

Intelligenza artificiale: il pensiero giudiziario e l'utilizzo dei sistemi di ia tra teorie euristiche, profili etici e principi nazionali e sovranazionali

Autore: Alessandro Zummo

In: Diritto civile e commerciale

SOMMARIO

1.1 Cos'è l'intelligenza artificiale 1.2 Intelligenza artificiale debole (weak IA) e forte (strong IA) 1.3 Machine learning 1.4 Deep learning 1.5 Le reti neurali 1.6 Le reti neurali biologiche come modello di riferimento 1.7 Limiti 1.8 Intelligenza artificiale e mondo del diritto 1.9 Formazione del pensiero giudiziale e teorie euristiche 1.10 Come decidono i giudici 1.11 Euristica della rappresentatività 1.12 Euristica dell'accessibilità 1.13 Euristica dell'ancoraggio e dell'aggiustamento 1.14 Euristica affettiva 1.15 Intelligenza artificiale ed emozioni 1.16 Progresso ed etica 1.17 Etica dell'intelligenza artificiale 1.18 Il bisogno di costruire un'IA etica 1.19 Questioni etiche sollevate dalle caratteristiche dell'IA 1.20 Il quadro normativo europeo 1.21 Il principio di conoscibilità 1.22 Il principio di non esclusività 1.23 Il principio di non discriminazione

1.1 Cos'è l'intelligenza artificiale

L'intelligenza artificiale è, in modo semplicistico, la capacità di un sistema tecnologico di risolvere problemi o svolgere compiti o attività tipiche dell'essere umano attraverso una capacità di pensiero che si basa su algoritmi di apprendimento progressivo, ciò vuol dire che l'algoritmo in questione è capace di insegnare ad una macchina come fare qualcosa (ad esempio come giocare a scacchi) ma anche come svilupparne il pensiero in base all'esperienza acquisita (quale sarà la mossa migliore per correggere un errore fatto in precedenza).

Sotto questo punto di vista bisogna pensare a quali limiti potrebbe avere l'intelligenza artificiale e a quanto potrebbe essere l'uomo incapace di comprendere totalmente un risultato raggiunto dall'IA a seguito di una crescita consequenziale che potrebbe anche nascere da errori e risoluzione di questi.

Occorre quindi chiedersi quale sia il bilanciamento tra paura dell'errore e conoscenza.

Nel dettaglio, la nascita dell'IA risale al 1943, quando due ricercatori Warren McCulloch e Walter Pitts

proposero al mondo scientifico il primo neurone artificiale, nel 1949 poi Donald Olding Hebb analizzò nel dettaglio per la prima volta i collegamenti tra i neuroni artificiali e i modelli complessi del cervello umano.

Dapprima l'IA ebbe grande utilizzo e diffusione nel mondo della matematica per poi arrivare negli anni '90 allo sviluppo di "chip neuromorfici" in grado di emulare le funzioni sensoriali e cognitive del cervello umano.

Il primo modello di rete neurale risale agli anni '50, il cosiddetto **perceptrone** proposto da Frank Rosenblatt, una rete con uno stato d'ingresso e uno d'uscita costruito su un algoritmo di minimizzazione degli errori.

Questo perceptrone si presentò molto limitato poiché ragionava in base agli input inseriti e in base alla scelta degli algoritmi da utilizzare per dare/modificare le sinapsi (in senso digitale) dell'IA e quindi veicolare l'output (l'informazione in uscita), si evince quindi come in questa neonata accezione l'IA veniva usata come un calcolatore in grado di arrivare a determinati risultati in maniera più veloce rispetto all'essere umano ma pur sempre soggetta all'errore cui si presta un ragionamento veicolato da determinate premesse (input) che potrebbero risultare errate, oppure ad un errato metodo di ragionamento applicato al caso di specie (algoritmo). Si denota in questo senso una forte dipendenza dell'IA all'uomo che qui è più che mai base e nutrimento dell'evoluzione di quest'ultima.

Già dalla seconda metà del XX secolo due matematici, Minsky e Papert intuirono che sarebbe stato opportuno costruire una rete a più livelli di perceptroni per riuscire a risolvere problemi più complessi, idea che successivamente sarà sviluppata in quello che conosceremo come deep learning, ma a quel tempo la tecnologia e i mezzi a disposizione di questa non erano ancora a passo con l'ingegno umano e quindi si dovette aspettare la fine degli anni '70 in cui con lo sviluppo delle GPU si assistette per tutto il decennio successivo a un forte sviluppo dell'IA.[1]

Partendo dall'idea dell'intelligenza artificiale come metodo di ragionamento informatico e quindi accostandolo al cervello umano possiamo distinguere due tipi di intelligenza artificiale, IA forte ed IA debole, in base alla capacità di svolgere alcune funzioni quali:

- **agire umanamente**, quindi in maniere indistinta rispetto ad un essere umano;
- **pensare umanamente** risolvendo problemi usando funzioni cognitive;
- **agire razionalmente**, cioè ottenendo i migliori risultati in base alle informazioni a disposizione.

1.2 Intelligenza artificiale debole (weak IA) e forte (strong IA)

Quando parliamo di intelligenza artificiale debole facciamo riferimento a quell'intelligenza artificiale che è in grado di simulare le capacità di ragionamento dell'uomo senza però riuscire a raggiungere le reali capacità del pensare umano (la riconduciamo ad una mera capacità di problem solving, causa-effetto).

Quando parliamo di intelligenza artificiale forte invece facciamo riferimento a veri e propri sistemi sapienti in grado di sviluppare una propria intelligenza senza emulare il pensiero umano, capaci di arrivare a conclusioni diverse rispetto a quelle che potrebbe raggiungere un uomo e di conseguenza arrivare a soluzioni giuste rispetto ad errori umani o a soluzioni migliori, alcuni scienziati definiscono questo tipo di IA come intelligenza artificiale cosciente di sé.

1.3 Machine learning

Il **machine learning** o **apprendimento automatico** è un insieme di metodi di intelligenza artificiale che permettono alle macchine di apprendere lavorando, correggendo gli errori, migliorando continuamente le proprie prestazioni nel tempo.

Il machine learning costituisce una grande svolta nel campo dell'intelligenza artificiale intesa come metodo di ragionamento informatico poiché tramite degli algoritmi che supportano l'IA sin dalla sua programmazione, questa è in grado di sviluppare e migliorare, imparando dall'esperienza, la propria capacità di agire e decisionale proprio come accade ad un essere umano.

Questi algoritmi sono in grado di imparare dai propri errori in contesti che non erano stati a priori previsti dai programmatori, arrivando a rendere l'IA capace di riuscire a fronteggiare situazioni per le quali non era preparata.

Le modalità di apprendimento della macchina sono più di una e differiscono le une dalle altre essendo comunque tutte efficaci. Sono tutte efficaci perché se da un lato è vero che portano a diversi tipi di apprendimento e l'esperienza ci insegna che il tipo di apprendimento si rivela cruciale a seconda della situazione che affrontiamo, è pur vero che il fine ultimo, l'apprendimento della macchina, viene comunque realizzato.

Queste modalità differiscono non solo per gli algoritmi utilizzati ma soprattutto per lo scopo per cui sono realizzate le macchine stesse.

A seconda della modalità con cui la macchina impara ed accumula dati e informazioni si possono distinguere tre tipi di apprendimento automatico:

- **Apprendimento supervisionato** che consiste nel fornire al sistema informatico della macchina una serie di nozioni specifiche e codificate, ossia di modelli ed esempi che permettono di costruire un vero e proprio database di informazioni e di esperienze.

In questo modo, quando la macchina si trova ad affrontare un problema dovrà semplicemente attingere alle esperienze inserite nel proprio sistema, analizzarle e decidere sulla base di esperienze già codificate.

In questo tipo di apprendimento la macchina deve essere solo in grado di scegliere quale sia la risposta migliore allo stimolo che le viene dato.

- **Apprendimento non supervisionato**, in questo caso alla macchina viene permesso di attingere a determinate informazioni senza avere alcun esempio del loro utilizzo e quindi senza avere conoscenza dei risultati attesi a seconda della scelta effettuata.

Sarà compito della macchina catalogare tutte le informazioni in proprio possesso, organizzarle ed impararne il significato, l'utilizzo e il risultato a cui portano.

Questo tipo di apprendimento offre maggiore libertà di scelta alla macchina che dovrà organizzare le informazioni in maniera intelligente e imparare quali sono i risultati migliori per le differenti situazioni che si presentano.

- **Apprendimento per rinforzo** che rappresenta il sistema più complesso, il quale prevede che la macchina sia dotata di sistemi e strumenti in grado di migliorare il proprio andamento e di comprendere le caratteristiche dell'ambiente circostante.

In questo caso alla macchina vengono forniti una serie di elementi di supporto, ad esempio sensori, che permettono di rilevare quanto avviene nell'ambiente circostante ed effettuare scelte per un migliore adattamento a questo.[2]

Volume consigliato

1.4 Deep learning

Il Deep learning è una sottocategoria del Machine learning ed ha come scopo quello di creare modelli di apprendimento su più livelli.

Immaginiamo di esporre più nozioni in sequenza, quello che fa il nostro cervello è apprendere e sintetizzare tutto quello che in precedenza abbiamo appreso per poi passare a nuovi apprendimenti ed esposizioni e così via, si crea quindi un meccanismo a catena che porta all'arricchimento di tutto quello prima appreso attraverso apprendimenti successivi (su più livelli quindi) ed è proprio questo lo scopo del deep learning (apprendimento approfondito).

Scientificamente possiamo definire l'azione del deep learning come l'apprendimento di dati che non vengono forniti direttamente dall'uomo ma che la macchina è in grado di apprendere ex post grazie all'utilizzo di algoritmi di calcolo statistico.

Lo scopo di questi algoritmi è comprendere il funzionamento del cervello umano e come questo riesca ad interpretare le immagini e il linguaggio, riuscendo ad emularlo. La forma dell'apprendimento è piramidale poiché i concetti più alti sono appresi a partire dai livelli più bassi.

A mio avviso tuttavia non si tratta di una vera e propria emulazione, piuttosto di una creazione in un altro campo di qualcosa che non era stato concepito prima, che non è nato spontaneamente ma che è stato frutto dell'evoluzione scientifica e della comprensione di quanto la tecnologia possa arrivare lontano. Non si sta emulando qualcosa, lo si sta creando dall'inizio. Non si sta emulando, si sta partendo da premesse forse uguali ma non è ancora stabilito né tanto meno anticipabile dove arriverà l'IA.

Il deep learning si basa sostanzialmente sulle reti neurali artificiali, che si ispirano alle reti neurali biologiche che nell'uomo funzionano grazie ai neuroni, cellule in grado di raccogliere ed inviare impulsi nervosi.

In questo ambito il primo passo fatto da McCulloch e Pitts è stato proprio quello di assimilare, nel 1943, il comportamento del neurone umano al calcolo binario (presenza o assenza di segnale).

Successivamente Rosenblatt con il perceptrone introdusse un primo schema di rete neurale artificiale basato su input e output attraverso una regola di apprendimento intermedia. Tuttavia questo sistema si dimostrò inefficace a risolvere problemi complessi.

Il passo decisivo fu compiuto da Rumelhart (psicologo americano che lavorò molto nell'ambito della psicologia matematica) con l'introduzione del terzo strato di reti neurali chiamato hidden (nascosto) ed è proprio in questi livelli che la rete neurale si avvicina, finalmente, al sistema nervoso umano.

Ad oggi, il deep learning ha compiuto passi da gigante, ottenendo risultati che fino a qualche decennio fa erano impensabili.

Tale successo è dovuto alle numerose conquiste in campo informatico relative soprattutto alla sfera dell'hardware, poiché se fino a qualche tempo fa il quantitativo di tempo da impiegare per portare il calcolatore a fare esperienza su un quantitativo sempre maggiore di dati sensibili fosse elevato, oggi, grazie all'introduzione delle GPU, ossia nuove unità che concorrono all'elaborazione dei dati, il tempo richiesto è molto inferiore grazie a procedimenti sempre più veloci.

Inoltre oggi è molto più semplice trovare numerose collezioni di dati (c.d. dataset) fondamentali per allenare l'IA.

Il deep learning è in grado di darci una rappresentazione di dati a livello gerarchico e soprattutto a livelli diversi tra loro, riuscendo ad elaborare e trasformare detti dati. Questa trasformazione ci permette di assistere ad una macchina che riesce a classificare gli input e gli output, evidenziando quelli utili alla risoluzione del problema e scartando quelli inutili.

La rivoluzione raggiunta con il deep learning sta nella capacità di elaborare dati e conoscenze a livelli che non sono affatto lineari tra di loro, capacità simile a quella del ragionamento umano, ed è proprio grazie a questa capacità che l'IA apprende e perfeziona capacità sempre più complesse.

1.5 Le reti neurali

“Tutto ciò che qualcuno può immaginare, altri potranno trasformarlo in realtà.” (Jules Gabriel Verne, Nantes 8 febbraio 1828 - Amiens 24 marzo 1905)

Se desiderassimo ricercare un cuore nell'organismo del deep learning lo identificheremmo certamente nelle reti neurali.

Avendo già considerato IA forte e debole è possibile giungere alla conclusione che una macchina, a prescindere dal livello di apprendimento della stessa raggiunto e potenzialmente raggiungibile, non sarà mai veramente intelligente se non riuscirà a riprodurre un sistema di ragionamento che sia biologicamente ispirato al cervello umano.

La macchina deve essere in grado di offrire un ragionamento simile a quello che l'uomo sviluppa grazie ai propri neuroni.[3]

Il neurone umano diventa così il paradigma computazionale che nutre il deep learning e lo fa tramite le Reti Neurali Artificiali che riproducono il funzionamento del neurone umano, quindi tutti quei processi che

avvengono nel cervello durante la fase di apprendimento e quella successiva di riconoscimento.

È necessaria la logica per comprendere? O quanto meno è sempre necessaria?

Naturalmente no, basti pensare all'apprendimento di un bambino in cui piuttosto che la logica è l'esperienza che guida l'apprendimento.

Nei software neuronali tale processo di apprendimento attraverso l'esperienza avviene tramite il programmatore che immette dei dati conosciuti nella macchina, inserisce quindi l'esperienza nella macchina in modo tale da consentirle di trovare delle risposte anche quando si trova dinanzi a situazioni totalmente nuove.

La rete neurale apprende grazie all'esperienza, ed è questo il punto fondamentale che spiega il concetto di apprendimento approfondito (deep learning) e proprio grazie all'esperienza riesce a costruire architetture gerarchiche fornendo livelli avanzati di output (risposte).

A questo punto è doveroso affermare che l'IA non viene programmata a priori ma piuttosto addestrata all'adattamento tramite i vari tipi di apprendimento a disposizione (supervisionato, non supervisionato e per rinforzo).

I circuiti neurali artificiali sono la base di sofisticate forme di intelligenza artificiale sempre più evolute e in grado di apprendere tramite l'utilizzo di meccanismi simili a quelli dell'intelligenza umana, riuscendo a raggiungere e anche a superare le capacità della mente umana trovando soluzioni per questa impossibili e rendendo l'IA tra i possibili candidati ad essere un protagonista di spessore del mondo ipertecnologico che sta arrivando.

1.6 Le reti neurali biologiche come modello di riferimento

Le reti neurali artificiali prendono spunto dalle reti neurali biologiche e dal funzionamento di queste.

Le reti neurali del cervello umano sono la sede della nostra capacità di comprendere l'ambiente, i mutamenti che in questo avvengono e di conseguenza di sviluppare una capacità di adattamento idonea alle esigenze che si presentano.

Le reti neurali biologiche sono formate da una moltitudine di cellule nervose in connessione tra loro:

- I **somi neuronali**, cioè i corpi dei neuroni, il loro compito è di ricevere e processare informazioni e sono capaci di generare impulsi che si propagano nella rete nervosa qualora il potenziale di azione in ingresso superi certi valori.

- I **neurotrasmettitori**, (ammine, peptidi, amminoacidi) responsabili della modulazione degli impulsi nervosi.
- Gli **assoni o neuriti**, sono la via di comunicazione in uscita da un neurone.
- I **dendriti**, sono la via di comunicazione in ingresso e sono molteplici in ogni neurone.
- Le **sinapsi**, sono i siti nei quali avviene il passaggio di informazioni fra neuroni, ognuno di essi ne possiede a migliaia.

Le sinapsi svolgono funzioni diverse a seconda dell'azione esercitata dai neurotrasmettitori (funzione eccitatoria o inibitoria).

Un singolo neurone è capace di misurare il potenziale elettrico proveniente da diversi segnali inviati da diverse sinapsi generando un proprio impulso nervoso una volta raggiunta la soglia di attivazione.

Quest'ultima attività neuronale è copiata ed implementata anche nelle reti artificiali, come anche la capacità di fornire risposte adattive più calibrate, emulando l'attività biologica creata dal maggior numero degli stimoli forniti dalle maggiori connessioni sinaptiche.

1.7 Limiti

Naturalmente le reti artificiali neurali presentano dei limiti:

1. Il funzionamento delle reti neurali artificiali avviene a scatola chiusa, questo vuol dire che non è analizzabile in modo completo, quindi anche quando queste siano in grado di fornire output corretti (e non è certo che gli output forniti dall'IA siano corretti), noi non siamo in grado di tracciare e studiare il percorso che è stato fatto dall'IA;
2. Anche quando gli output forniti sono corretti non è certo che si sia trovata la soluzione ideale;
3. Il periodo di apprendimento può diventare molto lungo quando, per esempio, l'IA viene allenata tramite l'immissione di input numerosi e complessi o in base all'algoritmo che viene utilizzato;
4. Le reti neurali non sono capaci di risolvere certi tipi di problemi.[4]

1.8 Intelligenza artificiale e mondo del diritto

Dall'accostamento tra intelligenza artificiale e diritto possono nascere forti dubbi e perplessità, ed è normale che sia così considerando che ci troviamo dinanzi a nuove entità le cui decisioni e risultati:

- non sono totalmente l'effetto di azioni umane ma derivano da una serie di processi che a seguito dell'istruzione della macchina hanno autonoma capacità decisionale e che non sono sempre giustificabili a posteriori;
- come già detto non è possibile tracciare e comprendere come il sistema sia arrivato alla propria decisione.

Come lo stesso Massimiliano Nicotra dice: "Il tema dell'intelligenza artificiale non è nuovo al mondo del diritto e dell'etica. Parallelamente al confronto che nasceva nel settore della ricerca scientifica emergevano anche studi giuridici che si ponevano l'obiettivo di esaminare il funzionamento e le implicazioni di quelli che allora erano denominati "sistemi esperti". Dagli anni '90 e fino a quelli più recenti, il tema della regolazione dei sistemi di intelligenza artificiale non è stato più oggetto di analisi significative, fino a quando dal 2014 in poi, vari Paesi hanno cominciato ad occuparsi di questa tematica soprattutto interrogandosi sui presupposti per un utilizzo etico di queste tecnologie, cercando contemporaneamente di individuare un quadro giuridico nell'ambito dei vari sistemi per la loro regolazione".[5]

Lo stesso Nicotra ha sottolineato come sia stata l'Unione Europea ad interrogarsi per prima riguardo l'eticità dell'intelligenza artificiale e quale sia il piano regolatore di questa dal punto di vista dei meccanismi di imputazione della responsabilità in ambito civile.

Già nella Risoluzione del Parlamento del 16 febbraio 2017, concernente norme di diritto civile sulla robotica, il Parlamento europeo aveva posto la sua attenzione su temi relativi ai risvolti etico-sociali sottolineando che l'intelligenza artificiale dovrebbe coadiuvare le capacità umane e mai sostituirle.

Quando ci esprimiamo sulla condotta, magari responsabile, del soggetto agente di norma noi ci chiediamo come avremmo agito in quella determinata circostanza, ci chiediamo dunque quale condotta avremmo posto in essere nel mondo fisico, per riuscire a capire, giustificare o non la condotta del soggetto agente.

L'intelligenza artificiale è in grado di sostituirsi al soggetto agente chiedendosi cosa avrebbe fatto in quel caso?

Al momento purtroppo no, perché il sistema di intelligenza artificiale è programmato e quindi nel

momento in cui arriva a delle decisioni a seguito del proprio apprendimento autonomo, a queste c'è comunque arrivata essendo prima programmata da un essere umano, quindi per quanto autonome possano essere le decisioni della macchina saranno sempre e comunque delle decisioni che partono da una programmazione, sono quasi delle soluzioni guidate.

È facile a questo punto pensare che i sistemi di intelligenza artificiale richiama la responsabilità del produttore che tuttavia deve adeguarsi al fatto che l'intelligenza artificiale acquisisce capacità decisionale autonoma.

Inoltre non solo i produttori ma anche i programmatori hanno partecipato alla costituzione di quell'intelligenza artificiale, quindi a quale soggetto dovremmo ricondurre la responsabilità di un insuccesso dell'intelligenza artificiale?

È la paura dell'errore che se da una parte stimola lo sviluppo di questo tipo di intelligenza, dall'altra la limita nell'utilizzazione e quindi nella progressione di quest'ultima, d'altronde l'evoluzione neuronale dell'intelligenza artificiale nasce proprio dell'errore e dalla correzione di questo.

1.9 Formazione del pensiero giudiziale e teorie euristiche

Passando ora a una trattazione che si distacchi dalla natura prettamente tecnica dei sistemi di IA e si avvicini al mondo del diritto, occorre analizzare una delle figure più importanti di tutto il sistema di giustizia.

Al di là della figura che ognuno di noi, nell'immaginario comune, ha del giudice, dobbiamo sicuramente prendere atto e renderci conto di come, prima di tutto, egli sia una persona e quindi suscettibile a condizionamenti e variabili legate al livello di esperienza, cultura, morale, ad eventi passati che ha vissuto in prima persona, a proprie idee coltivate e sviluppate e magari influenzate con il passare del tempo.

Questi fattori non dovrebbero palesarsi durante il giudizio ma dobbiamo ammettere che quando la distanza tra verità storica e verità processuale diventa significativa il giudice tenderà ad affidarsi in maniera automatica ai fattori poco prima citati che guideranno, a seconda dei casi più o meno, le trame del ragionamento che sfocerà nella decisione finale.

Detto ciò, è chiaro che il giudice si troverà spesso ad affrontare difficili problematiche durante un procedimento giudiziario, difficoltà che potrebbero portare il giudice a prendere una decisione non del tutto corretta.

La scienza stessa dimostra come dietro giudizi apparentemente razionali e ponderati possano nascondersi errori cognitivi di difficile individuazione e proprio ciò, spontanea, potrebbe nascere una riflessione:

potremmo ipotizzare che, considerata la crescente fiducia che in maniera particolare nell'ultimo ventennio abbiamo sempre di più accordato all'IA e sistemi informatici in generale, un giorno una macchina possa sostituire, in un'ipotesi estrema, o aiutare, in un'ipotesi molto più razionale, il giudice?

A proposito del processo decisionale, di particolare rilevanza si dimostra uno studio condotto da due psicologi israeliani, Amos Tversky (Haifa, 16 marzo 1937 - Stanford, 2 giugno 1996) e Daniel Kahneman (Tel Aviv, 5 marzo 1934), che dimostra come ogni individuo nell'adozione di decisioni complesse fa ricorso a precise strategie definite euristiche, strategie che aiutano il decisore ad arrivare ad una conclusione ma che possono anche far cadere quest'ultimo in errori gravi e sistematici.

Teniamo inoltre conto del fatto che gli essere umani sono soggetti ad emozioni, fatica, distrazione a cui le macchine non sono soggette; da ciò, ancora una volta spontanea, potrebbe nascere la considerazione che una macchina non sottoposta a questo tipo di pressioni potrebbe rendere i giudici sempre meno necessari e sostituirli nel loro operato.[6]

D'altro canto possiamo anche evidenziare quanto gli esseri umani siano facilmente influenzabili da notizie riportate su social media, come accade ora con Facebook che è diventato lo strumento di propaganda più utilizzato e dove le notizie che arrivano al lettore/ascoltatore siano a questo offerte in modo da catturarne l'attenzione e di ottenerne l'approvazione, influenza questa a cui un programma di intelligenza artificiale non sarebbe sottoposto.

"I messaggi pubblicati in rete sono ovviamente manipolabili e ottengono effetti prevedibili sugli essere umani, il che significa che, in realtà, spesso ci comportiamo con automatismi simili a quelli di una macchina."[7]

1.10 Come decidono i giudici

Le decisioni dei giudici sono spesso più meccaniche di quanto possa sembrare.

Durante il loro operato, i giudici non si trovano a dover decidere e a dover studiare le cause una alla volta, bensì si trovano ad affrontare più giudizi che nel tempo si vanno accumulando, in una situazione del genere il giudice sarà portato a decidere casi simili in maniera simile, redigendo motivazioni che sicuramente saranno diverse per ogni giudizio ma che prenderanno le basi da casi già risolti quando questi presentano dei tratti in comune con la vicenda da decidere.[8]

Naturalmente questo non deve suscitare pensieri circa il completo distacco del giudice dalle proprie ideologie durante un giudizio, lo stesso Posner dice che spesso i giudici sono guidati da motivazioni di natura ideologica[9], tuttavia più il carico di lavoro in capo al giudice aumenta e più il giudice si distaccherà dal suo pensiero in una determinata materia, a meno che questa non susciti particolarmente

la sua attenzione.

Quanto appena detto ci permette di spostare l'attenzione sul ricorso alle euristiche di pensiero, ampiamente descritte da Tversky e Kahneman[10].

Quando si parla di euristica si fa riferimento ad una scorciatoia di pensiero, ad un modo di pensare che solitamente viene utilizzato quando bisogna prendere una decisione, ricordando delle situazioni simili che si sono presentate in passato ed agendo in modo paragonabile quando il modo di agire in passato si è rivelato quello corretto o non errato.

L'euristica ci permette quindi di prendere quotidianamente delle decisioni che si presentano a noi come corrette, facendoci pensare di fare la cosa giusta.

Attraverso questo meccanismo cognitivo riusciamo a classificare le vicende vissute secondo grandi schemi che possono tornarci utili al momento di decidere[11].

Va tenuto in considerazione come questo meccanismo cognitivo sia essenzialmente basato sui nostri ricordi, i quali potrebbero anche essere distorti o sbiaditi per il passare del tempo, e della nostra percezione della vicenda passata. Sicuramente l'IA in scenari simili potrebbe rivelarsi estremamente utile.

1.11 Euristica della rappresentatività

Secondo l'euristica della rappresentatività[12] "quando una persona prende una decisione di solito ricorda ciò che è successo in situazioni analoghe passate, ovviamente con riferimento a quello specifico contesto"[13].

"Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that an object A belongs to a class B? What is the probability that event A originates from process B? What is the probability that process A will generate an event B? In answering such questions people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, i.e., by the degree of similarity between them. When A and B are very similar, e.g., when the outcome in question is highly representative of the process from which it originates, then its probability is judged to be high. If the outcome is not representative of the generating process, probability is judged to be low."**[14]**

Anche il giudice, quando sorge un contenzioso, fa uso di questa euristica, ad esempio, nel momento in cui decide sulla propria competenza.

In questo processo inizialmente fa uso delle proprie conoscenze e dei casi concreti che esso stesso si è

trovato a trattare.

Tuttavia per il giudice a volte possono sorgere dei problemi, ad esempio una mancanza di conoscenza in materia perché non ha mai affrontato o studiato casi simili.

In tale ipotesi il giudice fa affidamento all'IA nel momento stesso in cui va a consultare le banche dati per cercare e spesso trovare giurisprudenza in materia[15], tuttavia prima non era così, prima questo compito doveva essere svolto dal giudice andando a consultare degli archivi e correndo anche il rischio di non riuscire a trovare qualcosa di utile per il caso.

Possiamo già da ora notare un miglioramento che l'IA ha portato al metodo di ricerca migliorando di fatto le prestazioni del giudice.

In questo sistema vediamo chiaramente come sia la macchina a fare un uso più preciso ed affidabile dell'euristica della rappresentatività e proprio in quest'ottica l'euristica vista come scorciatoia di pensiero diventa invece un calcolo statistico che viene fatto dall'IA autonomamente una volta che a questa vengono dati i giusti input per svolgere tale compito di ricerca, ottenendo prestazioni migliori rispetto a quelle ottenute dal giudice.

Come Jordi Nieva-Fenoll afferma "si può quindi immaginare un' applicazione che legga gli scritti delle parti e identifichi l'argomento della questione da decidere, e che, conseguentemente, cerchi la giurisprudenza applicabile per determinare una o più possibili soluzioni alternative, fornendo anche dati percentuali del rapporto di frequenza di ciascuna di tali alternative. Si tratterebbe di un programma che, pur non imponendo al giudice la sentenza, sgraverebbe il giudice stesso di una parte importante del proprio lavoro"**[16]**.

Inoltre sull'efficienza dell'IA si può anche dire che questa, a differenza del giudice, sia sempre distaccata dal fattore umano, fattore che certamente rende la macchina più fredda e meno legata ai problemi della società in un determinato periodo storico, problemi di cui il giudice tiene conto durante il suo giudizio curandosi, non in maniera dominante ma sicuramente non trascurandolo, del fattore di accettazione della sentenza dal punto di vista dei cittadini, diverso da quello del giurista, rimanendo comunque consci che mai la giurisprudenza deve alienarsi nell'accettazione sociale.

1.12 Euristica dell'accessibilità

L'euristica dell'accessibilità[17] è un meccanismo di pensiero attraverso cui una persona stima la possibilità che un evento si verifichi a seconda della maggiore o minore facilità con cui questa persona tende a ricordare quell'evento, "There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences could be brought to mind" [18].

Un esempio di questo metodo di pensiero, come ci ricorda anche Jordi Nieva-Fenoll, è quello di associare uno sguardo fisso o penetrante ad una persona psicopatica o che soffre di qualche disturbo mentale[19].

L'esperienza piuttosto ci dice tutt'altro poiché spesso accade che stupratori, pedofili o comunque persone che soffrono di disturbi mentali non manifestano nessun segnale a livello espressivo ma anzi sembrano persone normalissime fin quando non si accerta la colpevolezza di questi.

Ciò chiarisce come spesso un metodo del genere possa trarci in inganno, errore che può verificarsi anche durante un processo, ad esempio nell'adozione di misure cautelari dove uno dei requisiti principali è il *periculum in mora*, cioè una situazione di pericolo in cui è centrale la percezione del rischio.

Poniamo l'esempio di un giudice che non abbia adottato una misura cautelare e che proprio questa omissione abbia permesso agli autori del reato la reiterazione dell'illecito.

Ponendo un caso del genere il giudice in un processo successivo e simile a quello che ha già trattato sarà sicuramente più propenso all'applicazione di una misura cautelare per evitare che il reato venga reiterato.

Tutto ciò ci permette di comprendere come il giudice possa perdere naturalezza proprio a causa di eventi passati; un programma di intelligenza artificiale potrebbe essere facilmente impostato per seguire un ragionamento del genere ma ancora di più un programma di intelligenza artificiale potrebbe invece essere programmato per superare situazioni del genere e quindi non essere sottoposto a certe debolezze.[20]

1.13 Euristiche dell'ancoraggio e dell'aggiustamento

Dell'euristiche dell'ancoraggio e dell'aggiustamento ne è stato riconosciuto l'influsso sulle decisioni giudiziarie[21].

"In many situations, people make estimates by starting from an initial value which is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or else it may be the result of a partial computation. Whatever the source of the initial value, adjustments are typically insufficient. That is, different starting points yield different estimates, which are biased towards the initial values. We call this phenomenon anchoring"[22].

Secondo tale euristica le persone tendono spesso a farsi un'idea della soluzione di un caso specifico sin dal momento iniziale in cui questo si presenta [23], creando di fatto un pregiudizio, su quella che sarà la soluzione finale, difficile da superare poiché si sa che quando siamo convinti di qualcosa cerchiamo sempre di vedere tutti i nuovi elementi in modo tale che questi facciano da supporto alla nostra idea, cercando addirittura di limitare il valore distruttivo che questi potrebbero avere contro la nostra teoria.

Questo errore viene denominato “bias di conferma”[24].

È ovvio come questa euristica sia in grado di semplificare il nostro lavoro tutte le volte in cui dobbiamo trovare una soluzione o di come sia fondamentale tutte le volte in cui portiamo avanti un progetto dandoci la convinzione, il supporto giusto per farci pensare che lo completeremo e che sarà un successo.

Allo stesso tempo però questa euristica si dimostra come una lama a doppio taglio poiché ci rende mano inclini a cambiare idea su una determinata questione, un po' com'è per una macchina che sfrutta l'IA che come già detto funziona in base ad algoritmi che sono già stati impostati e la porteranno ad una determinata soluzione.

Sotto quest'ottica l'uomo si dimostra superiore alla macchina grazie alla capacità di cambiare idea, cosa che la macchina non può fare autonomamente, infatti, tutte le volte in cui si riscontrerà un errore di programmazione della macchina, per risolverlo bisognerà riprogrammarla.

Quanto detto nasconde tuttavia un grande pregio dell'intelligenza artificiale poiché l'intenzion non è tanto precisare che il programma di intelligenza artificiale non cambierà mai e relegarlo nella collezione degli insuccessi, ma piuttosto precisare che la macchina avrà bisogno dell'immissione di determinati dati per cambiare il proprio pensiero, cosa che è possibile fare ogni qual volta lo si desidera.

La macchina può quindi funzionare benissimo come ausiliario del giudice in queste circostanze poiché sicuramente si presenta meno incline a certi cedimenti cui un essere umano potrebbe essere vulnerabile, ad esempio pregiudizi o continue reiterazioni di richieste o di risposte durante l'esame di un testimone, tuttavia, alla luce di quanto detto sopra, questo è terreno che può presentarsi pieno di insidie e quindi bisogna fare molta attenzione durante la programmazione e l'uso dell'Intelligenza Artificiale, difatti come afferma Jordi Nieva-Fenoll ne “Intelligenza Artificiale nel Processo” “l'imitazione di questa euristica presenta strane limitazioni che devono essere prese in considerazione durante la programmazione degli algoritmi”[25].

1.14 Euristica affettiva

Euristica, questa, legata alla capacità degli essere umani di essere condizionati da variabili emotive introdotte dalle espressioni, dal linguaggio o dall'apparenza[26].

“Affective reactions to stimuli are often the very first reactions, occurring automatically and subsequently guiding information processing and judgment”[27].

Anche il giudice è soggetto a lasciarsi trasportare, almeno in parte, dalle emozioni o dall'idea di sé che una persona è in grado di suscitare durante un giudizio, perdendo in parte l'imparzialità che proprio deve

essere del giudice .

Naturalmente il lasciarsi trasportare dalle emozioni è una cosa di tutti gli essere umani, quindi non credo questo venga visto dalla società come una perdita di imparzialità da parte del giudice, tuttavia se proprio dovessi classificarla sicuramente sarei portato a dire che questa rientra in una delle componenti che fanno venire meno, in misura da definire caso per caso, l'imparzialità del giudice stesso.

Sicuramente un algoritmo potrebbe essere programmato per seguire questa euristica, d'altronde la capacità di pensiero della macchina è frutto della programmazione e quindi a questa possono essere dati tutti i connotati che si desidera, tuttavia non bisogna dimenticare che al pari di questo la macchina potrebbe anche essere programmata per non seguire questa euristica.

Questa euristica, dice bene Jordi Nieva-Fenoll, è una delle più efficaci armi di retorica, così che di fronte a un discorso percepito dalla maggior parte degli essere umani come convincente, anche se in realtà privo di vere ragioni, la sua approvazione di solito è diffusa, mentre il disaccordo è fortemente contrastato, tanto da essere etichettato come mancanza di umanità, di pietà o di misericordia[28].

La retorica è certamente una delle migliori armi a disposizione di un avvocato. Una macchina non soggetta alle emozioni che la retorica può suscitare, provocherebbe uno stravolgimento totale del meccanismo di condanna sottoposto a quello che il giudice ritiene essere giusto, d'altronde la cosa logica sarebbe aspettarsi dai programmi di intelligenza artificiale decisioni oggettive e quindi non influenzate da giochi retorici.

“La retorica è l'arte di governare le menti degli uomini (Platone)”.

“La retorica un veleno micidiale (Giovanni Giolitti)”.

1.15 Intelligenza artificiale ed emozioni

A questo punto penso sia legittimo chiedersi se mai l'intelligenza artificiale riuscirà ad emulare l'uomo e il suo essere soggetto alle emozioni.

La domanda tuttavia potrebbe essere opposta, poiché sappiamo benissimo che a volte sono proprio le emozioni umane, le emozioni del giudice ad essere d'intralcio al raggiungimento della giustizia, infatti quando un giudice ha un coinvolgimento in causa dettato da amicizia, odio, inimicizia o da legami di parentela, il giudice stesso viene ricusato poiché perde l'imparzialità e la terzietà che dovrebbero contraddistinguere.

La soluzione giusta potrebbe così apparire quella di programmare una macchina che sia totalmente

indifferente alle emozioni umane, così da non lasciarsi influenzare da queste.

Tuttavia progettare una macchina del genere potrebbe essere un errore poiché proprio la mancanza di empatia potrebbe portare a conseguenze peggiori di quelle di una sentenza in parte errata, basti pensare ad una delle funzioni della motivazione del giudice che è il controllo sociale sulle decisioni giudiziarie, oppure basti pensare che tutte le volte in cui un governo dimostra empatia verso i governati, migliorandone lo stile di vita, il numero dei reati decresce.

Il controllo sociale sulle decisioni giudiziarie appare allora come una preoccupazione di grande interesse quando pensiamo ad una macchina che possa svolgere il compito del giudice nell'ottica del senso di giustizia socialmente percepito, mi riferisco al senso di giustizia colto dalla società a seguito di una decisione giudiziale e badiamo bene che questo senso di giustizia è molto diverso rispetto a quella che viene accettata da un giurista a seguito di una decisione giudiziale. Spesso la società si dimostra insoddisfatta delle decisioni prese dai giudici, questo accade per esempio perché i cittadini non conoscono tutti i fatti di causa di cui invece il giudice è a conoscenza durante un giudizio.

Spesso la società desidererebbe pene più severe o che il giudice stesso spiegasse il perché di quella decisione, il perché di quella pena che dai cittadini è stata percepita come lieve o troppo severa.

Pensiamo allora all'accettazione che i cittadini avrebbero di una decisione presa da una macchina, pensiamo al livello di accettazione di un cittadino che non ha mai mostrato interesse per giustizia, intelligenza artificiale, progresso, un cittadino del genere come potrebbe mai accettare una decisione presa da una macchina, specie quando la decisione riguarda proprio lui? A quali conseguenze potrebbe portare un malcontento generale nei confronti di una giustizia troppo perfetta, troppo giusta e quindi troppo soggetta a critiche a causa della mancanza di empatia che a volte può portare a chiudere un occhio quando la questione non è di grande rilevanza?

Questi timori penso che accompagnino molti quando si parla di intelligenza artificiale applicata al mondo del diritto, tuttavia io credo fermamente che queste paure siano superabili prestando maggiore attenzione durante questa fase del progresso e cercando di prevedere sempre con largo anticipo quali saranno le conseguenze di ogni nostro passo avanti, rendendo i cittadini partecipi di questo progresso.

Volendo riferirmi a F. Viola, G. Zaccaria e B. Pastore ne "Le ragioni del diritto" in cui la forza normativa di una regola o un atto giuridico è fortemente legato alla sua forza persuasiva, allo stesso modo più l'intelligenza artificiale sarà in grado di persuadere ed essere accettata dai cittadini e più sarà veloce e controllato lo sviluppo dei sistemi di IA grazie alla partecipazione dei cittadini in questa fase del progresso

Riporto inoltre le parole di Jordi Nieva-Fenoll che anche questa volta mi hanno molto colpito "Grazie all'influenza della letteratura, pensiamo sempre all'intelligenza artificiale come a

qualcosa che potrebbe farci sparire. È interessante notare che potrebbe finire per accadere esattamente

l'opposto"[29].

1.16 Progresso ed etica

L'IA sta già dimostrando il suo potenziale nel business, nei trasporti, nelle tecnologie di ultima generazione e in molti altri settori avendo un forte impatto sullo sviluppo della società.

Possiamo già considerare come un gran numero di posti di lavoro potrebbe negli anni a venire essere destinato alla applicazione di intelligenze artificiali, basti pensare ai veicoli a guida autonoma che potrebbero impattare l'intero mercato dei trasporti [30].

Inoltre, si parla di applicazione dell'intelligenza artificiale anche nell'ambito giuridico e proprio riguardo a questo ambito di applicazione bisogna parlare di programmazione etica dell'intelligenza artificiale e di IA etica.

Cos'è l'etica?

"L'etica può essere definita come i principi morali che governano i comportamenti o le azioni di un individuo o di un gruppo di individui" (Nalini, 2019).

Appare giusto considerare l'etica come un insieme di principi o regole che ci permettono di distinguere, o ancora meglio determinare, ciò che è giusto e ciò che è sbagliato.

Tutti noi, fin da bambini, abbiamo sempre avuto confidenza con l'etica insegnataci dai nostri genitori e dai nostri insegnanti come regola di condotta, lo stesso Aristotele associava la felicità della persona alla conformità con la virtù.

L'etica può essere quindi vista come un principio morale che aiuta le persone a prendere delle decisioni morali e le teorie etiche, con l'aumentare delle interazioni tra esseri umani e animali, macchine o tra le stesse macchine, sono state applicate a situazioni di vita reale come l'etica animale, la bioetica e l'etica delle macchine.[31]

"Ethics is the capacity to think critically about moral values and direct our actions in terms of such values" (Larry R. Churchill).

Nell'ambito dell'intelligenza artificiale l'etica si compone degli obblighi e dei doveri morali non solo dell'IA, ma anche dei suoi creatori poiché l'intelligenza artificiale deve innanzitutto essere programmata, ad esempio attraverso l'utilizzo di quadri etici quali possono essere quelli descritti da Blanchard & Peale, 2011 o dal Markkula Center for Applied Ethics, 2015.

1.17 Etica dell'intelligenza artificiale

L'etica dell'intelligenza artificiale, che riguarda tanto la roboetica quanto l'etica della macchina, fa parte dell'etica della tecnologia avanzata che si concentra sugli agenti artificialmente intelligenti.

Per quanto riguarda la roboetica, questa si occupa non dei robot ma dei comportamenti degli esseri umani mentre progettano, realizzano, usufruiscono ed interagiscono con gli agenti di intelligenza artificiale, considerando anche l'impatto che questi possono avere sull'umanità e sulla società tenendo in considerazione le questioni etiche che sorgono al momento della progettazione, quali ad esempio pregiudizi umani e trasparenza nell'operato della macchina e questioni etiche che potrebbero essere causate dagli agenti di intelligenza artificiale, quale ad esempio l'aumento del tasso di disoccupazione.

Inoltre, tenendo conto che quello dell'intelligenza artificiale è un settore in via di evoluzione ad una velocità sempre maggiore, basti pensare alle scoperte fatte negli ultimi venti anni nel campo della tecnologia, non possiamo escludere che le macchine possano arrivare ad acquisire una coscienza, dovendo a questo punto considerare i diritti della macchina stessa, alla stregua di quanto facciamo con i diritti degli animali.

Se oggi l'idea è quella di progettare un robot da mandare in guerra per evitare che un essere umano muoia, tra qualche decade o alla fine di questo secolo potremmo chiederci se sarà giusto mandare un robot a "morire" in guerra, oppure potremmo trovarci a dover garantire ad un robot i diritti alla libertà, alla libertà di espressione, all'uguaglianza, al libero sviluppo delle proprie emozioni, diritti che appartengono a qualsiasi essere pensante e cosciente.

L'etica della macchina si occupa dei comportamenti morali degli agenti di intelligenza artificiale.

"Machine ethics is concerned with giving machines ethical principles or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making".**[32]**

Se quanto detto in precedenza sulla acquisizione di sempre maggiore intelligenza e coscienza da parte della macchina ci fa intuire comportamenti che l'uomo dovrebbe tenere nei confronti di questa, è anche vero che allo stesso tempo nascono aspettative dell'uomo verso il comportamento della macchina stessa, aspettative di ordine morale ed etico.

Attualmente se volessimo fare un esempio di regole proposte per governare gli agenti di intelligenza artificiale potremmo considerare le tre leggi della robotica di Isaac Asimov.

Prima Legge, un robot non può ferire un essere umano o, attraverso l'inazione, permettere a un essere

umano di subire un danno.

Seconda Legge, un robot deve obbedire agli ordini dati da esseri umani, tranne quando tali ordini sarebbero in conflitto con la prima legge.

Terza Legge, un robot deve proteggere la sua esistenza fino a quando tale protezione non è in conflitto con la prima o la seconda legge.[33]

Nonostante queste regole siano legate più ai racconti di Asimov che a vere e proprie sperimentazioni, potrebbero aiutarci a comprendere quali aspettative dal punto di vista etico possono nascere nei confronti degli agenti di intelligenza artificiale, chiedendoci ad esempio come l'intelligenza artificiale dovrebbe operare eticamente nella società, oppure come dovrebbe interagire eticamente con gli esseri umani e con altre intelligenze artificiali.

1.18 Il bisogno di costruire un'IA etica

Non sappiamo ancora cosa aspettarci dall'intelligenza artificiale, ma è sicuro che questa cambierà il mondo ed entrerà sempre di più a far parte della vita degli esseri umani.

Già oggi dell'intelligenza artificiale se ne fanno molti usi, che potrebbero non essere etici.

Esistono tecnologie vocali basate sull'IA capaci di impersonare voci di esseri umani, ad esempio sarebbe possibile impersonare la voce di un amministratore delegato per porre in essere una frode, oppure le note "deepfake" attraverso cui dei volti umani possono essere sovrapposti a contenuti di vario tipo e per vari scopi, ad esempio potrebbero essere usate per influenzare gli elettori.

Oppure pensiamo alla sempre maggiore applicazione che si fa dell'IA nel sistema giudiziario penale dove un software di valutazione del rischio di recidiva ha mostrato chiari episodi di discriminazione basati su razza, genere ed etnia.

Da questi timori nasce quindi l'esigenza di creare un'intelligenza artificiale che sia capace, attraverso l'applicazione di regole etiche, attraverso l'autonoma comprensione di ciò che è giusto e ciò che è sbagliato, di evitare certi problemi che, dall'utilizzo degli agenti artificiali, possono essere causati.

Naturalmente con l'intelligenza artificiale verranno un guadagno e un progresso commerciale senza precedenti, tuttavia i leader politici dovranno misurarsi con un gran numero di problemi etici quali la privacy dei dati, problemi di disoccupazione, problemi per la sicurezza pubblica e proprio per questo motivo si sta già cercando di individuare dei quadri etici da seguire[34].

Il Future of Life Institute nel 2017 ha proposto un piano etico che si basa sul principio di non ledere, richiedendo lo sviluppo di IA a beneficio della società, promuovendo cooperazione ed evitando corse competitive.[35]

The Public Voice nel 2018 ha proposto delle linee guida che mirano al miglioramento nella progettazione e nell'uso dell'intelligenza artificiale perseguendo la protezione dei diritti umani in primis, promuovendo delle linee guida che, secondo The Public Voice, dovranno essere incorporate nelle norme etiche e adottate nel diritto nazionale e negli accordi internazionali.[36]

1.19 Questioni etiche sollevate dalle caratteristiche dell'IA

La prima delle questioni che sorgono quando trattiamo di etica nell'intelligenza artificiale è sicuramente quella legata alla trasparenza di questa.

Il processo di evoluzione dell'IA avviene attraverso un meccanismo brillante chiamato machine learning (apprendimento automatico), che è alquanto difficile da comprendere poiché si tratta di un processo di elaborazione interno all'intelligenza artificiale, questo tipo di elaborazione viene definito "black box"[37].

Questo processo rende gli algoritmi incomprensibili anche per i programmatori limitando di fatto la capacità delle persone di comprendere la tecnologia e di conseguenza portandole a sviluppare un senso di sfiducia verso gli agenti di intelligenza artificiale.

Inoltre l'incapacità di interpretare l'evoluzione interna dell'IA rende incapaci di monitorare e guidare l'intelligenza artificiale potendo anche portare a conseguenze totalmente indesiderate.

Gli esseri umani sicuramente preferirebbero avere il controllo assoluto sull'intelligenza artificiale in modo che questa faccia esattamente i compiti che gli vengono affidati, tuttavia un controllo assoluto e privo di vincoli, quali potrebbero essere quelli etici, porterebbe l'IA ad agire in maniera estremamente fredda, pensiamo ad esempio ad un veicolo a guida autonoma al quale ordinassimo di portarci al porto nella maniera più veloce possibile senza specificare altro. Un'IA che tenesse conto solo dell'ordine impartitogli senza avere nessuna capacità discrezionale, seppur minima, inizierebbe a violare tutte le regole stradali che si porrebbero in contrasto con il comando da seguire mettendo a rischio anche la vita di altre persone.

Ebbene, per esempio, un'IA etica sarebbe invece in grado di bilanciare l'ordine da seguire con il grado di sicurezza da garantire alla popolazione civile e non.

Altro problema riguarda l'enorme quantità di dati personali e privati cui l'IA è già in possesso, quantità che non può che aumentare con il passare del tempo, basti pensare all'enorme quantità di dati cui Apple Siri e Google Home hanno già accesso.

La possibilità che si possa abusare di dati personali è, quindi, da tenere in considerazione.

Le cartelle cliniche contengono sempre dei dati personali e sensibili che se non tenuti adeguatamente protetti potrebbero creare abusi da parte di malintenzionati dal punto di vista sia personale che economico.

Da qui sorge l'esigenza di una gestione seria e sicura dei dati personali in mano agli agenti di intelligenza artificiale, cercando di indirizzare l'IA stessa su una gestione dei dati che questa riesca a ritenere sicura ed opportuna.

Condivido inoltre il pensiero di J. P. Sullins riguardo i tre criteri che gli agenti di intelligenza artificiale devono soddisfare per essere visti come degli agenti morali: autonomia, intenzionalità e responsabilità.

Sono autonome le IA che non sono sotto il diretto controllo di altri.

Le macchine agiscono in modo intenzionale quando agiscono in modo dannoso o benefico solo a seguito di valutazioni e calcoli che la macchina ha in maniera indipendente effettuato.

Responsabilità poiché le macchine svolgono un ruolo sociale che porta con sé responsabilità assunte.

Questi criteri sono difficili da rispettare e sono difficili da raggiungere, infatti, come lo stesso Sullins dice, ad oggi non siamo nemmeno sicuri che certe caratteristiche le possiedano gli stessi esseri umani.[38]

"Autonomy" is a difficult concept to pin down philosophically. I am not suggesting that robots of any sort will have radical autonomy; in fact I seriously doubt human beings have that quality" (J. P. Sullins).

1.20 Il quadro normativo europeo

Avendo raschiato la superficie della significativa tematica dell'etica accostata all'intelligenza artificiale, tematica affrontata per permetterci di tenere sempre presente una delle questioni più rilevanti che oggi riguardano il mondo della robotica, urge ora affrontare come l'Europa e le nazioni intendano approcciarsi all'intelligenza artificiale.

Ad oggi, volendo parlare del più rilevante corpus normativo in materia, non possiamo fare altro che riferirci alla protezione dei dati personali trattata nel diritto europeo.

L'intelligenza artificiale si basa su un elevato numero di calcoli che da questa vengono svolti, consistente nella trattazione di informazioni che possono essere personali o non. A ben vedere, quindi, la disciplina sul trattamento dei dati personali è stata fra le prime ad essere posta all'attenzione dell'allora Comunità

Europea con la Convenzione n. 108/1981 sul trattamento di dati automatizzati di carattere personale, seguita in Italia dalla legge 21 febbraio 1989 n. 98.

Dopodiché è seguita la Direttiva 46/1995 della Comunità Europea, cui l'Italia ha dato esecuzione con la legge 675/1996.

Tutto questo complesso corpus normativo è stato di recente riformato e sostituito dal nuovo Regolamento in materia di privacy, Reg. 679/2016, entrato in vigore nel maggio 2018, seguito dal decreto legislativo di attuazione 101/2018, entrato in vigore il 19 settembre 2018 e modificativo del D.lgs 196/2003 che aveva introdotto il c.d. Codice della Privacy.

Al Regolamento 679/2016 va poi affiancata la coeva Direttiva 680/2016 relativa al trattamento dei dati personali da parte delle autorità competenti con finalità di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione dei dati.

Questo complesso di norme prende in considerazione il tema delle decisioni adottate sulla base di trattamenti automatizzati dei dati personali e che riguardino i titolari di tali dati.

Il quadro che emerge ci permette di arrivare a tre principi fondamentali in tema di intelligenza artificiale: conoscibilità, non esclusività e non discriminazione.[39]

1.21 Il principio di conoscibilità

A seguito di norme immediatamente applicative del diritto europeo, artt. 13, 2 lett. f) e 14, 2 lett. g) Reg. GDPR anticipato comunque dalla Direttiva 46/1995 Comunità Europea art. 12, esiste nel diritto italiano un principio, formulato in maniera generale e quindi applicabile sia a decisioni prese da soggetti privati che da soggetti pubblici[40], che garantisce ai cittadini di conoscere l'esistenza di processi decisionali svolti in maniera automatizzata che li riguardano e a ricevere informazioni sul processo decisionale stesso. Pretesa, questa, giuridicamente riconosciuta e tutelata.

Parlando di sistema decisionale non si fa altro riferimento se non all'algoritmo utilizzato dalla macchina in fase decisionale, ricevendo informazioni significative sulla logica utilizzata.

Un principio del genere tiene sicuramente in considerazione la logica utilizzata dalla macchina e quindi del nesso di causalità fra informazioni, ragionamento e deduzioni, tuttavia la conoscenza dell'algoritmo non basta da sola per comprendere la logica della macchina e quindi come questa sia arrivata a determinate conclusioni, d'altronde negli ultimi anni abbiamo assistito ad una vera e propria evoluzione dell'intelligenza artificiale considerato che i sistemi di oggi non si limitano a dedurre ma attraverso il meccanismo dell'apprendimento automatico, meccanismo che porta con sé sviluppo e arcani, le macchine

producono esse stesse criteri di inferenza che in molti casi non sono comprensibili neanche ai programmatori dell'intelligenza artificiale stessa.

Occorre a questo punto chiedersi se il principio di conoscibilità, principio riconosciuto a livello sia europeo che nazionale, possa trovare applicazione nel caso di algoritmi che seguono logiche di causa-effetto (determinismo), quali ad esempio quelle seguite da intelligenze artificiali che tengono conto non solo dei criteri di verità/falsità ma anche dei criteri di parziale verità e parziale falsità[41], casi questi in cui non avrebbe senso richiedere informazioni rilevanti sulla logica poiché la logica la capirebbe solo la macchina stessa che l'ha applicata.

Come il professore Andrea Simoncini dice, "Un algoritmo per essere oggetto di un effettivo diritto alla conoscenza ed alla comprensibilità dovrebbe essere, non tanto ragionevole... quanto ragionevole (dalla *rationabilitas canonica*[42]) cioè intellegibile secondo criteri logico-razionali.[43]

L'idea di una decisione sfornita di motivazioni logiche comprensibili si presenta inconciliabile con l'autorevole e, a mio avviso, pienamente condivisibile dottrina giuridica descritta da F. Viola, G. Zaccaria e B. Pastore ne "Le ragioni del diritto", secondo cui la forza normativa di una regola o di un atto giuridico consiste nella sua forza persuasiva.

Se è vero che l'obbligatorietà del diritto nasce dalle ragioni che questo esprime per agire in un determinato modo, sarà anche vero che nel momento in cui la forza persuasiva viene meno a causa della impossibilità di conoscerne le ragioni, viene meno anche la capacità di intendere il diritto come obbligatorio.

1.22 Il principio di non esclusività

Altro principio fondamentale in tema di decisioni prese in maniera automatizzata è definibile come principio di non esclusività [44], nel senso che qualora una decisione automatizzata produca effetti giuridici che riguardino o quanto meno incidano su una persona, questa ha diritto a che tale decisione non sia basata unicamente su tale processo automatizzato[45] e questo vale tanto per il diritto europeo quanto per il diritto interno.

Viene in questo senso data rilevanza ed importanza al contributo umano dato alla decisione in termini di controllo, convalida o smentita della decisione automatizzata.

Questo principio si basa su quella che da H.J. Wilson e P.R. Daugherty viene definita come "collaborative intelligence", che punta alla collaborazione tra l'intelligenza della macchina e l'intelligenza dell'essere umano, il pensiero quindi dei due diversi poli.

In base al vigente diritto europeo e nazionale ai sensi dell'art. 22,2 reg. GDPR una decisione automatizzata può incidere sui diritti della persona, anche senza l'intervento di un controllore umano, quando:

- vi sia il consenso esplicito dell'interessato[46];
- sia necessaria per concludere o eseguire un contratto tra interessato e titolare del trattamento;
- sia autorizzata dal diritto dell'Unione o dallo stato interessato e titolare del trattamento.

Il titolare del trattamento deve naturalmente adottare tutte le misure idonee per tutelare i diritti, le libertà e gli interessi dell'interessato, ottenendo quantomeno un intervento umano che sia in grado di contestare la decisione presa dall'agente di intelligenza artificiale.

Chiarito quale sia il campo di applicazione del principio di non esclusività, bisogna ora valutare quale sia la portata effettiva dell'applicazione di tale principio, poiché sappiamo benissimo che una volta introdotto un sistema automatizzato come ausiliario dell'uomo, questo sistema tenderà con il tempo a veicolare le decisioni, dapprima veicolandone la formazione per andare a sostituirla o catturarla con il passare del tempo.

O ancora, pensiamo a quanto sarebbe difficile dimostrare che la decisione non sia stata presa esclusivamente sulla base di un agente di intelligenza artificiale ausiliario dell'uomo, posto che chi ha preso la decisione può comunque sostenere di averla presa autonomamente e che questa, per quanto possa coincidere con la soluzione proposta dal sistema di IA, sia libera da qualsiasi tipo di influenza capace di snaturarne il carattere personale.

Sul punto, in mancanza di giurisprudenza, dovrebbe considerarsi rispettato il principio di non esclusività tutte le volte in cui il decisore umano è in grado di esprimere una propria motivazione che giustifichi la coincidenza o adesione alla valutazione effettuata del sistema di intelligenza artificiale, chiarendone la logica utilizzata[47].

1.23 Il principio di non discriminazione

Dalla lettura del Regolamento 679/2016 si evince un altro principio che viene chiamato, dal professore Andrea Simoncini, principio di non discriminazione algoritmica.

Cito testualmente: "Al fine di garantire un trattamento corretto e trasparente nel rispetto dell'interessato, tenendo in considerazione le circostanze e il contesto specifici in cui i dati personali sono trattati, è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la

profilazione, metta in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e impedisca, tra l'altro, effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero un trattamento che comporti misure aventi tali effetti” **[48]**.

Ciò che si intende sottolineare è la mancanza di discriminazione nella decisione presa dall'agente di intelligenza artificiale, decisione che qualora presentasse dei tratti discriminatori sarebbe sicuramente contraria all'art. 3 della nostra costituzione.

Il pensiero è lecito perché, come è già stato detto, essendo l'algoritmo frutto di una programmazione, questo potrebbe, a seguito della programmazione stessa, ereditare dei caratteri discriminatori che magari involontariamente sono stati sottovalutati.

Tutto questo è già capitato riguardo il conosciuto sistema COMPAS, a seguito di uno studio condotto da ProPublica che ha rilevato problemi discriminatori nella valutazione del rischio di recidiva[49] dando una valutazione più severa sulle persone afro-americane rispetto ai bianchi.

Non siamo ancora giunti ad un utilizzo regolare dell'intelligenza artificiale all'interno del nostro sistema processuale, tuttavia problemi del genere meritano di essere posti e risolti qualora nascessero. Nel caso di specie sarebbe possibile risolvere il problema discriminatorio, con la complicità del programmatore, attraverso la correzione degli input per evitare effetti discriminatori nella decisione della macchina.

Tuttavia per far sì che i consociati si fidino dell'intelligenza artificiale bisogna certamente evitare che questi errori si presentino del tutto, poiché quello dell'intelligenza artificiale applicata al processo è un terreno ancora inesplorato per il cittadino comune e nonostante qualsiasi persona viva a stretto contatto con le intelligenze artificiali già da ora, spesso non ne conosce il funzionamento e ancora più spesso la fiducia che tutti noi abbiamo già concesso all'intelligenza artificiale è una fiducia appunto concessa, che non ci è stata domandata, semplicemente abbiamo iniziato a farne uso.

Lo stesso non potrà dirsi per il mondo del diritto ed in particolare per il processo, dove l'intelligenza artificiale la fiducia dovrà guadagnarsela.

Volume consigliato

Bibliografia:

AMOS TVERSKY AND DANIEL KAHNEMAN, Judgment under Uncertainty: Heuristics and Biases, The Hebrew University, Jerusalem p. 3

ANDREA SIMONCINI (2019), L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà.

BEN-ARI, FRISH, LAZOVSKI, ELDAN, GREENBAUM (2017), Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment.

CHIARA MINELLI (2015), Rationabilis e codificazione canonica: alla ricerca di un linguaggio condiviso.

GARRIDO MARTIN E., HERRERO ALONSO C. (1997), Influence of the Prosecutor's Plea on the Judge's Sentencing in Sexual Crimes: Hypothesis of the Theory of Anchoring by Tversky and Kahneman.

ISAAC ASIMOV (1941) Liar.

ISAAC ASIMOV (1942) Runaround.

1. P. SULLINS (2006), When is a robot a moral agent, Machine ethics.

JORDI NIEVA-FENOLL "Intelligenza Artificiale e Processo"

KAHNEMAN D., SLOVIC P., TVERSKY A. (1982), Judgment under Uncertainty: Heuristics and Biases, Cambridge.

KENG SIAU, WEIYU WANG (2020), Missouri University of Science and Technology, Rolla, USA, Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI, Journal of Database Management Volume 31, Issue 2, April-June 2020

MASSIMILIANO NICOTRA (2019), "Intelligenza artificiale aspetti etici e giuridici":
<https://www.zerounoweb.it/analytics/cognitive-computing/intelligenza-artificiale-gli-aspetti-etici-e-giuridici/>

MELISSA L. FINUCANE, ALI ALHAKAMI, PAUL SLOVIC, STEPHEN M. JOHNSON (2000), The Affect Heuristic in Judgments of Risks and Benefits.

MICHAEL ANDERSON, SUSAN LEIGH ANDERSON (2011), Machine Ethics.

NICOLETTA BOLDRINI (2020), Intelligenza Artificiale(AI):cos'è, come funziona e applicazioni 2020:
<https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-cose/>

POSNER (2011), Como deciden los jueces.

UMBERTO SANTUCCI (2018), Atlante di strumenti di problem solving, Machine Learning:
<http://www.umbertosantucci.it/atlante/machine-learning/>.

Note

[1] Cfr. NICOLETTA BOLDRINI, Intelligenza Artificiale(AI):cos'è, come funziona e applicazioni 2020:
<https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-cose/>

[2] Cfr. UMBERTO SANTUCCI, Machine Learning:
<http://www.umbertosantucci.it/atlante/machine-learning/>

[3] Cfr. <http://www.intelligenzaartificiale.it/deep-learning/>

[4] Cfr. <http://www.intelligenzaartificiale.it/reti-neurali/>

[5] Cfr. MASSIMILIANO NICOTRA (2019), "Intelligenza artificiale aspetti etici e giuridici":
<https://www.zerounoweb.it/analytics/cognitive-computing/intelligenza-artificiale-gli-aspetti-etici-e-giuridici/>

[6] BEN-ARI, FRISH, LAZOVSKI, ELDAN, GREENBAUM, Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment, pp. 35-36; JORDI NIEVA-FENOLL Intelligenza Artificiale e Processo, p. 31

[7] JORDI NIEVA-FENOLL Intelligenza Artificiale e Processo, p. 31

[8] Cfr. JORDI NIEVA-FENOLL, Intelligenza Artificiale e Processo, p.32

[9] POSNER, Como deciden los jueces

[10] KAHNEMAN D., SLOVIC P., TVERSKY A., Judgment under Uncertainty: Heuristics and Biases, Cambridge, 1982.

[11] Cfr. KAHNEMAN D., SLOVIC P., TVERSKY A., Judgment under Uncertainty: Heuristics and Biases, Cambridge, 1982.

[12] KAHNEMAN, SLOVIC, TVERSKY, Judgment under Uncertainty: Heuristics and Biases, Cambridge 1982, pp. 33 ss.

[13] JORDI NIEVA-FENOLL, Intelligenza Artificiale e Processo, p. 34

- [14] AMOS TVERSKY AND DANIEL KAHNEMAN, Judgment under Uncertainty: Heuristics and Biases, The Hebrew University, Jerusalem p. 3
- [15] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*.
- [16] Cfr. JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*, p. 35
- [17] KAHNEMAN, TVERSKY, Judgment under Uncertainty: Heuristics and Biases, Cambridge 1982.
- [18] AMOS TVERSKY AND DANIEL KAHNEMAN, Judgment under Uncertainty: Heuristics and Biases, The Hebrew University, Jerusalem p.15
- [19] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*, p.36
- [20] Cfr. JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*.
- [21] GARRIDO MARTIN E., HERRERO ALONSO C., Influence of the Prosecutor's Plea on the Judge's Sentencing in Sexual Crimes: Hypothesis of the Theory of Anchoring by Tversky and Kahneman.
- [22] AMOS TVERSKY AND DANIEL KAHNEMAN, Judgment under Uncertainty: Heuristics and Biases, The Hebrew University, Jerusalem, p.20
- [23] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*.
- [24] JORDI NIEVA FENOLL, *Intelligenza Artificiale nel Processo*; TVERSKY, KAHNEMAN, Judgment under Uncertainty and Biases, cit., p.16
- [25] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*, p.40.
- [26] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*
- [27] MELISSA L. FINUCANE, ALI ALHAKAMI, PAUL SLOVIC, STEPHEN M. JOHNSON, The Affect Heuristic in Judgments of Risks and Benefits, J. Behav. Dec. Making, 13: 1-17 (2000), p.2
- [28] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*, cit. p. 41.
- [29] JORDI NIEVA-FENOLL, *Intelligenza Artificiale nel Processo*, cit. p.44.
- [30] KENG SIAU, WEIYU WANG, Missouri University of Science and Technology, Rolla, USA, Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI, Journal of Database Management Volume 31, Issue 2, April-June 2020

[31] KENG SIAU, WEIYU WANG, Missouri University of Science and Technology, Rolla, USA, Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI, p. 75.

[32] MICHAEL ANDERSON, SUSAN LEIGH ANDERSON, Machine Ethics, p. 1.

[33] ISAAC ASIMOV, le tre leggi della robotica, Liar, Runaround.

[34] KENG SIAU, WEIYU WANG, Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI, p 77.

[35] FUTURE LIFE INSTITUTE, 2017

[36] THE PUBLIC VOICE, 2018

[37] KENG SIAU, WEIYU WANG, Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI, p. 79.

[38] J. P. SULLINS, When is a robot a moral agent, Machine ethics.

[39] ANDREA SIMONCINI, L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà, p.77.

[40] Ritengo opportuno, per quanto riguarda le decisioni prese da soggetti pubblici, menzionare l'art. 42 della Carta Europea dei Diritti Fondamentali, affermando che tutte le volte in cui una Pubblica Amministrazione intenda adottare una decisione che può avere effetti non desiderati riguardo la sfera personale o giuridica della persona stessa, intendendosi annesse anche considerazioni di carattere patrimoniale, essa ha l'obbligo di sentire la persona interessata prima di agire, di consentire l'accesso ad archivi e documenti interessanti la stessa e di dare ragione della propria decisione.

[41] Sul punto vedi logica fuzzy.

[42] Sul punto vedi CHIARA MINELLI, Rationabilitas e codificazione canonica: alla ricerca di un linguaggio condiviso.

[43] ANDREA SIMONCINI, L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà, p.79.

[44] ANDREA SIMONCINI, L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà.

[45] Cfr. art. 22 Reg. GDPR

[46] Sul punto si consideri che spesso il consenso dell'interessato non è così libero quanto possa sembrare, basti pensare al consenso fornito in fase di utilizzazione di applicazioni quali: Facebook,

Instagram, Google, Google maps, che sebbene gratuite richiedano il consenso da parte dell'interessato a pena di inutilizzabilità del servizio stesso.

[47] ANDREA SIMONCINI, L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà, p.81. Sul punto chiaro è il richiamo all'art. 42 della Carta Europea dei Diritti Fondamentali che obbliga a chiarire le ragioni della decisione quando questa possa essere lesiva di un diritto.

[48] Cfr. Considerando n. 71 reg. 679/2016

[49] COMPAS è un sistema di IA che calcola il rischio di recidiva di soggetti arrestati in via preventiva o definitiva.

<https://www.diritto.it/intelligenza-artificiale-il-pensiero-giudiziario-e-lutilizzo-dei-sistemi-di-ia-tra-teorie-euristiche-profilo-etici-e-principi-nazionali-e-sovraneazionali/>