nel resto della mia vita, ciò che all'inizio dei miei vent'anni è importante è che ci credo e che mi piacciono.

Il futuro dell'intelligenza artificiale appare connesso coi futuri intrecci di poteri tra le nazioni civilizzate, giocherà probabilmente un ruolo molto rilevante nella vita di noi individui, più di quanto possiamo oggi immaginare. Una sua regolamentazione etica è il presupposto per scegliere liberamente in che futuro vogliamo vivere.

Nella presente tesi si è parlato molto dell'aspetto tecnico-giuridico dell'IA, ma ciò che è il vero fulcro della questione sono le scelte di politica legislativa che verranno fatte. Come si è visto gli approcci sono diversissimi e spaziano dal controllo totalitario delle tecnologie informatiche della Repubblica Comunista cinese fino alla totale deregolamentazione statunitense. Quale sia la verità e dove si trovi la giustizia sono questioni destinate a trovare risposta nelle vicende future.

Quel che oggi è possibile dire è poco, così come quel che ci è concesso vedere, ma sufficiente per individuare alcuni punti cardinali in grado di guidarci nel futuro che ci attende: vita, libertà, persona, sono questi i valori che la civiltà occidentale è chiamata a tutelare, tutti valori suoi propri e costituenti, messi a repentaglio da queste tecnologie e dagli avvenimenti futuri.

Il lockdown ci ha allontanato dalle persone e ci ha privato di alcune libertà, ma era stato fatto per tutelare la vita. I controlli biometrici tramite IA ci privano di alcune libertà invocando una maggiore tutela dell'utente. I movimenti LGBT invocano maggiori libertà ma categorizzando le persone sulla base dell'orientamento sessuale rischiano di dimenticarsi dell'individuo.

Nella sfida che ci attende, e che prende il nome di Globalizzazione, questi sono alcuni dei temi affrontati su cui nei prossimi anni saremo chiamati a trovare delle risposte. Spero che le troveremo e che, come hanno fatto le generazioni che ci hanno preceduto, riusciremo a far vivere ancora tramite noi i valori dell'Occidente: vita, libertà, persona.

## Riferimenti Bibliografici:

ALAN MATHISON TURING,

*On computable numbers, with an application to the Entscheidungsproblem*, 1936.

*Computing machinery and intelligence*, *Mind* 49: 433-460, 1950.

ANDREA BERTOLINI, *Artificial intelligence and Civil Liability*, Policy Department for Citizens' Rights and Constitutional Affairs, European Parliament, 2020.

ANDREA ERRERA, *Lineamenti di epistemologia giuridica medievale*, G. Giappichelli editore, Torino, 2006.

ANTONIO LA TORRE, *La finzione nel diritto*, Facoltà di Giurisprudenza dell'Università di Catanzaro p. 315-334, 8.11.1999.

BURKEMAN OLIVER, *La libertà è un'illusione*, *The Guardian*, Regno Unito, in *Internazionale* n.1416, p. 44-51, 2.07.2021.

CARL GUSTAV JUNG,

*L'analisi dei sogni, gli archetipi dell'inconscio, la sincronicità*, trad. di Lucia Personeni, Silvano Daniele, Elena Schanzer, Antonio Vitolo, Bollati Boringhieri, 2016.

*Psicologia e Alchimia*, trad. di Roberto Balzen, Lisa Baruffi, Bollati Boringhieri, 1944.

CARLO ROVELLI, *Helgoland*, Adelphi, Milano, 2020.

COMMISSIONE EUROPEA,

*L'intelligenza artificiale per l'Europa*, COM(2018) 237 final, Bruxelles, 25.4.2018.

*Coordinated Plan on Artificial Intelligence*, COM(2018) 795 final, Brussels, 7.12.2018.

*Member States and Commission to work together to boost artificial intelligence "made in Europe"*, Brussels, 7.12.2018.

*Creare fiducia nell'intelligenza artificiale antropocentrica*, COM(2019) 168 final, Bruxelles, 8.4.2019.

*Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*, COM(2020) 64 final, Bruxelles, 19.2.2020.

*White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, Brussels, 19.2.2020.

*Regulation of the European Parliament and of the Council, laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM(2021) 206 final, Brussels 21.4.2021.

DAVID COLE, *The Chinese Room Argument*, The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), 2020,

URL = <https://plato.stanford.edu/archives/win2020/entries/chinese-room/> .

ERWIN SCHRÖDINGER, *Che cos'è la vita?*, trad. di Mario Ageno, Adelphi, Milano, 1944.

EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES- NEW TECHNOLOGIES FORMATION, *Liability for Artificial Intelligence and other emerging digital technologies*, European Union, 2019.

FEDERICA PAOLUCCI, *Il blocco dei social di Trump e la libertà di espressione online*, Ius in Itinere, 09.4.2021, URL = <https://www.iusinitinere.it/il-blocco-dei-social-di-trump-e-la-liberta-di-espressione-online-35567>

FRANCESCO ANTOLISEI, *Manuale di Diritto Penale, parte generale*, Giuffrè, ottava edizione, 1989.

FRITIJOF CAPRA, *Il Tao della fisica*, trad. di Giovanni Salio, Adelphi, Milano, 1982.

GIORGIO MARINUCCI, EMILIO DOLCINI, GIAN LUIGI GATTA, *Manuale di Diritto Penale, parte generale*, Giuffrè, settima edizione, 2018.

HANS KELSEN, *Lineamenti di dottrina pura del diritto*, trad. di Renato Treves, Piccola Biblioteca Einaudi, 1934.

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE,

*A definition of AI: main capabilities and disciplines*, 8.4.2019.

*The assessment list for trustworthy artificial intelligence (ALTAI)*, 8.4.2019

HUBERT LEDERER DREYFUS, *Alchemy and artificial intelligence*, The Rand Corporation, Santa Monica, California 10.1965.

IDA MAGLI. *Dopo L'Occidente, Lo strapotere della finanza, la fine della politica, il tramonto della Chiesa. Come possiamo riprendere in mano il nostro futuro, prima che i banchieri ce lo comprino a prezzi stracciati*. BUR Rizzoli, Milano, 2012.

JAMES LOVELOCK,

*Gaia, nuove idee sull'ecologia*, trad. di Vania Bassan Landucci, Bollati Boringhieri, Torino, 1981.

*Novacene, l'età dell'iperintelligenza*, trad. di Allegra Panini, Bollati Boringhieri, Torino, 2020.

JOHN ROGERS SEARLE,

*Minds, brains and programs*, Behavioral and Brain Sciences, 3: 417–57, 1980.

*The Intentionality of Intention and Action*, Berkeley, University of California, Cognitive science, 4: 47-70, 1980.

*Mente, coscienza, cervello: un problema ontologico*, in E. CARLI (a cura di), *Cervelli che parlano. Il dibattito su mente, coscienza e intelligenza artificiale*, Mondatori, Milano 1997.

*What your computer can't do*, (revisione critica delle opere: *The 4h Revolution, how the infosphere is reshaping human reality* di LUCIANO FLORIDI e *Superintelligence: paths, danger, strategies* di NICK BOSTROM), The New York Review of Books, vol. LXI n. 15, 9.10.2014.

*Freedom and neurobiology, reflections on free will, language and political power*, Columbia University Press, New York, 2007.

LEO STILO, *L'algoritmo giuridico e le sue algo-regole*, La Nuova Procedura Civile, 16.2.2021.

LEON MAX LEDERMAN, CHRISTOPHER T. HILL, *Fisica quantistica per poeti*, trad. di Luigi Civalleri, Bollati Boringhieri, 2016.

LUCIANO FLORIDI, *Roman law offers a better guide to robotrights than sci-f*, Artificial Intelligence and Robotics, 22.02.2017.

MARCO MAVALDI, *La direzione del pensiero, matematica e filosofia per distinguere cause e conseguenze*, Raffaello Cortina Editore, Novara, 2020.

MAX PLANCK, Paolo Benanti (a cura di), *Libero arbitrio*, Castelvevchi, 2018.

NASSIM NICHOLAS TALEB,

*Il Cigno nero, come l'improbabile governa la nostra vita*, trad. di Elisabetta Nifosi, ilSaggiatore, Milano, 2007.

*Antifragile, prosperare nel disordine*, trad. di Daniela Antongiovanni, Marina Beretta, Francesca Cosi, Alessandra Repposi, ilSaggiatore, Milano, 2012.

NICOLA PICARDI, *Manuale del processo civile*, terza ed., Giuffr  Editore, Milano, 2013.

NORBERT WEINER,

*Cybernetics: or Control and Communication in the Animal and the Machine*, The M.I.T Press, Cambridge, Massachusetts, 1948.

*The human use of human beings*, Boston, 1950.

RICHARD WILHEM (a cura di), *I-Ching, Il Libro dei Mutamenti*, trad. di Bruno Veneziani, A.G. Ferrara, Adelphi, Milano, 1991.

ROGER PENROSE, *La mente nuova dell'imperatore, la mente, i computer e le leggi della fisica*, trad. di Libero Sosio, BUR Rizzoli, Milano, 1992.

SELMER BRINGSJORD, NAVEN SUNDAR GOVINDARAJULU, *Artificial Intelligence*, The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), 2020, URL = <https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/> .

STEFANO NASINI, *Il primo teorema di incompletezza di Gödel*, Dept. of Statistics and Operations Research, Universitat Politècnica de Catalunya.

STEFANO QUINTARELLI (a cura di), FRANCESCO COREA, CLAUDIA GIULIA FERRAUTO, FABIO FOSSA, ANDREA LOREGGIA, SALVATORE SAPIENZA, *Intelligenza artificiale, cos'è davvero, come funziona, che effetti avrà*, Bollati Boringhieri, Torino, 2020.

STUART RUSSELL, PETER NORVING, *Artificial intelligence, a modern approach*, Pearson, third edition, 2010.

UNIONE EUROPEA, Regolamento 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati.

YUVAL NOAH HARARI, *Sapiens, Da animali a dei, breve storia dell'umanità*, Bompiani, trad. di G. Bernardi, 2014.

## Siti Web:

<https://www.ali.org/rivista/dettaglio/animali-o-res-codice-civile-o-codice-del-consumo-oppure-altro-ancora/>

<https://www.diritto.it/le-finzioni-giuridiche-nel-diritto-romano/>

<https://ec.europa.eu/digital-single-market/en/artificial-intelligence>

<https://www.europarl.europa.eu/news/it/headlines/society/regolamento-sull-intelligenza-artificiale-cosa-vuole-il-parlamento-europeo>

<https://www.ilpost.it/2020/12/09/gebru-google-intelligenza-artificiale>

<https://www.treccani.it/vocabolario/autonomia/>

## YOUTUBE:

LUCIANO FLORIDI, *Il futuro prossimo dell'intelligenza artificiale*, Teatro Franco Parenti, 2020, <https://www.youtube.com/watch?v=0E1kNRpNW10&t=2635s>.

SIEGFRIED ZIELINSKI, *Artificial extelligence: Transformation and Projection Alchemy as method & Form of Thinking/Creating*, European Graduate School, 22.9.2019, <https://www.youtube.com/watch?v=mVBuvKWqLSE>

## WIKIPEDIA:

[https://it.wikipedia.org/wiki/Ada\\_Lovelace](https://it.wikipedia.org/wiki/Ada_Lovelace)

[https://it.wikipedia.org/wiki/Cambridge\\_Analytica](https://it.wikipedia.org/wiki/Cambridge_Analytica)

[https://it.wikipedia.org/wiki/Causa\\_\(filosofia\)](https://it.wikipedia.org/wiki/Causa_(filosofia))

[https://it.wikipedia.org/wiki/Esperimento\\_della\\_doppia\\_fenditura](https://it.wikipedia.org/wiki/Esperimento_della_doppia_fenditura)

[https://it.wikipedia.org/wiki/Esperimento\\_di\\_Davisson\\_e\\_Germer](https://it.wikipedia.org/wiki/Esperimento_di_Davisson_e_Germer)

[https://it.wikipedia.org/wiki/Great\\_Firewall](https://it.wikipedia.org/wiki/Great_Firewall)

[https://it.wikipedia.org/wiki/Homo\\_faber\\_fortunae\\_suae](https://it.wikipedia.org/wiki/Homo_faber_fortunae_suae)

[https://it.wikipedia.org/wiki/Metodologia\\_agile](https://it.wikipedia.org/wiki/Metodologia_agile)

[https://it.wikipedia.org/wiki/Nesso\\_di\\_causalita](https://it.wikipedia.org/wiki/Nesso_di_causalita)

[https://en.wikipedia.org/wiki/Newcomb\\_paradox](https://en.wikipedia.org/wiki/Newcomb_paradox)

[https://it.wikipedia.org/wiki/Responsabilit%C3%A0\\_oggettiva](https://it.wikipedia.org/wiki/Responsabilit%C3%A0_oggettiva)

[https://it.wikipedia.org/wiki/Sistema\\_di\\_credito\\_sociale](https://it.wikipedia.org/wiki/Sistema_di_credito_sociale)

[https://it.wikipedia.org/wiki/Teoria\\_della\\_complessita](https://it.wikipedia.org/wiki/Teoria_della_complessita)

[https://it.wikipedia.org/wiki/Teoria\\_della\\_complessita\\_computazionale](https://it.wikipedia.org/wiki/Teoria_della_complessita_computazionale)

[https://it.wikipedia.org/wiki/Teoria\\_della\\_imputazione\\_oggettiva\\_dell'evento](https://it.wikipedia.org/wiki/Teoria_della_imputazione_oggettiva_dell'evento)

[https://it.wikipedia.org/wiki/Paradosso\\_di\\_Russell](https://it.wikipedia.org/wiki/Paradosso_di_Russell).