

*Dedico questa tesi a me stessa, in primis,
ai miei sacrifici, alla mia tenacia, al mio carattere un tantino irruento
che in fondo mi piace
e che mi ha permesso di arrivare fin qui.
La dedico poi alle tante persone che si sono succedute nella mia vita
e sono sparite quando hanno capito che, per una volta, ero io quella
che aveva bisogno di conforto e sostegno,
sì, la dedico anche a loro, credo se lo meritino perché pur non
volendo
mi hanno dispensata dell'insegnamento più importante
e ammetto lo faccio nella speranza che leggano l'unica cosa che la
gente legge
quando regali loro una copia della tesi, la dedica.*

*E poi la dedico a te, Francesco, che sei più di un amico, meno di un
fidanzato e simile a un angelo,
ci sei stato sempre,
mi hai appoggiata anche quando sceglievo di fare l'opposto dei tuoi
consigli,
hai rispettato il mio essere libera perché hai capito che solo così so
essere,
accompagnandomi nel mio volo, senza paura di cadere.
E' anche grazie a te che ho terminato la mia tesi dottorale e di questo
ti sarò grata per sempre.*

Indice

Introduzione.....6

Capitolo I

Il rapporto tra umano e artificiale.....11

- 1.1 Sviluppo storico e definitorio dell'intelligenza artificiale.....16
- 1.2 Gli agenti artificiali: tassonomie e classificazioni.....29
 - 1.2.1 L'interazione agente/ambiente.....36
 - 1.2.2 L'interazione agente/agente.....39
 - 1.2.3 L'interazione agente/utente.....42
- 1.3 Dai primordi alle odierne applicazioni di AI: la robotica civile.....43
- 1.4 Dalla struttura delle reti neurali al machine learning e deep learning.....54

Capitolo II

Il diritto e l'intelligenza artificiale73

- 2.1 Il diritto visto dall'AI: diritto e logica.....78
 - 2.1.1 I primi sistemi di intelligenza artificiale applicati al diritto.....85
- 2.2 L'intelligenza artificiale vista dal diritto.....96
 - 2.2.1 Il caso della *privacy by design*.....111
 - 2.2.2 La natura giuridica degli agenti artificiale e la responsabilità per le loro azioni.....130
- 2.3 Una questione di coscienza.....168

Capitolo III

L'intelligenza artificiale applicata al diritto: il giudice robot.....175

- 3.1 Giustizia, intelligenza artificiale, Unione Europea e Italia: La Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari

e il Programma Strategico di Intelligenza Artificiale 2022-2024 del Governo italiano.....	180
3.2 Alle origini della motivazione delle decisioni pubbliche.....	213
3.3 La giustizia digitale tra modello induttivo e modello deduttivo.....	233
3.4 I profili ontologici della decisione robotica tra (im)prevedibilità e certezza del diritto.....	253
3.5 La sentenza come sillogismo matematico: un giudice senza empatia	269

Capitolo IV

La giustizia predittiva: problemi tecnici, procedurali ed epistemologici.....	280
--	------------

4.1 La decisione presa da algoritmi: problemi e prospettive.....	281
4.2 Causalità e correlazione.....	292
4.3 Discontinuità tra il fatto e il procedimento e una inevitabile de- ritualizzazione del processo.....	300
4.4 Verso l'ultra-soluzione: problemi di politica del diritto e il rischio di un giudice (ir)responsabile.....	310

Conclusioni.....	319
-------------------------	------------

Bibliografia.....	332
--------------------------	------------

Ringraziamenti

Desidero ricordare e ringraziare tutti coloro che mi hanno aiutata nella stesura della tesi di dottorato, avvenuta in uno dei periodi più complicati della mia vita e, forse, della vita dell'intera umanità nata dopo il secondo conflitto mondiale. Ho iniziato la stesura di questa tesi nel marzo 2020 a Sao Paulo, in Brasile, dove mi trovavo a trascorrere i 4 mesi di soggiorno all'estero previsti dal corso dottorale e all'alba dell'epidemia da Sars-CoV-2 e l'ho terminata quando ancora l'epidemia mieteva poco più di un centinaio di vittime al giorno solo sul suolo nazionale. Il dolore della perdita di persone che conoscevo bene, di una morte percepita ma non vista, d'un saluto troppo distante, di un addio non dato a pezzi della mia vita, della mia infanzia e in fondo a pezzi di me. Voglio ringraziare innanzi tutto loro.

Desidero poi esprimere un profondo ringraziamento a coloro che, con critiche ed osservazioni, mi hanno fatta crescere in questo percorso: a loro va la mia gratitudine.

Ringrazio anzitutto il Prof. Ugo Mattei, il quale è stato per me guida e fonte di preziosi consigli, per la sua professionalità e disponibilità; da lui ho imparato l'arte del dubbio, a pormi domande anche e soprattutto quando le risposte sembrano scontate e questo lo ritengo uno dei doni più grandi che una persona possa ricevere.

Desidero inoltre ringraziare i Prof.ri Roberto Cavallo Perin, Alberto Oddenino Lucilla Gatt, Oreste Pollicino, Marco Orofino, Pier Giuseppe Monateri, Alessandro Morelli, Luca Nivarra e Andrea Pradi, eccellenti costituzionalisti e civilisti per le loro preziose osservazioni.

Grazie all'amico Enrico Labella, professore mancato ma eccellente funzionario all'Autorità Garante della Concorrenza e del Mercato per la pazienza, l'affetto e le cene romane processual civilistiche.

Un grazie speciale al mio amico Francesco, Lino, Augusto, Domenico Peroglio Carus, i cui genitori, a ragion veduta gli diedero tanti nomi perché per me vale come 4 amici e col quale ho condiviso gioie e dolori della mia vita digitale e analogica.

Grazie al dott.ri Stefano Leucci e Fabio Chiusi per gli aperitivi digitali durante il periodo del lockdown: grazie per avermi fatta sentire meno sola.

Desidero ringraziare, poi, i due Marco più noti della politica riformista italiana attuale per aver condiviso con me le loro battaglie ed esperienze nel settore del progresso tecnologico.

Grazie di cuore agli amici dell'emisfero sud, Virginia, Gregory, Jose, Rafael, Ludai, Anna, Renato, Talita; ai Professori della FMU di Sao Paulo, Jose Marcelo Vigliar, Greice Fuller, Roberto Senise Lisboa, Vicente Bagnoli e Andrea Naccache, ai Prof.ri Fernando Rister e Pedro Buck della Mackenzie Presbyterian University e alla giudice federale Isabela Rossi Cortes Ferrari.

Grazie ai miei genitori, dai quali ho sempre ricevuto amore e sostegno in ogni scelta della mia vita.

Da ultimo, ma non ultimo, un ringraziamento particolare desidero rivolgerlo al mio ex compagno di vita, Roberto, per avermi insegnato che l'amore è libertà, passione, dedizione, poesia, e che il mondo dell'accademia, del lavoro così come quello della politica è un mare in burrasca dove le onde vanno surfate con destrezza e determinazione senza perdere di vista l'obiettivo ma vivendo ogni momento perché alla fine "la discesa annoia, è la salita che carica".

INTRODUZIONE

Il tema della presente tesi di dottorato concerne l'ambito dell'intelligenza artificiale applicata al diritto e, più in particolare, lo sviluppo massiccio, soprattutto sull'altra sponda dell'oceano atlantico, degli algoritmi programmati per svolgere ruoli considerati dall'origine dei tempi prettamente umani, come l'attività del giudicare, ossia del definire un controversia, sia essa civile o penale dalla quale discendono doveri ed obblighi da un parte e diritti e facoltà dall'altra.

Naturalmente la difficoltà nell'affrontare un argomento così nuovo e finora poco esplorato quale la giustizia digitale ha fatto sì che l'approccio col quale mi sono avvicinata a tale argomento sia stato di estrema prudenza. Durante la mia ricerca triennale, ho cercato di non utilizzare nessun approccio, conservatore o progressista e tantomeno di demonizzare una realtà che ancora conosciamo solo in parte. Mi sono avvalsa di una visione critica della nuova realtà digitale, ponendo l'accento tanto sugli aspetti positivi quanto su quelli negativi della giustizia robotica.

Durante la mia ricerca mi sono poi prontamente resa conto quanto il tema centrale della tesi dottorale, ossia l'utilizzo di software programmati mediante algoritmi per dare un ausilio al giudice nell'emettere il suo verdetto finale, in ogni grado del processo, necessitasse di una introduzione che ponesse l'intelligenza artificiale applicata alla giustizia in una più ampia visione di come la tecnologia stia radicalmente mutando ogni aspetto della società a partire dalle azioni fino a qui interamente gestite dalle reti neurali di cui è composto il cervello umano.

Pertanto, il primo capitolo della mia tesi affronterà il rapporto tra ciò che definiamo artificiale e ciò che definiamo umano. In particolare mi soffermerò sui tipi di interazioni che i due, mondi umano e artificiale, possono avere, sulle loro caratteristiche e difficoltà tecniche per poi approdare a quella parte dell'intelligenza artificiale, tra le più quotate e finanziate che è la robotica.

Una precisazione di metodo vorrei fare sin dal principio: come il lettore avrà notato dando un veloce sguardo all'indice, faccio esclusivo riferimento alla robotica civile. Non sviluppo, se non con succinti riferimenti il tema della robotica militare, non per questioni di preferenza o semantiche ma semplicemente perché, ai fini della mia ricerca, è la robotica utilizzata a fini civili che a me interessa; un giudice robot infatti potrà più semplicemente, nell'esercizio delle sue funzioni, commettere illeciti civili piuttosto che delitti o contravvenzioni. Il capitolo d'apertura terminerà poi con un'analisi delle tecniche di machine learning utilizzate per rendere i sistemi di AI sempre più autonomi mediante l'autoapprendimento.

Il secondo capitolo poi, avrà ad oggetto il complicato rapporto tra il diritto e l'intelligenza artificiale. Ho pensato di dividere questa seconda parte in due macro-aree: da una parte esamino il diritto visto dai sistemi di intelligenza artificiale, vale a dire come i programmatori di algoritmi necessitano della supervisione del giurista nel loro compito di programmazione algoritmica, di quale delicato compito la società affida loro e quali sono, allo stato attuale i principali usi di tali sistemi nel supportare l'attività del giurista, sia esso ricercatore in ambito giuridico, oppure avvocato, giudice o notaio. Nella seconda parte del capitolo analizzo poi come il diritto si rapporta all'intelligenza artificiale; esporrò sinteticamente il contenuto dei principali documenti normativi

europei e nazionali, per lo più di soft law, che cercano con difficoltà di delineare i confini di una branca del diritto totalmente affidata alla tecnica ed in continua evoluzione che mal si adatta ai canoni classici propri di istituzioni per lo più dei secoli passati mutuati e fatti proprio dal diritto odierno.

Terminerò questa seconda parte con l'approfondimento della normativa italiana in tema di responsabilità giuridica degli agenti artificiali. Le domande che mi porrò saranno, ad esempio, se un giudice robot può essere ritenuto responsabile a seguito di un verdetto che risulta fondato su fatti discriminatori o su dati falsi, chi debba rispondere in tal caso, il suo programmatore, o il giudice che lo utilizza come suo assistente, e ancora, l'intelligenza artificiale è neutra o risente di pregiudizi propri di qualsiasi essere umano? Può essa rispondere di fronte alla legge di atti pregiudizievoli nei confronti di terzi pur essendo priva di coscienza, considerata essa *conditio sine qua non* per dover rispondere di azioni proprie. Infine, esiste davvero un solo tipo di coscienza oppure la stessa può essere declinata in differenti tipologie? In riferimento a tali quesiti di difficile soluzione cercherò non di dare una soluzione ma di esprimere il mio punto di vista circa la possibile chiave di lettura per interpretare siffatti dilemmi ontologici.

Il terzo e centrale capitolo della tesi si occupa dell'AI e del mondo della giustizia e, quindi, a cosa cambia o stia cambiando nel modo in cui i giudici decidono, in casi sia civili che penali. Le domande alle quali dare risposta sono: i giudici utilizzano tecniche di AI? Se sì, quali? E i giudici italiani? In che modo tutto ciò incide sulle decisioni dei giudici? E sulla motivazione dei loro atti? I giudici sono parti delle istituzioni dello Stato: in che modo questi problemi si pongono nella pubblica amministrazione in generale? E negli apparati di polizia? Preferiamo

essere giudicati da un robot o da un essere umano (a proposito del caso Loomis)?

Si farà poi doverosamente riferimento all'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, document approvato nel 2018 dalla CEPEJ dove, per la prima volta, un documento ufficiale europeo accosta le parole "giudice" e "intelligenza artificiali". Tale documento, come è ovvio, non si rivolge solo ai giudici ma anche a tutte le istituzioni incaricate di prendere decisioni nel campo legislativo e regolamentare dell'intelligenza artificiale.

Verrà poi analizzata una delle questioni filosofiche ma dai risvolti più tangibili della prevedibilità della decisione robotica con il suo conseguente fattore di rafforzamento di uno dei capisaldi della giustizia occidentale moderna ossia la certezza del diritto o come la chiamano gli autori del *common law: the rule of law*.

La certezza del diritto è un fattore imprescindibile che deve necessariamente coesistere, direi forse preesistere in uno stato di diritto, è ciò che sta alla base dei rapporti tra gli uomini, le cui azioni sono guidate non solo dalla legge che ci prescrive ciò che possiamo o non possiamo fare, ciò che dobbiamo fare per raggiungere un certo risultato ma è anche l'attività dell'interprete che ci chiarisce, con l'attività appunto di interpretazione del diritto ciò che la legge prescrive o vieta. Il giudice, quale interprete più accreditato, deve non solo decidere entro i confini talvolta stringenti dell'atto normativo ma anche, qualora possibile, sulla base dei fatti concreti per cui si decide, dare un'interpretazione il più coerente possibile con quella data dai giudicanti prima di lui. Solo così, il cittadino sarà veramente libero di

orientare il proprio comportamento, consapevole delle conseguenze giuridiche in cui potrà imbattersi.

Il capitolo si conclude poi con una analisi orientata a dar luce ai costi e benefici di una giustizia resa da un giudice algoritmo. Basti dire in questa sede che il mito del giudice robot come espressione massima di neutralità, così come scritto anche nel Libro Bianco dell'Agenzia per l'Italia Digitale, in realtà è, per l'appunto, un mito, in quanto, come ormai noto, i programmi di giustizia predittiva si basano su dati presi da sentenze passate, elaborate da umani e che in tal modo finiscono per incorporare e a volte anche ampliare pregiudizi tipici delle persone fatte di carne e ossa.

Si farà poi riferimento ad altre problematiche più tecniche quali i *bias* di progettazione ma anche alla difficoltà tecnica di incorporare in una macchina concetti aleatori quali la ragionevolezza e l'equità, concetti che nel diritto non sono solo lodevoli assunti linguistici, ma spesso si ergono a metro di decisione nella redazione delle sentenze e quindi nelle decisioni di merito finali.

Nel capitolo quarto che è il capitolo conclusivo della tesi, infine, analizzeremo in modo più dettagliato quanto esposto nel terzo sotto forma di politica del diritto, vale a dire, rispettivamente nella prima parte tenterò di mettere in evidenza le problematiche tecniche di una giustizia robotica, vale a dire il problema dei bias informativi, del mito della neutralità dell'algoritmo e via discorrendo, per concludere poi l'elaborato con questioni che attengono alla politica del diritto ed alla sociologia: ci domanderemo se veramente un giudice umano è sempre necessariamente migliore di un giudice privo di coscienza (da intendersi essa più come capacità di empatia che come etica delle relazioni umane) e come questo può, ed anzi deve, nel caso, mutare

radicalmente la tipica sovranità statale del diritto, che vedrà, nel caso di uno sviluppo massivo dell'uso dell'AI nella giustizia, perdere la sua tipica centralità a favore di una giustizia che rischia di essere privatizzata.

Capitolo 1

Il rapporto tra umano e artificiale

L'intelligenza è la cosa che più di ogni altra distingue l'uomo tra tutti gli esseri viventi. L'intelligenza è lo strumento che consente di elaborare ciò che gli organi di senso recepiscono dall'ambiente circostante trasformandolo in qualcosa di estremamente più complesso. A moltissimi essere viventi sarà capitato di osservare un tronco o un sasso rotolare da un pendio, ma solo l'uomo ha saputo elaborare questa immagine ed utilizzarla ai fini della propria esistenza. Così l'uomo scopre l'utilità della ruota che gli facilita il trasporto delle cose, scopre come costruire utensili, armi adatte per uccidere prede grandi e pericolose, capisce quanto sia utile il fuoco o come generarlo.

Ha poi creato il cannocchiale per scorgere pericoli in lontananza, ha inventato le macchine che hanno segnato l'era della prima rivoluzione industriale e che hanno in parte aiutato l'uomo nei lavori faticosi e fisicamente usuranti e gli hanno reso possibile gli spostamenti anche per lunghe percorrenze, penso ai treni innanzi tutto, alle macchine poi, alle navi e infine agli aerei.

Tuttavia tali macchine erano create per dare ausilio agli essere umani i quali ne avevano il totale controllo e non era certamente richiesta loro alcuna minima variazione del lavoro che svolgevano e per le quali erano

state create. Col passare del tempo l'uomo si rese conto che tale aiuto era divenuto ormai imprescindibile per la sua vita e che anzi ne voleva di più, di più qualificato e di più vario.

Il problema pertanto divenne non solo lo sviluppo di macchine più potenti ma quello della loro capacità a svolgere mansioni fino ad allora interamente devolute alla mente umana.

Tali macchine dovevano quindi, per soddisfare pienamente le richieste del proprio datore di lavoro umano, essere "intelligenti" ossia riprodurre anche se solo una rudimentale architettura cognitiva della mente umana. L'idea era quella di agente intelligente, anche se non necessariamente "creature artificiali che in contesti appropriati appaiono come persone".

Qui naturalmente vengono in gioco le due definizioni di AI forte e AI debole di cui si parlerà più ampiamente nel prosieguo. Basti dire fin da qui che prima e piuttosto che imbarcarsi nella diatriba "i computer sono più o meno intelligenti degli umani" pare interessante porsi il problema: "a che punto siamo nella comprensione del funzionamento del cervello umano?" Secondo Joshua Greene, uno dei massimi esperti nel campo delle neuroscienze applicate all'intelligenza artificiale e professore ad Harvard, gli umani hanno una capacità il cui funzionamento non è ancora noto, e cioè la capacità di considerare tutti i tipi di idee, pur senza credere nel fatto che esse siano vere, infatti, ci spiega il summenzionato ricercatore "una funzione fondamentale nella pianificazione delle azioni e nell'immaginare possibilità che ancora non esistono nella realtà". Cercare quindi di capire come gli uomini facciamo ciò è un passaggio essenziale per tentare di costruire un'intelligenza artificiale generale anche definita *human-level AI* oppure intelligenza artificiale forte.

Sostiene Greene che quello che si sta cercando di fare è capire come i nostri cervelli siano capaci di una cognizione di alto livello. In termini tecnici si può parlare di una semantica compositiva o di una semantica compositiva multimodale. Questo significa, con parole più semplici, cercare di capire come il cervello prende i concetti e li mette insieme in modo da comporre un pensiero, così che si possa leggere una frase come “il cane ha inseguito il gatto”, e si capisca che essa significa qualcosa di diverso da “il gatto ha inseguito il cane”. Gli stessi concetti sono coinvolti, “cane”, “gatto”, “inseguire”, ma il nostro cervello è capace di mettere le cose insieme in modi diversi al fine di produrre un diverso significato.

Inoltre molto successo l'apprendimento automatico (*machine learning*) lo ha avuto sul versante della percezione e del riconoscimento di oggetti.

Greene prosegue poi affermando:

Non si può realmente comprendere qualcosa se non si prendono certe parole e le si collega significativamente alle cose che si possono vedere o toccare o sperimentare in modo sensoriale. Così pensare è qualcosa tra le immagini e tra le parole. [...] quello che il mio laboratorio sta cercando di fare è capire come questo aspetto centrale, ma veramente poco compreso, dell'intelligenza umana funziona. Come noi combiniamo i concetti in modo da formare pensieri. Come può lo stesso pensiero essere rappresentato in termini di parole contro cose che si possono vedere o sentire con gli occhi e con le parole contro cose che si possono vedere o sentire con gli occhi e con le orecchie della tua mente? [...] Pensi che non abbiamo ancora capito il caso umano, e la mia ipotesi è che ovviamente è qualcosa che per intero sono i neuroni a fare, ma queste capacità non sono ancora ben

catturate dagli attuali modelli di reti neurali. [...] Se si vuole veramente costruire un'intelligenza generale artificiale, si deve iniziare con le conoscenze consolidate sulla cognizione, e non solo cercando di costruire qualcosa che possa, ad esempio, leggere frasi e dedurre cosa da quelle frasi.

Il riconoscimento delle immagini è un settore, così come il riconoscimento degli oggetti di cui si è detto sopra, nel quale vi sono stati i maggiori progressi, grazie a innovazioni algoritmiche e investimenti in infrastrutture, ad esempio, nell'hardware utilizzato per addestrare il sistema o nel software utilizzati per connettere l'hardware. La velocità di addestramento di un algoritmo di ImageNet¹, a novembre 2018, è stata 16 volte maggiore rispetto a quella di giugno 2017. Il tasso di errore dell'etichettatura automatica di ImageNet è diminuito dal 28% nel 2010 a meno del 3% nel 2016, mentre le prestazioni umane, che sono il punto di riferimento, sono circa al 5% di errore.

Questi dati, raccolti dall'*AI Index 2018 Report* dell'Università di Stanford, ci dicono di un settore in grande espansione, con risultati che superano quelli umani e che ha grandi implicazioni a livello sociale e giuridico, come possibilità di controllo delle pubbliche autorità (e non solo) su tutti i cittadini.

Vi sono, allo stesso tempo, questioni di grande importanza tuttora aperte: per esempio, se si vuole che l'intelligenza artificiale aumenti l'intelligenza umana anziché sostituirla, è importante passare da ambienti nei quali le macchine sono opposte agli umani a ambienti dove

¹ ImageNet è un'ampio database di immagini realizzato per l'utilizzo, nell'ambito del *machine learning* per il riconoscimento di oggetti. In sostanza fornisce all'algoritmo un'enorme quantità di immagini e poi gli chiede di riconoscerle sulla base, appunto, dell'apprendimento automatico delle macchine.

umano e macchine lavorano insieme. Inoltre, i sistemi di domande/risposte basati sull'intelligenza artificiale non hanno la capacità di sostenere un dialogo. La profonda comprensione del linguaggio naturale è ancora una sfida, così come le capacità di ragionamento del buon senso. Ancora, l'apprendimento automatico ha molto successo in molte applicazioni, poiché può scoprire correlazioni nascoste all'interno di enormi quantità di dati. Ciò consente di fare previsioni molto accurate, ma non dà informazioni circa i rapporti di causalità.

Questo significa che, in settori come il diritto, di cui si dirà meglio nel secondo e terzo capitolo, la strada da compiere è ancora lunga per arrivare a usi diffusi da parte di cittadini e professionisti.

Vi sono poi i timori di vari personaggi, anche autorevoli, secondi i quali le macchine supereranno gli umani. L'idea che l'intelligenza artificiale possa significare la fine della razza umana e che un robot possano rimpiazzare l'umanità, perché capaci di un'evoluzione più celere di quella biologica è stata espressa anche scienziati del calibro di Stephen Hawking in più occasioni, tra le quali una molto nota intervista alla BBC. La preoccupazione è quella dell'avvento di macchine dotate di un 'intelligenza superiore a quella umana, che possono migliorare autonomamente la loro architettura e realizzare quella che Vernor Vinge già nel 1993 aveva chiamato *singularity* e che nel film *Transcendence* con Johnny Depp e Morgan Freeman prende il nome di, appunto, *transcendence*.

Il commento di un altro noto studioso matematica britannico su queste posizioni è *tranchant*: non sono le macchine di per sé ad essere pericolose, più pericoloso della macchine è il pensiero che esse possano essere migliori di noi.

Intanto pare utile segnalare tre ragioni che fanno ritenere infondato questo timore, almeno allo stato delle cose e per un lungo periodo futuro.

In primo luogo, lo sviluppo dell'apprendimento automatico richiede la disponibilità di una grandissima quantità di dati, che siano selezionati con l'aiuto di un umano che confermi la corretta identificazione di un oggetto. E anche quando, per esempio, un sistema venisse alimentato con libri di progettazione e ingegneria delle auto, potrebbe evidenziare e riassumere le informazioni chiave, ma non potrebbe mai progettare una nuova auto.

In secondo luogo, ad oggi la potenza di calcolo delle macchine non può avvicinarsi a quella di un umano. La situazione potrebbe cambiare con i computer quantistici, che non è dato prevedere quando saranno disponibili per l'uso corrente. Intanto, è certo che, sulla base dei principi di ingegneria, "l'attuale tecnologia dei circuiti integrati che alimenta le capacità computazionali dei computer convenzionali non sarà mai in grado di replicare il cervello umano, tanto meno superarlo nelle capacità intellettuali. Infine, non possiamo progettare macchine che superino ciò che non comprendiamo ancora del tutto, cioè l'essere umano; la scienza moderna infatti, non comprende ancora fino in fondo come funzioni il cervello umano.

1.1 Sviluppo storico e definitivo dell'intelligenza artificiale

"L'intelligenza artificiale è un divorzio, non un matrimonio; il divorzio tra la capacità di agire per un fine, il porre in essere un'azione per risolvere un problema e la necessità, nel fare ciò, di essere intelligenti.

Queste due cose, quello che in inglese si direbbe la “*agency*” cioè la “gentità” cioè la capacità di agire in modo intelligente noi oggi, sempre più spesso la paragoniamo alla mera capacità di agire nel senso di svolgere un compito anche senza usare per nulla l’intelligenza.

Si crea così un baratro tra la capacità di agire e quella di essere intelligente nell’agire”.

Quanto detto, sono le parole pronunciate dal noto professore di etica e filosofia dell’informazione dell’Università di Oxford, Luciano Floridi durante un convegno svoltosi al teatro Parenti di Milano qualche anno fa.

Il termine intelligenza artificiale, da un punto di vista linguistico è propria del vincitore del Premio Turing nel 1971, l’informatico statunitense John McCarthy che pronunciò il termine nel lontano 1955 in occasione della creazione di un gruppo di lavoro che avrebbe dovuto incontrarsi al Dartmouth College nell’estate del ’56.

Tuttavia il padre per così dire biologico, usando un ossimoro è stato il matematico e filosofo britannico Alan Mathison Turing. Il suo lavoro ebbe una vasta influenza sulla nascita della disciplina dell’informatica, grazie alla formalizzazione dei concetti di algoritmo e calcolo, mediante la sua nota macchina che a sua volta costituì un significativo passo in avanti nell’evoluzione verso il moderno computer.

Era l’anno 1950, quando ai primi di ottobre la rivista *Mind* pubblica un articolo che passerà alla storia: *Computer Machinery and Intelligence*².

² A. Turing, in “*Mind*”, 1950, vol.59, n. 236, pp. 433 ss. Sulla complessa e controversa figura di Alan Turing, scienziato che lavora per lo spionaggio inglese nella Seconda Guerra Mondiale, sul famoso programma Enigma, venendo poi arrestato a causa della sua omosessualità e poi suicidatosi, si veda la presentazione della Stanford Encyclopedia of Philosophy.

L'intelligenza artificiale (d'ora in poi AI) s'intreccia con lo sviluppo odierno delle scienze, soprattutto cognitive, per cui negli ultimi cinquant'anni lo studio della mente umana è andata approfondendosi in rapporto alla rapida evoluzione tecnologica, che ha reso necessario porsi il problema del raffronto tra il cervello umano ed il computer.

Il termine intelligenza artificiale chiama innanzitutto in causa il concetto di intelligenza nelle due classiche accezioni di: (1) *intus legere*, cioè comprendere in profondità, nel rimando alla capacità della mente di concepire e rappresentare la realtà circostante; e (2) *inter legere*, che rappresenta l'attitudine della nostra mente di scegliere in modo razionale tra ipotesi alternative possibili.³

Alla luce delle precedenti considerazioni e dopo aver sottolineato le dipendenza concettuale della nozione di intelligenza artificiale da quella di intelligenza *tout court* occorre precisare il significato dell'AI. Con questa espressione si intende fare riferimento alla scienza volta a sviluppare modelli computazionali intelligenti, cioè programmi in grado di simulare le attività del cervello umano, ossia di svolgere compiti che richiederebbero "intelligenza" da parte dell'uomo, fino alla prospettiva di poter creare un'intelligenza senza corpo.⁴

L'intelligenza artificiale, occorre precisarlo, non è mai intesa come mera elaborazione di dati o meccanica esecuzione di compiti impartiti da umani, ma come applicazione di tecnologie in grado di svolgere prestazioni complesse, paragonabili a quelle di cui sono capaci le menti umane.

³ Riprendo la distinzione da M. Durante, *L'intelligenza artificiale*, da M. Durante e U. Pagallo (a cura di), *Manuale d'informatica giuridica e diritto delle nuove tecnologie*, UTET, Torino 2012, p.92.

⁴ Si v. ad esempio R. Kurzweil, *La singolarità è vicina*, tr. It. Apogeo, Milano, 2008.

La duplicità concettuale che deriva dallo studio della mente in rapporto alla capacità di realizzare comportamenti intelligenti, concerne ora, come detto, la scienza cognitiva, che studia l'intelligenza al fine di rappresentarla in modelli che possano essere trasferiti in applicazioni informatiche, ora invece si occupa prettamente delle tecnologie ausiliarie per tali applicazioni. In sostanza, occorre necessariamente e congiuntamente occuparsi dell'acquisizione della base di conoscenza necessaria al programma informatico e dell'esecuzione dei compiti realizzati dall'elaboratore in base all'informazione ottenuta e formalizzata, cioè tradotta nel "linguaggio del computer"⁵.

A questo proposito anche se ne ripareremo più nello specifico in seguito, si può far riferimento a due ulteriori nozioni di intelligenza artificiale: in senso forte ed in senso debole. La prima costituisce un programma di ricerca che assumendo come possibile la piena assimilabilità degli elaboratori alla mente umana, adotta un approccio sostanzialmente funzionalista che aspira a creare computer dotati di vera intelligenza e quindi di riprodurre gli stati cognitivi della mente umana.

L'AI in senso debole invece, rimanda all'antico sogno della creazione di autonomia di hobbesiana memoria,⁶ individuando un programma di ricerca che mira a creare computer in grado di eseguire prestazioni assimilabili a quelle determinate dall'intelligenza umana, ma senza dotare l'elaboratore di un'intelligenza equivalente a quella umana, nel senso che il computer simula ma non comprende gli stati cognitivi della

⁵ F. Casa, *Dalle scienze cognitive alle applicazioni giuridiche dell'intelligenza artificiale*, in U. Pagallo (a cura di), *Prolegomeni d'informatica giuridica*, Cedam, Padova, 2003, pp. 67-131.

⁶ Parla di "automata" Thomas Hobbes nella prime righe dell'Introduzione del *Leviatano*, Tr. It. A cura di A. Pacchi, Laterza Roma-Bari, 1992.

mente. Esso si limita a riprodurli, senza essere in grado di coglierli nella “loro piena complessità”.⁷

Questo ci rimanda al discorso di Floridi che fa da apertura al presente capitolo, vale a dire al fatto che siamo spinti, erroneamente a ritenere che una macchina che simula un’azione umana, per il solo fatto di saperla fare come l’uomo, o addirittura meglio sia necessariamente intelligente.

Un esempio: si pensi ad un algoritmo addestrato a riconoscere, in molteplici figure che gli vengono poste dinnanzi, una figura di gatto; essa dopo aver visto miriadi di gatti ed addestrata a fare esclusivamente questo lavoro, sarà sicuramente impeccabile nel riconoscere gatti di ogni genere e specie in qualsiasi altra figura che gli verrà presentata. Ciò nonostante questo non significa che tale algoritmo sia intelligente in quanto esso, nella sua attività non utilizza in alcun modo la logica ma semplicemente la statistica.

Per comprendere meglio quanto detto Turing si inventò un test, noto appunto come test di Turing. Tale test, nasce da una domanda che il matematico inglese si pone nel suo scritto del 1950 citato poc’anzi e cioè se le macchine pensano. Nell’articolo, in fondo, in quelle pagine che in effetti molti non leggono, lui risponde alla domanda affermando che essa è semplicemente idiota, in quanto non sappiamo cosa sia una macchina e cosa significhi pensare. Sosteneva infatti che per avere una definizione di macchina occorressero tutte le condizioni necessarie e sufficienti di modo tale che qualunque cosa sia una macchina soddisfi la definizione e chi soddisfa la definizione sia una macchina, ebbene a parer suo questo non è possibile da definire.

⁷ Si v. D. Tiscornia, *Intelligenza artificiale e diritto*, in R. Nannucci (a cura di), *Lineamenti d’informatica giuridica*, ESI, Napoli, 2002.

Il test in sostanza consiste nel prendere un essere umano ed un computer e metterli ciascuno in una stanza e poi avere un pubblico che gli sottopone delle domande; se in base alla risposta data le persone che interrogano l'umano e macchina non riescono a capire chi è chi, allora il computer avrebbe passato il test. Ad oggi il test non lo passa nessuno. Anzi, la cosa interessante è che a chiunque di noi sarà capitato di provare ad accedere ad un programma dove gli è stato richiesto di dimostrare di non essere un robot. Le cose si sono ribaltate ma a ben vedere se ci viene chiesto di dimostrare, tramite quadrati costituiti da pezzi di foto con semafori per esempio, di dimostrare di non essere un software ciò significa che un software non sarebbe, allo stato, in grado di bypassare tale "test". Lo sarà in futuro? Molto probabilmente sì ma a quale punto basterà spostare di poco l'asticella del test.

La chiave di lettura è pertanto quella di smettere di pensare che se computer e umano arrivano ad uno stesso risultato allora il procedimento che li ha portati lì sia lo stesso.

Oggi l'AI esiste perché abbiamo molta statistica e molto machine learning, ormai il mondo si sta adattando alle macchine, si pensi alle automobili autonome, esse sono fin dal principio forgiate in modo tale da esserlo e costruite per essere indipendenti e non sono, se nate per essere autonome, in alcun modo, guidabili da un essere umano.

Vi sono poi, come detto nell'introduzione, due principali tipologie di intelligenza artificiale, quella forte e quella debole le quali, nel corso degli anni sono state chiamate anche usando altre terminologie, ad esempio, l'AI "riproduttiva" e l'AI "produttiva". La prima cerca di ottenere con mezzi non biologici l'esito (chiamiamolo *output*) del nostro comportamento intelligente, cioè risolvere problemi o svolgere compiti con successo in vista di un fine. Si pensi a un robot che taglia

l'erba del prato come o meglio di noi. Non ci interessa se il processo che esegue è identico o anche solo simile al nostro, ci interessa soltanto il risultato. E in effetti, se una persona tagliasse l'erba come fa il mio robot, la prenderei per matta: quasi tre ore per un giardino di medie dimensioni di lenti giri casuali, su e giù per un piccolo prato, sbattendo qua e là, per fare un lavoro che ad un umano richiederebbe venti minuti. L'ovvio vantaggio è che i venti minuti io me li risparmio (si veda in seguito a proposito della macchine che ruberebbero il lavoro agli esseri umani).

Poi vi è il secondo tipo di AI, ossia quella "produttiva" che cerca di ottenere l'equivalente non biologico della nostra intelligenza, indipendentemente dal maggiore o minore successo applicativo del risultato. Usando lo stesso esempio del robot tagliaerba: l'AI produttiva vorrebbe creare un robot che taglia l'erba non girando per ore a caso, ma in modo intelligente ed efficace, per esempio, "riconoscendo" dove l'erba deve ancora essere tagliata, facendo attenzione agli angoli più difficili da raggiungere, fermandosi sui punti dove l'erba è cresciuta di più, evitando magari un gruppo di margherite che è bello lasciare in una parte del giardino, e così via. Questa AI non esiste e gli esempi della sua esistenza allo stato sono solo cinematografici: la falsa Maria in *Metropolis* (1927), HAL 9000 in *2001: Odissea nello spazio* (1968), C-3PO in *Star Wars* (1977), Rachael in *Blade Runner* (1982), Data in *Star Trek: The Next Generation* (1987), Agente Smith in *Matrix* (1999), Samantha in *Lei* (2013), o Ava in *Ex Machina* (2014).

L'AI riproduttiva come branca dell'"ingegneria" interessata alla riproduzione dei "risultati" (*output*) di un comportamento o processo intelligente (non del comportamento o del processo stesso), ha avuto un successo incredibile, ben oltre le più rosee aspettative. Al giorno d'oggi,

grazie anche alla straordinaria quantità di dati disponibili (se quantifichiamo tutti i dati a nostra disposizione dalle prime tavolette di creta a oggi , quasi tutti sono stati prodotti dalla nostra generazione) a una potenza di calcolo che segue la nota legge di Moore⁸, e quindi al passaggio da una base logico-matematica (se A allora B) a una statistica (B è correlato spesso ad A), ci affidiamo sempre di più ad applicazioni di AI per eseguire compiti non solo tediosi o che richiederebbero ma anche che potrebbero essere difficili o impossibili se dovessero essere eseguiti da un'intelligenza umana non assistita o aumentata. L'AI riproduttiva supera e sostituisce regolarmente l'intelligenza umana in un numero sempre più ampio di contesti. Si pensi al recente successo ottenuto da AlphaFold, un sistema di AI disegnato da DeepMind in grado di prevedere come le proteine si piegano nello spazio in forme tridimensionali.

Si tratta di un processo estremamente complesso, alla base del meccanismo biologico della vita, dato che la struttura tridimensionale delle proteine ne determina la funzione, per esempio come gli anticorpi combattono il coronavirus, Per cinquant'anni il problema è rimasto insormontabile, con soluzioni inefficienti e inaccurate, fatte "a mano" da scienziati in laboratorio al lavoro su poche strutture (si sa che esistono oltre 200 milioni di proteine, ma le loro strutture sono largamente ignote). Oggi AlphaFold permette di modellare queste strutture automaticamente, in breve tempo, e con una accuratezza

⁸In elettronica e informatica viene indicata come legge di Moore il seguente enunciato: "La complessità di un microcircuito, misurata ad esempio tramite il numero di transistor per chip, raddoppia ogni 18 mesi (e quadruplica quindi ogni 3 anni). Vale a dire che la velocità di calcolo dei computer cresce e crescerà nel futuro esponenzialmente.

superiore al 90%. Un lavoro praticamente impossibile se fatto da esseri umani.

D'altra parte, l'AI riproduttiva, come branca della scienza cognitiva interessate alla produzione di intelligenza umana (o magari superiore), è un fallimento completo. I sistemi che sappiamo costruire hanno un'intelligenza di un tostapane e non abbiamo, allo stato, la minima idea di come migliorare la situazione, non fosse altro per il fatto che sappiamo veramente pochissimo sulla stessa intelligenza umana. Basti pensare che già una vecchia indagine elencava 53 definizioni scientifiche di "intelligenza", ciascuna delle quali può essere i linea di principio "artificiale", e 18 definizioni di AI, e il numero è in crescita. Non è che l'AI "produttiva cognitivista" funzioni meno bene di quella "riproduttiva ingegneristica", o abbia un rendimento inferiore rispetto all'intelligenza umana: è che non è proprio parte della stessa competizione. Il famoso commento di Edsger Wybe Wijkstra secondo cui "la domanda se un computer può pensare non è più interessante della domanda se un sottomarino può nuotare" è indicativo dell'approccio ingegneristico condiviso dall'AI riproduttiva. Mentre il commento deluso di John McCarthy sulla vittoria di Deep Blue contro Kasparov è sintomatico del tipo di AI produttiva che disapprova l'AI riproduttiva, pensando che quest'ultima sia una sorta di tradimento dell'ideale iniziale.

E' perciò ovvio che "l'intelligenza artificiale" è un'etichetta vaga, che fa riferimento a una famiglia di discipline, tecniche e soluzioni ingegneristiche. Il resto è fantascienza.

Detto questo le due anime dell'AI, quella ingegneristica riproduttiva e quella cognitivista produttiva si sono spesso scontrate in faide fratricide per il predominio intellettuale e per le risorse finanziarie. Ciò è in parte

dovuto al fatto che entrambe rivendicano antenati comuni e un ‘unica eredità intellettuale: Alan Turing, la sua macchina con i suoi limiti computazionali, il suo famoso test e il fatto che una simulazione può essere usata per controllare sia e il comportamento o le prestazioni della fonte sono stati riprodotti o addirittura superati (AI riproduttiva o ingegneristica) sia se la fonte simulata è stata prodotta (AI produttiva o cognitivista). Questi due tipi di AI sono stati chiamati, come detto in vari modi, ed a parte quelli esposti, altri celebri nomenclature sono Good-Old Fashioned Artificial Intelligence (GOFAI) *vs* New o Nouvelle AI, o “AI universale”. Il disallineamento dei loro obiettivi e risultati ha causato infinite e inutili diatribe. I difensori dell’AI indicano i successi pratici dell’AI ingegneristica e riproduttiva per difendere gli obiettivi di quella cognitivista e riproduttiva. Mentre i detrattori dell’AI indicano gli insuccessi pratici dell’AI cognitivista e produttiva per attaccare gli obiettori di quella ingegneristica e riproduttiva.

In tutto ciò esiste un punto fondamentale che rischia di sfuggire: l’avvento dell’AI rappresenta una rivoluzione non nelle forme dell’intelligenza, ma nella forme dell’agire. Stiamo cioè inscrivendo nuove forme dell’*agere* nel libro dell’universo, non nuove forme dell’*intelligere*.

Come espresso ampiamente in precedenza, l’equazione giusta è *AI = AGERE ARTIFICIALE* e non *AI = INTELLIGERE ARTIFICIALE*.

Senza addentrarci in riflessioni che faremo a seguire sulle abilità straordinarie dell’AI ma anche sulle sue incapacità, avendo riguardo, per esempio, al tema dell’ambiente che circonda gli oggetti capaci di agire ma non intelligenti nel senso umano del termine ed anche alle regole che gli stessi seguono e sono in grado di comprendere (nel senso di servirsene per agire non nel senso di comprenderne la ragione della

loro esistenza) a discapito di quelle attività umane che invece seguono regole post-stabilite (ovvero nate dopo l'attività medesima) e quindi, allo stato, totalmente incomprensibili ad un software, ci poniamo adesso il problema di chi si adatterà a chi, vale a dire se siamo noi che dovremmo cambiare la nostra realtà per far coesistere tali oggetti dotati di AI o se inevitabilmente sono loro a dover adattarsi al mondo naturale per “avere successo”.

I sistemi di AI avranno sempre più successo, e perciò saranno sempre più utili, fino al punto di diventare indispensabili, tanto più avanza nella digitalizzazione dei nostri ambienti e nell'espansione dell'infosfera. Il mondo diventa infatti sempre più complesso da gestire e quindi sempre più bisognoso di AI per funzionare bene. Si tratta di due tendenze fondamentali che andranno irrobustendosi sempre di più. Da qui segue che anche l'avvolgimento (*enveloping*) del mondo è una tendenza robusta, cumulativa, e che si raffina progressivamente. Non ha nulla a che fare con la *singolarità* futura, in cui si oltrepassa un punto oltre il quale l'AI avrà uno sviluppo incontrollato, si gestirà da sola e finirà per schiavizzarci. Il processo e i problemi che ho delineato in questa lezione non si basano su speculazioni fantasiose relative a una qualche superintelligenza che conquisterà il mondo del prossimo futuro. Nessun Superman guiderà una rivolta digitale. Tuttavia, avvolgere il mondo a misura dell'AI e facilitare così l'emergere di agenti artificiali e il successo dei loro comportamenti è un processo che solleva sfide concrete e urgenti.

Immaginiamo due agenti: A e U. Sono una coppia e desiderano davvero far funzionare bene la loro relazione. Purtroppo A, che fa sempre di più in casa, è inflessibile, testarda, intollerante verso gli errori e difficilmente cambierà. Mentre U è esattamente l'opposto: è

intelligente, flessibile, malleabile ma sta anche diventando progressivamente più pigra e dipendente da A. Il risultato è una situazione di squilibrio, in cui A finisce per plasmare la relazione e distorcere i comportamenti di U, praticamente, se non di proposito. Il rischio è che, se la relazione funziona, è perché U si adatta ad A. La loro relazione, accuratamente ritagliata su A, diventa interpretabile nei termini della dialettica hegeliana padrone-schiavo. Ora, le “tecnologie intelligenti” svolgono il ruolo di A, mentre i loro utenti umani sono chiaramente U. Il rischio che stiamo correndo è che, disegnando un mondo digitale a misura dell’*agere* artificiale e non dell’*intelligere* umano, le nostre tecnologie e in particolare l’AI potranno finire con il modellare i nostri ambienti fisici e concettuali e “costringerci” o perlomeno invitarci ad adattarci a loro perché questo è il modo più semplice o migliore, o talvolta l’unico, per far funzionare le cose. Dopotutto, dato che l’AI è il partner stupido ma laborioso e inflessibile, e l’umanità quello intelligente ma flessibile e malleabile, chi si adatterà a chi? Il lettore probabilmente ricorderà molti episodi della vita reale in cui qualcosa non si è potuto fare affatto, o è stato fatto in modo laborioso o sciocco, perché quello era l’unico modo per far fare al sistema computerizzato ciò che si vuole ottenere. “Computer says no” è una famosa frase che il personaggio di Carol Beer, nella serie televisiva comica inglese *Little Britain*, usa per rispondere a qualsiasi richiesta del cliente.

Ma per chi non conosce le serie comiche inglesi, pensiamo al robot che pulisce i pavimenti di casa di nome Roomba. Il rischio è che potremo finire per costruire case con pareti curve e mobili con gambe sufficientemente alte per adattarci alle capacità di Roomba in modo molto più efficace.

Sarà certamente capitato a molti di noi di andare all'Ikea e comprare un divano o un letto pensando di prenderlo magari di un'altezza tale per cui Roomba possa passarci sotto e togliere la polvere che ivi si annida. Se ci pensiamo bene sono tanti i luoghi "rotondi" in cui viviamo, dagli igloo alle torri medievali, agli edifici pubblici, dove gli angoli delle stanze sono arrotondati per motivi igienico-sanitari. Pensandoci bene, se trascorriamo la maggior parte del nostro tempo all'interno di luoghi squadrati è a causa di un'altra serie di tecnologie legate alla produzione di massa di mattoni e infrastrutture in calcestruzzo, e alla facilità di tagli diritti del materiale da costruzione.

Sta di fatto che luoghi squadrati o rotondi sono stati costruiti seguendo le tecnologie predominanti piuttosto che attraverso le scelte dei loro potenziali abitanti. Sulla base di questo esempio, è facile vedere come l'opportunità rappresentata dal potere riontologizzante del digitale si presenti in tre forme: rifiuto, accettazione critica e progettazione proattiva. Diventando più criticamente consapevoli del potere riontologizzante del digitale e dell'AI, potremmo essere in grado di evitare le peggiori forme di distorsione (rifiuto) o almeno essere consapevolmente tolleranti nei loro confronti (accettazione), specialmente quando non è importante o comunque fondamentale o quando questa è una soluzione temporanea, in attesa di un design migliore. In quest'ultimo caso, essere in grado di immaginare come sarà il futuro e quali richieste adattive rischiano di imporre ai loro utenti umani il digitale in generale, e la sempre più comune presenza di agenti artificiali, aiuta a ideare soluzioni tecnologiche che potranno abbassarne i costi antropologici e aumentarne i benefici ambientali. In poche parole, il design umano intelligente dovrebbe svolgere un ruolo essenziale nel plasmare il futuro delle nostre interazioni con gli artefatti

intelligenti attuali e futuri e gli ambienti che condividiamo con loro. Dopotutto, è un segno di intelligenza far funzionare la stupidità a nostro vantaggio.

1.2 Gli agenti artificiali: tassonomie e classificazioni

Abbiamo finora posto in evidenza come nasce l'intelligenza artificiale e le sue più immediate implicazioni nel mondo reale e quali prospettive si hanno per il futuro del complesso rapporto tra umano e artificiale.

Occorre ora soffermarsi, più specificatamente, su quali siano i soggetti "protagonisti di questo mondo emergente, vale a dire cosa può essere rinchiuso all'interno della definizione di agente, quale sia il passaggio logico e semantico tra un agente e un agente artificiale, quanti tipi di agenti sono oggi riscontrabili in azione e quali sono le principali finalità per cui tali agenti vengono predisposti.

Un agente si intende un ente che pone in essere un comportamento in base ad una determinazione assunta in modo autonomo. Ancora, un agente, secondo l'UCITA⁹ (Uniform Computer Information Act)

⁹ L'UCITA fu redatta tra il 1996 e il 1999 dalla National Conference of Commissioners on Uniform State Law (NCCUSL). Si tratta di un ente composto da legislatori, magistrati, professori di diritto e avvocati nominati da ciascuno stato dell'Unione. La NCCUSL ha elaborato tutte le leggi commerciali uniformi, ivi incluso lo Uniform Commercial Code L'UCITA fu redatta tra il 1996 e il 1999 dalla National Conference of Commissioners on Uniform State Law (NCCUSL). Si tratta di un ente composto da legislatori, magistrati, professori di diritto e avvocati nominati da ciascuno stato dell'Unione.

La NCCUSL ha elaborato tutte le leggi commerciali uniformi, ivi incluso lo Uniform Commercial Code (UCC). L'UCITA incontrò una forte opposizione, in quanto si riteneva che limitasse i diritti dei consumatori, per questo fu adottata solo in due stati per poi essere definitivamente abbandonata nel 2002. Essa costituisce tuttavia un importante punto di riferimento per la regolamentazione delle attività degli agenti software.

statunitense sarebbe “un programma informatico, elettronico o altro mezzo informatico, usato in modo indipendente per avviare un’azione o per rispondere a messaggi o esecuzioni elettroniche, per conto della persona senza revisione o azione da parte di un individuo al momento dell’azione o della risposta al messaggio o all’esecuzione¹⁰.

Vi sono tre aspetti che vengono quindi presi in considerazione per definire la nozione di agente: 1) *l’aspetto del comportamento*, dell’azione concretamente posta in essere. Tale aspetto punta sull’elemento materiale dell’azione, ossia il comportamento dell’agente come realizzazione di un medesimo posto in essere da un uomo; 2) *l’aspetto della decisione*, cioè di quel processo volitivo che porta alla determinazione dell’agire, laddove l’attenzione è puntata sull’aspetto della razionalità della decisione, in quanto tale assimilabile a quella umana; 3) ed infine, *l’aspetto dell’autonomia*: tale aspetto indica che tanto l’azione quanto la decisione non sono eterodeterminate ma trovano origine nel soggetto agente. Qui ciò che conta è la qualità dell’azione e della decisione, come portatrici di risultati non totalmente prevedibili.

E’ a questo punto evidente come tali definizioni e analisi dei comportamenti degli agenti abbiano necessariamente implicazioni pratiche nel mondo dell’AI formalmente intesa e nel mondo della tecnologia tout court.

Un esempio a tal fine è la definizione di agenti software, cioè di sistemi informatici in grado di agire con autonomia, senza il controllo diretto

¹⁰ La traduzione è da ricondurre a Cevenini in G. Sartor, C. Cevenini e c. Di Cocco, *Lezioni di informatica giuridica*, Gedit, Bologna, 2004.

del loro utilizzatore. Si vuole come sempre sottolineare la capacità di un agente, nel caso del software costruito da un programma informatico, d'interagire in un ambiente complesso per svolgere un compito che richiede l'adozione di scelte alternative che si protrae per un certo tempo.

Per non complicare ulteriormente il discorso ci limiteremo ad analizzare in questa sede le caratteristiche degli agenti software partendo dal loro possibile inquadramento in due gruppi: con il primo descriveremo le qualità che definiscono lo statuto dell'agente e la sua capacità di eseguire compiti e con il secondo la loro capacità di interagire con l'ambiente, l'utente e altri agenti.

Le caratteristiche riferibili al primo gruppo sono le seguenti: 1) *l'autonomia*, cioè la facoltà per cui l'esecuzione di un compito trova origine nell'agente che lo realizza; 2) *la persistenza*, l'attitudine a permanere nel proprio stato intero anche dopo aver eseguito il compito per cui è predisposto; 3) *la vitalità*, cioè la capacità di individuare ed affrontare situazioni di anomalia; 4) *l'abilità*, ovvero la capacità di organizzazione "personale" del lavoro, suddividendolo e comunicando alle parti di cui è composta la procedura di esecuzione; 5) *la proattività*, la capacità di generare, anche insieme ad altri agenti, nuove situazioni ed eventi nell'ambiente in cui opera.

Circa il secondo gruppo, ovvero quello concernente le caratteristiche di tali agenti, esse sono così enucleabili: 1) *la reattività*, che indica la capacità dell'agente di interagire con l'ambiente, modificando o adattando i propri comportamenti in dipendenza degli stimoli provenienti dall'esterno; 2) *la comunicatività*, la capacità di entrare in comunicazione con l'utente o altro agente, al fine di porre in essere comportamenti collaborativi, di condividere risorse, di apprendere

dall'esterno ecc.; 3) *la mobilità*, la capacità dell'agente di muoversi all'interno della rete o di altro ambiente, allo scopo di mutare gli elementi di partenza o di variare i partners con cui è entrato in comunicazione, laddove si renda necessario reagire alle modificazioni o difficoltà intervenute nelle condizioni di esecuzione del compito; 4) *la benevolenza*, vale a dire l'attitudine dell'agente a compiere, esattamente, tutto ciò che occorre per l'esecuzione del compito per il quale è stato programmato, senza esserne deviato da stimoli provenienti dall'ambiente o da altri agenti; 5) *la veridicità*, cioè il non comunicare informazioni false nell'interazione con altri agenti suscettibili di fare affidamento sulle informazioni immesse e condivise.

Su queste basi proviamo ora a fare una breve classificazione di questi agenti software.

Possiamo individuare almeno due classi principali di agenti, utilizzando quali criteri distintivi di classificazione: 1) il sostrato sul quale l'agente viene implementato; 2) le finalità per cui l'agente è predisposto.

E' chiaro che tali classi prevedono l'esistenza di elementi comuni, classificati sulla base di diversi criteri distintivi, così come prevedono l'esistenza di varie sottoclassi di cui farò cenno.

In base al primo dei due criteri di classificazione è possibile distinguere tra agenti: 1) biologici; 2) robotici; 3) computazionali.

Gli agenti biologici sono quelle forme di organismi in cui sono emerse le capacità di adattarsi all'ambiente e di interagire con esso secondo il meccanismo di stimoli e risposte, dando luogo così ad una rappresentazione dello spazio fondata sulla distinzione tra interno ed esterno.

Gli agenti robotici sono agenti fisici che eseguono compiti manipolando il mondo entro un ambiente fisico. Sono agenti che non operano dentro la realtà simulata e connessa in rete, bensì dentro un ambiente fisico, solo parzialmente osservabile ed in continuo mutamento, in cui diviene più difficile per l'agente apprendere ed interagire con esso; tali agenti sono per ora caratterizzati da maggiori forme di automaticità e minori forme di intelligenza, rispetto agli agenti biologici (in particolare gli umani).

Gli agenti computazionali, infine, sono agenti autonomi¹¹ costituiti su base computazionale da sistemi informatici, avvalendosi di linguaggi di programmazione (tra cui Java), al fine di realizzare compiti, senza che ciò avvenga sotto il controllo diretto del loro utilizzatore.

Gli agenti computazionali, di cui si dirà ampiamente in seguito a proposito del loro impiego nel mondo del diritto e della giustizia in particolare, possono a loro volta essere classificati in base ai seguenti criteri distintivi: 1) il grado di autonomia; 2) lo statuto della base computazionale.

In base al primo criterio (grado di autonomia) gli agenti computazionali si suddividono in : 1) *agenti software*¹², sono quegli agenti caratterizzati da un ampio grado di autonomia, che sussiste non solo in

¹¹ L'autonomia risulta caratteristica essenziale degli agenti ed indica, con specifico riferimento agli agenti software, la capacità di originare decisioni o azioni, laddove designa, invece, con riferimento agli agenti generalmente intesi, la capacità di persistere, vale a dire di continuare ad esistere, dopo aver realizzato un determinato compito.

¹² Essi si descrivono come "agenti autonomi utili", che possono essere suddivisi in agenti reattivi ed agenti deliberativi, a seconda del loro grado di autonomia nell'esecuzione del compito del rilievo posto sul fare o sul decidere. La principale categoria degli agenti software è costituita dagli agenti intelligenti, che sono caratterizzati da un insieme di capacità, quali capire problemi e risolverli, pianificare, acquisire e rappresentare la conoscenza; apprendere, ricercare e condividere informazioni.

relazione all'ambiente in cui operano ma anche alle istruzioni ricevute; ciò significa che tali agenti assumono iniziative, vale a dire eseguono compiti non solo in seguito a richieste esterne ma anche sulla base del loro stato interno; 2) *artificial life agents*, : sono agenti caratterizzati da un minore grado di autonomia, che sussiste solo in relazione a richieste che provengono dall'ambiente. Non sono in grado di assumere iniziative nell'esecuzione dei compiti richiesti.

In base al secondo criterio (statuto della base computazionale) gli agenti computazionali si suddividono in agenti dotati di : *b1) sistema computazionale forte*: si tratta di agenti che, oltre ad essere dotati della sensibilità dell'ambiente, della capacità di modificarlo e di agire in seguito a richieste dipendenti dal proprio stato interno, sono conformati in modo tale da essere visti come entità antropomorfe, che riproducono, cioè, caratteri propri dell'essere umano, quali credenze, desideri, intenzioni, emozioni; *b2) sistema computazionale debole*: sono agenti che possiedono della precedente categoria le stesse caratteristiche funzionali, ad eccezione delle caratteristiche antropomorfe; *b3) sistema computazionale assistente*: sono agenti, invece, che si limitano ad aiutare l'utente nella sua interazione con il calcolatore. L'assistenza può assumere la forma della mediazione, della compagnia o dell'assistenza vera e propria dipendendo dal compito che l'utente deve svolgere.

Per quanto invece concerne la finalità per cui è predisposto, è conveniente una suddivisione in agenti: 1) intelligenti; 2) di informazioni; 3) di intrattenimento; 4) di interfaccia.

- 1) Gli *agenti intelligenti* sono destinati alla risoluzione di problemi. Essi pongono in essere una pluralità di operazioni allo scopo di soddisfare le richieste provenienti dagli utenti ovvero da altri agenti. Nell'eseguire tali compiti sono in grado di

interagire non solo con l'ambiente ma anche con le istruzioni ricevute, attraverso l'utilizzo di risorse d' intelligenza, secondo il grado di sofisticazione del loro stato interno.

- 2) Gli *agenti di informazione*: essi hanno accesso a diverse fonti d'informazione o per meglio dire agenti tramite i quali diverse fonti d'informazione divengono osservabili; essi sono in grado, inoltre, di recepire e manipolare le informazioni allo scopo di offrire una selezione di informazioni per ovviare al crescente problema del sovraccarico informativo (*information overload*). Tali agenti sono in grado di interagire con l'utente per specificare le sue richieste al fine di costruire una documentazione più rispondente alle sue reali esigenze.
- 3) Gli *agenti di intrattenimento*: essi hanno la caratteristica peculiare di simulare mondi, in particolare allo scopo di intrattenimento. Hanno subito recentemente grandi sviluppi in quanto, oltre agli scopi ludici perseguiti dagli utenti, costituiscono piattaforme tecnologiche per la sperimentazione di applicazioni proprie del mondo virtuale. Inutile ribadire il successo che l'evoluzione di tali piattaforme ha avuto nei tempi più recenti e di cui *Second Life* è certamente l'illustrazione più celebre. Tali piattaforme, infatti, sono suscettibili di conoscere uno sviluppo che va ben oltre gli scopi ludici di intrattenimento, dal momento che, per esempio, in un colloquio scientifico realizzato da *Second Life* è possibile raccogliere la stessa quantità di materiale di cui si può disporre in un colloquio organizzato nel mondo fisico, ma a costi decisamente inferiori.
- 4) Gli *agenti di interfaccia*: sono agenti che offrono un'ampia tipologia di assistenza agli utenti, tramite l'interazione con altri

strumenti computazionali. Il sistema di interfaccia di cui l'agente dispone permette allo stesso di svolgere un compito in correlazione con l'utente, che ne percepisce l'esecuzione, o in parallelo con l'utente, che ne visualizza i risultati. Tali agenti sono dunque programmati per collaborare con l'utente al fine di coadiuvarlo nella realizzazione di un lavoro o di assisterlo nella sua esecuzione secondo gradi e modalità d'interazione ed autonomia. In tal modo si ha un sistema di interfaccia sempre più *user friendly*, in grado di suggerire il modo più semplice di eseguire un comando da parte dell'utente ovvero di sostituirsi ad esso nell'esecuzione del comando.

1.2.1 L'interazione agente/ambiente

Occorre ora una breve premessa prima di descrivere più nello specifico i tre tipi di interazioni possibili ad un agente.

L'analisi svolta nel precedente paragrafo, relativa alla struttura ed alle finalità degli agenti, tende a mostrare come l'azione autonoma, posta in essere da un agente individuale o collettivo, sia sempre più percepita come parte di un tutto che si definisce nei termini di una trama o rete di interazioni. Il termine interazione indica precisamente che non è né la parte ad assumere un significato a partire dal tutto né questo a partire dalle parti ma vi è uno scambio reciproco determinato dalla capacità dell'uno di produrre adattamenti o modificazioni al variare delle condizioni dell'altro. Tale trama presuppone la molteplicità di relazioni che avvengono tra utente ed agente, tra agenti, tra agente ed ambiente in cui si trova ad operare. Se il termine di trama appare qui preferibile

a quello di rete è perché con esso si intende suggerire, alla stregua di un impianto normativo, che non tutti gli elementi in relazione sono suscettibili di rivestire il medesimo ruolo e la medesima importanza. Infatti, nella configurazione reticolare, ovvero nella società di parti i nodi della rete o le parti del tutto sono suscettibili di avere un ruolo differente, non in relazione a un ordinamento gerarchico della configurazione reticolare o della società di parti, ma, alla rilevanza ed affidabilità delle informazioni immesse e condivise in rete, da cui dipende, per esempio, la capacità di un agente di eseguire, in modo autonomo ed efficiente, il compito per cui è previsto.

L'efficacia nell'esecuzione di tale compito dipende non solo dalle qualità del singolo agente, ma anche dalla sua capacità d'interagire con l'ambiente, gli altri agenti e l'utente finale, vale a dire dal grado di decentramento della trama di interazioni. L'agente è, infatti, tanto più efficace quanto più riesce a muoversi entro l'ambiente. Inoltre, la sua efficacia si accresce condividendo informazioni con altri agenti oppure tenendo comportamenti collaborativi con l'utente. Requisiti come mobilità (delocalizzazione), comunicazione (condivisione delle risorse remote), e collaborazione (apprendimento per esperienza) sottolineano che la società di agenti è basata su una trama di interazioni in cui assume un valore essenziale la parte, l'apporto individuale non considerato in quanto tale ma globalmente in vista dell'obiettivo comune, si tratti di agente individuale o collettivo.

Per chiarire ulteriormente la consistenza di tale trama di interazioni, nel campo di analisi di agenti software, occorre precisare le tre direzioni in cui l'agente è suscettibile di interagire: 1) nei confronti dell'ambiente; 2) di altri agenti; 3) dell'utente.

Inutile dire che tali possibilità di interazioni si sovrappongono e si sommano tra loro.

Ora, venendo all'interazione agente/ambiente, l'ambiente in cui gli agenti oggetto del nostro esame operano si definisce, dal punto di vista operativo (*task environment*), secondo i dati e le informazioni che l'agente riceve tramite specifici sensori ed ai quali reagisce tramite specifici attuatori.¹³

L'interazione dell'agente software consiste in tal senso in una riduzione della complessità dell'ambiente esterno, tramite la selezione dei dati e delle informazioni rilevanti per l'esecuzione del compito. La capacità d'interazione con un'ambiente viene a dipendere, dunque, da tre fattori: 1) la *conformazione* dei sensori e degli attuatori di cui è provvisto l'agente: tale conformazione è tanto più efficace quanto più l'agente è in grado di interagire con le istruzioni di partenza (e non soltanto con l'ambiente in cui si trova ad operare); 2) la *mobilità* dell'agente entro l'ambiente operativo che gli consente di decentrare la propria azione di acquisire informazioni a dati laddove sono prodotti, con il risparmio del traffico in rete e del carico computazionale; 3) la *conformazione* dell'ambiente operativo, che può essere configurato secondo proprietà rispondenti a quelle degli agenti, le quali rendono più facile ed efficiente il grado d'interazione.

Si tratta dunque di indicare quali siano tali proprietà dell'ambiente, definite dalla specifica del *task environment*, in grado di influenzare l'azione degli agenti: 1) *l'osservabilità*: indica la predisposizione dell'ambiente a rendere osservabili le risorse e i dati che sono necessari

¹³ Per la nozione di sensori ed attuatori si veda l'analisi PEAS (*Performance, Environment, Actuators, Sensors*).

per l'esecuzione del compito da parte dei sensori dell'agente; 2) la *dinamicità*: indica l'attitudine dell'ambiente a mutare conformazione in dipendenza dell'azione dell'agente; 3) la *determinatezza*: indica l'attitudine dell'ambiente ad indurre l'agente a determinare lo stato interno successivo (che costituisce l'insieme delle informazioni necessarie all'azione) in base a quello attuale; 4) *l'episodicità*: indica l'attitudine dell'ambiente a conformarsi secondo episodi o moduli, in modo tale che l'azione dell'agente viene a dipendere dalla consistenza del singolo episodio o modulo (ciò limita la possibilità di propagazione dell'errore derivante da un regime sequenziale, in cui ogni azione influenza quella successiva); 5) la *discretezza*: indica l'attitudine dell'ambiente a fornire solo un numero limitato di dati ed informazioni definiti con chiarezza, cui è associato un numero limitato di percezioni ed azioni da parte dell'agente.

L'interazione agente/ambiente è ordinata dalla logica del decentramento che permette ad ogni agente di massimizzare la ricerca efficiente di informazioni, che costituiscono la base di conoscenze, corrispondente allo stato interno degli agenti, necessarie per l'esecuzione del compito previsto: la logica del decentramento opera, nell'ambito della trama di interazioni agente/ambiente, tramite la mobilità dell'agente, l'osservabilità e la dinamicità dell'ambiente.

1.2.2. L'interazione agente/agente

Di fondamentale importanza appare anche l'interazione agente/agente; essa rispetto a quella agente/ambiente tende ad ovviare a due possibili problemi: nel primo ambito di interazioni si tratta, in primo luogo, di

acquisire le informazioni necessarie nell'esecuzione di un compito; talora è però necessario poter disporre di informazioni ulteriori allo scopo di superare un ostacolo, di procedere ad una scelta alternativa, di modificare un comportamento, a causa della conformazione stessa dell'ambiente: per fare tutto ciò occorre acquisire informazioni da altri agenti. Inoltre, nell'ambito delle interazioni con l'ambiente, l'acquisizione delle informazioni avviene solitamente sulla base di una regola: per la risoluzione di determinati problemi o l'esecuzione di taluni compiti occorre, invece, poter estrarre conoscenza dai metadati del web semantico. A tal fine occorre costruire piattaforme comuni, che rendano possibile la comunicazione tra agenti e la suddivisione di risorse: tale problema si definisce al duplice livello della costruzione di architetture condivise e dell'elaborazione di protocolli standard di linguaggio.

E' possibile indicare una tipologia di architetture, ordinate secondo il grado di condivisione delle risorse basato sulla capacità di estrarre conoscenza da metadati. Il primo livello è definito dalle *architetture reattive*, entro le quali l'acquisizione di informazioni da parte degli agenti è guidata dalla reazione a stimoli provenienti dall'ambiente: la condivisione di risorse è, dunque, ancorata allo stato attuale della situazione in cui versa l'agente, senza che questo sia in grado di predire stati futuri del mondo, ovvero di generare piani per affrontare casi non previsti. Il secondo livello è definito dalle *architetture ibride*, nelle quali l'acquisizione di informazioni sulla base di stimoli provenienti dall'ambiente è, almeno in parte, associata alla capacità di accedere a conoscenze fornite da altri agenti allo scopo d'elaborare piani per ovviare a difficoltà non previste. In tal caso il grado di adattabilità ai mutamenti è maggiore, senza tuttavia dare luogo ad una vera e propria

evoluzione dello stato interno dell'agente. Il terzo livello è definito da *architetture basate sulla conoscenza*, entro le quali gli agenti assumono la conformazione di agenti deliberativi, caratterizzati cioè dalla capacità di rappresentarsi la modificazione dell'ambiente e delle condizioni di esecuzione del compito in base alla previsione degli effetti delle proprie azioni e di quelle di altri agenti. Tali agenti intendono simulare la risposta fornita da un essere umano, sulla base di un ragionamento simbolico che tiene conto, per la sua scelta, degli stati possibili del mondo.

L'esigenza di comunicazione tra agenti e di condivisione delle risorse richiede, inoltre, la definizione di protocolli standard di linguaggio, da implementare secondo le architetture consuete *client/server* ovvero tramite architetture in via di costruzione che siano in grado di mettere in comunicazione diretta gli utenti. La definizione di un protocollo di linguaggio si fonda su una serie di decisioni relative alla sintassi, semantica e pragmatica del linguaggio al fine di specificare il suo standard d'applicazione. Oltre ai requisiti di forma e contenuto, che costituiscono le proprietà comuni di ogni protocollo di linguaggio, devono essere previsto, in vista della piena implementazione del protocollo, secondo la logica del decentramento, due importanti ulteriori requisiti: il *networking* che esprime l'attitudine del protocollo ad adattarsi alle tecnologie della rete, utilizzando tutti i tipi di connessione di base (sincrono e asincrono); l'*affidabilità*, che definisce l'attitudine del protocollo a generare affidamento, dal punto di vista della privacy degli scambi, dell'autenticità dell'agente della comunicazione, del funzionamento del sistema comunicativo in rete.

1.2.3 L'interazione agente/utente

L'acquisizione delle conoscenze nella trama di interazioni procede secondo un'ulteriore direzione: quella che pone in relazione l'utente e l'agente.

Dobbiamo richiamare alla mente, a tal proposito, una considerazione svolta nell'ambito dell'analisi dei sistemi intelligenti: un'agente intelligente dispone di una base di conoscenza che può progressivamente arricchire (nel senso di un 'accumulazione di conoscenze), attraverso la comunicazione con altri agenti come si è visto nel paragrafo precedente, ma non disponendo di un contesto di esperienza, che è fondamentale per la selezione delle informazioni. Solo su queste basi è infatti possibile definire un profilo dell'utente, apprendere dagli errori, fare tesoro delle indicazioni ricevute, tenere comportamenti collaborativi finalizzati a precisare le esigenze dell'utente, generalizzare risultati e trattarli alla stregua di esempi indicativi per l'adozione di scelte future.

In tale prospettiva, l'esperienza indica la capacità di trattenere qualcosa di soggettivamente significativo del passato: tale capacità risulta essenziale per l'applicazione di una logica decentrata, perché fa sì che l'utente (ovvero l'agente) possa accedere all'informazione utile e remota, senza dover percorrere l'intero tragitto compiuto durante la prima selezione; è sufficiente per l'utente (o per l'agente) memorizzare il profilo della ricerca o apprendere dall'errore ovvero ancora prendere nota di un risultato conseguito (che diviene parte della base di conoscenza).

A tale scopo è necessario riprodurre, tramite simulazione, un contesto di esperienza per l'agente, che trovo fondamento nella trama di

interazioni agente/utente e consente all'agente di acquisire nuove ed ulteriori conoscenze. La simulazione del contesto di esperienza si avvale di una serie di metodologie, di cui indichiamo di seguito le principali: 1) *user direct feedback*: l'agente acquisisce conoscenze rivolgendosi direttamente all'utente ed interrogandolo su determinate scelte fatte, che vengono in tal modo chiarite; 2) *user indirect feedback*: l'agente interagisce con l'utente proponendo suggerimenti in vista dell'esecuzione di un compito o della risoluzione di un problema, memorizzando i suggerimenti trattenuti; 3) *learning by example*: l'agente dispone di esempi, forniti dall'utente, per fondare su di essi le proprie scelte; 4) *user looking*: l'agente acquisisce indicazioni osservando le scelte dell'utente entro un periodo di tempo e modificando il profilo utente in ragione delle scelte effettuate; 5) *agent asking*: l'agente acquisisce conoscenze, di cui non dispone circa un dato problema, facendone richiesta ad altri agenti (tale metodologia è suscettibile d' accrescere il contesto di esperienze dall'agente, ma si avvale dell'interazione con altri agenti).

1.3 Dai primordi alle odierne applicazioni di AI: la robotica civile

Uno dei settori che maggiormente ha interessato, anche e soprattutto da un punto di vista economico e di investimenti il settore dell'intelligenza artificiale è senza dubbio la robotica.

Una premessa: l'oggetto delle presenti tesi dottorale concerne il giudice robot, ossia quella branca dell'intelligenza artificiale applicata alla giustizia ma non si poteva non fare un seppur conciso riferimento alla

robotica civile tout court, ossia a quel settore che attiene alla costruzione e progettazione di un complesso variegato ed eterogeneo di macchine, che vanno dai robot chirurgici, ai sistemi automatizzati di trasporto aereo, terrestre e marittimo, fino alle implicazioni industriali nell'ambito manifatturiero o nell'agricoltura, la robotica di servizio e molto altro.

Già dal 1824, il famoso poeta di Recanati, Giacomo Leopardi, prendendo spunto dai versi dell'amico poeta Giovanni Battista Casti in "Gli animali parlanti" e "dei fortunati secoli in cui viviamo", sosteneva che questi "fortunati secoli" fossero proprio "l'età delle macchine" e perciò, ironicamente, sottolineava come queste ultime avrebbero contribuito al progresso e alla felicità degli uomini¹⁴.

Più di un secolo dopo, nel 1942, un altro tra i massimi esponenti della letteratura, questa volta fantastica, dei suoi tempi, stiamo parlando di Isaac Asimov, conia, nel suo breve racconto intitolato "Runaround" il termine di questo capitolo: robotica. Da quell'anno il tema dei robot è stato sviscerato da matematici, informatici, giuristi e studiosi vari con la dovuta serietà analitica. La robotica i fatti, fin dalla sua nascita, ha interessato i campi più vari della nostra vita, si tratta cioè di un settore per eccellenza interdisciplinare, nel quale convergono ricerche di informatica e cibernetica, matematica e meccanica, elettronica e neuroscienza, biologia e scienze umane, psicologia, economia e *last but not least* il diritto.

La grande varietà, illustrata sopra, di ambiti in cui i robot operano ha fatto sì che si discuta ancora, dopo più di mezzo secolo dalla loro nascita, non solo lessicale, sulla definizione di "robot" e su alcune sue

¹⁴ Riprendo le citazioni da U. Pagallo, *Robotica*, in M. Durante e U. Pagallo, *Manuale d'informatica giuridica*, p.141.

proprietà quali “l’autonomia”, “l’adattatività” e i gradi di “interattività”.

Senza cadere nell’eccesso di formalismo abbiamo voluto, per completezza riportare di seguito la definizione che uno dei più autorevoli dizionari italiani dà del termine “robot”. Esso viene così definito: “apparato meccanico ed elettronico, programmabile, impiegato nell’industria in sostituzione dell’uomo per eseguire automaticamente ed autonomamente lavorazioni od operazioni ripetitive, complesse, pesanti o pericolose”¹⁵.

La definizione appare ellittica e manchevole in molti punti che nel corso del discorso saranno resi evidenti.

Studiosi del settore informatico definiscono la robotica come l’area dell’intelligenza artificiale volta alla costruzione di macchine in grado di “sentire, pensare e agire”¹⁶.

Altri, come il direttore dei Laboratori di intelligenza artificiale presso l’Università di Stanford, Sebastian Thrun, presentano i robot come macchine con l’abilità di “percepire alcunché di complesso e prendere decisioni appropriate”¹⁷. Mentre ulteriori definizioni insistono sulla capacità di apprendimento e adattamento dei robot all’evoluzione dell’ambiente, nell’ambito delle applicazioni industriali può rinviarsi alla definizione ISO 8373 richiamata dalle Nazioni Unite e dal rapporto della Commissione economica per l’Europa e la Federazione internazionale di robotica, per cui il robot è un “manipolatore multiuso,

¹⁵ Riprendo la definizione del dizionario Treccani online.

¹⁶ G. A. Bekey, *Autonomous Robots: From Biological Inspiration to Implementation and Control*, The Mit Press, Cambridge, Mass., Londra, 2005.

¹⁷ P. Singer, *Wired for War. The Robotics Revolution and Conflict in the 21 Century*, Penguin, Londra 2009, p. 77.

controllato automaticamente, riprogrammabile e programmato in tre o più assi, che possono essere mobili o fissi in un luogo”¹⁸.

Da questo tipo di definizioni tuttavia sorgono ulteriori problemi.

Basti pensare alla Nota sulla dottrina unitaria del Ministero della Difesa del Regno Unito sui “sistemi aerei senza pilota” del 30 marzo 2011, secondo cui il concetto di autonomia va declinato come “capacità di comprendere un più alto livello di intento e direzione”. Per giungere al nodo del problema occorre, a parere di chi scrive, semplicemente precisare un livello minimo, ma ben definito, delle nozioni. Anche ammettendo cioè che i robot possiedono l’intelligenza di un frigorifero, bisogna concedere che alcune di queste macchine non solo sentono ma agiscono (ovviamente questo non significa che siano intelligenti come ampiamente detto sopra).

Riprendendo la tripartizione di Allen, Varner e Zinser nel saggio sullo status degli agenti morali artificiali, si può dire che un ente umano, animale o artificiale “agisce” allorché esso sia interattivo, autonomo e capace di adattarsi al proprio ambiente. Questo in sostanza significa che:

- a) L’agente risponde agli stimoli dell’ambiente attraverso il mutamento degli stati interni o valori delle sue proprietà (interattività)
- b) L’agente è in grado di cambiare detti stati indipendentemente da stimoli esterni (autonomia)

¹⁸ UN World Robotics, *Statistics, Market Analysis, Forecast, Case Studies and Profitability of Robot Investment*, a cura di UN Economic Commission for Europe and coauthored by the International Federation of Robotics, UN Publication, Ginevra, 2005.

- c) L'agente è capace di accrescere o migliorare le regole attraverso cui tali stati cambiano (adattatività)

Su queste basi il problema oggi non è tanto quello di determinare se e in che modo gli agenti artificiali, e, in particolare, i robot eventualmente "agiscano". Piuttosto la questione verte sulla circostanza che l'interattività, l'autonomia e l'adattatività dei robot comporta l'imprevedibilità delle loro azioni, sia nei confronti dei programmatori e costruttori di tali agenti artificiali sia dei loro stessi proprietari o utilizzatori.

Secondo il parere di Curtis Karnow, ad esempio, la capacità di agire dei robot comporterebbe "il fallimento del nesso di causalità giuridica", in quanto sarebbe arduo predire il comportamento della macchina e, di qui, i possibili danni derivanti dalle sue azioni o decisioni.¹⁹

A sua volta un esperto del settore come Gianmarco Veruggio avverte che "non è necessario immaginare scenari di fantascienza in cui i robot dispongono di coscienza, libero arbitrio ed emozioni, per ammettere che tra pochi anni saremo destinati a convivere con robot dotati di auto-coscienza ed autonomia"²⁰.

Assodata quindi la capacità di agire dei robot che dire della loro imputabilità?

Per chiarire il punto utilizzeremo un esempio derivante dal mondo economico²¹.

Si pensi ad c.d. mercato a "doppia asta", vale a dire il sistema in cui tanto i venditori quanto gli acquirenti immettono simultaneamente le

¹⁹ Si v. C. Karnow, *Liability for Distributed Artificial Intelligence*, in "Berkeley Technology and Law Journal", 1996, 11, pp. 147-183.

²⁰ G. Veruggio, *Euron Roboethics Roadmap*, in "Proceedings Euron Roboethics Atelier", 2006, Genova, p. 27.

²¹ Traggo spunto da U. Pagallo, *Robotica*, cit., p. 144.

proprie domande e offerte in competizione sul mercato; in alternativa quindi, al sistema dei titoli non quotati in borsa, in cui gli scambi sono negoziati. Dai risultati dei primari esperimenti realizzati presso il prestigioso Santa Fe Institute del New Mexico, nel marzo del 1990, è stato dimostrato che un mercato popolato da soli robot commercianti, (dal significativo nomignolo di Zero Intelligence Agent), è di per sé in grado di raggiungere ciò che gli economisti chiamano “equilibrio competitivo”. Questi agenti “non intelligenti” non solo superano gli esseri umani inesperti nel gioco della domanda e dell’offerta di mercato, ma hanno la capacità di raggiungere un alto livello di efficienza distributiva nello stabilire il prezzo medio e quantità dei beni scambiati sul mercato.

Questo equilibrio è stato spiegato sostenendo che esso non dipende tanto dall’abilità degli agenti nel processare informazioni e prendere di conseguenza decisioni razionali, bensì dall’intelligenza implicita nelle stesse regole del mercato²². Tutto questo per affermare e sostenere che la divergenza tra efficacia operativa e intelligenza dei robot dimostra come mai nella loro capacità d’agire non debba necessariamente far seguito la loro imputabilità e personalità giuridica.

Inoltre a ben vedere non ha molto senso discutere innanzi ad un giudice il grado di imputabilità di un robot dato che esso non dispone (ancora e forse mai disporrà) di proprietà quali la coscienza o l’intenzionalità delle proprie azioni che fungono da requisiti indispensabili per stabilire la responsabilità dell’agente per il suo comportamento.

²² Si v. S. Sunder, *Market as Artifacts: Aggregate Efficiency from Zero Intelligence Traders*, in M. Augier e J. Marsch (a cura di), *Models of a Man: Essay in Memory of Herbert A. Simon*, Mit Press, Cambridge, Mass. 2004, pp. 501-519.

Come suggerito da più esperti di responsabilità civile, la consapevolezza risulta caratteristica essenziale per determinare l'imputabilità, morale e giuridica, relativa a specifiche azioni, in quanto se è pur evidente che i robot possono essere fonte di danni e perfino di comportamenti moralmente riprovevoli, non avrebbe tuttavia senso porsi il problema se essi possono essere riconosciuti come soggetti "colpevoli" non essendo, per ora, capaci di "volere" o "desiderare" di fare un qualcosa.

Peraltro tutte le teorie sulla pena ruotano attorno alla rieducazione del soggetto penalmente sanzionato e/o all'eventuale risarcimento che esso deve come ristoro del danno fatto patire al terzo. Appaiono anche qui di improbabile applicazione le due sanzioni ad un robot che non può essere rieducato e tanto meno possiede redditi propri.

Tuttavia il problema dell'efficacia operativa dei robot e la loro imprevedibilità nelle azioni che pongono in essere concerne non solo la contrattualistica (se decido di utilizzare un robot o un software stipulerò necessariamente un contratto con il suo costruttore o venditore) o la responsabilità civile (qui intesa come responsabilità civile extracontrattuale) ma anche il diritto penale.

Prima di approfondire il tema legato alla responsabilità civile e penale dei nostri agenti artificiali occorre ancora segnalare l'aspetto connesso ai vari tipi di robot esistenti ed operanti oggi.

Semplificando il discorso, per non incorrere in micro-tipologie che a nulla servono ai fini della mia ricerca possiamo distinguere la robotica civile e la robotica militare. Tuttavia occorre sin d'ora rilevare che la robotica civile andrebbe ulteriormente distinta al suo interno in altre due macro-categorie: la robotica industriale e la robotica di servizio.

Nondimeno per non incorrere nell'ormai celebre parossismo della "legge del cavallo", espressione coniata dal giudice Franck Easterbrook aprendo i lavori di un convegno sulla *cyberlaw* nel lontano 1996 bisognerebbe mettere in guardia dagli eccessi di classificazione: se infatti molti casi giuridici interessano questi nobili animali, i cavalli pongono problemi molto diversi e quindi, a seconda del caso che il giudice ha di fronte, egli deve saper vederli come fossero animali ogni volta differenti²³.

La morale da trarsi è che un'astratta classificazione dei robot, anche solo in una delle sotto-categorie come quella dei robot di servizio, con connessa normativa di riferimento, rischierebbe di rivelarsi un esercizio inutile ancorché dannoso: per quanto possano individuarsi parametri oggettivi, infatti, la malleabilità logica delle tecnologie digitali, la continua evoluzione delle stesse e la multi-funzionalità potenziale di ogni macchina renderebbe ogni tentativo destinato a durare l'espace du matin.

Il fatto che un drone dedicato alla video sorveglianza di un parco si sposti volando genera evidentemente problematiche diverse rispetto ad un robot che si muove con le stesse finalità lungo i sentieri del medesimo parco, ma questo non esclude che molte criticità di sicurezza o di privacy siano per le due macchine concettualmente simili.

Per non parlare del grado di interazione tra macchina e uomo, che per i robot industriali è definito come grado di collaboratività

²³ Così si è espresso Easterbrook: "Any effort to connect these strands into a course on "The Law on the Horse is doomed to be shallow and to miss anything principles". In *Cyberspace and the Law of the Horse*, in "University of Chicago Legal Forum", 1996, pp. 207 ss. Le tesi di Easterbrook, naturalmente, hanno destato un vivace dibattito. Basti v. A.D. Murray, *The Regulation of Cyberspace: Control in the Online Environment*, Routledge- Cavendish, New York, 2007, p. 9 ss., e, prima ancora, L. Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, in "Harvard Law Review", 1999, 113, pp. 501 ss.

nell'esecuzione dei compiti, è evidentemente un parametro potenzialmente fondamentale, ma una classificazione basta sul grado di autonomia nell'output della macchina, qualora realizzabile, non potrebbe generare categorie omogenee utili se non a meri fini speculativi.

Da quanto detto emerge quindi un modo di vedere, in chi scrive, orientato verso un approccio pragmatico, cioè verso orientamenti e decisioni (almeno in partenza) date caso per caso, senza cioè che i giudici si lascino imbrigliare da categorie e vincoli anche mentali che difficilmente possono appartenere ad un mondo così vario e che in buona parte ancora non esiste.

Per concludere, senza mettere in discussione quanto detto, come rilevato più sopra, l'unica classificazione, che a fini espositivi conviene fare è tra la robotica civile e la robotica militare; tuttavia ai fini della presente tesi dottorale mi soffermerò solo sulla prima.

Le applicazioni della robotica civile, come detto, sollevano un plesso di questioni che, in parte, si sovrappongono all'ambito delle applicazioni in campo militare e, dall'altro però se ne differenziano notevolmente²⁴.

La progettazione e la conseguente produzione sino alla vendita di robot, con le relative clausole, termini e condizioni di utilizzo, sono ambiti regolati dalla materia contrattual-privatistica, vale a dire che nel caso in cui, per esempio, sorgano problemi inerenti all'utilizzo improprio, in violazione cioè a norme, civili o penali, responsabili saranno evidentemente le autorità civili e militari che hanno autorizzato l'uso, mentre nel caso in cui la macchina non agisca secondo le corrette

²⁴ Prendo spunto da U. Pagallo, *Robotica*, cit., pp.148 ss. E anche dal documento Nexa su "Il diritto della robotica di servizio", disponibile sul sito web del Centro per le ricerche d'internet e società del Politecnico di Torino.

(conformi a legge) istruzioni impartite, la responsabilità ricadrà sul produttore del robot. La questione si fa più sofisticata però se si prendono come esempio i famosi robot da Vinci, quegli agenti cioè, massicciamente utilizzati negli ospedali per le operazioni alla prostata. Qui il grado di responsabilità eventualmente attribuibile all'ingegnere-costruttore sarà calibrato in base al grado di prevedibilità e al rischio che questi si è assunto nel costruirli, vale a dire egli non risponderà dell'eventuale danno subito dal paziente dovuto al malfunzionamento della macchina, solo se riesce a dimostrare che un caso fortuito è occorso, cioè se era totalmente imprevedibile e irrilevabile ad un occhio esperto (in quanto giustamente è questo il grado richiesto dai produttori di tali agenti, non cioè di quello del cosiddetto "buon padre di famiglia", ossia dell'uomo medio) il guasto.

Tornando poi alla classe dei robot commercianti segnalata nel paragrafo precedente, anche gli Zero Intelligence Agents possono essere rischiosi e pericolosi: il loro spirito eccessivamente utilitarista ha infatti suggerito analogie con l'avidità degli speculatori in borsa e con le bolle dei mercati, dato che, come gli uomini, anche questi robot possono essere sopraffatti dalla complessità dell'ambiente e apparire dilettoni allo sbaraglio. La difficoltà per il legislatore è dunque quella di trovare una normativa che sia in grado di prevenire che le persone facciano a meno d'impiegare tali agenti per paura di incorrere in causa di responsabilità dovute all'imprevedibilità degli stessi.

Al riguardo alcuni studiosi hanno proposto di adottare meccanismi assicurativi come sistemi di autenticazione dei robot o forme di responsabilità limitata²⁵. Questa faceva sì che gli schiavi potessero

²⁵ Cfr. U. Pagallo, *The Law of Robots: Contracts, Crimes, and Torts*, Dordrecht, Springer, 2013, cap. 3.4.1.

acquistare merci e porre in essere il commercio in quanto in grado di garantire alle proprie controparti che le obbligazioni da loro assunte, sarebbero state soddisfatte proprio mediante il *peculium*. Questo per far comprendere al lettore che una simile assicurazione per i robot potrebbe rivelarsi vincente sotto due aspetti: il primo incentivando l'utilizzo e la messa in circolazione di tali artefatti in quanto i proprietario, avendo il robot il suo *peculium*, non temerebbe di essere rovinato da strambe decisioni della macchina poniamo nel fare affari azzardati in borsa, e secondo anche coloro che entreranno in contatto con tali agenti sarebbero sicuri che in caso di danni da questi cagionati, gli stessi potrebbero rispondere col loro *peculium* ed entro i limiti di questo; ma del resto anche gli umani non potranno mai essere chiamati a rispondere (in questo caso a pagare) oltre i limiti consentiti dal proprio budget.

Per concludere il discorso (che riprenderemo a conclusione del capitolo 2) e prima di addentrarci nel complesso mondo che racconta di come gli agenti artificiali possono entrare in contatto con noi, lavorare per noi, vale a dire il metodo di apprendimento automatico definito *machine learning* o se si preferisce *deep learning* (anche se le due terminologie non sono propriamente sinonimi), possiamo notare come l'attuale legislazione, italiana e non solo, non sia, allo stato, in grado di far fronte a tali nuove problematiche.

Occorre pertanto un legislatore sensibile, in grado di comprendere i cambiamenti della società che l'uso di questi nuovi agenti comporta e di conseguenza porre in essere una serie di tutele senza scoraggiarne però l'utilizzo. Si potrebbe pensare a forme di assicurazione obbligatoria come per i veicoli tradizionali e a forme di responsabilità limitata quali il già citato *peculium digitale*. Il problema in cui sarebbe sconveniente incorrere è il già menzionato "vizio del giurista" di

classificare ed etichettare ogni fattispecie esistente in griglie predefinite. Con gli agenti robotici occorre che ci si renda conto fin dai primordi che essi includono tipologie molto differenti tra loro che vanno dai robot preposti alla cura della casa, a quelli utilizzati nelle sale operatorie, dai robot educatori a veri e propri robot badanti e quindi umanoidi.

Un atteggiamento fiducioso nei confronti dei parlamenti è necessario in quanto, non è certo la prima volta che essi sono chiamati a dettare regole per la disciplina di costrutti artificiali, occorre soltanto che ne comprendano il loro potenziale.

1.4 Dalla struttura delle reti neurali al machine learning e deep learning

In questo sottocapitolo, che funge da parte conclusiva del primo capitolo in cui ci siamo occupati di umano e artificiale e del loro complesso rapporto, nonché delle loro differenze ontologiche, ci occuperemo delle modalità attraverso le quali gli agenti artificiali entrano in contatto con gli esseri umani e come riescano in taluni casi a risolvere problemi meglio degli stessi umani che li hanno creati. Per fare ciò è però necessario occuparci di cosa c'è di artificiale nell'intelligenza artificiale (il gioco di parole è voluto) e poi di come funziona il cervello umano, in quanto è proprio di lì che gli scienziati sono partiti per studiare il modo attraverso il quale una macchina possa interagire con un essere dotato di intelligenza, ossia un essere umano. Coloro i quali creano intelligenza artificiale hanno bisogno di avere una visione, anche solo con un atto di immaginazione, dell'umanità e

dell'umano. Questo significa che è necessario che diverse visioni per genere, età, etnia e cultura, concorrano tra loro. Questa ispirazione è al centro di un istituto fondato nel 2019 dall'Università di Stanford, in California, l'istituto per un'intelligenza artificiale centrata sull'uomo (Stanford Institute for Human-Centered Artificial Intelligence)²⁶. Rappresentare cosa sia umano significa rendere visibile o manifesto lo scopo perseguito da chi lavora nella ricerca e negli investimenti e anche ciò che è invisibile, perché potrebbe essere implicito nelle tecnologie usate o costruite.

Quella di Stanford non è una congregazione di scettici delle tecnologie, che cercano di imbrigliare lo sviluppo tecnico con norme "esterne", etiche o giuridiche, ma piuttosto un gruppo di ricercatori di avanguardia (con collegamenti anche con Alphabet-Google), che, ai livelli più avanzati, ha sviluppato la consapevolezza di quanto grande sia la partita in corso e, perciò, la necessità di una visione globale.

E' interessante, poi, che la necessità di avere una visione dell'umanità (*need to represent humanity*) venga sollecitata da Fei-Fei Li, una scienziata che ha dedicato tutta la sua ricerca proprio alla visione computerizzata (*computer vision*), uno dei settori di avanguardia. Vi è come un'assonanza tra l'esperienza della catalogazione di un'enorme quantità di cose oggetto di visione e lo sviluppo della consapevolezza che la più completa lista di cose (che possono essere riconosciute da una macchina) non ci dice ancora nulla sulle relazioni tra quelle cose. E' necessario, a tal fine, che si sviluppino parimenti la capacità di analisi

²⁶ Se si visita il sito web dello Stanford Institute for Human-Centered Artificial Intelligence, si trovano le dichiarazioni dei due co-direttori, Fei-Fei Li e Etchemendy. In particolare Fei-Fei Li dichiara: "The creators of AI need to represent humanity. This requires a true diversity of thought across gender, age, and ethnicity and cultural background as well as a diverse representation from different disciplines".

di un insieme di dati sugli attributi delle cose e sulle relazioni tra le stesse e la collaborazione di neuroscienziati ed esperti di processi cognitivi umani, che dicano come ciò accade esattamente negli uomini. Quando i ricercatori californiani sostengono che non vi sia nulla di artificiale nell'intelligenza artificiale lo dicono intendendo: "che essa è ispirata da persone, è creata da persone e, cosa più importante, incide sulle persone". E' uno strumento potente che si sta solo ora cominciando a capire come funzioni, e questa è una profonda responsabilità".

John Etchemendy, l'altro co-direttore del centro di Stanford, enuncia così i tre principi fondamentali che guidano l'attività del centro: una scommessa che il futuro dell'AI sia ispirato dalla nostra comprensione dell'intelligenza umana; la tecnologia deve essere guidata dalla nostra comprensione di come si sta influenzando la società umana; le applicazioni di intelligenza artificiale dovrebbero essere progettate per migliorare e aumentare ciò che gli umani possono fare.

Si può dire che non vi è niente di artificiale, nel senso di non umano, nell'AI, così come non vi è niente di artificiale nell'energia elettrica, nel motore a vapore, nella stampa a caratteri mobili e via discorrendo, nel senso che nessuno direbbe mai "stampa artificiale", "elettricità artificiale", o simili. E allora bisognerebbe dire intelligenza delle macchine, o capacità computazionale delle macchine, piuttosto che AI. E' come se per la prima volta si fosse usciti dalla fase sperimentale della computazione digitale, iniziata con Turing, e si cominciasse solo ora a svilupparla e renderla materiale e fruibile.

Tuttavia per ora la visione di Fei-Fei- Li si collega alle riflessioni di Joshua Greene, quando dice che finché non capiremo meglio come funzioni la cognizione umana e finché non faremo dei passi decisivi su

questo punto, non faremo neanche dei passi decisivi sull'intelligenza "derivata" o artificiale.

Il problema pertanto non è solo lo sviluppo di macchine più potenti ma quello dell'architettura cognitiva. Se l'idea è quella di agente intelligente, anche se non necessariamente "creature artificiali che in contesti appropriati appaiono come persone" è importante il modello di intelligenza che si assume come riferimento.

Prima e piuttosto che imbarcarsi nella diatriba "i computer sono più o meno intelligenti degli umani" pare interessante porsi il problema: "a che punto siamo nella comprensione del funzionamento del cervello umano". Secondo Joshua Greene, professore di psicologia a Harvard, gli umani hanno una capacità il cui funzionamento non è ancora noto, e cioè la capacità di considerare tutti i tipi di idee, pur senza credere nel fatto che esse siano vere (una funzione fondamentale nella pianificazione delle azioni e nell'immaginare possibilità che ancora non esistono nella realtà).

Cercare di capire come gli uomini facciano ciò è un passaggio essenziale per tentare di costruire un'intelligenza artificiale generale anche chiamata forte, ossia quella che riprodurrebbe gli stati cognitivi della mente umana e la vera sfida per il futuro dell'intelligenza artificiale.

Sostiene Greene che quello che stiamo tentando di fare è capire come i nostri cervelli siano capaci di una cognizione di alto livello. In termini si può parlare di una semantica composizionale, o di una semantica composizionale multimodale. Questo significa, in parole più semplici, cercare di capire come il cervello prende i concetti e li mette insieme in modo da comporre un pensiero, così come si possa leggere una frase come "il cane ha inseguito il gatto", e si capisca che esso significa

qualcosa di diverso dal “il gatto ha inseguito il cane”. Gli stessi concetti sono coinvolti, “cane”, “gatto”, “inseguire”, ma il nostro cervello è capace di mettere le cose insieme in modi diversi al fine di produrre un significato diverso.

Molto di quello nel quale l'apprendimento automatico (*machine learning*) ha avuto successo è stato sul versante della percezione, riconoscendo oggetti, o quando si tratta di andare dalla visione al linguaggio, semplicemente etichettando scene che sono già familiari, così che si possa mostrare un'immagine di un cane che insegue un gatto e sia possibile che essa dica qualcosa come “cane che insegue gatto”. Greene prosegue affermando:

Non si può realmente comprendere qualcosa se non si comprendono certe parole e le si collega significativamente alle cose che si possono vedere o toccare o sperimentare in modo sensoriale. Così pensare è qualcosa tra le immagini e tra le parole. [...] quello che il mio laboratorio sta cercando di fare è capire come questo aspetto centrale, ma veramente poco compreso, dell'intelligenza umana funziona. Come noi combiniamo i concetti in modo da formare pensieri. Come può lo stesso pensiero essere rappresentato in termini di parole contro cose che si possono vedere o sentire con gli occhi e con le orecchie della tua mente? [...] Penso che non abbiamo ancora capito il caso umano, e la mia ipotesi è che ovviamente è qualcosa che per intero sono i neuroni a fare, ma queste capacità non sono ben catturate dagli attuali modelli di reti neurali. [...] Se si vuole veramente costruire un'intelligenza generale artificiale, si deve iniziare con le conoscenze consolidate sulla cognizione, e non solo cercando di costruire qualcosa che possa, ad esempio, leggere frasi e dedurre cose da quelle frasi.

Il problema è, quindi, duplice. Il cervello umano è certamente complicato e lungi dall'essere compreso, mentre l'intelligenza

artificiale, che si sta cercando di replicare nelle macchine, esiste solo negli esseri umani e ciò, per la scienza attuale crea non pochi problemi posto che, allo stato viene considerato “non etico” fare ricerca sugli umani a questo livello. E’ probabile che lo sviluppo della ricerca su umanoidi cerebrali (sviluppati con materiale umano ma fuori dal corpo umano) possa aprire nuove strade di ricerca.

Il riconoscimento delle immagini, come detto, è il settore in cui vi sono stati i maggiori progressi, grazie a innovazioni algoritmiche e investimenti in infrastrutture, ad esempio, nell’hardware utilizzato per addestrare il sistema o nel software utilizzato per connettere questo hardware. La velocità di addestramento di un algoritmo di ImageNet poniamo, a novembre 2018, è stata 16 volte maggiore rispetto a quella di giugno 2017. Il tasso di errore dell’etichettatura automatica di ImageNet è diminuito vertiginosamente negli ultimi 10 anni fino ad arrivare ad un tasso di errore pari oggi al 3%, mentre le prestazioni umane, che sono il punto di riferimento, sono circa al 5% di errore.

Questi dati, raccolti dall’*AI Index 2018 Report*, dell’Università di Stanford, ci dicono di un settore in grandissima espansione, con risultati che superano quelli umani e che ha grandi implicazioni a livello sociale e giuridico, come possibilità di controllo delle pubbliche autorità (e non solo) su tutti i cittadini.

Vi sono poi, questioni di grande importanza aperte tutt’ora e segnalate nel summenzionato *Index* quali, per esempio, se si vuole che l’intelligenza artificiale aumenti l’intelligenza umana anziché sostituirla, si sostiene qui che sia importante passare da ambienti nei quali le macchine sono opposte agli umani a ambienti dove umano e macchina lavorano insieme. Inoltre, i sistemi di domande/risposte basati sull’intelligenza artificiale non hanno poi, la capacità di sostenere

un dialogo; la profonda comprensione del linguaggio naturale è ancora una sfida, così le capacità di ragionamento e di prendere decisioni basate sul buon senso. Ancora, l'apprendimento automatico ha molto successo in molte applicazioni, poiché può scoprire correlazioni nascoste all'interno di enormi quantità di dati. Ciò consente di fare previsioni molto accurate, ma non dà alcuna informazione circa i rapporti di causalità (che come abbiamo già avuto modo di dire ma che sottolineeremo anche in seguito sono altra cosa).

Questo significa che, in settori come il diritto, la strada da compiere è ancora lunga per arrivare a usi diffusi da parte di cittadini e professionisti.

Vi sono poi i timori di vari personaggi anche piuttosto autorevoli secondo i quali le macchine supereranno gli umani. L'idea che l'intelligenza artificiale possa significare la fine della razza umana e che i robot possano rimpiazzare l'umanità, perché capaci di un'evoluzione più celere di quella biologica è stata espressa anche da Stephen Hawking in più occasioni, tra le quali una nota intervista alla BBC. (che si può trovare agevolmente in rete).

La preoccupazione è quella dell'avvento di macchine dotate di un'intelligenza superiore a quella umana, che possano migliorare autonomamente la loro architettura e realizzare quella che Vernor Vinge già nel 1993 aveva chiamato *singularity* e che nel film *Transcendence*²⁷ prende appunto il nome del titolo del film. Il

²⁷ Transcendence, film del 2014. Il breve film racconta la storia dello sviluppo incontrollato di un'intelligenza artificiale con sensibilità umana, vessata dalla minaccia di terroristi che alla fine hanno la peggio sul senso di umanità che comunque è stata immessa nell'intelligenza artificiale che riesce a replicarsi e così a salvare il pianeta dalla minaccia di terroristi invece pronti ad usarla per scopi meno nobili.

commento di Roger Penrose²⁸ su queste posizioni è *tranchant*: non sono le macchine di per sé ad essere pericolose, più pericoloso delle macchine è il pensiero che esse possano essere migliori di noi.

Intanto, pare utile segnalare tre ragioni che fanno ritenere infondato questo timore, almeno per un (non sappiamo quanto) lungo periodo.

In primo luogo, lo sviluppo dell'apprendimento automatico richiede la disponibilità di una grandissima quantità di dati, che siano selezionati con l'aiuto di un umano che confermi la corretta identificazione dell'oggetto (come, per esempio, un 'automobile o un umano). E anche quando, per esempio, un sistema venisse alimentato con libri di progettazione e ingegneria della auto, potrebbe evidenziare e riassumere le informazioni chiave, ma non potrebbe mai progettare una nuova auto.

In secondo luogo, ad oggi la potenza di calcolo delle macchine non può avvicinarsi a quella di un umano. La situazione potrebbe cambiare con i computer quantistici, che non è dato prevedere quando saranno disponibili per l'uso corrente. Intanto, è certo che, sulla base dei principi di ingegneria, "l'attuale tecnologia dei circuiti integrati che alimenta le capacità computazionali dei computer convenzionali non sarà mai in grado di replicare il cervello umano, tanto meno superarlo nelle capacità intellettuali.

Infine, non possiamo progettare macchine che superino ciò che allo stato non siamo ancora stati in grado di comprendere del tutto, ossia noi stessi. La matematica computazionale, che ha alimentato gli attuali progressi dell'AI, si è avvalsa di reti neurali, ispirate al cervello umano.

²⁸ R. Penrose, matematico britannico, nel 2020 gli viene assegnato il Premio Nobel per la fisica per aver scoperto che la formazione dei buchi neri è la robusta previsione della teoria generale della relatività.

Ma la scienza moderna non comprende ancora fino in fondo come funzioni il cervello, questione decisiva.

Per concludere questa prima parte per poi giungere alla parte più tecnica, e forse anche più noiosa di cosa sia effettivamente il *machine learning* e di come funzioni, occorre sottolineare come negli ultimi anni si sia affermato un approccio *AI for good*. Ne sono prova McKinsey *Tech for good: Using technology to smooth disruption well-being* o anche l'AI for Good Global Summit 2018, svoltosi a Ginevra il 15-17 maggio 2018 e, in qualche modo, anche l'iniziativa della Stanford University di cui si è detto sopra. Sullo sfondo è sempre presente l'idea della neutralità della tecnologia che l'AI è all'origine di gravi pericoli ma può al tempo stesso essere la soluzione. (sulla neutralità della tecnologia si veda ampiamente nel prosieguo del nostro lavoro).

Tess Posner²⁹ indica una via differente. Non si preoccupa solo degli algoritmi non etici e dell'AI che prende una cattiva strada, ma lavora per aumentare la diversità e l'inclusione nell'intelligenza artificiale con un motto che dovremmo tenere a mente: "l'intelligenza artificiale cambierà il mondo; chi cambierà l'intelligenza artificiale".

E' giunto il momento di occuparci più nel dettaglio degli aspetti tecnici e semantici del funzionamento del *machine learning*.

La maggior parte dei sistemi di intelligenza artificiale moderni funzionano sulla base del *machine learning*, ovvero dell'apprendimento automatico. Le prestazioni di questi sistemi risultano sempre più efficienti ed accurate, questo grazie alla possibilità di usufruire sempre più di grandi quantità di dati. Il *machine learning* permette alle macchine e ai computer di svolgere attività e funzioni sulla base dell'apprendimento derivante dall'esperienza. In pratica, gli

²⁹ T. Posner è il CEO di A14ALL.

algoritmi di *machine learning* usano procedimenti matematico-computazionali per acquisire informazioni direttamente dai dati senza l'utilizzo di equazioni o modelli matematici predeterminati. Tali algoritmi migliorano le proprie prestazioni in modo "adattivo", cioè in maniera graduale all'aumento degli esempi che vengono forniti.

Se dovessimo tradurre in italiano il termine *machine learning* bisognerebbe farlo con "apprendimento automatico", che indicherebbe quella particolare abilità dei computer di apprendere senza essere stati preventivamente ed esplicitamente programmati in tal senso.

La definizione di *machine learning* ad oggi più accreditata è quella di T.M. Mitchell, direttore del dipartimento *machine learning* della Carnegie Mellon University, essi afferma:

"Si dice che un programma apprende dall'esperienza E con riferimento ad alcune classi di compiti T e che con misurazione della performance P, se le sue performance nel compito T, come misurato da P, migliorano l'esperienza E".

Cercando di utilizzare un linguaggio più semplice, il *machine learning* consente ai computer di imparare dall'esperienza. Vi è apprendimento quando le prestazioni del programma migliorano dopo lo svolgimento di un compito o il completamento di un'azione, anche se sbagliata. Anziché scrivere il codice di programmazione tramite il quale, di volta in volta, si dice al computer cosa fare, vengono forniti alla macchina solo un set di dati inseriti in un generico algoritmo che sviluppa una propria logica per svolgere l'attività ed i compiti richiesti.

L'apprendimento automatico opera sulla base di diversi approcci che permettono di differenziarlo in sottocategorie: 1) apprendimento

supervisionato; 2) apprendimento non supervisionato; 3) apprendimento per rinforzo; e 4) apprendimento semi supervisionato.

La prima tipologia di apprendimento, vale a dire quello supervisionato fornisce al computer esempi completi da utilizzare come direttrice da seguire per eseguire il compito richiesto. Nello specifico si parla di apprendimento supervisionato quando al computer vengono forniti sia dei set di dati come input sia le informazioni relative ai risultati auspicati, con l'obiettivo che il sistema trovi una regola generale che colleghi i dati in ingresso con quelli in uscita³⁰ in modo tale da poterla riutilizzare per altri compiti simili. In questo modello il lavoro di risoluzione viene lasciato nelle mani del computer e, una volta compresa la funzione matematica che ha portato a risolvere un insieme specifico di problemi, sarà possibile riutilizzare la funzione per rispondere a qualsiasi altro problema simile.

L'apprendimento non supervisionato, ossia quello in cui viene lasciato lavorare il software senza alcun tipo di aiuto esterno, avviene quando al sistema vengono forniti solo set di dati senza alcuna indicazione del risultato desiderato. Il fine di questo approccio è quello di poter risalire ai modelli e schemi nascosti, ossia identificare negli input una struttura logica senza che questi siano preventivamente qualificati.

Poi abbiamo la tipologia dell'apprendimento per rinforzo, altrimenti detto rinforzato, tramite il quale il sistema deve interagire con un ambiente dinamico che gli fornisce i dati di input e raggiungere un

³⁰ Gli vengono forniti degli esempi di input e di output in modo che impari il nesso tra loro.

obiettivo imparando anche dagli errori; nel momento in cui raggiunge l'obiettivo il sistema riceve una ricompensa, mentre quanto commette un errore riceve una punizione. Le prestazioni ed il comportamento del sistema vengono determinate da una routine di apprendimento basata su ricompensa e punizione. In questo modo il computer impara ad esempio a battere un avversario in un gioco concentrando gli sforzi nello svolgere un determinato compito al fine di raggiungere il massimo valore della ricompensa. Il sistema così impara degli errori commessi, migliorando le sue future prestazioni in funzione dei risultati che ha ottenuto in precedenza.

Esiste poi un modello ibrido di machine learning, ovvero l'apprendimento semi supervisionato: in questo caso al computer viene fornito un set di dati incompleti per l'apprendimento/allenamento; alcuni di questi input sono dotati dei rispettivi esempi di output, come nell'apprendimento supervisionato, mentre altri ne sono privi, come nell'apprendimento non supervisionato. L'obiettivo è quello di identificare funzioni e regole per la soluzione dei problemi e di strutture e modelli di dati utili al raggiungimento di determinati obiettivi.

Ancora, esistono poi delle sottocategorie di machine learning che appaiono utili a fornire una sorta di classificazione pratica perché identificano degli approcci pratici di applicazione degli algoritmi di machine learning da cui si possono desumere delle categorie di apprendimento dei sistemi.

Gli "alberi delle decisioni", ad esempio, sono sistemi basati su grafici attraverso i quali vengono sviluppati modelli predittivi con cui è possibile scoprire le conseguenze di determinate decisioni.

Un altro esempio proviene dal “clustering” , ossia da modelli matematici che consentono di raggruppare informazioni, dati, oggetti simili; è un’applicazione pratica del machine learning dietro al quale esistono differenti modelli di apprendimento che vanno dall’identificazione delle strutture (quali caratteristiche ha un cluster e qual è la sua natura) al riconoscimento degli oggetti che devono far parte di un raggruppamento piuttosto che di un altro.

Esiste inoltre la sottocategoria dei “modelli probabilistici” che basano l’apprendimento del sistema sulla base del calcolo delle probabilità³¹. Infine fanno parte di questi sotto insiemi le reti neurali artificiali, sistemi che utilizzano per l’apprendimento certi algoritmi ispirati al funzionamento, alla struttura e alle connessioni delle reti neurali dell’essere umano. Queste reti neurali possono anche essere multi-strato ed in questo caso si parla del campo del *deep-learning* o apprendimento profondo.

Il *deep learning* è quell’area del machine learning che si basa su un particolare tipo di apprendimento dei dati caratterizzato dalla creazione di un modello di apprendimento automatico a più livelli nel quale i livelli più profondi prendono come input le uscite dei livelli precedenti, trasformandoli e astraendoli sempre di più; essi sono in sostanza ispirati al funzionamento ed alla struttura del cervello umano e vengono comunemente chiamati “reti neurali profonde” perché hanno, oltre ai tre livelli basilari (input layer, hidden layer, output layer), un numero di hidden layer che può arrivare anche oltre i 150. Tali reti neurali profonde riescono, grazie al loro numero maggiore di strati intermedi,

³¹ La più nota è la rete Bayes, un modello matematico probabilistico che rappresenta in un grafico l’insieme delle variabili causali e le relative dipendenze condizionali.

a costruire più livelli di astrazione che permettono alle reti di riuscire a risolvere problemi complessi.

Il deep learnign consente ai computer di imparare con l'esempio, ed è tra l'altro, il tipo di tecnologia che più di altre si sta sviluppando in quanto riesce a raggiungere risultati che prima non sembravano possibili. Esempi di questa tecnologia li troviamo nelle automobili a guida autonoma o nel controllo vocale in dispositivi quali telefoni cellulari, tablet, tv e altoparlanti vivavoce.

Tale tipo di apprendimento è stato teorizzato per la prima volta negli anni '80 ma ha avuto un forte sviluppo solo di recente per due ragioni principali:

- 1) Il *deep learning* richiede una grande quantità di dati etichettati. Processo possibile solo grazie alle recenti tecnologie.
- 2) E' necessaria una notevole potenza elaborativa per processare tali dati.

I software in questione, infatti, apprendono attraverso dei "training set" ossia un insieme di dati su cui esso si allenerà, andando a paragonare i dati ricevuti e apprendendone le caratteristiche secondo le impostazioni iniziali indicate dall'algoritmo di programmazione.

Durante l'apprendimento poi, la rete confronta il risultato ottenuto con quello che sarebbe stato corretto indicato nel set di esempi inserito; se la risposta differisce la rete si corregge cambiando la propria configurazione dei pesi sulle connessioni.

Ripetendo così la procedura più volte, si riconfigura casualmente minimizzando l'errore. Le reti neurali seguono una funzione di apprendimento definita "*back propagation*" che consiste nel

metabolizzare alla rete neurale una quantità enorme di esempi svolti in modo da impostare i suoi pesi e i suoi neuroni. Si procede dando alla rete in input uno di questi esempi svolti che viene fatto passare attraverso tutti i suoi livelli. Arrivati alla fine del processo si verifica la risposta che è stata ottenuta in modo quasi casuale. Molto probabilmente questa risposta è sbagliata, ma dato che si tratta di un esempio svolto la rete ha la possibilità di vedere “quanto” ha sbagliato e di correggere l’output. La rete neurale fa poi il processo inverso, risalendo di nuovo tutti i livelli. Dopo un buon numero di esempi metabolizzati e propagati all’indietro, la rete inizierà a fare le associazioni giuste, minimizzando sempre di più l’errore e inizierà così a dare risposte corrette con sempre maggiore probabilità.

Dopo un ingente addestramento, la rete acquista l’abilità di dare risposte corrette non solo nei casi contenuti nell’insieme degli esempi, ma anche in casi analoghi.

Questo tipo di struttura di apprendimento permette alle reti neurali di:

- a) Essere in grado di trovare una soluzione pur partendo da informazioni incomplete o parziali;
- b) Generalizzare, cioè reagire ad un input mai incontrato, in maniera analoga alla risposta data ad un insieme di input già conosciuti;
- c) Apprendere dall’esperienza;
- d) Memorizzare le informazioni e utilizzarle in relazione al significato;

- e) Elaborare nuove regole partendo dall'osservazione di regole costitutive (su questo tema ci soffermeremo tra poco alla fine di questo capitolo)

Le reti, inoltre, non ragionano secondo uno schema formale “se-allora”, ma secondo uno schema probabilistico legato al contesto specifico sul quale sono costruite. Sono adatte a trovare informazioni nuove laddove non si conosca tutto il dominio da rappresentare, ma sfruttando i dati noti con una elaborazione di tipo statistico in modo tale da renderli coerenti con un fine da raggiungere. Basandosi su regole di tipo statistico si differenziano dal modello deterministico dei sistemi esperti, le reti non sono quindi la migliore soluzione nel caso in cui si conoscano tutti i dati, ma risultano vincenti quando è necessaria un'attività integrativa, simile a quella che compie il cervello umano. Le reti, dall'analisi dei dati ricevuti, enucleano le regole sulle quali si fonda la correlazione tra input e output, e tale tipo di apprendimento dinamico richiede che la riconfigurazione della rete avvenga più volte finché non si raggiunga il pattern di riferimento. Una volta raggiunta una regolarità però, la rete potrà essere applicata a nuove situazioni purché esse non siano dissimili da quelle precedentemente analizzate. Attraverso il *machine learning* quindi partendo da un algoritmo predeterminato, la rete riesce, una volta accumulato e confrontato enormi quantità di dati a “comprendere” la realtà circostante e a capire come rispondere e comportarsi. Un esempio lo abbiamo, ancora una volta, con il riconoscimento di immagini da parte di un calcolatore. Un computer non è in grado di “vedere” la differenza tra un cane e un gatto; ciò però se esso non è stato a ciò addestrato. Se però si fornisce alla macchine un numero elevato di immagini di cani e di gatti e la si addestra a comprendere i tratti differenziali, allora esso alle fine sarà in grado di

distinguere se la foto che gli si prospetta davanti ritragga un cane o un gatto.

Un altro esempio eclatante e col quale ci avviamo alla conclusione del presente paragrafo, è stato quello di AlphaGo che ha sconfitto il campione mondiale di Go, un gioco da tavola cinese molto più complesso degli scacchi. Le prime versioni del gioco partivano dall'analisi di partite effettuate da esseri umani, che venivano elaborate analizzando l'efficacia di ogni mossa in vista dei suoi possibili sviluppi predetti dal calcolatore. Attraverso algoritmi di *machine learning* il software sta continuando a migliorarsi ed ormai è diventato praticamente imbattibile.

L'ultima versione già citata in precedenza, ossia "AlphaGo Zero" inoltre è stato programmato per allenarsi da solo giocando contro se stesso (senza cioè che ci sia dall'altra parte un essere umano e questo ovviamente è un grandissimo salto qualitativo in termini di *machine learning* posto che fino ad allora ad allenare una macchina o un software c'era pur sempre dietro un essere umano), senza caricare alcuna partita giocata ma semplicemente fornendogli le regole del gioco. In soli ventuno giorni il programma è riuscito a superare il livello della precedente versione, cioè quella che aveva sconfitto il campione del mondo.

E arriviamo qui, di nuovo, alla teoria del già citato Prof. Floridi circa la fondamentale differenza che sussiste, in tema di apprendimento automatico delle macchine, tra un apprendimento basato su attività costituite da regole costitutive e attività che invece si basano su mere regole vincolanti.

Un esempio: il gioco del tennis, possiede delle regole che, se non seguite portano a tirare la pallina da una parte all'altra del campo senza

un fine ben preciso ma comunque non sono necessarie per far divertire con due racchette e una pallina, poniamo, due bambini. Esse in quanto tali, si possono definire regole vincolanti, ovvero regole che non preesistono al gioco medesimo ma semplicemente che se non seguite non consentono di giocare.

Abbiamo poi una seconda tipologia di attività sportiva, per restare in tema di sport, si pensi al gioco degli scacchi, o della dama, o a scopa; qui le regole poste a fondamento del gioco sono di tipo costitutivo, in quanto preesistono al gioco medesimo e in assenza delle stesse non vi è alcun tipo di gioco né di passatempo.

Ebbene la riflessione che a questo punto ci offre il filosofo di Oxford è la seguente: le macchine apprendono molto bene le attività che si fondano su regole costitutive mentre sono ancora del tutto incapaci ad apprendere a svolgere attività basate invece su regole vincolanti, le quali ovviamente rappresentano la stragrande maggioranza delle attività umane.

Concludendo, la vera difficoltà è quella di descrivere un determinato problema in forma di algoritmo in quanto lo stesso non ha mai regole precise, analitiche, inequivoche, astratte, ma ogni problema ha le sue di regole.

Per esempio, sembra banale ma far ridere in modo consapevole un robot è al momento impossibile, e si badi, non per questioni di coscienza, etica, morale, che allo stato un robot sappiamo non avere, ma proprio per una questione matematica: non esiste un algoritmo in grado di spiegare come questo succede. Quello che invece si potrebbe già fare è l'applicazione di un "modello di comportamenti" cioè una sorta di algoritmo che dia al robot il comando di ridere quando, per esempio, tutti coloro che lo circondano ridono. Ma la sfida di oggi è come far

ridere un robot quando si trova da solo davanti ad una scena divertente? Una soluzione potrebbe essere quella di caricare nella memoria del robot migliaia di copioni teatrali, cinematografici, indicandogli i momenti in cui la gente ha riso ad una battuta. In questo modo, attraverso le tecniche di apprendimento automatico il robot sarebbe in grado di ridere anche senza avere nessuno intorno.

Il meccanismo è esattamente lo stesso che utilizza AlphaGo Zero, una volta che ha metabolizzato le regole costitutive del gioco del Go, registra ogni mossa in un determinato contesto fino a capire quando la stessa può portare alla vittoria e quando alla sconfitta. Gli stessi computer attraverso il *machine learning* potrebbero aiutarci a trovare un modo per elaborare algoritmi in grado di rendere calcolabili varie categorie di problemi ancora oggi “incompresi” dalle macchine stesse.

Capitolo II

Il diritto e l'intelligenza artificiale

Il diritto costituisce un ricco banco di prova e un importante campo per lo sviluppo di un'intelligenza artificiale basata sulla logica, in particolare per quanto riguarda i modelli logici di argomentazione giuridica. Secondo un'opinione autorevole, l'attività di creazione e applicazione del diritto comporta l'elaborazione di informazioni, il prendere decisioni e la comunicazione delle stesse. A sostegno dell'importanza della logica occorre notare che il diritto è parte della società e che questo rende la logica particolarmente rilevante per il diritto, poiché esso deve essere compreso da chi è il destinatario di una norma che, poniamo, imponga particolari doveri o anche diritti a fronte

di una controprestazione e la sua applicazione deve essere spiegata e giustificata. Sono, quindi importanti la chiarezza dei significati e la coerenza dei ragionamenti, e quindi, si può dire, l'uso appropriato della logica.

Naturalmente quanto detto non significa che il diritto possa agevolmente essere rappresentato con formule matematiche o in un linguaggio formale e che, una volta fatto questo passaggio, le conseguenze giuridiche possano scaturire automaticamente per deduzione. Ciò non è possibile in quanto gli studiosi, tra i quali il giurista bolognese Giovanni Sartor, sono concordi nel dire che il diritto non è soltanto un sistema concettuale e assiomatico, ma ha obiettivi ed effetti sociali, e poi perché, anche nel caso di produzione legislativa in senso proprio, la norma viene nel tempo chiamata ad essere applicata in circostanze non previste al momento della sua approvazione, senza contare le formule astratte ed imprecise come la "buona fede", l' "abuso di segreti industriali" , la "correttezza", la "ragionevolezza", tutte formule che richiedono di essere tradotte in norme aventi un contenuto concreto.

Quindi, il giurista, di fronte alla tecnologia applicata al suo pane quotidiano, ossia il diritto, appare spesso perplesso in termini non soltanto di efficacia ma anche di tutela di quei diritti o di quelle posizioni di fatto che come detto, non possono essere rappresentate matematicamente, ossia nell'unico modo che può compreso da una macchina.

Pertanto, l'orientamento del diritto al futuro e a situazioni non previste, la tensione tra i termini generali nei quali è formulata la legge e i casi concreti ai quali va applicata, nonché la natura ontologicamente conflittuale delle procedure, tutto ciò fa sì che il ragionamento giuridico vada oltre il significato letterale delle norme di legge. Appare quindi

necessario uno studio dell'argomentazione giuridica (la materia viene già studiata in alcune facoltà oltreoceano pioniere nel settore dell'AI applicata al diritto), un settore nel quale vi sono non solo da applicare le tecniche di AI, ma anche da sviluppare esperienti e ricerche che possano rappresentare un avanzamento proprio di quelle tecniche.

Il tema come si può evincere è complesso e comprende numerosi profili sia teorici, sia giuridico pratici. A livello teorico, vi è per esempio, il rapporto tra linguaggio formale, visto da alcuni come lingua universale del diritto del futuro, e il perdurare del diritto espresso nei molteplici linguaggi naturali. E si deve anche tenere a mente il fatto che “la logica possa essere un utile strumento in ambito giuridico non significa sposare una logica monotonica, secondo la quale se si accettano le premesse di un ‘inferenza conclusiva, se ne debbano necessariamente accettare anche le conclusioni, mentre sono ormai prevalenti, specie in ambiti ad alta incertezza, le cosiddette logiche non monotoniche, come per esempio la *defeasible logic*, secondo la quale, posta una certa premessa, dalla quale discende una certa conseguenza, si può accettare di dover cambiare la conclusione se una nuova informazione rivela l’erroneità dell’inferenza prima ottenuta”³²

L’esempio divenuto classico nel campo dell’AI è il seguente: se io dico che Tweety è un uccello, l’ascoltatore deduce che Tweety può volare, ma se io poi lo informo che Tweety è un pinguino, l’inferenza evapora di conseguenza, come infatti dev’essere.

La logica è rilevante, a livello giuridico e pratico, a proposito dell’obbligo di motivazione dei provvedimenti giurisdizionali e della pubblica amministrazione, per esempio.

³² G. Sartor in *Logica e diritto*, in *Le direzioni della ricerca logica in Italia*, a cura di Hykel Hosni, Gabriele Lolli e Carlo Toffalori, Edizioni della Normale, Pisa, 2015.

La questione dei rapporti tra diritto e logica si intreccia con le tradizioni giuridiche e con gli orientamenti presenti nel dibattito sull'AI.

Quanto alle tradizioni giuridiche sia sufficiente ricordare i due principali orientamenti, di tipo razionalistico e di tipo pragmatico o storico, i cui massimo esponenti sono Sir Edward Coke, secondo il quale “reason is the life of the law, nay the common law itself is nothing else but reason”, e Oliver Wendell Holmes Jr. , che apre il suo volume *The Common Law*, con l'affermazione provocatoria “the life of the law has not been logic”. Inglese il primo, americano il secondo, in pieno Seicento il primo e Ottocento il secondo, sono diventati l'emblema dei due principali orientamenti circa il ruolo della razionalità del diritto: orientamenti che segnano un discrimine che, per certi versi, è più importate di quella tra *common law* e *civil law*.

Sul fronte tecnologico va ricordato il ruolo dell'approccio logicista nelle famiglie dell'AI.

Nell'ambito della ricerca sull'AI sono presenti tre principali orientamenti: un'impostazione logicista e una non-logicista. La seconda possiede al suo interno un simbolico, ma non logicista, e altri approcci connessionisti e neurocomputazionali. L'impostazione logicista dell'AI, nel contesto della creazione di un agente intelligente, è stata così sintetizzata: “un agente intelligente riceve percezioni dal mondo esterno sotto forma di formule in alcuni sistemi logici, e deduce, sulla base di queste percezioni e della sua base di conoscenze, quali azioni dovrebbero essere eseguite per garantire gli obiettivi dell'agente.”.

Ma, se si prova ad applicare il modello di tipo logistico non solo a set limitati di dati e decisioni, ma ad ampi settori di attività cognitiva, la

complessità di calcolo diventa tale da rendere difficile da capire come potrebbe funzionare.

All'interno dell'approccio non logicista neurocomputazionale si collocano poi i ricercatori che utilizzano i network neurali artificiali (*artificial neural networks*):

The main distinction is between feed-forward and recurrent networks. In feed-forward networks like the one pictured immediately above, as their name suggests, links move information in one direction, and there are no cycles; recurrent networks allow for cycling back, and can become rather complicated. [...] The backpropagation method for training multilayered neural networks can be translated into a sequence of repeated simple arithmetic operations on a large set of numbers.

L'aspetto interessante è che questi orientamenti, per quanto presenti sin dalle origini della moderno AI e pur partendo da presupposti e impostazioni diverse l'uno dall'altro, sono sempre più spesso usati congiuntamente, in modo da essere interoperabili, ed è acquisito il concetto che, se si vogliono costruire dispositivi funzionanti in modo efficace e conveniente, bisogna usare tutti gli approcci e le tecniche disponibili. Per esempio, è questo il tipo di impostazione seguito con l'architettura di DeepQA di Watson (IBM) e da Google DeepMind AlphaGo. Quello che forse è ancora più interessante è che questo approccio con più sistemi è alla base di quello adottato nella recente ripresa della cosiddetta AI di livello umano (*human-level AI*).

Quindi la domanda che sorge adesso è: una combinazione dei tre approcci può essere usata anche dal diritto? Si può adempiere all'obbligo di motivazione se si usano i *neural networks*? L'AI può organizzare anche attività che non appaiono spiegabili in termini

razionali? L'approccio logicista tende a escluderlo, perché non è spiegabile, così come sembra escluderlo tutto quel movimento che attualmente punta sulla spiegabilità delle scelte fatte o proposte da sistemi di AI.

Ci siamo fin qui soffermati su uno degli aspetti a parere di chi scrive più complicato ed interessante del rapporto tra il diritto e l'intelligenza artificiale, ossia come immettere la logica naturalmente sottesa ai principi giuridici in un algoritmo di AI.

Nel prosieguo ci soffermeremo invece sui settori forensi che maggiormente hanno subito l'evoluzione delle tecnologie di AI.

Nel 1963 l'American Bar Association pubblica un articolo in cui, per la prima volta, si pone il quesito: " Will computers revolutionise the practice of law and the administration of justice, as they will in almost everything else?"

Nell'articolo veniva esaminata l'incidenza delle nuove tecnologie sulla società concludendo che queste avrebbero inevitabilmente rivoluzionato tutti i settori compreso ovviamente quello giurisprudenziale.

La giurisprudenza in un primo momento viene chiamata a dare ordine a tutte le nuove problematiche scaturite dal rapido sviluppo tecnologico che si intersecano sempre più frequentemente con la vita di ogni individuo.

Gli operatori del diritto si trovarono a dover esaminare e disciplinare questioni di natura totalmente nuova. Le problematiche nascono e si trasformano di pari passo con l'inarrestabile progresso scientifico. Il primo ostacolo che si oppone sulla strada di chi tenta di dettare norme sul nuovo fenomeno è quello di dover acquisire competenze informatiche, seppur basilari, indispensabili per poter capire il

funzionamento e i problemi relativi a questi nuovi sistemi, per poi poterli disciplinare in maniera ottimale.

Col passare del tempo, il giurista infatti ha iniziato sempre più ad utilizzare questi sistemi imparando così nuove modalità di esecuzione di un lavoro che storicamente tutto era fuorché informatico ed informatizzato.

2.1 Il diritto visto dall'AI: diritto e logica

Anche le università, le scuole di formazione forensi e tutto il mondo delle scuole di specializzazione orientate a formare i funzionari della Pubblica Amministrazione di domani hanno iniziato a dare importanza all'informatica creando corsi di diritto dell'informatica, diritto delle nuove tecnologie, informatica giuridica e via discorrendo, in cui si inizia ad insegnare all'operatore del diritto anche ad utilizzare gli strumenti informatici che possono fungere da ausilio al loro lavoro e soprattutto, da ultimo, si è cercato sempre di più la strada per far acquisire consapevolezza dei rischi dei mezzi informatici utilizzati oltre che delle evidenti potenzialità.

L'informatica del diritto può avere diversi ruoli nell'attività e nella cultura del giurista; in primo luogo, come accennato, può accrescere l'efficienza del suo lavoro attraverso l'utilizzo di strumenti che gli permettano di svolgerlo in tempi più brevi e, a volte, raggiungendo risultati migliori: si pensi ad un avvocato che utilizzi una grande banca dati online e a quanto sia per lui, posto che sia in grado di utilizzarla al meglio e senza difficoltà più semplice e veloce ma soprattutto sia migliore la ricerca effettuata ai fini, per esempio, della ricerca di

giurisprudenze recente della Suprema Corte di Cassazione a sostegno della sua tesi difensiva. Ma ovviamente la tecnologia per un giurista può servire ad altri mille aspetti del suo lavoro come per esempio la redazione di atti (di cui diremo ampiamente nel capitolo a seguire), accesso ad ogni tipo di documenti e non solo alla giurisprudenza e anche, per esempio, alla tenuta della contabilità.

L'utilità che può apportare l'informatica rimane molto limitata se le attività giuridiche continuano a svolgersi secondo norme pre-informatiche, una piena potenzialità dei sistemi informatici richiede una riorganizzazione delle varie attività giuridiche e del loro modo di esecuzione.

Un esempio emblematico lo possiamo vedere in relazione al processo civile telematico: affinché l'attività giudiziale possa sfruttare in maniera piena le nuove tecnologie non è sufficiente fornire agli uffici giudiziari dei calcolatori più potenti e software migliori, ma bisogna determinare dei protocolli sulla comunicazione processuale telematica, sulle forme che devono avere gli atti e realizzare un'infrastruttura tecnologica che possa garantire efficienza e sicurezza.

L'utilizzo di sistemi informatici è ormai largamente diffuso nelle professioni legali e non solo, basti pensare all'uso delle banche dati, di cui ormai nessuno pensa di poter fare a meno.

Le aree di maggiore crescita di tecnologia nella pratica forense sono le seguenti:

- 1) Ricerca legale: in particolare per quanto riguarda le tecniche di recupero delle informazioni legali e giuridiche per fornire elementi utili ai fini della decisione; in un ordinamento, come quello italiano, che vede un enorme proliferare di leggi e di giurisprudenza, la creazione di banche dati in grado di darti il

testo di una legge con la relativa giurisprudenza appare fondamentale. Di queste banche dati la più famosa è quella curata dalla Corte di Cassazione.

- 2) Big Data Law: con ciò si intendono le varie tecniche di elaborazione di NLP³³ e *machine learning* che permettono di prevedere risultati e creare modelli di casi, contratti e ogni tipo di documentazione legale.
- 3) Contract automation: si può con “automazione del contratto” e con ciò si intende l’intervento dell’AI in materia contrattuale che comprende i vari ambiti di:
 - *Drafting* (redazione): cioè la redazione automatizzata dei contratti standard;
 - *Contract Management* (gestione del contratto): permette di tenere sotto controllo tutti i contratti attivi con scadenze, adempimenti e pagamenti;
 - *Contract Review* (revisione del contratto): permette la revisione automatizzata dei contratti;
 - *Extraction* (estrazione): permette l’estrazione di informazioni mirate nei documenti contrattuali.
 - *Due diligence* (una sorta di *compliance* contrattuale): software in grado di rilevare anomalie o previsioni specifiche in una determinata mole di contratti.

³³Natural Language Processing (NLP) è un campo di ricerca che abbraccia l’informatica, l’intelligenza artificiale e la linguistica, il cui scopo è quello di sviluppare algoritmi in grado di analizzare, rappresentare e quindi “comprendere” il linguaggio naturale, scritto o parlato, in maniera simile o addirittura più performante rispetto agli esseri umani.

- 4) Expertise Automation (competenza automatizzata): sistemi che aiutano l'esperto del diritto nello svolgimento della propria professione;
- 5) ODR (Online Dispute Resolution): sistemi che permettono di risolvere le controversie online. La Gran Bretagna insieme ai Paesi Bassi e la Lettonia, sono i paesi europei più all'avanguardia sotto questo profilo, a livello europeo, per la risoluzione delle controversie transfrontaliere grazie al regolamento n. 524/2013.

Questi servizi online di risoluzione delle controversie si stanno gradualmente estendendo e vengono, sempre più integrati nel processo giudiziario. L'ambito di applicazione non è più solo la controversia di modesto valore, ma anche il contenzioso in materia tributaria, quello relativo ai servizi della previdenza sociale e le cause di divorzio. Il ruolo principale di questi sistemi di ODR è chiaramente quello di contribuire all'applicazione di servizi stragiudiziali di conciliazione, mediazione e arbitrato, al punto che tali servizi vengono anche utilizzati nel corso del procedimento giudiziario sotto la supervisione del giudice che dovrà decidere l'esito della controversia.

Nonostante la grande diffusione di queste ODR non bisogna mai dimenticare che in Europa è stato adottato un quadro normativo protettivo che all'art. 22 del GDPR vincola gli Stati membri a tutelare i cittadini dando la possibilità alle persone fisiche di rifiutarsi di essere sottoposte a decisioni basate esclusivamente su un trattamento automatizzato dei loro dati.

- 6) Giustizia predittiva (di cui naturalmente si dirà ampiamente nei capitoli a seguire della presente tesi dottorale): significa, in breve, la capacità di elaborare previsioni attraverso sistemi

probabilistici che utilizzano algoritmi operanti su base logico-statistica, in grado di determinare l'esito di un processo.

Gli algoritmi cosiddetti di giustizia predittiva sono svariati oramai ed incidono sui più vari settori della giustizia e alcuni addirittura sulla prevenzione del crimine attraverso l'analisi di dati estrapolati da denunce presentate alla polizia, relativi per esempio a rapine o furti che si sono verificati ripetutamente in zone specifiche e con modalità analoghe, determinati sistemi di AI sono pertanto utilizzati per prevedere e possibilmente anche prevenire in quali luoghi di una città e a quali orari potranno più facilmente verificarsi altri reati dello stesso tipo. Una sorta di *Minority Report* dove esiste un dipartimento di pre-crimine in cui la polizia grazie alle premonizioni di individui dotato di poteri extrasensoriali e precognitivi, riesce ad impedire gli omicidi prima che glie stessi si verificano e ad arrestare i colpevoli. Nella realtà odierna un software basato su questo tipo di algoritmo è già largamente utilizzato e non solo negli USA (sono più di 9 le prefetture italiane che utilizzano questo tipo di algoritmo. Si chiama XLAW e attualmente è utilizzato in città come Napoli dove ha un'accuratezza che si aggira tra l'87% e il 93%, a Venezia tra il 92% ed il 93% e a Prato intorno al 94%³⁴.

Tale programma permette di elaborare i dati di tutti gli scippi, rapine, borseggi e furti che vengono desunti dalle denunce effettuate

³⁴ Si tratta di un software ideato e sviluppato dal 2004 dall'Ispettore Capo della Polizia di Stato Elia Lombardo e reso in comodato d'uso alla Questura di Napoli, che permette di isolare le zone di caccia e fare un'analisi previsionale dei reati predatori (vale a dire furti, rapine) onde individuare tempi, luoghi, modalità, possibili vittime nonché possibili autori targettizzandoli sulla base di caratteristiche di razza, stato sociale, culturale ed economico.

dai cittadini e dalle informazioni raccolte dalla polizia giunta poco dopo sul luogo del delitto.

Mediante questo software in uso in Italia, si è desunto che i crimini predatori urbani si commettono in luoghi particolari e devono richiedere, auspicabilmente per l'autore del reato, il minor tempo possibile per la loro realizzazione.

Il luogo in cui si verifica la suddetta tipologia di crimine ha quasi sempre due caratteristiche precise, una di tipo soggettivo ed una di tipo oggettivo: la prima richiede la presenza di "prede" o target appetibili, vale a dire soggetti fragili, donne, ragazzini, turisti che non conoscono bene i luoghi di particolare criminalità in una città e la seconda, la presenza di vie di fuga o rifugi di qualche genere nel caso di un arrivo imminente delle forze dell'ordine.

Se, insieme a queste dinamiche, che si verificano in maniera sequenziale, vengono aggiunte le caratteristiche specifiche dei vari crimini, le caratteristiche di coloro che li hanno commessi e i luoghi in cui si sono verificati appare piuttosto evidente che in una certa misura è possibile anche prevederli e prevenirli.

Attraverso l'utilizzo di un algoritmo euristico e non statistico infatti, tali dati vengono poi elaborati su una carta topografica digitale che è in grado di rappresentare alle forze dell'ordine, con un tempo di aggiornamento solitamente tarato ogni mezz'ora, le aree della città in cui è più probabile che si verifichi un reato e anche all'incirca l'ora.

Una volta terminata la sua analisi il software si esprime attraverso un'interfaccia WebApp che permette, come detto, alla polizia di usufruire in tempo reale, sui loro dispositivi elettronici queste notizie sempre aggiornate. In questo modo viene favorita non sola

la predisposizione dei presidi nel luogo e nei tempi giusti ma vengono anche messe in atto strategie operative che rendono gli aggressori meno efficaci o più vulnerabili.

Usando questi strumenti è stato possibile attuare la migliore azione d'intervento in ogni occasione cosa che ha ridotto in modo considerevole il rischio anche di operazioni pericolose per le altre persone che si trovano sul luogo in quel momento e per gli stessi poliziotti chiamati ad intervenire.

Senza soffermarsi nuovamente su percentuali che poco importano ai fini della nostra analisi, basti dire che i crimini nelle città italiane sopra menzionate che utilizzano questi algoritmi sono notevolmente diminuiti.

L'intelligenza artificiale poi, finisce per giocare un ruolo molto importante anche per quanto riguarda la prova penale sia in relazione alla sua rappresentazione che per quanto riguarda la sua valutazione. La progressiva automazione della vita di ognuno di noi permette infatti di acquisire facilmente dati che sono governati ed elaborati da strumenti intelligenti.

Si pensi, uno tra i tanti, al campo della domotica, cioè l'applicazione di strumenti informatici ed elettronici alla gestione dell'abitazione. Sono ormai ampiamente utilizzati elettrodomestici che si attivano solo quando viene raggiunta una determinata temperatura interna nella stanza permettendo così al software di fornire informazioni sulla presenza o meno di persone nell'ambiente in cui si è svolto un delitto. Si comprende certamente come tutto ciò incide sul lato processuale penalistico dei mezzi di ricerca della prova e come il diritto ed in tal caso il processo orale basato sulle prove addotte in giudizio finisca per divenire, con un uso particolarmente disinibito

degli strumenti informatici, un resoconto di prove precostituite e avverso le quali è difficile se non impossibile provare il contrario. Un altro aspetto dell'applicazione dell'intelligenza artificiale e di come questa infici sul diritto riguarda poi la valutazione della prova orale e documentale.

Molti sono i settori in cui l'AI si incontra e talvolta scontra con il diritto. Ora però ci occuperemo più nel dettaglio, nel capitolo a seguire, di come invece il settore dell'intelligenza artificiale applicato al diritto in senso stretto, sia invece di grande aiuto all'operatore del diritto.

2.1.1 I primi sistemi di intelligenza artificiale applicati al diritto

Nonostante i limiti sopra rilevati degli attuali sistemi di AI, ad oggi la loro applicazione al mondo del diritto produce risultati che sono apprezzabili sia a livello teorico che a livello pratico.

Teoricamente parlando, l'AI può contribuire alla riflessione giuridica sotto tre principali aspetti:

- 1) Lo sviluppo degli studio di AI debole si è intrecciato, eminentemente, con lo sviluppo delle scienze cognitive, i cui strumenti di analisi costituiscono oggi, un'importante risorsa al fine di decodificare la complessità della realtà del diritto e la struttura del ragionamento giuridico:
- 2) La segnalata necessità di elaborare modelli di rappresentazione della conoscenza giuridica ai fini della programmazione dei sistemi di intelligenza artificiale ha permesso di correlare i modelli della

teoria giuridica con i modelli di AI ed ha perciò favorito un'ulteriore profonda riflessione su tali modelli e sulla struttura della conoscenza di cui sono permeati così come sul contenuto medesimo di ciò che, in una data società, possa essere riconosciuto come diritto; ed infine

3) L'esigenza di formalizzazione, che presuppone in ambito informatico la costruzione di un'ontologia³⁵, cioè la delimitazione dei contenuti di un dominio di conoscenza e la definizione delle caratteristiche proprie di tali contenuti, accompagnate da un formalismo di rappresentazione, solleva l'esigenza correlativa di elaborare un'ontologia formale volta a definire nel modo il più possibile certo o concetti ed il linguaggio con cui viene rappresentata ed espressa la realtà giuridica.

Dal punto di vista pratico si possono rinvenire molteplici applicazioni di AI al diritto. Tali applicazioni vengono solitamente classificate in quattro tipologie che esamineremo nei sotto-paragrafi che seguono.

I primi di cui parliamo sono i cosiddetti sistemi di aiuto nelle decisioni: essi sono sistemi informatici, intelligenti, sviluppati soprattutto nei paesi di civil law, creati per fornire supporto alla decisione di un caso pratico attraverso l'analisi della base di conoscenza costituita dalla fonti del diritto con la conseguente individuazione di possibili soluzioni³⁶.

³⁵ Cfr. Il numero monografico "Ontologia2, in Networks, Rivista di filosofia dell'intelligenza artificiale e scienze cognitive, 6, 2006. Nell'ambito del diritto si v. G. Ajani e M. Ebers, *Uniform Terminology for European Contract Law*, Baden-Baden, 2005.

³⁶ Riprendo M. Durante, *L'intelligenza artificiale*, cit., pp.104-105.

Stiamo parlando di sistemi giuridici esperti, ideati tramite la descrizione di una normativa di settore, che individua la base di conoscenza per la formulazione automatica di possibili soluzioni al caso concreto; la base di conoscenza giuridica viene integrata da conoscenze relative a concetti giuridici e a casistiche giurisprudenziali. In questo modo il sistema giuridico esperto riproduce la struttura condizionale, basata su regole, del ragionamento deduttivo nell'ambito dell'applicazione del diritto ad un caso di specie. Il limite più volte rilevato è dato dalla difficoltà della macchina di accedere alla base di conoscenza contestuale, implicita nel caso che si esamina, così come dalla difficoltà di tradurre una valutazione prudenziale delle circostanze di fatto e degli elementi di diritto che caratterizzano il caso ed incidono sulla decisione della controversia.

Tra gli esempi di tali sistemi di aiuto alla decisione, si pensi "all'Automa Infortunistico", sistema realizzato nel 1975 presso l'Istituto per la documentazione giuridica del CNR di Firenze. Con questo sistema si intendeva fornire supporto al cittadino nell'ambito della valutazione di un incidente e della conseguente entità ai fini del risarcimento del danno: si trattava di un sistema in cui il profilo di intelligenza artificiale era minimo, in quanto si privilegiava l'aspetto automatico di mera applicazione della norma al caso di specie. Veniva così fornita una sorta di consulenza giuridica che consentiva al cittadino di orientarsi meglio circa l'effettiva opportunità di agire in giudizio e di verificare con maggiore consapevolezza l'operato dell'esperto del diritto, la cui attività non era tuttavia resa superflua, ma anzi richiesta, dall'utilizzo del sistema informatico.

“Remida” è un altro esempio di sistema utilizzato nell’ambito delle pratiche giuridiche commerciali al fine di calcolare, a seconda del titolo dell’obbligazione alla base del diritto al risarcimento, la capitalizzazione degli interessi semestrali, sia sul tasso legale che sui tassi convenzionali nonché la rivalutazione monetaria calcolata sulla base degli indici Istat sulla scala mobile ovvero in base al costo della vita. Ancora in ambito commerciale, particolarmente significativa è stata la creazione di Taxman, nelle sue varie applicazioni. Il programma, elaborato nel corso degli anni settanta del secolo scorso, era stato messo a punto allo scopo di verificare se una data riorganizzazione dell’assetto di una società commerciale dovesse o meno essere classificata all’interno di una delle forme di riorganizzazione societaria esente da imposte, secondo quanto previsto dall’ *International Revenue Code* degli USA.

Tale sistema, come detto, è stato sviluppato in più generazioni (Taxman I, Taxman II ed una terza fase basata su un modello di logica deontica), per sopperire alle difficoltà che tali sistemi incontravano nella loro applicazione al diritto. Per Taxman tale difficoltà era quella di formalizzare la nozione di titolo azionario e di costruire un meccanismo di inferenza capace di rendere conto delle trasformazioni giuridiche dei titoli azionari e dell’applicazione della fattispecie astratta al caso concreto.

I sistemi di aiuto alla decisione appena esemplificati, intendono riprodurre il ragionamento del giurista chiamato a individuare la fattispecie astratta che deve essere applicata al caso concreto e a suggerirne una soluzione: tale meccanismo di ragionamento non è in realtà unidirezionale (dalla fattispecie astratta al caso concreto o dalla fattispecie concreta alla norma) ma è guidato da una

valutazione prudenziale che mette in comunicazione le due fattispecie e tenta di stabilire un rapporto di adeguamento dell'una all'altra al fine di tener conto delle circostanze del caso nonché dell'esperienza stessa del giurista nell'apprezzamento degli elementi di diritto e di fatto che rilevano per risolvere la controversia. Per ovviare ad suddetto problema sono stati inventati sistemi basati sui casi, che assumono direttamente come base di conoscenza giuridica sistemi di decisioni, vale a dire una vera e propria casistica.

Tali sistemi, basati sui casi sono sistemi intelligenti sviluppati in particolare nei paesi di Common Law³⁷ che muovendo dall'analisi degli elementi che caratterizzano un determinato caso, rintracciano in una base di dati, cioè di conoscenze indicizzate, il precedente giurisprudenziale più adatto a supportare la soluzione del caso in esame.

Tali sistemi, ovviando alla lacuna di esperienza della macchina, riproducono artificialmente il rapporto tra presente e passato fondandolo sulla regolarità di determinate connessioni (all'evento X normalmente segue l'evento y e da cui ne consegue la decisione z). Tale modalità racchiusa e formulata in un sistema di decisioni giurisprudenziali, che fonda una vera e propria casistica, viene assunta come base di conoscenza giuridica del programma.

³⁷ D, Tiscornia. *Intelligenza artificiale e diritto*, cit. , p.137: "I sistemi basati sui precedenti giudiziari sono stati sviluppati soprattutto nei paesi anglosassoni; ovviamente la diversità dei sistemi giuridici influenza i campi di sperimentazione, per cui nei paesi di Civil Law l'interesse si è concentrato sul trattamento della legislazione e delle fonti scritte, ove il numero delle leggi è limitato ed il diritto applicato si fonda sulle decisioni giudiziarie, si lavora sul trattamento delle sentenze (cases).

Vari studiosi hanno però sottolineato che questa connessione astratta tra presente e passato può tradursi in due forme di ragionamento che sono talora assimilate ma che restano concettualmente distinte: il ragionamento basato su esempi ed il vero e proprio ragionamento basati sui casi.

Il primo costituisce il meccanismo inferenziale di quei sistemi intelligenti che adottano come base di conoscenza una serie di esempi su uno stato di cose, definendo così un dominio, cioè un campo giuridico. In sostanza il ragionamento procede per via induttiva, dal particolare al generale.

Il secondo modello di ragionamento, invece, ha a che fare con un meccanismo inferenziale di quei sistemi intelligenti che adottano come base di conoscenza un insieme di casi che individuano i precedenti da applicare ai casi analoghi nuovi: qui il ragionamento procede in via analogica, dal particolare al particolare, tramite l'estensione di una soluzione adottata in un caso ritenuto analogo alla nuova controversia.

Il meccanismo inferenziale riproduce la struttura del ragionamento analogico: si tratta della ricerca e selezione di un caso analogo sulla base della descrizione degli elementi rilevanti del caso nuovo, confrontati con le conoscenze indicizzate con cui sono descritti i casi assunti nella base di conoscenza.

Questi sistemi intelligenti basati sui casi fungono da valido supporto per l'attività dell'avvocato, in quanto vengono analizzati molti precedenti e viene consentito di individuare, nei sistemi di Civil Law la norma generale ed astratta da applicare al caso nuovo e, nei sistemi di Common Law, la decisione da assumere per risolvere la controversia nuova.

L'esempio tra i più considerevoli di quanto detto è rappresentato dal sistema *Hypo*, costruito negli anni ottanta, al fine di offrire soluzione per dirimere controversie giuridiche relative alla tutela dei software e del diritto dei contratti.

Possiamo inoltre citare alcuni altri progetti tra cui il sistema *Cato* utilizzato per educare gli studenti nell'analisi dei precedenti e nella preparazione di argomentazioni giuridiche, tramite la soluzione di casi ipotetici; il sistema *LDS* ed il sistema *Sal* utilizzati rispettivamente per determinare l'entità di un giusto risarcimento per il caso di responsabilità derivante da prodotti difettosi o per il caso di danni dovuti all'asbestosi: questi sistemi, soppesando l'insieme di elementi giuridici e fattuali da tenere in considerazione per la valutazione del caso, suggerivano ipotesi transattive per un'equa risoluzione della causa.

Abbiamo poi la terza tipologia che concerne i sistemi per il reperimento di informazioni giuridiche, i quali aiutano il giurista nella ricerca di documenti; cioè materiali che raccolgono le fonti del diritto formulate testualmente come leggi, sentenze, voci, massime ed indicizzate nella base di conoscenza del programma informatico.³⁸

I sistemi qui in esame costituiscono, con tutta evidenza, una delle più riuscite applicazioni dell'intelligenza artificiale applicata al diritto e si affiancano ai tradizionali sistemi di reperimento dei dati e delle informazione, siano esse cartacei o informatici. Il loro fondamentale scopo è quello di aiutare l'utente ad interrogare il

³⁸ Si v. H. Turtle, *Text Retrieval in the Legal World*, in "Artificial Intelligence and Law", 1995, 3, pp. 5-54.

sistema in modo tale da consentire a quest'ultimo di indicare, in risposta, il documento ricercato e comunque più idoneo a soddisfare le esigenze dell'utente. Per fare ciò il sistema deve sviluppare un meccanismo di rappresentazione della conoscenza che tenga conto della dimensione semantica della ricerca, ovverosia del contenuto concettuale delle informazioni, dei documenti, delle voci che l'utente ricerca, ed avere un meccanismo di interfaccia che consenta al sistema di interagire con l'utilizzatore in modo tale da permettergli di specificare la sua domanda ed al sistema di acquisire esperienza, vale a dire di acquisire elementi di conoscenza in relazione alla domanda proposta.

Tale peculiarità viene realizzata servendosi del Web semantico, un'estensione del web tradizionale³⁹, formato da un insieme di software che dispongono di strumenti che consentono la descrizione concettuale dei documenti, in modo tale da favorire una ricerca non meramente classificatoria delle parole chiave, ma che interroghi e selezioni i dati a livello concettuale. Tutto ciò è reso possibile dalle ontologie formali che hanno acquisito recentemente un'importanza crescente nell'ambito del diritto e costituiscono la rappresentazione concettuale di un dominio della conoscenza, realizzata con la costruzione di un vocabolario di termini (concept names), la definizione analitica dei concetti in gioco, i metadati e le tassonomie, cioè l'individuazione delle relazioni tra i concetti

³⁹ Il richiamo va doverosamente al "padre del web": T. Berners-Lee, *L'architettura del nuovo web. Dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, tr. It. Milano 2001. Per un approccio filosofico e giuridico a questi temi si v. M. Durante, *Il futuro del web: etica, diritto, decentramento*, Torino, 2007.

affinché i vincoli che definiscono le modalità di relazione concettuale siano definiti in modo univoco e consistente.

Tra i più noti sistemi di reperimento intelligente d'informazioni giuridiche non può non farsi menzione di Italgire: tale sistema, promosso già qualche anno fa oramai dal Centro di elaborazione dati della Suprema Corte di Cassazione, consente la ricerca, basata sui termini descrittivi, indicati dall'utente nella domanda rivolta al sistema, di documenti inerenti ogni settore e fonte del diritto. Esso viene poi progressivamente arricchito con funzioni di ricerca che consentono una selezione di dati a livello dei concetti, il cui percorso di reperimento costituisce una forma di acquisizione di esperienza da parte del sistema basato sull'interazione con l'utente.

Infine, l'ultima categoria di sistemi che in questa sede analizzeremo, sono i programmi che aiutano il giurista nella redazione di documenti o che producono automaticamente essi stessi documenti. Essi sono con tutta evidenza di ausilio nel produrre documenti giuridici o atti amministrativi.

Occorre rilevare però come i sistemi di aiuto non vengono allo stato classificati tra i sistemi intelligenti, in quanto si limitano talora ad aggregare esclusivamente le informazioni che occorrono per completare la redazione di un documento, quindi informazioni provenienti dal produttore stesso del documento per assicurare in tal modo l'accuratezza.

I sistemi per la redazione automatica di documenti sono invece programmi più complessi, che si avvalgono, solitamente, di tecniche di AI per l'assemblaggio e la redazione di un'ampia tipologia di documenti giuridici. Tra i più significativi esempi di tali

sistemi si può menzionare Noemi, che si occupa della produzione di documenti notarili, il sistema *Lexeditor*, che redige testi legislativi, il progetto Iri-AI concepito per integrare l'accesso a banche dati con l'elaborazione di testi ai fini della redazione assistita dei documenti giuridici e se ne potrebbero menzionare molti altri.

Quest'ultima tipologia di sistemi ha una notevole possibilità di sviluppo ed applicazione, in particolare nell'ambito della legimatica per l'elaborazione di testi che siano più facilmente traducibili nei linguaggi di programmazione, nell'ambito della attività della Pubblica Amministrazione per la produzione automatica di atti e certificati nell'ambito civilistico per la redazione assistita di atti e schemi contrattuali.

Gli esperti di intelligenza artificiale applicata al diritto stanno studiando come mettere in pratica, in un futuro non troppo lontano, sistemi che possano operare in ambiti non solo prettamente giuridici, ma anche ad essi affini, come per esempio la psicologia cognitiva, la teoria delle reti, l'etica del computer o la teoria del valore.

Due aspetti fondamentali occorre che il lettore comprenda fin da subito: il primo aspetto ha carattere prevalentemente epistemologico, legato alla possibile e forse necessaria riconcettualizzazione dei tradizionali strumenti teorici utilizzati, l'altro, di carattere più pratico, legato all'analisi delle conseguenze derivanti dalle concrete applicazioni di tali strumenti.

Innanzitutto il richiamo va al dibattito tra i fautori della tesi continuista e quelli della discontinuista, per cui i primi sostengono che le categorie tradizionali possano essere adattate a scenari futuri,

mentre i secondi sostengono che dette categorie richiedano un ripensamento in profondità. Dobbiamo infatti chiederci se è ancora possibile applicare determinate nozioni proprie della tradizione filosofica moderna degli attuali agenti artificiali. A titolo esemplificativo si pensi al concetto di autonomia di kantiana memoria, quella proprietà, cioè, della volontà di dare a se stessa una legge e se essa possa ancora essere applicata ai nuovi sistemi intelligenti oppure occorra affrontare la questione in termini nuovi. A mio giudizio ritengo piuttosto evidente la necessità di mutare concettualmente prerequisiti oramai obsoleti, in quanto tutto l'ambito della computer science e del diritto debbono sapersi concepire in modo rinnovato ed epistemologicamente adeguato alla rivoluzione informatica in corso. A ben vedere una moderna scienza giuridica non può che essere costruita se non avvalendosi delle risorse che vengono messe a disposizione dall'informatica e dall'AI in particolare, per renderla capace di dar conto del flusso di dati ed informazioni che caratterizzano la società attuale. La separazione tra diritto e tecnologia non è pertanto auspicabile né dal punto di vista teorico né da quello etico: “ il difetto di comunicazione tra forme diverse del sapere è una delle cifre negative della cultura contemporanea”⁴⁰.

Come sostenuto a sua volta da Sartor, il confronto tra tradizione teorico-giuridica e proposte dell'intelligenza artificiale può consentire di elaborare modelli di linguaggio e ragionamento che offrano non solo una migliore comprensione dei fenomeni giuridici,

⁴⁰ M. Durante, e U. Pagallo, *Manuale d'informatica giuridica*, cit.

ma anche gli strumenti per un intervento fattivo nella pratica del diritto.⁴¹

2.2 L'intelligenza artificiale vista dal diritto

L'argomento è arduo e occorre subito una nota metodologica. È evidente quanto l'intelligenza artificiale, a contatto col diritto lo plasmi, talvolta lo crei, e comunque ne ha radicalmente mutato il modo di percepirlo da parte dei cittadini, di studiarlo da parte degli studenti delle facoltà di legge e perfino di spiegarlo e "utilizzarlo" da parte degli operatori del diritto siano essi professori in materie giuridiche, avvocati, giudici ma anche si pensi ai funzionari amministrativi che quotidianamente lavorano con leggi, regolamenti e normative varie.

Qui noi, ai fine della nostra ricerca intendiamo l'intelligenza artificiale vista dal diritto in due sensi fondamentali e solo quelli: il primo riguarda il modo in cui il diritto regola o non regola, per accidente o per scelta, le varie applicazioni dell'intelligenza artificiale. Il secondo, invece, riguarda il modo in cui è il diritto stesso a essere soggetto a trasformazioni per il fatto di esistere in una realtà sociale ampiamente tecnologizzata, e in via di ulteriore tecnologizzazione, come quella attuale.

Il secondo profilo appena accennato, ossia quello del diritto tecnologizzato, vede il diritto quasi in una posizione riflessiva, d'interrogazione su sé stesso, su come esso sia cambiato, stia

⁴¹ Si v. G. Sartor, *Intelligenza artificiale e diritto*, cit. p. 134.

cambiando ed è prevedibile che cambi ancora nel prossimo futuro. E' un profilo di grande complessità teorica e pratica. Cercare di capire se i vecchi sistemi di produzione di regole (nazionali, generali e astratte) funzionino in una società intrisa di tecnologia come la nostra.

A un primo livello, si potrebbe pensare che il diritto debba cambiare collocazione: non essere più il diritto che dall'esterno regola i fenomeni sociali, ma un diritto coesenziale ai fenomeni stessi da regolare. Si pensi al caso della *privacy by design* di cui diremo meglio dopo proprio in relazione a diritto e tecnica (o per meglio dire diritto nella tecnica), nozione introdotta dal regolamento europeo sulla protezione dei dati personali (GDPR), e cioè la prescrizione secondo la quale, già in sede di progettazione, si debbono rispettare vincoli di progettazione di determinati prodotti. In tal modo la norma giuridica appare coesenziale al dispositivo tecnico e al modo in cui è progettato. Rimane, ovviamente, una differenza tra il piano tecnico e quello giuridico, ma vi è una prossimità concettuale e operativa. Qualcosa del genere accade anche negli *smart contract*, dove l'ambiente tecnologico nel quale si colloca l'accordo tra le parti non è più "ambiente" di qualcosa sorto altrove ma "linguaggio del contratto" e possibilità di una sua esecuzione e di far fronte alle fasi patologiche dello sviluppo della relazione economica (inadempimento, violazione dei patti o altro).

In altri casi il rapporto tra diritto e tecnologia assume connotazione anche diverse. Vi è però un momento in cui, tirando le somme di tutti questi cambiamenti, legati all'intelligenza artificiale, ma già avviati nel corso degli ultimi decenni per effetto dei cambiamenti sociali che sono addirittura prevedenti alle stesse tecnologie di tipo

informatico, ci si deve chiedere come il diritto stesso stia cambiando e sia cambiato nella sua intrinseca natura e nel modo in cui può essere concepito. Qui si pone un problema teorico che riguarda la concezione stessa del diritto cosa esso sia e come viva.

E' necessario chiedersi quale debba essere la risposta alla generale invocazione di regole. Quando l'UNESCO formula auspici per una regolazione internazionale, anche se di base o minima, dell'intelligenza artificiale, formula una richiesta che appartiene al primo genere di rapporto tra intelligenza artificiale e diritto, così come quando l'Unione Europea auspica regolamentazione a proposito dei dati personali che, in alcuni articoli, regola fenomeni che sono collegati alle applicazioni dell'intelligenza artificiale. Alla stessa categoria è da ascrivere anche la definizione di Zuckenberg, il quale sostiene che sia il momento che anche negli USA venga recepita la normativa europea in materia di protezione dei dati. A Singapore, per esempio, ci si è posti il problema della protezione dei dati personali in questi termini:

A model framework for AI by Singapore's Personal Data Protection Commission (PDPC), launched by Minister for Communications and Informations, emphasizes the importance of human-centricity and transparency. In essence, the individual consumer or customer should be the focus of AI design and deployment, and decisions made by AI should be explainable and fair.

Che siano necessarie regole, e possibilmente universali, per l'intelligenza artificiale è il mainstream del momento, tanto che qualcuno arriva a dire che "regolamentare l'intelligenza artificiale è più urgente che affrontare il problema del *global warming*" perché

“la dipendenza dagli algoritmi sta distruggendo la fiducia dei cittadini nei loro governi e nella democrazia”. L’algoritmo, ossia un codice costruito per risolvere un problema, risente delle visioni del mondo e dei pregiudizi di chi l’ha sviluppato, ma non si può rinunciare alla cultura degli algoritmi, visto che “anche gli umani sono tutt’altro che perfetti: i magistrati in carne e ossa hanno i loro pregiudizi e l’intelligenza artificiale diagnostica i cancro meglio di un oncologo”. E’ necessaria però la trasparenza e che chi disegna gli algoritmi sia consapevole delle conseguenze sociali del loro impiego. Anche il rapporto *10 imperativi per l’Europa nell’era dell’AI e dell’automazione*, pubblicato dal McKinsey Global Institute nell’ottobre 2017, richiama in più parti la necessità di un chiaro quadro normativo che consenta alle aziende del settore di svilupparsi anche oltre i confini nazionali dei singoli Stati europei. Per altro verso, vi sono segnali di una applicazione di sistemi sanzionatori già esistenti. Ricorda Giovanni Pitruzzella, in un articolo, tre fatti che meritano una riflessione:

la decisione della Federa Trade Commission di multare Facebook per la violazione della privacy nel caso Cambridge Analytica, la legge approvata dall’Assemblea nazionale francese in prima lettura contro la diffusione dei discorsi d’odio su internet e le conclusioni dell’avvocato generale Szpunar nella causa davanti la Corte di giustizia dell’Unione Europea sulla responsabilità di Facebook per la diffusione di contenuti diffamatori.

Gli esempi possono continuare a conferma dell’orientamento oggi prevalente nel senso di introdurre regole per l’AI. In realtà è facile a dirsi, ma meno facile a farsi, perché introdurre regole è tutto tranne

un'operazione meramente tecnica e neutrale. Molti sono gli aspetti: cosa regolare dell'AI? Cosa regolare e cosa lasciare all'evoluzione spontanea di tecnologia ad alta velocità di trasformazione? Quale tipo di regole adottare: tecniche, professionali, etiche, giuridiche? Emesse da chi?

In modo schematico possiamo dire che gli aspetti da regolare riguardano a) cos'è l'AI; b) cosa fa l'AI; c) quali effetti ha l'AI. Circa la prima domanda vanno considerati i *bias* e la correttezza, la responsabilità e la riparabilità, la trasparenza, l'interpretabilità e la spiegabilità.

Su cosa faccia l'AI emergono problemi della sicurezza, dell'interazione uomo-intelligenza artificiale, dell'uso dannoso, della privacy. Infine, circa la terza domanda, ossia gli effetti dell'AI le questioni riguardano l'automazione, la perdita di posti di lavoro e le tendenze del lavoro, l'impatto sulla democrazia, sui diritti civili e sull'interazione uomo-uomo. La questione degli usi militari delle macchine poi è particolarmente delicata e meno regolata.

Tim Berners-Lee, il padre del web così come lo conosciamo oggi, sta lavorando a un progetto (Solid), che punta esattamente a ribaltare i termini tecnici del problema e che lo porta a dire: “ We could end up with a really interesting world in which the privacy questions s turned upside down. So there you go! This is Solid. That's the idea!”⁴²

Nel dilemma se puntare su regole poste dall'esterno oppure sulla creatività dei cultori di Internet e sullo sviluppo tecnologico, Tim Berners-Lee è un chiaro esempio del secondo approccio, che

⁴² Discorso fatto al Berkman Klein Center, *The future of the decentralized web*, disponibile sul sito del centro di ricerca di Harvard.

cambia l'impianto del web, lo decentralizza radicalmente, e potrebbe così ridurre il potere di quei grandi soggetti economici che attentano ai diritti degli utenti della rete e dei sistemi di AI e "restituire le informazioni degli utenti nelle mani degli utenti".

Vi è anche chi fa notare che è troppo presto per tentare un sistema generale di regolazione dell'AI e che la trasparenza richiesta ex ante può limitare l'uso di alcune tecnologie come i neural networks, concludendo che "una saggia inattività nel porre regole è probabile che produca un miglior risultato a lungo termine rispetto a una corsa a regolare qualcosa di sconosciuto".

Nel solito dilemma "regolare-non regolare", come si inserisce poi il lavoro del ricercatore, dello studioso del diritto dell'informatica? A mio avviso, in generale e specie in questo campo, il confine tra ricerca e applicazioni non è sempre facile da tracciare. Si pensi, tra i tanti esempi possibili, alla ricerca sulle reazioni emotive dei consumatori rispetto ai modi di presentazione di alcuni prodotti nei negozi. Tuttavia, non si può non ribadire che la ricerca nel campo dell'intelligenza artificiale gode della protezione della libertà di ricerca scientifica, apprestata sia dalle costituzioni, in particolare quelle europee e italiana, sia da documenti internazionali, come il Patto internazionale sui diritti economici, sociali e culturali, siglato a New York nel 1966.

Quindi, tornando al dilemma relativo a quali tipi di regole il diritto deve adottare rispetto alle evoluzioni dell'intelligenza artificiale, la scelta appare strettamente collegata alla natura dell'oggetto tecnico della regolazione e ha numerose e complesse sfumature. Come è stato osservato in un intervento pubblicato sulla rivista "Proceedings of the IEEE" (la rivista dell'Institute of Electrical and

Electronic Engineers) la velocità dei cambiamenti tecnologici è il problema cruciale, tanto che “se un sistema normativo riuscisse in qualche modo a introdurre nuove regolamentazioni per una tecnologia emergente, esse saranno probabilmente obsolete prima che l’inchiostro si asciughi sull’atto di promulgazione [...] Questo porta i sistemi regolatori tradizionali a produrre nessuna regolamentazione o una cattiva regolamentazione”⁴³.

Propongono, quindi, un sistema agile che comprende, tra l’altro, l’introduzione di strumenti di *soft law* che vengono sperimentati e poi, eventualmente, trasfusi in regole giuridiche vincolanti.

Urs Gasser, giurista e direttore del Berkman Klein Center, in un articolo scritto con Carolyn Schmitt, ricercatrice del medesimo centro, parte dalla considerazione che i sistemi di AI e le tecnologie basate sull’AI sono regolati da un insieme sempre più complesso di norme giuridiche, etiche, sociali e di altri tipi, che emergono da varie fonti e contribuiscono a ciò che potrebbe essere descritto un “patchwork di norme”. Non manca chi propone il vecchio modello dei comitati etici. Molti sono poi gli studiosi che non nascondo i rischi, anche gravi, che queste tecnologie possono avere, come, per esempio, la creazione di un algoritmo in grado di registrare ogni nostra azione, di sorvegliare ogni istante della giornata, e quindi, proprio per evitare queste distorsioni, si propone la creazione di comitati etici per studiare norme e leggi a tutela delle libertà degli individui e nello stesso tempo tracciare limiti invalicabili nell’interesse generale dell’umanità.

A cavallo tra le norme tecniche e quelle etiche si colloca la ricerca sull’etica delle macchine (*machine ethics*), un campo relativamente

⁴³ Wallant e Marchant 2019. Si veda anche Winfield- Others 2019.

nuovo al quale la rivista “Proceedings of the IEEE” ha dedicato un numero monografico uscito nel marzo del 2019.

La *machine ethics* affronta il tema di come i sistemi automatici possono essere impregnati di valori etici:

Ethical autonomous systems are needed because, inevitably, nearfuture systems are moral agents; consider driveless cars, or medical diagnosis AIs, both of which will need to make choices with ethical consequences.

This special issue includes papers that describe both implicit ethical agents, that is machines designed to avoid unethical outcomes, and explicit ethical agents: machines which either encode or learn ethics and determine actions based on those ethics. [...] an ethical machine is guided by an ethical rule, or set of rules, in deciding how to act in a given situation. It follows that we are concerned here with autonomous machines: either software AIs, or their physically embodied counterparty, robots, which determine how to respond to input without direct human control.

Naturalmente vi sono implicazioni regolatorie anche per l’etica delle macchine, in quanto una governance etica è necessaria al fine di sviluppare “standards and processes that allow us to transparently and robustly assure the safety of ethical autonomous systems and hence build public trust and confidence”.

E’ interessante notare come il dibattito sull’etica delle macchine richiami quella prossimità tra dispositivo tecnico e regola di cui abbiamo detto sopra. Anche nell’etica delle macchine il dispositivo tecnologico non solo “ospita” una regola etica, ma fa vivere nella macchina un elemento disomogeneo rispetto alla mera tecnicità, che può avere la forza di trasformare la macchina stessa.

Si arriva quindi al dilemma se optare per regole giuridiche o regole etiche.

L'Organizzazione per la cooperazione e lo sviluppo economico, a proposito della regolamentazione delle tecnologie non usa il termine "ethics" ma "human rights", perché ritiene che, se è vero che sia l'etica sia i diritti umani proteggono lo stesso insieme di valori, è anche vero che per i diritti umani già esiste un 'infrastruttura giuridica ed è possibile azionare tutele giuridiche per ottenere rimedi effettivi. Per l'etica, invece, non vi è accordo su quale sia l'etica di riferimento (in presenza di molti tipi di etica si intende) e quale sia l'autorità che abbia titolo a prendere decisioni.⁴⁴

La questione dei diritti umani, in relazione all'etica, è alquanto elusiva e ricca di pezzi mancanti. Come poi avremo ancora modo di discutere, i diritti umani vanno a collocarsi in quella fascia alta dei diritti che in alcuni ordinamenti, come quello italiano, possono avere la forza di diritti costituzionali, e in altri quella di meri obblighi morali, che possono essere indistinguibili da posizioni etiche.

Per esempio, nell'Unione Europea i diritti umani sono stati istituzionalizzati nella *Carta dei diritti fondamentali dell'Unione Europea* che raccoglie e compendia una serie di altri strumenti normativi europei, per i quali può parlarsi di giuridicizzazione dei diritti umani. Al contrario, il quadro dei diritti umani nei Paesi extra-UE si presenta molto più problematico su fronti che riguardano diritti dei minori, tortura e pena di morte, violenze

⁴⁴ Posizione espressa dal rappresentante dell'OCSE alla "Conference On the Ethics of Science, Technology and Sustainable Development" nel corso dell' undicesima Session of the World Commission On the Ethics of Scientific Knowledge and Technology of Unesco, Bangkok, Thailandia.

perpetrate nei confronti di donne e ragazze, assistenza e monitoraggio delle elezioni da parte dell'UE, campi nei quali l'UE si impegna per l'osservanza del diritto internazionale umanitario.⁴⁵ L'Unione Europea si è mossa con la una Comunicazione della Commissione al Parlamento Europeo del 25.04.2018, che lancia un'iniziativa europea in tema di AI che ha tra i suoi tre obiettivi, accanto ai temi più strettamente economici e sociali, quello di “assicurare un quadro etico e giuridico adeguato, basato sui valori dell'Unione Europea e coerente con la Carta dei diritti fondamentali dell'UE. Ciò comprende futuri orientamenti sulle norme esistenti riguardanti la responsabilità per danno da prodotti difettosi, l'analisi dettagliata delle sfide emergenti e la collaborazione con i portatori di interessi, attraverso l'Alleanza europea per l'AI, per lo sviluppo di linee guida etiche riguardo all'AI.

Nella primavera del 2019, lo High-Level Expert Group on Artificial Intelligence dell'Unione europea ha reso pubblica la versione finale delle *Linee etiche per un'intelligenza artificiale affidabile (Ethics Guidelines For Trustworthy AI)*.

Il documento ovviamente meriterebbe un'analisi ben più approfondita per la quale non vi è spazio in questa sede. Mi limiterò, pertanto, ad alcune notazioni con un occhio particolare ai rapporti tra etica e diritto. Le linee guida hanno tre basi: una giuridica, una etica ed una tecnica.

Testualmente affermano:

⁴⁵ Per informazione ulteriormente dettagliate si veda il sito EUR-lex, nella sezione *Diritti umani*.

L'AI affidabile ha tre componenti, che dovrebbero essere soddisfatte durante l'intero ciclo di vita del sistema:

1. Dovrebbe essere legale, cioè conforme a tutte le leggi e i regolamenti applicabili;
2. Dovrebbe essere etica, garantendo l'adesione a principi e valori etici;
3. Dovrebbe essere robusta, sia dal punto di vista tecnico che sociale, poiché, anche con buone intenzioni, l'AI i sistemi possono causare danni involontari.

L'aspetto etico prende forma in quattro principi, che costituiscono "imperativi etici" ai quali gli operatori devono sforzarsi di aderire: i) rispetto per l'autonomia umana; ii) la prevenzione dei danni; iii) la giustizia; iv) la spiegabilità. Fatto salvo per quest'ultimo principio, che è strettamente legato all'attualità dei sistemi di AI, non è necessario essere esperti di storia della bioetica per sentire l'eco dei principi del *Belmont Report* (USA, 1974).

Qui si apre un'interessante questione circa il modo in cui il documento stabilisce i rapporti tra diritto ed etica. Le linee guida, in modo dichiarato e fedelmente al titolo, non affrontano i problemi giuridici, dando tuttavia alla dimensione giuridica un'importanza di primo ordine. Precisano, infatti, i redattori, in modo perfino meticoloso, il quadro giuridico generale europeo, dai documenti di rango sostanzialmente costituzionale, fino a norme ordinarie di tipo strettamente regolativo:

I sistemi di AI non operano in un mondo senza leggi. A livello europeo, nazionale e internazionale un corpus normativo giuridicamente vincolante è

già in vigore o è pertinente per lo sviluppo la distribuzione e l'utilizzo dei sistemi di AI. Le fonti giuridiche pertinenti sono, a titolo esemplificativo, il diritto primario dell'UE (ossia i trattati dell'Unione Europea e la sua Carta dei diritti fondamentali), il diritto derivato dell'UE (ad esempio il regolamento generale sulla protezione dei dati, le direttive antidiscriminazione, la direttiva macchine, la direttiva sulla responsabilità dei prodotti, il regolamento sulla libera circolazione dei dati non personali, il diritto dei consumatori e le direttive in materia di salute e sicurezza sul lavoro), ma anche i trattati ONU sui diritti umani e le convenzioni del Consiglio D'Europa (come la Convenzione europea dei diritti dell'uomo) e numerose leggi degli stati membri dell'UE. Oltre alle norme applicabili orizzontalmente, esistono varie norme specifiche per settore applicabili a particolari applicazioni di AI (ad esempio il regolamento sui dispositivi medici nel settore sanitario).

In un punto successivo viene poi precisato che il quadro giuridico europeo sopra descritto ha alcune norme che sono giuridicamente vincolanti, mentre altre riconoscono diritti a persone "in virtù del loro status morale di essere umani, indipendentemente dalla loro forza giuridica". A questo punto viene precisato che i diritti fondamentali, se intesi come diritti legalmente riconosciuti, ricadono nel primo componente di un'AI affidabile, se, invece, intesi come "diritti di ciascuno, radicati nello status morale intrinseco degli esseri umani", essi rientrano anche nel secondo componente che riguarda norme etiche che non sono necessariamente giuridicamente vincolanti.

Da ciò si desume quanto il quadro normativa sia davvero poco chiaro. Se si vuol dire che tra i diritti fondamentali ve ne sono alcuni non riconosciuti come *legal rights* e che essi possono essere riconosciuti su base etica, il tutto è chiaro: fino ad un certo punto arriva il diritto e oltre

può esservi l'etica, in attesa che il diritto arrivi, quando e se arriva. Ma, se si colloca la distinzione sul concetto di *human being*, tutto diventa meno chiaro, perché l'ordinamento giuridico può ben riconoscere diritti sulla base della sola qualità di essere umano, come accade in tutti i casi in cui la disposizione usa (o lascia intendere) il termine "individuo" e non "cittadino" (caso emblematico il diritto alla salute che la Costituzione italiana, all'art. 32, riconosce esplicitamente come diritto fondamentale dell'individuo). Ma, per altro verso, è ben possibile che i cittadini e/o *human beings* si vedano negati diritti che, almeno secondo una certa visione etica, sono fondamentali (si pensi ai diritti per le coppie costituite da persone dello stesso sesso).

In generale sembra aleggiare una visione positivista dell'etica, che risulta inscritta nel perimetro del diritto e che si presenta come oggettivamente e immediatamente comprensibile, perdendo di vista il patrimonio di elaborazioni sulla pluralità delle etiche, che probabilmente è la maggiore acquisizione di decenni di dibattito bioetico. La preoccupazione sembra essere quella di definire un'etica pubblica, da affiancare strettamente al diritto positivo come una sorta di *soft law*.

Regole giuridiche europee, (ma che con una forza espansiva direi quasi senza eguali per alte normative europee), sono quelle contenute nel GDPR 679/2016 (diventato legge in tutti i Paesi dell'Unione Europea il 25 maggio 2018) e che toccano aspetti essenziali delle applicazioni di AI. Tra questi basti citare l'art. 22, che stabilisce il diritto dell'interessato "di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che indica in modo analogo significativamente sulla sua persona" e l'art. 24 che, sotto la

rubrica “responsabilità del titolare del trattamento”, introduce il principio di *accountability*, che responsabilizza maggiormente le aziende al loro interni:

L’art. 24 cita testualmente al suo primo paragrafo: “Tenuto conto della natura, dell’ambito di applicazione, del contesto e delle finalità del trattamento, nonché dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche, il titolare del trattamento mette in atto misure tecniche e organizzative adeguate per garantire, ed essere in grado di dimostrare, che il trattamento è effettuato conformemente al presente regolamento”.

Abbiamo poi l’art. 25, che introduce il ben noto principio della *privacy by design*, di cui diremo tra poco.

A questo punto occorre chiedersi quali sono gli organi deputati a decidere se e come introdurre nuove regole e quale natura dare alle stesse.

Le linee guida dell’High- Level Expert Group sono state criticate proprio sul punto del chi decide sulle regole per l’AI. La questione di che abbia la legittimazione e la forza per definire e dettare queste regole si è posta nel 2019, proprio in coincidenza con la diffusa richiesta di regole.

Ho fatto scalpore l’articolo di Mark Zuckerberg, fondatore e capo di Facebook, pubblicato il 30 marzo 2019 sul Washington Post, con il tono perentorio di chi individua un problema, ossia la necessità di nuove regole e indica anche la strada da seguire (come il lettore potrà notare soggetto che avrebbe giusto qualche conflitto di interesse nel “consigliare al legislatore come muoversi”), ma sta di fatti che affermò: “The internet needs new rules. Let’s start in these four areas”.

L'intervento, nel caso di specie, affrontava una pluralità di questioni che, oltre l'intelligenza artificiale, riguardano anche la rete.

Chiede, sostanzialmente, un ruolo più attivo dei governi e dei regolatori e, in primo luogo, sollecita un approccio più standardizzato ai contenuti violenti e offensivi, che vada oltre iniziative di singole entità, come il corpo indipendente creato da Facebook per porre reclami contro l'operato della società e la pubblicazione di un rapporto trimestrale sulla trasparenza circa l'effettiva rimozione di quei contenuti offensivi. In secondo luogo, Zuckerberg afferma che una legislazione appropriata è importante per proteggere le elezioni che verranno (qui ovviamente il riferimento va al supposta intrusione della Russia che attraverso un mare di fake news in rete avrebbe condotto alla vittoria Donald Trump nel 2017). In terzi luogo, la protezione della privacy e dei dati personali dovrebbe essere armonizzata a livello globale, e propone che venga adottata in tutto il mondo la legislazione europea regolativa della privacy, ossia il GDPR 679/2016. In tal modo si eviterebbe anche la frammentazione di internet a livello globale. Sono necessarie regole globali chiare su quando un'informazione può essere usata nel pubblico interesse e su come regolare nuove tecnologie come l'intelligenza artificiale. Sollecita infine regole conformi al principio di portabilità dei dati, nel momento in cui un cliente passa da un fornitore agli altri.

Tanto attivismo e tanto zelo hanno insospettito alcuni commentatori, che fanno notare che sarebbe il caso che ognuno facesse la propria parte e, quindi, che i legislatori facessero i legislatori, perché è evidente che se il legislatore chiede al capo di Facebook come regolare la materia è come “se chiedesse alla volpe come meglio custodire il pollaio”⁴⁶.

⁴⁶ Questa è l'opinione di Corynne McSherry, legal diret of the Electronic Frontier Foudation, in CONDLIFFE, 2019.

In un articolo pubblicato l'1 maggio 2019 sulla rivista "Nature", Yochai Benkler, professore di diritto e co-direttore del Berkman Klein Center for Internet & Society di Harvard, critica l'influenza che le industrie tecnologiche stanno esercitando per indirizzare la ricerca e la creazione di regole conformi ai loro interessi e chiede alla società di respingere questo tentativo. Benkler si riferisce esattamente alle linee guida europee, di cui si è detto sopra, e alla composizione del gruppo di esperti di alto livello, ma anche al caso del Comitato etico formato da Google e dissolto in una settimana, o al fatto che Facebook ha investito 7,5 milioni di dollari americani in un centro di etica e AI alla Technical University di Monaco di Baviera.

In effetti molti sono gli esperti che hanno creato linee guida e che provengono o erano molto in linea con gli interessi del settore tecnologico da Microsoft a Apple, passando per Facebook.

La proposta di alcuni studiosi è stata quella formulata in sede di IEEE, di creare un gruppo internazionale che operi come una sorta di direttore d'orchestra nell'elaborare e ipotizzare sistemi di regolazione dell'AI, ma allo stato questo gruppo ancora non esiste.

2.2.1 Il caso della *privacy by design*

Uno dei settori in cui più si può assistere alla ostinata resilienza del diritto di fronte ai cambiamenti tecnologici in corso è senza dubbio il settore della privacy e della protezione dei dati personali.

Il Regolamento Europeo sulla protezione dei dati personali infatti, dopo anni di ricerca e studio effettuati da esperti tecnici in seno alla Commissione ed al Consiglio europei, ha messo a punto ed elevato a

livello normativo il principio della *privacy by design*. Tale principio, come noto, obbliga il titolare, nel momento in cui determina le finalità e le modalità del trattamento, tenuto conto dei rischi di trattamento, dello stato dell'arte delle tecnologie e dell'ambito di applicazione, di mettere in atto misure tecniche e organizzative adeguate per integrare nel trattamento le necessarie garanzie per tutelare i diritti degli interessati. "Integrare le tutele nel trattamento". E' questo il passaggio che appare come la chiave, al quale possono essere attribuiti due significati: o trattare i dati in modo da minimizzare l'uso per perseguire una finalità, al punto da non ritenersi più necessario un trattamento di dati personali; ovvero trattarli in modo da incrementare la sicurezza, rafforzando la confidenzialità del dato. Sono due distinte forme di tutela integrate nel trattamento, che richiedono un approfondimento, in ragione delle conseguenze significativamente diverse a cui la loro applicazione conduce.

Dopo aver spiegato cos'è tecnicamente la *privacy by design*, occorre ora tornare al ragionamento che abbiamo cominciato all'inizio di questo capitolo, ossia come il diritto vede la tecnologia e quindi come il diritto può, attraverso la *privacy by design* regolare la tecnologia quanti e quali modi ha per farlo.

Come noto il diritto alla privacy nasce negli Stati Uniti. Partiti inizialmente dal tema della intercettazioni, del video-voyerismo, delle leggi sui paparazzi e del diritto di cronaca che, su scala prevalentemente statale, più che federale, regolano l'uso di fotografie, video e audio potenzialmente intrusivi per la privacy. In ragione dell'approccio, o vocazione settoriale del modello statunitense della privacy, il problema è stato inizialmente affrontato con varie leggi, sia a livello federale ma

soprattutto statale, che vietano le fotografie fatte per esempio dai droni senza il consenso degli interessati nei luoghi ritenuti di privata dimora. In Europa, a vocazione garantista, ogni qualvolta qualcuno, sia esso umano o agente artificiale, tratta dei dati, insieme con la loro raccolta ed eventuale uso, deve sottostare ad alcune regole che sono ben fissate dal legislatore e ben più stringenti di quelle statunitensi.

In un recente scritto su *Laws on Robots, Laws by Robots, Laws in Robots*⁴⁷, gli autori Ronald Leenes e Federica Lucivero, suggeriscono di cogliere l'intento regolativo del diritto nell'ambito della robotica, secondo una quadripartizione.

In primo luogo, si pensi alla disciplina dei progettisti e costruttori di robot attraverso la legge, come nel caso degli standard di sicurezza ISO o le norme sulla responsabilità civile e penale per produttori e utenti di robot; in secondo luogo, il richiamo va alla regolazione del comportamento degli utenti e/o operatori dei robot, tramite il design di questi ultimi, vale a dire progettando queste macchine in modo tale che non sia consentito alcun comportamento illecito degli essere umani; in terzo luogo, si può pensare alla disciplina legale degli effetti dei comportamenti robotici per il tramite delle leggi approntate dal legislatore; è il caso della contrattualistica e negoziazione a mezzo di agenti software; infine in quarto e ultimo luogo, la legge può mirare alla disciplina del comportamento robotico tramite il suo design, ossia immettendo direttamente i dettami della legge nel software dell'agente robotico.

⁴⁷ Si v. R. Leenes e F. Lucivero, *Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design*, in "Law, Innovation and Technology", 2014, vol. 6(2), 193-220.

Lasciando qui da parte i metodi tradizionali di regolamentazione giuridica il punto è capire l'intento regolativo della legge tramite il principio di cui al titolo di questo paragrafo per gli agenti artificiali.

Coniata alla fine degli anni novanta dal Commissario per la privacy dell'Ontario, in Canada, Ann Cavoukian, la formula "privacy by design" sarebbe stata ripresa poco più tardi in alcuni testi ufficiali, tra cui il *Privacy Design Principles for an Integrated Justice System*, presentato, nell'aprile del 2000, dal Dipartimento della giustizia nordamericano e, appunto dal Commissario per la privacy dell'Ontario. Più recentemente, la formula è stata significativamente ripresa anche dalle linee guida del più volte richiamato studio "Robolaw", secondo il quale, in sostanza, "l'emergente settore della privacy by design può servire a rendere e mantenere il comportamento degli agenti artificiali, qualora fungano da titolari, nei limiti delle leggi in materia di protezione dei dati"⁴⁸.

L'obiettivo, in linea di massima, è che una macchina sia in grado di processare e intendere l'informazione giuridica occorrente per le sue operazioni, tramite la modellizzazione di nozioni tradizionalmente impiegate dai giuristi, come ciò che è "dovuto", "proibito", o "permesso".

La formalizzazione delle norme, diritti e doveri, nei diversi settori del diritto civile, penale o amministrativo, richiede di distinguere i requisiti dai vincoli ontologici. Nel primo caso, si tratta di quella parte dell'ontologia che, tramite l'uso di tassonomie, contiene tutti i concetti rilevanti del settore. Nel secondo caso, i vincoli ontologici corrispondono all'insieme delle regole e dei limiti con cui quegli stessi concetti vanno rappresentati. "Un sistema esperto deve di qui

⁴⁸ Robolaw, *Guidelines on Regulating Robotics*, p.19.

processare l'informazione giuridica e risolvere i problemi di applicazione della normativa vigente, mediante la concettualizzazione di classi, relazioni, proprietà e istanze che appartengono a quello specifico settore. Tanto più quest'ultimo si presta a un approccio "dall'alto verso il basso", ossia, tramite l'applicazione dei concetti normativi di base, quali "validità", "obbligazione", "proibizione", ecc., tanto più l'automazione del ragionamento giuridico sarà possibile con la quantificazione dei dati"⁴⁹.

Nel caso degli agenti artificiali o dei droni, pertanto, alcune disposizioni in tema di sicurezza dei dati, ad esempio, possono implementarsi nel software e interfaccia della medesima macchina tramite dispositivi di criptazione e sistemi di controllo per l'accesso dei dati. Del pari, riprendendo le linee guida di Robolaw, alcuni requisiti come il consenso informato possono essere implementati nella progettazione del sistema, ad esempio tramite l'interazione con gli utenti mediante schermi e dispositivi di caricamento (input). In fondo, questo è ciò che accade già con alcuni sistemi operativi in commercio, come nel caso di Android, che richiedono il consenso dell'utente ogni qualvolta un'applicazione intenda fare uso dei loro dati personali. Inoltre, possiamo anche pensare ai droni configurati in modo tale che l'insieme dei dati da raccogliere e processate sia ridotto al minimo, in omaggio al principio della finalità del trattamento.

Per cui, da un lato, sempre secondo le linee guida di Robolaw "come corollario della valutazione di impatto, dovrebbe implementarsi un meccanismo di controllo che accerti quando le tecnologie sono costituite nel modo più consono per la tutela della privacy e

⁴⁹ U. Pagallo, *Il diritto nell'età dell'informazione*, p. 26.

compatibilmente con altre istanze quali il bisogno di informazioni o la sicurezza.

D'altro canto, "l'adozione di misure di sicurezza lato privacy aggiornate non dovrebbe essere considerata una semplice scelta dell'utente, bensì uno specifico obbligo giuridico. Appare chiaro che è improbabile che il trattamento illecito dei dati sia considerato una responsabilità dei costruttori di robot, ma piuttosto dell'utente, che può essere ritenuto "custode dei dati personali".

Naturalmente l'applicazione dei concetti giuridici dipende, molto spesso, dal contesto in cui tali concetti devono essere interpretati. Un esempio è dato proprio dal settore della protezione dei dati personali, in cui già solo quest'ultima nozione, al pari di altre, come "misure di sicurezza" o "controllore dei dati", non possono essere del tutto decontestualizzate e, quindi, sorge il sospetto che ciò che può diventare arduo decodificare a un umano esperto nel settore giuridico, a maggior ragione diventi un'impresa impossibile da compiere per un drone.

In aggiunta, sembra discutibile la tesi di Robolaw che sia sempre e solo l'utente ad essere il "custode" dei dati personali e, per ciò stesso, responsabile per ogni possibile uso o trattamento illecito di dati personali. In fondo, basterebbe pensare a quanto accade già al giorno d'oggi con le questioni relative alla connettività su internet, con le applicazioni per i servizi mobili, come i nostri smartphone, oppure con i sensori. Ci sono molte circostanze in cui il trattamento illecito di dati personali ben può dipendere dal modo in cui il drone è stato disegnato o costruito, dalla negligenza del fornitore di connettività o di coloro i quali sviluppano determinati applicativi. Il trattamento illecito dei dati personali potrebbe, per esempio, farsi risalire al cattivo funzionamento del drone.

Inoltre, un punto fondamentale da evidenziare è che permangono, tra i diversi sistemi normativi, differenze sostanziali, come nel caso di USA e UE, oppure differenze anche nel modo di interpretare alcune nozioni chiave di questi sistemi, come per lo stesso concetto di *privacy by design* nell'ordinamento dell'Unione Europea.

Nel primo caso, potremmo pensare proprio di ricorrere alle vie della tecno-regolazione giuridica per superare i diversi approcci alla privacy e alla protezione dei dati di molteplici ordinamenti. Si presti attenzione allo scenario proposto dal progetto nato grazie al lavoro di alcuni ricercatori del Centro Nexa for Internet & Society del Politecnico di Torino sui “beni comuni ubiqui” (*ubiquitous commons*)⁵⁰.

L'idea base concerne il modo in cui gli utenti possono modellare il sistema di collegamento per la navigazione del proprio robot. Dati e informazioni possono essere criptati prima ancora di essere smistati presso i server informatici del fornitore di servizi, ora dipendendo dalle scelte del titolare dei dati, ora invece sulla base delle misure di sicurezza stabilite, poniamo, dal legislatore.

Questo sistema di collegamento per la navigazione del drone prevede la possibilità di condividere con altri, ossia “l'utente B” dello schema, le chiavi d'accesso per la decodifica dei dati su una rete peer-to-peer esterna al fornitore di servizi.⁵¹

Dati e informazioni, ospitati nei server informatici del fornitore di servizi, diventano utilizzabili soltanto qualora i soggetti o enti che mirano a far uso di tali dati e informazioni, sono identificati. Dati e

⁵⁰ S. Iaconesi, *Ubiquitous Commons: A Browser Plugin to Enable User-generated Peer-to-peer-enacted Licensing Models for Online Data*, 2014.

⁵¹ Per un'introduzione a questo tipo di rete si v. U. Pagallo e G. Ruffo, *Peer-to-peer*, in M. Durante e U. Pagallo, *Manuale di informatica giuridica e diritto delle nuove tecnologie*, op. cit. pp.487-508.

informazioni, in altri termini, possono essere letti dall'utente B a condizione che B sia autorizzato in ragione del contesto e dell'uso dei dati che s'intende fare. D'altro canto, la banca dati del fornitore di servizi appare come uno spazio di raccolta di dati criptati accessibili solo alle condizioni e per i soggetti definiti dalla chiave di decodifica disponibile sulla rete peer-to peer.

Naturalmente, altri schemi in tema di sicurezza e protezione dati possono immaginarsi accanto, o in alternativa, a quello ora esposto.

Sebbene il principio della *privacy by design* vada ricondotto a uno dei settori più affascinanti e dinamici dell'odierna ricerca nel settore della protezione dei dati, rimane il fatto che anche in quegli ordinamenti, come quello dell'UE, in cui è stato adottato, persistono alcune incertezze cruciali circa il modo d'intendere il principio stesso.

Pertanto alcuni studiosi suggeriscono di intendere il diritto come "meta-tecnologia", ossia ricondurre l'intento tecno-regolativo del diritto alla tradizionale condizione di "tecnica di controllo sociale che si attua mediante la minaccia di misure coercitive".⁵²

Nel caso in cui l'oggetto di questa regolazione sia la tecnologia, possiamo dunque pensare al diritto come a una sorta di meta-tecnica o, meglio ancora, di una meta-tecnologia.⁵³

Lasciando per ora da parte il modo in cui dobbiamo cogliere il nesso tra fattispecie astratta ("A"), e conseguenza giuridica ("B"), vale a dire ora sul piano tradizionale del dover essere, ora su quello odierno dell'automatismo tecnico-giuridico, bisogna però distinguere

⁵² Basto solo accennare a H. Kelsen, *Dottrine pure del diritto*, tr. It. Einaudi, Torino, 1952.

⁵³ Così in U. Pagallo, *The Laws of Robots: Contracts, Crimes and Torts*, Dordrecht, Springer, 2013.

ulteriormente le finalità che il diritto può avere in questo modo.⁵⁴ Alcuni studiosi, come Bert-Jaap Koops, propongono quattro diversi obiettivi, vale a dire il raggiungimento di particolari effetti come la sicurezza nel trattamento dei dati; l'equivalenza funzionale tra attività online e offline; la non discriminazione tra tecnologie con effetti equivalenti e, infine, il fatto che, almeno in linea di principio, la legge non dovrebbe arrestare il progresso tecnologico, oppure richiedere di essere rivista continuamente per via di questo stesso progresso.⁵⁵

Altri studioso, come ad esempio Chris Reed, suggeriscono a loro volta di cogliere la natura meta-tecnologica del diritto secondo una tripartizione. In particolare, occorrerebbe in primo luogo vagliare l'indifferenza tecnologica della legge, ossia il fatto che la finalità perseguita dall'ordinamento valga al di là della tecnologia in oggetto, come ad esempio nel caso dell'autorizzazione a comunicare le proprie opere al pubblico nell'ambito del diritto d'autore. In secondo luogo, il richiamo va alla neutralità d'implementazione e, cioè, al fatto che pur essendo l'oggetto della disciplina giuridica specifico di una data tecnologia, tuttavia, la legge sia indifferente alle sue possibili implementazioni (come nel caso della firma elettronica). Infine, il riferimento va riportato alla potenziale neutralità della legge, nel senso che quest'ultima specifica i requisiti di legittimità delle applicazioni tecnologiche ma poi, come nel caso delle firme biometriche, consente che queste ultime possano essere modificate per ottemperare alle finalità disposte dall'ordinamento.

⁵⁴ M. Durante, *Normativity, Constructionism, and Constraining Affordances*, in "Ethics and Politics", 2011, XIII (2): 180-200.

⁵⁵ Si v. B-J. Koops, *Should ICT Regulation Be Technology-neutral?*, in *Starting Points for ICT Regulation: Deconstructing Prevalent Policy One-liners*, pp. 77-108, a cura di B-J. Koops e al. TMC Asser, The Hague, 2006.

All'interno delle diverse finalità che il diritto può perseguire attraverso la disciplina della ricerca e del progresso tecnologico, trova dunque spazio quell'ulteriore carattere tecno-regolativo del diritto. Ad esempio, la prima finalità della meta-tecnologia del diritto nella versione di Koops, vale a dire il raggiungimento di particolari effetti come la sicurezza nel trattamento dei dati, può essere raggiunto attraverso l'immissione di limiti e condizioni giuridiche all'interno stesso del disegno, o progetto tecnologico. Naturalmente, nel rapporto tra il diritto come meta-tecnologia e l'intento tecno-regolativo del diritto, il piano d'intervento della legge può mutare radicalmente. Per restare all'esempio della sicurezza dei dati, da un lato, tanto il diritto come meta-tecnologia quanto la natura tecno-regolativa del diritto possono registrarsi sul piano tradizionale, "kelseniano", del dover essere delle cose, ossia come una prescrizione del legislatore, alla quale, in caso di violazione, dovrà seguire il rigore sanzionatorio della legge. D'altro canto, però, le cose cambiano, quando il rigore sanzionatorio segue in automatico per via delle istruzioni immesse nella tecnologia stessa. In questo caso, è evidente che non ci si colloca più sul piano del dover essere, ma dell'essere delle cose. Tra gli esempi possibili di quanto è stato riassunto nei termini dell'applicazione automatica della legge, tramite l'impiego della tecnologia a "controllo totale", basti far cenno all'impiego del DRM nel campo della proprietà intellettuale, o all'uso di sistemi di filtraggio in rete per fini di sicurezza.

Tornando qui al tema del diritto e di come esso tenti di regolare la tecnologia, diventa importante capire quale sia la finalità che il design potrà avere in questi casi. Riprendendo ancora una volta l'esempio delle misure di sicurezza nel trattamento dei dati, si può infatti pensare a queste ultime misure secondo un triplice punto di vista, vale a dire, ora

al fine di invogliare il soggetto a mutare il proprio comportamento, ora al fine di attenuare l'impatto di comportamenti dannosi, ora infine con lo scopo di prevenire del tutto l'eventualità che detti comportamenti dannosi si verifichino.⁵⁶

Ci sono poi svariati esempi di misure digitali di sicurezza, il cui fine è quello di attenuare l'impatto dei comportamenti dannosi ora sulle persone, ora sui sistemi tecnologici dell'informazione e comunicazione. Basti pensare alla configurazione di base degli interfacce e dei sistemi informatici, cui si è fatto cenno sopra riprendendo *Laws on Robots*, *Laws by Robots*, *Laws in Robots* di Ronald Leenes e Federia Lucivero. Qui, il disegno dell'applicazione robotica può essere impostato in modo da essere appropriato all'uso anche da parte dei meno esperti e, tuttavia, senza che per questo il sistema smetta di aumentare la propria efficienza. Al pari di altre modalità per la sicurezza, come la configurazione dell'interfaccia dei sistemi informatici, le copie di riserva o backup, l'obiettivo di questa modalità del design non è dunque quello di impedire che un evento dannoso si verifichi ma, piuttosto, nel caso in cui quest'ultimo evento si materializzi, il fine è che il sistema sia preparato a tutelarsi. Ecco perché si è suggerito un parallelismo tra questo tipo di design e gli airbag delle automobili. L'idea non è né di invogliare gli automobilisti a cambiare direttamente stile di guida, né di impedire che si verifichino incidenti. Piuttosto, nel caso d'incidente, l'intento è appunto quello di attutirne, in tutto o in parte, gli effetti. Questa regolazione della tecnologia può naturalmente essere disegnata in automatico e riportata alle finalità del diritto come meta-tecnologia

⁵⁶ Si v. K. Yeung, *Towards an understanding of regulation by design*, in *Regulating Technologies: legal futures, regulatory frames and technological fixes*, pp. 79-108, a cura di R. Brownsword e K. Yeung, Hart, Londra, 2007.

illustrate poc'anzi, per cui, ora con le tesi di Koops, si può pensare a raggiungere determinati effetti di sicurezza, con equivalenza funzionale tra online e offline, e senza discriminare eventualmente tecnologie con effetti equivalenti; oppure, con le tesi di Reed, in termini dell'indifferenza tecnologica o di potenziale neutralità della legge. Quale che sia il caso, tuttavia, l'intento tecno-regolatorio del diritto ha a che fare più con l'affidabilità e meticolosità del disegno delle misure di sicurezza in questione, che con le sue ricadute sull'autonomia dei soggetti. Si tratta di una differenza critica con l'accennata finalità del design d'invogliare il mutamento di comportamento dei soggetti e, soprattutto, con l'altra finalità del design, quello a controllo sociale, su cui ora brevemente ci soffermeremo.

L'altro intento tecno-regolativo del diritto in automatico riguarda le finalità del design di prevenire del tutto che comportamenti dannosi si verificino. Oltre agli esempi già ricordati dell'impiego di DRM nel campo della proprietà intellettuale, o l'uso di sistemi di filtraggio in rete ai fini della sicurezza nazionale, questo può essere il caso della privacy by design nell'ambito della robotica, dei droni ma in generale di ogni agente artificiale. Come riferito in precedenza, riprendendo le tesi di Ronald Leenes e Federica Lucivero⁵⁷, l'obiettivo può riguardare sia la regolazione del comportamento degli utenti e/o operatori di agenti artificiali, tramite il design di questi ultimi, in modo tale cioè che non sia consentito alcun comportamento illecito degli esseri umani, sia la disciplina del "comportamento artificiale" tramite il loro design, ossia immettendo direttamente i dettami della legge nel software dell'agente artificiale, in modo tale da prevenire atti indesiderati. Si tratta di un

⁵⁷ Cfr. R. Leenes e F. Lucivero, *Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour. By Design*, cit.

approccio al principio della privacy by design reso popolare proprio dalla sua stessa inventrice, il già citato Commissario per la privacy dell'Ontario, Ann Cavoukian, sulla base dei sette punti cardine con cui lei stessa ha provveduto a riassumere il principio⁵⁸:

- i) La protezione dei dati personali deve avvenire in forma proattiva, più che reattiva, nel senso che l'approccio della privacy tramite design ha uno scopo preventivo e non, semplicemente, correttivo o curativo;
- ii) I dati personali devono essere automaticamente protetti in ogni sistema informativo (ICT), come configurazione base, o di default, del sistema;
- iii) La protezione dei dati deve per ciò essere implementata nella progettazione o design dei sistemi informativi;
- iv) La piena funzionalità del principio che discende da (ii) e (iii) mira a conseguire una tutela che sua vantaggiosa per tutti e, comunque, renda superfluo, o inutile, qualsiasi consueto tipo di compromesso o bilanciamento tra gli interessi in gioco: ad esempio, il classico scambio tra privacy e fini di sicurezza;
- v) La protezione dei dati deve coprire l'intero ciclo vitale dell'informazione, ossia dalla culla alla tomba, per cui le misure a tutela dei dati devono essere predisposte ancor prima che un singolo dato, o pezzo d'informazione, sia stato raccolto;

⁵⁸ Si v. A. Cavoukian, *Privacy by Design: The Definitive Workshop*, cit. La sintesi dei sette punti chiave la traggio da U. Pagallo, *Il diritto nell'età dell'informazione*, cit., p. 288.

- vi) A prescindere dai tipi di tecnologia o pratica imprenditoriale, il disegno e la programmazione dei sistemi deve rendere i meccanismi di protezione dei dati visibili e trasparenti sia per gli utenti dell'ICT sia per i loro fornitori;
- vii) Il principio richiede infine che “architetti e operatori abbiano innanzitutto a cuore gli interessi dell'individuo, nell'offrire misure come rigorose impostazioni predefinite [default] del sistema, informativa appropriata e opzioni che siano amichevoli per l'utente e la rafforzino”⁵⁹. In altre parole, la privacy tramite design richiede d'incentrarsi sul rispetto per l'individuo e a tutela della privacy dell'utente.

Concentrando l'attenzione soprattutto sui punti (ii), (iii) e (v) dello schema di Cavoukian, tuttavia, l'obiettivo del design a controllo totale presenta una serie di problematiche. La prima è di natura tecnica e riporta, in parte, alla difficoltà già descritte in precedenza a proposito delle ontologie giuridiche, ossia l'obiettivo difficoltà di formalizzare alcuni concetti cardine della disciplina in materia di privacy e protezione dei dati, spesso dipendenti dal concetto preso in esame e soggetti peraltro ad evoluzione. Come ricorda Karen Yeung, “non solo è inevitabile il rischio di fallimenti operativi, ma la finalità di disegnare standard che siano in grado di raggiungere l'obiettivo desiderato del regolatore in forma precisa e accurata, non può che essere, con ogni evenienza, un'impresa improba.”⁶⁰ Dal punto di vista strettamente tecnologico, la ragione la spiega esaurientemente uno dei massimi

⁵⁹ A. Cavoukian, *Privacy by Design: The Definitive Workshop*, p. 248.

⁶⁰ K. Yeung, *Towards an Understanding of Regulation by Design, in Regulating Technologies: legal futures, regulatory frames and technological fixes*, a cura di R. Brownsword e K. Yeung, Hart, London, 2007.

esperti del settore, Eugene Spafford: “L’unico sistema [informatico] realmente sicuro è quello che sia stato spento, messo dentro a un blocco di cemento e sigillato a piombo in una stanza attornata da guardie giurate, e anche così avrei i miei dubbi⁶¹.

Una seconda ragione di perplessità è poi suggerita da un’ampia e comprovata letteratura scientifica che evidenzia il modo in cui particolari configurazioni di valori sociali, etici o politici, siano direttamente e sistematicamente realizzati, o soppressi, dal disegno tecnologico⁶².

Avendo a mente le differenze, anche radicali, appurate nel corso del precedente capitolo, a proposito delle diverse forme che il diritto alla privacy e con la tutela della protezione dei dati personali assume negli Stati Uniti e nell’Unione Europea, c’è da sospettare che questa forma di design a controllo totale finisca per aggravare, piuttosto che risolvere, queste differenze. Basti pensare alla diatriba sulla configurazione dei sistemi e interfaccia informatici, tra il tradizionale approccio *opt-in* europeo e il consueto modello di *opt-out* invalso invece in America.

Infine, pesano ragioni di ordine giuridico. Da un lato, di fronte alla presunta infallibilità di una tecnologia automatizzata, bisogna ricordare l’ammonimento di Lawrence Lessig, secondo cui “i controlli sull’accesso ai contenuti non saranno controlli ratificati dalle corti; i controlli sull’accesso ai contenuti saranno controlli codificati dai programmatori. E laddove i controlli che sono immessi nel nome della legge devono sempre essere sottoposti al controllo di un giudice, i

⁶¹ La cit. in S. Garfinkel e G. Spafford, *Web Security & Commerce*, O’Reilly, Sebastopol, 1997, p. 9.

⁶² M. Flanagan, D.C. Howe e H. Nissenbaum, *Embodyng Values in Technology: theory and practice*, in J. van den Hoven and J. Weckert (a cura di), *Information Technology and Moral Philosophy*, pp. 322-353, Cambridge University Press, New York, 2008.

controlli immessi nella tecnologia non hanno in se stessi un simile controllo⁶³. D'altro canto, per dirla con Jonathan Zittrain, “la perfetta applicabilità [della legge] fa svanire la pubblica comprensione del diritto in quanto la sua applicazione [automatica] elimina un'utile interfaccia tra i termini della legge e la sua imposizione. Parte di ciò che ci rende umani sono le cose che facciamo ogni giorno su quel che rappresenta alquanto di giusto o sbagliato (...). In un ambiente monitorato e sorvegliato del tutto, quelle stesse scelte, come è ovvio, svaniscono⁶⁴.”

Arriviamo infine, a conclusione del presente capitolo, a quella parte che potremmo definire dedicata ai dilemmi della *privacy by design*. Infatti, alla luce dei diversi, se non opposti, fini che il design può avere, sembra quasi paradossale che, più spesso, studiosi e legislatori si riferiscano al principio della *privacy by design* in forma neutra, quasi che l'intento di immettere il rigore sanzionatorio delle legge dentro la tecnologia possa essere imparziale o privo di implicazioni valoriali. Si prenda ad esempio l'emendamento 118 del Parlamento Europeo alla proposta della Commissione sul GDPR (prendo questo vecchio emendamento ad esempio ma per spiegare che se anche oggi, a Regolamento approvato ed entrato in vigore oramai da qualche anno, a parte alcune rare specifiche di organi deputati a produrre meri documenti di *soft law*, non siamo molto distanti da una normativa ancora troppo ampia nei contenuti), lo stesso afferma, facendo riferimento al principio di *privacy by design*: “onnicomprensive salvaguardie procedurali che

⁶³ L. Lessig, *Free Culture: the nature and future of creativity*, Penguin Press, New York, 2004, p. 152.

⁶⁴ J. Zittrain, *Perfect Enforcement on Tomorrow's Internet*, in *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes*, a cura di R. Brownsword e K. Yeung, Hart, London, 2007, pp. 125-156.

attengano all'accuratezza, confidenzialità, integrità, sicurezza fisica e cancellazione di dati personali", senza che sia affatto chiaro il fine dell'emendamento, ovvero avrebbe dovuto essere, impedire il verificarsi di eventi dannosi o anche di invogliare i soggetti titolari a mutare il proprio comportamento in uno più "privacy friendly".

Di fronte ai silenzi, o alle incertezze delle istituzioni il rischio è la caduta in una visione eccessivamente paternalistica, quale quella del controllo totale evidenziata sopra⁶⁵.

A prevenire fraintendimenti, nondimeno, la tesi che qui si vuole sostenere, non esclude di per sé il ricorso all'automatismo della legge. Abbiamo in fondo ricordato sopra l'odierno settore della ricerca nel campo dell'intelligenza artificiale e anche delle ontologie giuridiche il cui fine è di migliorare l'efficienza, consistenza, comprensibilità e finanche la prevedibilità dei sistemi giuridici; fini tutti encomiabili.

Bisognerebbe ricordare che le scelte del legislatore, anche e forse soprattutto nell'ambito del design a controllo sociale, non si danno in una sorta di vuoto normativo ma, al contrario, vengono strutturalmente definite dalla presenza di valori e principi. Per dirla con Jon Garthoff, sulla scia del pensiero del grande filosofo del diritto Herbert Hart, "strettamente parlando, gli individui non sono costretti a fare ciò che la legge dice loro di dover fare, ma piuttosto ciò che la prevalente convinzione sociale dice loro che essi sono obbligati a compiere (e questo con la pandemia da Covid-19 e le relative regole sul Green pass ne sono un perfetto esempio); il luogo normativo dell'autorità politica è per ciò la convenzione supportata dalla legge, più che la legge stessa, dato che è la convenzione a risolvere il problema del coordinamento

⁶⁵ Sul paternalismo tecnologico ha insistito di recente U. Pagallo, *Il diritto nell'età dell'informazione*, cit., p. 149.

morale, trasmettendo così l'autorità normativa morale alla struttura sociale in questione"⁶⁶. Il contesto normativo che deve conseguentemente essere valutato ai fini dell'adozione, o meno, di tecno-regolazioni ad effetto automatico tramite il disegno di meccanismi a controllo totale, metterà in mostra l'insieme di valori e principi in gioco e, soprattutto, se e fino a quale punto viga al riguardo una qualche forma di coesione cruciale. Infatti, le questioni inerenti all'automazione giuridica, come nel caso del principio della privacy by design applicata ai droni, non riguardano soltanto il grado di prevedibilità e affidabilità di tali automatismi ma anche, se non soprattutto, il grado di accordo, o disaccordo, che è dato registrare nel contesto normativo preso in esame.

Su queste basi, possiamo affrontare le sfide giuridiche dei droni, specialmente nell'ambito della protezione dati e della privacy, riportandole al quadro generale dell'odierna convergenza tecnologica dell'internet delle cose, il cloud computing, i big data e l'intelligenza artificiale. Il dato che siamo in un mondo ormai sempre più connesso, "onlife" per citare il nome di un vecchio progetto della Commissione Europea, sta a suggerire come il nuovo scenario incida sulla condizione umana, modificando il rapporto che gli uomini hanno con sé stessi, con gli altri e con il mondo in generale. Le sfide giuridiche dei droni vanno dunque comprese in un mondo, in cui si registrano le seguenti trasformazioni:

- a) L'erosione dei confini tra il reale ed il virtuale;
- b) L'erosione dei confini tra uomo, macchina e natura;

⁶⁶ J. Garthoff, *Legitimacy is not Authority*, in "Law and Philosophy", 2001, vol. 29(6), pp. 669-694. La citazione si trova a p. 669.

- c) Il rovesciamento della situazione nella sfera dell'informazione: dalla scarsità alla sovrabbondanza;
- d) La transizione dal primato del soggetto al primato dell'interazione⁶⁷.

Parlando di tali cambiamenti in riferimento alle tematiche attinenti alla privacy, abbiamo più volte ripetuto quanto stia mutando tale concetto sia in luoghi pubblici, sia in casa propria, sia in quello spazio dell'*infosfera* che può essere spiegato e rappresentato con gli “ambienti intelligenti” e “l'internet di tutto”.

Questi problemi sono stati spesso risolti a livello di diritti di accesso e consenso nel dare i propri dati ma, analogamente a quanto succede nel film *Blade Runner*, dove i droni volano indisturbati per le città, occorre porsi il problema di quale possa essere mai in futuro la nostra ragionevole aspettativa di privacy e quali saranno i mezzi per renderla possibile.

Nel corso del presente capitolo, abbiamo saggiato come il principio della privacy by design, se opportunamente calibrato, può rendersi utile per tutelare quell'insieme discreto e unitario di dati, avente strutture appropriate e dotato dell'insieme di operazioni, funzioni e procedure che qualificano, nella sua individualità ogni ente informazionale⁶⁸. La privacy by design può infatti rafforzare il plesso di principi, regole, decisioni e pratiche che caratterizzano il diritto alla tutela dei dati personali, come insieme di forze che si oppongono al flusso di

⁶⁷ In *Onlife Manifesto*, a cura di Luciano Floridi, 2004.

⁶⁸ Riprendo qui a pieno la terminologia utilizzata da L. Floridi nello spiegare la filosofia dell'informazione partendo dai soggetti visti come “enti informazionali” cioè costituiti di informazioni.

informazioni negli ambienti intelligenti, secondo ciò che Luciano Floridi ha definito con il termine di “frizione ontologica”, ossia “l’ammontare di lavoro e sforzi richiesto a un certo agente per ottenere, filtrare e/o bloccare informazioni su altri agenti in un ambiente determinato⁶⁹. La tutela della privacy degli individui, con la sfera dei dati personali, corrisponde di qui all’obiettivo di mantenere l’integrità di quel pacchetto di dati in base al quale è possibile intendere la stessa identità degli individui, per cui maggiore il grado di “frizione ontologica”, minore l’accessibilità all’informazione e quindi maggiore la loro privacy informazionale. Come questo processo possa essere reso, in tutto o in parte automatico, ho tentato di spiegarlo sopra. La questione, però, non è certamente solo tecnica. Si tratta di esaminare il grado di coesione sociale e il contesto normativo, entro il quale le scelte del design e tecno-regolazione giuridica andranno valutate. Maggiore il grado della coesione sociale, maggiore i rischi che saranno socialmente accettati attraverso il giudizio delle conseguenze non del tutto prevedibili degli agenti robotici e dei loro infiniti utilizzi.

2.2.2 La natura giuridica degli agenti artificiali e la responsabilità per le loro azioni

Questo paragrafo è dedicato alle conseguenze giuridiche, penalmente e civilmente rilevanti per il diritto, scaturenti dal comportamento degli agenti artificiali, in particolare degli agenti software. Una premessa: quanto verrà illustrato vale generalmente per qualsiasi agente artificiale

⁶⁹ Si v. L. Floridi, *Infosfera*, collana Digitalica, Giappichelli, Torino, 2006.

ma dalla nostra noi prendiamo ad esempio in particolare gli agenti software per fare una sorta di prima introduzione al tema che sarà illustrato nel capitolo terzo, ossia la giustizia che utilizza gli algoritmi al fine di tentare di predire il verdetto di una causa, la possibilità di recidiva e via discorrendo. Il lettore non rimanga deluso, sebbene giuristi abbiamo ben a mente la differenza che intercorre tra un agente software e un algoritmo, ma posto che per essere eseguito da un computer l'algoritmo deve essere codificato in un programma informatico cioè in un software appunto, allora conviene sicuramente interrogarci sulla sua natura giuridica e sulla responsabilità per le sue azioni.

Molte della azione poste in essere nel *cyberspazio* dall'utente, tramite un agente software ovvero ancora da un agente software quando agisce in autonomia, sono suscettibili di produrre conseguenze giuridiche rilevanti. Quanto più si allontana il legame tra utente e agente software, vale a dire tra l'agente software e una persona che possa fungere da centro di imputazione giuridica, tanto più diviene difficile ricondurre una determinata situazione entro gli schemi concettuali e le categorie proprie della dogmatica giuridica. L'impatto delle tecnologie rende difficile, per il diritto, rendere conto della realtà, non solo virtuale, che risulta trasformata in modo incessante dalle innovazioni e modificazioni tecnologiche. Il diritto non riesce (o per meglio dire non riesce in modo altrettanto efficace) a regolare la realtà che si trova ad essere già regolata dalla tecnica.

Ed il problema, tuttavia, non si pone solo in questi termini, cioè in un confronto tra un sistema statico (il diritto) ed un sistema dinamico (la tecnica). Nulla impedisce, infatti, di concepire il diritto come un sistema dinamico (Kosmos) che evolve, in quanto parte della cultura, nello

stesso modo in cui evolve la realtà regolata dalla tecnica. Il problema appare ancora più profondo, dal momento che la tecnica, non solo amplia il novero delle nostre domande, ma muta anche il senso di tali domande, creando così vuoti concettuali. Ciò significa che l'impatto tecnologico modifica il sostrato stesso di determinati concetti giuridici, quale per esempio la nozione di *persona*, mostrando il tal modo che tale nozione non è altro che un dispositivo concettuale (artificiale) inteso a produrre la differenza tra ciò che è (ritenuto come) *personale* e ciò che non lo è (*l'impersonale*). Che il diritto possieda natura concettuale non stupisce, colpisce, invece, il fatto che la tecnica mostri questa natura concettuale del diritto, a partire da una trasformazione che investe la decisione antropologica della persona, colta qui, nella sua consistenza di soggetto agente. E' rispetto a tale sostrato antropologico che si delinea la natura di dispositivo concettuale della nozione giuridica di persona: dispositivo, che a seguito dell'impatto tecnologico non scompare ma anzi è chiamato a riarticolare la differenza tra ciò che si ritiene personale e ciò che si ritiene impersonale. Riarticolazione della massima importanza, poiché da essa dipende la possibilità di riconoscere l'esistenza di un centro di imputazione responsabile delle conseguenze giuridicamente rilevanti che derivino dalle azioni poste in essere da un soggetto agente (*persona o software*).

Si tratta di altri termini di delineare, non tanto chi sia oggi, nel *cyberspazio*, una persona, quanto piuttosto cosa possa essere inteso come *personale*: termine che vale qui come attributo posto a fondamento del riconoscimento di responsabilità, di un'azione che genera in altri un ragionevole affidamento in base al grado di previsione delle conseguenze derivanti da tale azione. La responsabilità dev'essere, allora, ricostruita a partire dal nesso concettuale tra

previsione ed affidamento: nesso, il cui statuto concettuale è, in primo luogo, informatico. Si tratta cioè di comprendere cosa significhi prevedere e fare affidamento, dal punto di vista dell'epistemologia informatica, e misurare in che modo tale epistemologia è coerente o conciliabile con quella giuridica. Tale considerazione rafforza, ancora una volta, l'idea che, nello spazio dell'informazione, il diritto può svolgere un ruolo centrale nella determinazione del nesso che lega previsione ed affidamento, precisando i termini in cui tale nesso si proietta da mondo virtuale al mondo reale.

Tornando all'esempio degli agenti software, vedremo brevemente in che modo le loro azioni sono rilevanti per il diritto. Tali conseguenze si possono apprezzare in particolare nel campo civilistico, dal punto di vista negoziale, per la capacità dell'agente software di commerciare, di concludere contratti, di richiedere servizi, di avviare trattative, di generare affidamento in una controparte, ecc.

Il complesso insieme di relazioni giuridiche soggiacenti alla trama di interazioni che avvengono tra utenti, fornitori, agenti ecc., nell'ambito del cyberdiritto sono ordinabili lungo tre direttrici: 1) questioni attinenti allo status giuridico dell'agente: si tratta di un soggetto di diritto? E' dotato di capacità di agire? E' mosso da intenzioni? E' dotato di stati psicologici giuridicamente rilevanti (buona fede, coscienze e volontà dell'azione)? 2) questioni attinenti alla vita dell'agente (costruzione, cessione e funzionamento): si tratta di questioni relative al brevetto tecnologico di costruzione; al copyright sul software che implementa; al contratto di cessione della licenza per l'utilizzo; al contratto di manutenzione per il suo funzionamento, ecc. 3) questioni attinenti alle azioni realizzate dall'agente: si tratta di questioni relative alla negoziazione di trattative al fine di stipulare un determinato accordo

contrattuale, alla conclusione stessa del contratto; all'affidamento che l'utente può fare sull'esecuzione del compito; all'affidamento che altro utente o agente può fare sulle azioni o informazioni dell'agente poste in essere durante l'esecuzione del compito; a conseguenze giuridiche derivanti dalla inesatta ovvero mancata esecuzione dei compiti dell'agente.

La natura giuridica degli agenti software può essere messa a punto a partire da un'alternativa di fondo, relativa alla possibilità di considerare tali agenti come *persone* ovvero *cose*. L'ipotesi prospettata da Sartor suggerisce l'opportunità, dal punto di vista della qualificazione e della disciplina giuridica da applicare, di considerare gli agenti software alla stregua delle persone⁷⁰. Tale ipotesi riposa su un presupposto concettuale, che riguarda il modo stesso di percepire il fenomeno degli agenti: a tale proposito Sartor richiama la tripartizione con cui Dennet⁷¹ qualifica la prospettiva dalla quale possiamo guardare agli agenti: 1) prospettiva *fisica*, per la quale il comportamento di un ente è spiegato in base al suo stato fisico e alle leggi di natura; 2) prospettiva *progettuale*, per la quale il comportamento di un ente è spiegato in base alla struttura del programma con cui l'ente è stato progettato; 3)

⁷⁰ Si tratta di una ricostruzione che non presuppone di riconoscere all'agente software lo statuto di persona o di soggetto morale o di diritto, ma di riconoscere ad esso quei caratteri (stati morali o cognitivi intenzionali) che permettono di ricostruire la loro natura giuridica e responsabilità nei termini delle produzioni della differenza tra un elemento personale ed uno impersonale (soggetto o cosa). Per tale ipotesi si v. G. Sartor, "Gli agenti software e la disciplina giuridica degli strumenti cognitivi", cit., il quale sottolinea che, p. 18 "si osservi che la rilevanza degli stati cognitivi dell'as è la ragione sola per la quale la relazione tra un AS e il suo utilizzo può essere assimilata, a certi fini ed in certe circostanze, a quella tra un dipendente ed il suo datore di lavoro (ai fini della responsabilità vicario od oggettiva). Ciò non presuppone che l'AS sia classificato come una persona o un soggetto morale o giuridico".

⁷¹ D. Dennet, *Kids of Minds: Towards an Understanding of Consciousness*, Basic Books, New York, 1996, p. 28 ss.

prospettiva *intenzionale*, per la quale il comportamento dell'ente è spiegato in base agli stati mentali e cognitivi associati ai comportamenti dell'ente.

La possibilità di decodificare il fenomeno degli agenti software nei termini della nozione di persona riposa dunque sull'adozione di una prospettiva intenzionale, che rende possibile associare ai comportamenti degli agenti queglii stati cognitivi (che possono essere definiti, per traslato, mentali) in base ai quali prevedere, comprendere e valutare tali comportamenti. Tale possibilità è suggerita da Sartor in termini che ne rilevano l'orizzonte categoriale sottointeso:

Pensiamo a un'organizzazione per il commercio, in cui operino diverse componenti: il software intrattiene i contatti con i clienti, predisponendo le offerte di vendita, ricevendo e confermando le accettazioni dei clienti, controllando lo svolgimento del rapporto; i programmatori provvedono a stendere e aggiornare il software secondo le indicazioni della direzione; gli operatori provvedono a ritoccare i parametri del sistema quanto necessario; i funzionari studiano i dati aggregati e definiscono i compiti dei programmatori e degli operatori.

Nessuna componente del sistema ha una visione accurata di tutte le conoscenze in esso presenti [...] ma il sistema nel suo complesso appare operare razionalmente: esso persegue gli obiettivi che lo caratterizzano tenendo conto di tutte le informazioni a esso disponibili⁷².

Le ultime righe del passaggio citato stanno a sottolineare che se le componenti del sistema informatico non consentono una previsione totale delle operazioni poste in essere in base all'autonomia dell'agente software, il sistema considerato nel suo complesso agisce in base ad un

⁷² G. Sartor, "*L'intenzionalità dei sistemi informatici e il diritto*", cit., p. 25.

criterio che può essere valutato come *razionale* e in tal senso sinonimo della sussistenza di stati cognitivi intenzionali. Tale criterio consiste nel porre in stretta correlazione l'affidamento sulle informazioni disponibili con il perseguimento di obiettivi associati a tali informazioni. Diviene possibile decodificare così il comportamento di agenti teleologicamente orientati al conseguimento di un fine nei termini di stati cognitivi intenzionali a partire da una concezione della razionalità che non pone più in diretta relazione intenzione ed azione, ma affidamento e comportamento. Tale relazione non esclude ma al contrario rende ricostruibile l'azione degli agenti software nei termini di stati mentali diretti ad un fine:

La tutela dell'affidamento non rappresenta una negazione dell'aspetto intenzionale o psicologico delle azioni umane, ma intende anzi garantire la possibilità di interpretare intenzionalmente le azioni altrui. Io posso confidare nella mia attribuzione di certi stati intenzionali alla mia controparte (l'intenzione di vendere un certo bene, l'intenzione di impegnarsi a svolgere una certa attività, la credenza in quanto affermati), sulla base degli indizi forniti dalla stessa controparte, in conformità alle convenzioni in vigore, in quanto so che la mia attribuzione, anche se fosse errata, determinerebbe [...] conseguenze giuridiche ad essa corrispondenti.

L'affidamento è concepito, non solo come base necessaria per l'azione, ma anche e soprattutto come parte della trama di interazioni che si stabiliscono tra utenti e tra utenti ed agenti. E' proprio a partire da tale categoria che è possibile ricostruire un concetto di agire razionale, sul quale fondare il riconoscimento di un elemento intenzionale che dà in

tal modo contenuto ad una nozione di *persona* che permette di distinguere, anche nell'ambito informatico, tra ciò che si ritiene come *impersonale*.

L'ipotesi della razionalità del comportamento dell'agente software conferisce alla trama di interazioni in rete una dimensione negoziale, che costituisce, ad avviso di Sartor, la prospettiva che è in grado, in modo più idoneo, di fondare la disciplina giuridica degli agenti software e dare conto delle conseguenze giuridiche derivanti dalle loro azioni.

Ciò che preme sottolineare non è tanto la prospettiva negoziale che qualifica l'agente software come un rappresentante dell'utente dotato di personalità giuridica, vale a dire centro di imputazione di diritti e doveri, il cui collegamento con l'utente, in base ad un rapporto di rappresentanza giuridica, assicura anche agli eventuali creditori la possibilità di aggredire un patrimonio. Piuttosto, l'intento è di mettere in risalto l'idea che il riconoscimento di stati cognitivi intenzionali sia un modo, articolato e complesso, di ricostruire una nozione di *personalità*, che si iscrive nel solco della tradizione giuridica tracciato dal dispositivo (artificiale) della *persona* e lo amplia, al fine di comprendervi la realtà peculiare dell'autonomia d'azione degli agenti software e di evitare così una duplicazione di categorie, giuridiche da una parte ed informatiche dall'altra. Come abbiamo posto in rilievo, non si tratta di elaborare categorie che risultino inapplicabili al mondo del diritto e al mondo dell'informatica, quanto piuttosto di comprendere il modo in cui l'elemento categoriale (artificiale) del diritto viene portato alla luce ed aggiornato dalla rivoluzione tecnologica in corso, che l'iscrive entro una configurazione diversa dello spazio e del tempo. La capacità di autonomia e decentramento che connota la natura degli agenti software determina una concezione della razionalità fondata

sull'idea di affidamenti: vale a dire sul grado di affidamento che è ragionevole fare su informazioni che, come Sartor mette bene in evidenza, sono sempre più prodotti in modo automatico nella nostra società, in seguito alla possibilità di interagire in rete non solo tra utenti, ma tra utenti e agenti e tra agenti e agenti.

Il tema della responsabilità giuridica degli agenti artificiali è tema arduo e complesso e allo stato non si rinvengono leggo che parlano chiaro in merito pertanto qui tenteremo di fare una breve ricognizione sul tema al fine di acquisire qualche punto fermo.

Le azioni poste in essere da un agente software nell'esecuzione dei suoi compiti possono dare luogo a responsabilità giuridica sotto più profili: per aver ingenerato un affidamento legittimo nella controparte; per l'adempimento delle obbligazioni nate dalla conclusione di un contratto; per l'evento dannoso conseguente all'esecuzione del contratto, ecc. Rispetto a tali ipotesi occorre identificare il rapporto tra l'agente e l'utente, per cui l'agente opera, possiamo dire, dal punto di vista giuridico. D'altro canto è necessario verificare quali siano i termini in cui è possibile riconoscere in capo al soggetto utente una responsabilità per le azioni poste in essere dall'agente. Si tratta, come è ovvio, di una responsabilità vicaria per il fatto altrui che trova fondamento sul collegamento esistente tra utente ed agente.

Naturalmente ciò presuppone di aver già risolto una serie di problemi relativi all'identificazione dell'inadempimento, del danno, dell'illecito, ecc., ovvero dell'autore stesso di tale inadempimento, danno o illecito: la possibilità di tale identificazione si scontra con la dematerializzazione, la delocalizzazione e il decentramento delle azioni poste in essere in rete. In relazione *all'invisibilità* di tali azioni è difficile ricostruire di volta in volta una nozione di soggetto agente da

poter costituire come centri di imputazione giuridica delle conseguenze dannose o comunque illecite realizzate. Anche da questo punto di vista si fa strada la luce di un decentramento radicale: la logica del decentramento deve rendere conto dell'esigenza di conservare un riferimento possibile ad una certa misura di centralità, vale a dire alla definizione di centri di imputazione giuridica. Che tale definizione non proceda più dall'identificazione di un sostrato personale non costituisce ancora una volta una novità per il diritto che è già pienamente aduso dall'utilizzo di finzioni giuridiche in tal campo, qual è quella della persona giuridica.

Come più volte sottolineato, nel contesto informatico della rete, bisogna ricostruire un legame tra l'agente e il soggetto della responsabilità che muova dal riconoscimento di stati cognitivi intenzionali che sono rivelati dalla dimensione linguistica di dichiarazioni, promesse, accordi, ecc., e sono veicolati *online* dall'elaborazione automatica dell'informazione. La possibilità di identificare questo legame vale a fondare una soluzione giuridica per il problema in esame, che passa attraverso la riconfigurazione della dimensione concettuale (artificiale) del dispositivo giuridico della *persona*, ma non implica di riconoscere, certo, all'agente software lo statuto di persona o la sua responsabilità giuridica o morale: occorre "considerare se gli stati cognitivi di un AS (sigla che sta per agente software) possono essere rilevanti per stabilire e circoscrivere le possibilità derivanti da un danno causato da quel AS. [...] porsi questo quesito non significa chiedersi se l'AS sia giuridicamente o moralmente responsabile [...]. Tuttavia il fatto che la responsabilità dell'utilizzatore possa dipendere dagli stati cognitivi del suo AS, differenzia l'AS da altre cose o strumenti, e comporta

un'analogia con la responsabilità vicaria"⁷³. Ciò viene a ribadire il criterio per cui l'utilizzatore risponde a titolo di responsabilità per il fatto altrui, vale a dire per le azioni poste in essere dall'agente in base a stati intenzionali che hanno indotto in altri utenti o agenti un legittimo affidamento.

A prescindere dalla correttezza di tale soluzione, ci può tradursi nella situazione per cui le conseguenze delle azioni poste in essere dagli agenti sono inesorabilmente ricondotte alla responsabilità dell'utente sia che questi abbia voluto le operazioni automatiche dell'agente sia che questi non abbia voluto le operazioni conseguenti agli stati intenzionali dell'agente⁷⁴.

Tale criterio viene ad ampliare la sfera di responsabilità dell'utente o utilizzatore rispetto al principio per cui un soggetto è, di regola, responsabile per le conseguenze delle azioni (da lui o da altri poste in essere) che egli poteva ragionevolmente prevedere (che rientravano, cioè, entro una sfera ragionevole di prevedibilità). E' proprio questo il punto che ci preme sottolineare: la riconfigurazione necessaria, dal punto di vista dell'epistemologia informatica e giuridica, della nozione di previsione.

Dalla stessa definizione di agente software, sistema informatico che genera informazioni e compie azioni in modo autonomo rispetto alle istruzioni con cui è stato programmato, emerge per l'utente o l'utilizzatore l'impossibilità di prevedere la complessità delle azioni ed

⁷³ G. Sartor, *Gli agenti software e la disciplina giuridica degli strumenti cognitivi*, cit., p.16.

⁷⁴ In tal senso quanto afferma l'Uniform Computer Information Transaction Act (UCITA): "La persona che usa un agente elettronico che ha scelto per effettuare un'autenticazione, prestazione o accordo, inclusa la manifestazione del consenso, è vincolata dalle operazioni dell'agente elettronico, anche se nessun individuo era a conoscenza o verificava le operazioni dell'agente o i risultati delle operazioni".

informazioni generate, dai contesti in cui un agente è suscettibile di agire nella sua mobilità, dalle modificazioni stessa dell'agente se in grado di apprendere dell'esperienza e d'interagire con le proprie istruzioni. Si tratta, come è evidente, di un problema di portata più generale: l'imprevedibilità degli effetti derivanti dalle azioni ed informazioni generate al di fuori del controllo degli utenti della rete. La tesi che abbiamo avanzato è che, a fronte di tale problematica, giuridica e non solo, non sia sufficiente indicare un principio vicario di responsabilità, anche se necessario alla sensibilità del giurista: "L'imprevedibilità dell'attività dell'agente non esclude quindi che gli effetti giuridici di tale attività possano imputarsi all'utilizzatore. Ma tale imputazione dovrà in molti casi appoggiarsi a fondamenti diversi dal fatto che l'utilizzatore abbia voluto (o abbia previsto o potuto prevedere) quegli effetti: l'utilizzatore dovrà farsi carico di effetti in quanto ha scelto di inviare l'agente a contrattare per suo conto"⁷⁵.

L'esigenza di dare quindi fondamento giuridico al riconoscimento di responsabilità sulla base di criteri individuati in relazione alla peculiarità della situazione in esame non esime, a nostro avviso, dalla necessità di confrontarsi con il problema di fondo dello statuto della *previsione*: non si tratta più di compiere un giudizio prognostico sull'occorrenza di un comportamento futuro a partire dalla conoscenza del passato, quanto piuttosto di concepire delle aspettative razionali che trovano fondamento sulla conoscenza dei limiti cognitivi degli agenti software. La rappresentazione di ciò che è razionale attendersi da un agente che è teleologicamente programmato per conseguire un dato obiettivo (*razionalità strumentale*) è corretta, così, sulla base della capacità di previsione dei limiti peculiari di ogni singolo agente. In tale

⁷⁵ G. Sartor, *Gli agenti software: nuovi soggetti del cyberspazio*, cit., p. 79.

prospettiva la previsione viene ad essere concepita come la rappresentazione razionale di limiti cognitivi suscettibili di delimitare il campo della probabilità di un'azione differente o erronea rispetto a quella che è attesa sulla base di una concezione strumentale della razionalità.

Il senso di tali considerazioni emerge in modo ancor più chiaro, dalle analisi con le quali Sartor delinea tale punto in relazione alla problematica giuridica degli agenti software:

La conoscenza dei principi del funzionamento dell'agente e l'analisi del suo comportamento nel passato (l'agente può aver commesso errori) non offrono una base sufficiente per effettuare una previsione del suo comportamento futuro. Tale conoscenza può invece consentire di determinare i limiti delle capacità "razionali" dell'agente e, quindi di individuare le ipotesi nelle quali l'agente può comportarsi in modo inadeguato. Combinando l'ipotesi della razionalità dell'agente con la conoscenza dei limiti di tale razionalità, siamo in condizione di prevedere e spiegare il comportamento dell'agente con sufficiente accuratezza. Non è diversa la strategia che adottiamo di regola nei confronti dei nostri simili: la nostra conoscenza dei limiti delle idiosincrasie, delle attitudini di ciascuno non offre una base sufficiente per prevedere il comportamento altrui (tranne che nel caso di comportamenti dettati da riflessi automatici o da pulsioni incontrollabili). Tale conoscenza ci consente però di integrare e correggere l'ipotesi che ciascuno tenda a perseguire i propri scopi in modo razionale: possiamo spiegare e prevedere il comportamento umano combinando l'ipotesi generale della razionalità di ciascuno con la conoscenza dei limiti di tale razionalità nei singoli casi⁷⁶

⁷⁶ G. Sartor, *Gli agenti software: nuovi soggetti del cyberspazio*, cit., p. 81.

Dopo una succinta panoramica delle possibili allocazioni di responsabilità tra i soggetti coinvolti nell'uso degli strumenti informatici, in particolare software e agenti robotici, cercheremo di circostanziare, a seguire, il tema della responsabilità civile per danni cagionati da agenti non umani; in particolare analizzeremo la responsabilità dei robot (avendo a mente ovviamente che essa è molto differente a seconda del compito che è chiamato a svolgere).

Il tema della responsabilità civile e della sua gestione nell'ambito degli scenari che concernono le moderne applicazioni robotiche è uno degli aspetti cruciali e più dibattuti nell'ottica dell'implementazione concreta dei singoli progetti e della loro sostenibilità economica in vista di sviluppi orientati al mercato. Pertanto, analisi quanto più possibile accurata e lungimirante delle relazioni tra i vari soggetti e dei rispettivi profili di rischio è determinante per individuare gli accorgimenti e le cautele che consentono di ridurre il rischio a carico di attori del sistema, trasferendolo su altri e, dove possibile, neutralizzandolo.

L'analisi del tema della responsabilità civile, già di per sé controverso e ricco di vuoti normativi, diventa ancor più multiforme quando si ha a che fare con danni cagionati da agenti non umani e relativi risarcimenti da allocare di volta in volta a seconda di parametri non predeterminabili a priori, data l'evoluzione e quindi il raggio d'azione potenziale che questi agenti finiranno per avere.

Inoltre, occorre far notare fin da subito, che i dispositivi in esame e le relative attività, potenzialmente dannose, al pari di quelle umane, sono attività innovative e sono il frutto del lavoro congiunto e della collaborazione di numerosi soggetti, i quali, allo stato soggiacciono a regole di natura pubblicistica, in considerazione dei possibili utilizzi di tali applicazioni suscettibili di interferire con l'uso di spazi e luoghi

pubblici come lo spazio aereo, il demanio pubblico o se si pensa all'ambito militare, agli apparecchi e alle tecnologie *dual use*.

Il tema che qui di seguito verrà analizzato, cioè quello dell'allocazione più efficiente del rischio di danno cagionato da robot, per dirla, alla Posner in termini economici, cioè la responsabilità civile, rileva sotto una duplice veste: quella della responsabilità contrattuale e quella della responsabilità extracontrattuale (c.d. "responsabilità civile" in senso stretto). Data l'estensione della disciplina in esame, in questo capitolo ci limiteremo ad analizzare la responsabilità extracontrattuale, non senza sottolineare che la gestione delle ipotesi di responsabilità contrattuale sia almeno alquanto importante.

Un tipo di responsabilità civile può nascere in capo al produttore o all'utilizzatore di un agente artificiale.

Occorre rilevare fin da subito come esistano tre condizioni in cui, astrattamente, un dato agente, sia esso individuale, sociale o artificiale, viene a ritrovarsi in rapporto al principio di responsabilità⁷⁷.

In primo luogo si ha una condizione di assoluta esenzione di responsabilità. Sono molteplici le ragioni del principio nei vari campi, quale il criminale (*habeas corpus*), o il civile (*ad impossibilia nemo tenetur*), ma le sue radici possono essere riassunte con la vecchia formula che fu già di Hobbes, in base alla quale, tutto ciò che non è espressamente proibito è permesso. Si tratta di un tipico principio caro anche alla tradizione liberale, che vede tra i suoi massimi esponenti Popper e Hayek.

In secondo luogo, gli agenti possono essere ritenuti, all'opposto, sempre e comunque responsabili. Se escludiamo, come sembra oggi

⁷⁷ Riprendo U. Pagallo, *Sul principio di responsabilità giuridica in rete*, in "Il diritto dell'informazione e dell'informatica", XXV, 4-5, 2009, pp. 705-734.

ragionevole, tentazioni totalitarie, la ratio del principio di responsabilità oggettiva nasce da svariate esigenze sociali; i datori di lavoro che rispondono per gli illeciti dei propri dipendenti, i padroni di animali per danni cagionati da questi (salvo la prova del caso fortuito), i genitori per il comportamento dei figli minori e così via, in ipotesi tipizzate dalla legislazione civile di riferimento.

Infine, tra le due posizioni estreme, vi è quel tipo di responsabilità di gran lunga più frequente nella aule di tribunale, cioè la responsabilità che nasce per colpa o dolo e che può essere convenientemente ricondotta al broccardo dell'*alterum non laedere*. E' proprio questa terza tipologia di responsabilità che crea molti degli equivoci che affollano l'odierno dibattito, non solo giuridico, intorno alle nuove scoperte tecnologiche.

La progettazione, la costruzione, la commercializzazione e l'utilizzo dei robot e la fornitura di servizi attraverso gli stessi sono tutte attività suscettibili di costituire fonte di responsabilità civile. In particolare la costruzione, la commercializzazione e l'uso dei robot rilevano ai fini della responsabilità civile sia sul piano della c.d. responsabilità "indiretta" (di padroni e committenti, vedi art. 2049 c.c.) e sia su quella della responsabilità oggettiva o quasi oggettiva (c.d. danno imputabile a cose).

L'aspetto che qui più rileva per inquadrare le differenti modalità con le quali può concretizzarsi la responsabilità degli agenti è quello delle diverse forme di imputabilità accanto alla tradizionale imputabilità "soggettiva" per dolo o colpa; vi sono infatti, una serie di criteri di imputabilità cosiddetta "speciale" che prescindono dell'esistenza di uno stato soggettivo rilevante in capo all'agente e fondano l'imputabilità su diversi elementi come la relazione con l'oggetto/soggetto causatore del

danno, l'assunzione del rischio, la valutazione (in senso stretto e a priori) del rapporto socio-economico tra danneggiante e danneggiato e la conseguente imputazione del danno al soggetto che, sempre in astratto e aprioristicamente, appaia maggiormente in grado di sostenerlo economicamente⁷⁸.

Sarà quindi in quest'ottica che il legislatore dovrà trovare soluzioni per i casi nei quali l'imprevedibilità della azioni dei robot o l'incapacità di prevederle di chi li controlla, suggerisca meccanismi come, per esempio, una nuova tipologia di "responsabilità del robot" che (almeno ad oggi) non può trovare collocazione nel nostro ordinamento.

Sebbene scontato, prima di procedere ad un'analisi più dettagliata della responsabilità civile e poi a seguire della responsabilità penale degli agenti robotici, occorre che il lettore abbia chiaro che la mia analisi, almeno per il momento, e circa questo tema, sarà condotta con mero riferimento alla legislazione italiana.

Come dicevamo in qualche capoverso precedente le ormai differenti tipologia di robot sono suscettibili di configurare le più svariate forme e modalità in cui le ipotesi di responsabilità civile possono manifestarsi. Va peraltro rilevato che ciascuna di tali categorie di applicazione (robot di telepresenza, robot courier, quadricotteri operanti in spazi urbani, giudici robot di cui ampiamente si dirà nel prosieguo di questo lavoro) può presentare una serie di varianti costruttive e operative potenzialmente illimitate; di conseguenza, le problematiche giuridiche potenzialmente rilevanti coprono pressoché tutto l'ambito della responsabilità civile.

⁷⁸ Si v. sul punto P.G. Monateri, *La responsabilità civile*, collana diretta da Rodolfo Sacco, Utet, Torino, 1998, pp. 234 ss.

Provvediamo ora a dettagliare gli attori potenzialmente coinvolti nei vari progetti espressi in sintesi nella parentesi sopra, e gli scenari di maggior interesse in cui coloro che producono e utilizzano robot possono incorrere.

I soggetti che interagiscono a qualsivoglia titolo con detti agenti sono così classificabili:⁷⁹

- a) Soggetti che contribuiscono alla realizzazione del prodotto ed alla prestazione dei servizi offerti dai robot, per esempio fornendo singole componenti hardware, ovvero servizi propedeutici o funzionali alla produzione o all'utilizzazione dei robot. Tale categoria include dunque i costruttori di componenti meccaniche e componenti elettriche, i fornitori di componenti hardware e software, i progettisti del robot o di alcune sue parti, i fornitori di servizi cloud e di piattaforme di sviluppo, i fornitori di apparati ospitati a bordo dei robot (c.d. payload, come ad esempio apparecchiature fotografiche), i soggetti che addestrano, provvedono alla manutenzione dei robot, o li pilotano;
- b) Soggetti che commercializzano i robot e i relativi servizi nei confronti dei clienti finali: tali soggetti possono svolgere anche parte dell'attività di cui al punto a);
- c) Clienti finali, che possono essere sia imprese che consumatori: acquirenti dei robot o dei relativi servizi;
- d) Autorità pubbliche di controllo o di regolazioni del settore;
- e) Società che forniscono servizi di assicurazione contro i rischi derivanti dall'uso dei robot;

⁷⁹ Riprendo lo studio del Centro Nexa for Internet & Society, *Il diritto della robotica di servizio*, 2015

- f) Terzi estranei alle categorie sopra elencate, che possono essere danneggiati od ottenere benefici dall'uso dei robot, nonostante non siano entrati in relazione diretta con il fornitore di servizi di robotica.

I principali scenari nell'ambito dei quali si manifestano rischi di responsabilità civile in capo agli agenti coinvolti nella produzione, commercializzazione e utilizzo dei robot possono essere riassunti in cinque categorie:⁸⁰

- 1) Progettazione e produzione delle macchine;
- 2) Realizzazione e messa a disposizione di beni e servizi funzionali alla progettazione e produzione della macchine (es. fornitori dei componenti dei robot, fornitori di piattaforme di servizio da utilizzare per la progettazione);
- 3) Commercializzazione dei robot;
- 4) Prestazione di servizi attraverso i robot;
- 5) Prestazione di servizi di deposito, conservazione e manutenzione dei robot.

Per ciascuno di tali scenari rileveranno di volta in volta svariate ipotesi di responsabilità extracontrattuale, che qui di seguito esplicheremo.

E' rinvenibile però un tratto comune negli scenari sopra individuati, e cioè che l'intera materia della r.c. (acronimo di responsabilità civile) si presta a suggestivi scenari di imputazione della responsabilità alle macchine, in considerazione di una loro supposta capacità di autodeterminazione ed a considerazione *de jure condendo* sul

⁸⁰ Quadripartizione operata da C.M. Bianca in *Diritto civile*, vol. 5, La responsabilità, Giuffrè, 2012.

medesimo argomento. E' il tema legato, più in generale, ai principi del rapporto di causalità, dell'imputato del fatto dannoso (ex. Art. 2046 c.c.) e dell'elemento soggettivo (ex art. 2043 c.c.), quali presupposto dell'esistenza della responsabilità civile e della risarcibilità dei relativi danni, ai quali fanno da contrappunto il principio della causalità omissiva, le disposizioni che regolano la responsabilità di padroni e committenti e la responsabilità oggettiva.

Necessario, ai fini di questo discorso, è evidenziare che il nostro ordinamento individua quali possibili centri di imputazione della responsabilità civile unicamente i soggetti che ad oggi sono ritenuti tali dall'ordinamento italiano, cioè le persone fisiche (artt. 1 e ss. del c.c.) e le persone giuridiche (artt. 11 e ss. del c.c.); ciò a dire che l'imprevedibilità delle azioni dei robot che discenda dalla loro interattività ed autonomia, non può, *legibus sic stantibus*, fondare una "responsabilità delle macchine" nel nostro ordinamento per l'elementare ragione che tali entità non rientrano (ancora) tra i possibili centri di imputazione di responsabilità previsti dalla legge.

Pertanto le conseguenze giuridicamente rilevanti che derivano dall'imprevedibilità delle azioni dei robot saranno sussumibili nella categoria del caso fortuito/forza maggiore ovvero costituiranno fonte di responsabilità risarcitoria in capo a colui (persona fisica o giuridica) o a coloro ai quali potranno essere ricondotte sulla base delle regole di imputabilità (soggettiva o oggettiva) previste dall'ordinamento⁸¹.

⁸¹ Per l'analisi del seguente paragrafo faccio riferimento a: F. Bocchini ed E. Quadri, *Diritto privato*, quarta edizione, Giappichelli, Torino, 2011; G. Branca, *Sulla responsabilità oggettiva per danni cagionati da animali*, in Riv. Trim Dir. Proc. Civ. 1950, p. 225; G. Chinè, M. Fratini e A. Zoppini, *Manuale di diritto civile*, IV edizione, nel Diritto Editore, 2013; F. Galgano, *Diritto civile e commerciale*, vol. 2, Cedam, Padova, 2004; A. Torrente e P. Schlesinger, *Manuale di diritto privato*, Giuffrè Editore, 2011; e infine P. Zatti e V. Colussi, *Manuale di istituzioni di diritto privato*, Cedam, Padova, 2010.

Ai fini di un discorso maggiormente dettagliato sulla responsabilità civile, occorre innanzi tutto, come già accennato, distinguere tra responsabilità contrattuale e responsabilità extracontrattuale. La differenza di fondo che contrappone le due risiede nel fatto che la prima sanziona l'inadempimento di un'obbligazione già esistente, che trova la fonte in un contratto o in altro atto giuridico idoneo a far sorgere un obbligo giuridicamente rilevante in capo ad un soggetto ed un equipollente diritto in capo al destinatario a che tale obbligo venga effettivamente posto in essere; con la responsabilità extracontrattuale, invece, si sanziona un fatto illecito dannoso, il quale diventa per ciò stesso fonte di risarcimento. Secondo l'art. 1218 c.c.: "Il debitore che non esegue la prestazione è tenuto al risarcimento del danno, se non prova che l'inadempimento o il ritardo è stato determinato da impossibilità della prestazione derivante da causa a lui non imputabile". Si tratta della classica responsabilità in cui incorre il debitore. La prestazione viene adempiuta in modo inesatto, in maniera non adeguata alla previsione o non viene adempiuta affatto. Il creditore dovrà quindi provare che l'obbligazione inadempita era effettivamente esistente. Il debitore, per parte sua, dovrà fornire la prova che non ha avuto colpa nell'inadempimento.

Se avviene un ritardo nella prestazione l'inadempimento risulterà solo temporaneo: la prestazione verrà eseguita in un momento successivo rispetto a quanto pattuito, arrecando un danno in capo al creditore. Se viene provato che il debitore nell'impossibilità della prestazione non ha responsabilità, se non vi è quindi una causa imputabile nei suoi confronti, l'obbligazione si estingue e il debitore è liberato. Occorre però che l'impossibilità risulti assoluta, che non consenta cioè di adempiere in alcun modo: non basta una difficoltà nell'adempimento,

ma è necessaria una oggettiva impossibilità per chiunque, non solamente per il debitore. Sebbene a prima vista potrebbe sembrare che la responsabilità del debitore inadempiente derivi dalla mancata, inesatta o tardiva esecuzione della prestazione e che quindi il debitore sia responsabile in ogni caso, non è così in quanto occorre leggere l'art. 1218 c.c. alla luce dell'art. 1176 c.c.: "Nell'adempiere l'obbligazione il debitore deve usare la diligenza del buon padre di famiglia. Nell'adempimento di obbligazioni inerenti all'esercizio di un'attività professionale, la diligenza deve valutarsi con riguardo alla natura dell'attività esercitata". Questo articolo obbliga il debitore a comportarsi secondo una diligenza sommariamente prestabilita, senza imporgli però di fare qualsiasi cosa per evitare l'inadempimento. Il debitore non deve perciò "fare l'impossibile", ma agire nei limiti della diligenza.

Quindi il debitore sarà responsabile per inadempimento se è dimostrata la sua colpa.

Le due responsabilità differiscono anche quanto a disciplina.

- A) Mentre la responsabilità contrattuale non presuppone la capacità di intendere e di volere dell'obbligato (né tanto meno la sua capacità di agire), la responsabilità extracontrattuale richiede, di regola, la capacità di intendere e di volere dell'autore dell'illecito dannoso (art. 2046 c.c.)
- B) Mentre la responsabilità contrattuale importa la risarcibilità prevedibile nel tempo in cui è sorta l'obbligazione, salvo in ipotesi di dolo del debitore- danneggiante (art. 1225 c.c.), la responsabilità extracontrattuale importa la risarcibilità sia dei danni prevedibili, sia dei danni imprevedibili (art. 2056 co. 1 c.c.)

- C) Mentre il diritto al risarcimento del danno da inadempimento è soggetto, salvo eccezioni, (artt. 2946- 2952 c.c.) all'ordinaria prescrizione decennale (art. 2946 c.c.), il diritto al risarcimento del danno da illecito extracontrattuale è soggetto a prescrizione breve quinquennale (art. 2947 co. 1 c.c.)
- D) Mentre nella responsabilità contrattuale il creditore-danneggiato che agisce per il risarcimento ha l'onere di provare il suo credito, il danno di cui chiede il ristoro, il nesso causale tra danno lamentato e inadempimento allegato; mentre per il debitore dovrà provare di avere correttamente eseguito la prestazione o in alternativa che l'adempimento o il ritardo è stato determinato da impossibilità della prestazione derivante da causa a lui non imputabile (art. 1218 c.c.); nella responsabilità extracontrattuale il danneggiato che agisce per il risarcimento ha l'onere di provare non solo il danno di cui chiede il ristoro ed il nesso causale tra il danno lamentato ed illecito allegato, ma anche il fatto illecito da cui tale danno scaturisce, nonché il dolo o per lo meno la colpa del danneggiante.

Con riferimento alla responsabilità extracontrattuale, l'art. 2043 c.c., come noto recita: "Qualunque fatto doloso o colposo che cagioni ad altri un danno ingiusto obbliga colui che ha commesso il fatto a risarcire il danno". Essa è una clausola generale di responsabilità in quanto appare prescindere, almeno in prima battuta, da quelle che sono le caratteristiche soggettive del danneggiato e del danneggiante, ovvero dice che in qualunque caso un danno, se avente il requisito dell'ingiustizia, cioè se cagionato immeritariamente, deve essere fonte di ristoro, economico e/o non.

Con il termine “fatto” si intende il comportamento, la condotta della persona ed il danno cagionato. Il comportamento può concernere sia un’azione che il soggetto avrebbe dovuto astenersi dal compiere, sia un’omissione: occorre però specificare che l’omissione diventa fonte di risarcimento solo se l’autore aveva il dovere giuridico di agire e non l’ha fatto.

Mentre il dolo, poi, concerne la piena coscienza ed intenzionalità dell’atto da parte del soggetto, ai fini dell’individuazione della colpa il criterio di valutazione del comportamento dell’agente è costituito dalla diligenza propria del buon padre di famiglia; con tale concetto si intende la condotta che, con riferimento alla natura dell’attività esercitata, può essere richiesta al soggetto secondo una retta coscienza sociale. Il mancato rispetto da parte dell’agente di tale criterio costituisce la cosiddetta colpa lieve. Esistono tuttavia ipotesi in cui, in relazione a particolari tipi di attività, la responsabilità può derivare solo da atti commessi dolosamente o con colpa grave, ossia con negligenza, imprudenza ed imperizia di particolare rilevanza. Tale limitazione di responsabilità è prevista, ad esempio, per il libero professionista nel caso in cui la prestazione comporti la soluzione di problemi tecnici di particolare difficoltà.

Con il termine colpa lievissima si definisce tradizionalmente il discostarsi dalla diligenza particolare richiesta rispetto a quella media prevista.

In ipotesi particolari si parla di poi di “*culpa in vigilando*”, ravvisabile ad esempio nella colpa dei genitori rispetto ai comportamenti dei propri figli minori, o ancora di “*culpa in eligendo*”, cioè quella ravvisabile nella colpa nell’atto di scegliere, ad esempio, propri dipendenti o collaboratori. Mentre nel primo caso, il soggetto può essere liberato

dalla responsabilità se prova di non aver potuto impedire il fatto adempiendo correttamente all'obbligo di vigilanza, il cosiddetto caso fortuito, nel secondo si ravvisa una responsabilità oggettiva e diretta, prescindente da una particolare colpa personale del soggetto e fondata sul rischio che, per solidarietà sociale, deve gravare sul committente in ragione dell'utilità che lui stesso trae dall'attività affidata a terzi, o sembra, più semplicemente per ragioni di economia giuridica, ossia è il committente, in genere, fornito di più mezzi, latamente intesi, che appare nella condizione più agevolata di risarcire il danno.

Il comportamento, omissivo o attivo che sia, deve essere causa del danno ingiusto: il danno deve risultare come conseguenza immediata e diretta del comportamento.

Bisogna quindi, nel caso di specie, verificare che il danno non si sarebbe prodotto in assenza di tale comportamento e che quel comportamento sia perciò stato idoneo a produrre il danno lamentato. Infine il danno deve risultare come conseguenza prevedibile e naturale di tale comportamento.

Cosa si intende per danno ingiusto?

“L'evoluzione della responsabilità civile nel senso dell'accentuazione della sua funzione riparatoria ha portato a concentrare l'attenzione non più sulla condotta ingiusta ma sull'interesse leso che, in quanto comunque reputato rilevante per l'ordinamento, viene considerato meritevole di riparazione, ove la sua lesione sia ingiustificata. L'ingiustizia viene così riferita immediatamente al danno (danno ingiusto): ai fini della riparazione del danno allora diventa determinante stabilire se l'interesse leso sia effettivamente meritevole di tutela e se la relativa lesione diventi ingiustificata. Solo in tal caso (danno non iure), infatti, della riparazione potrà essere fatto carico ad altro soggetto,

identificato sulla base di un suo comportamento colpevole (secondo la regola generale dell'art. 2043 c.c.) o della peculiare situazione in cui si trovi in relazione all'evento produttivo del danno. E' venuto quindi meno l'univoco collegamento tra risarcimento del danno e violazione di un diritto soggettivo assoluto, risultando suscettibili di reintegrazione anche interessi (detti legittimi) considerati appunto meritevoli di tutela, ma non tutelati attraverso il riconoscimento al titolare di una posizione giuridica soggettiva di alcun tipo. Viene così indubbiamente esaltato il carattere di atipicità dell'illecito civile, dato che esso si considera suscettibile di tutela attraverso lo strumento risarcitorio, al di là della sfera dei diritti soggettivi.

Inoltre il tema di responsabilità extracontrattuale oltre gli essenziali elementi testé descritti di elemento soggettivo (dolo o colpa), comportamento dannoso e nesso causale tra i due, si rinvengono poi nel nostro codice una serie di articoli in tema di responsabilità oggettiva, particolarmente rilevanti ai fini del nostro discorso. Infatti molti danno attualmente causati dai droni, per esempio, sono oggi fonte di responsabilità oggettiva verso i loro costruttori fino all'imputare all'utilizzatore la responsabilità sull'ovvio presupposto che per ora, tali dispositivi difettano del primo dei requisiti cardine in tema di responsabilità dal dolo o dalla colpa, non essendo essi considerati (ancora!) capaci di intendere e volere e quindi di provocare colposamente o dolosamente un danno.

Pertanto, specifica applicazione troveranno anche le norme in materia di responsabilità per esercizio di attività pericolose (art. 2050 c.c.), che potrà, ad esempio, venir applicato ai robot sia in considerazione delle applicazioni utilizzate sia dell'uso specifico che ne viene fatto. E' altresì di immediata evidenza la rilevanza dell'art. 2054 c.c. in materia

di circolazione di veicoli senza rotaie, vizi di costruzione o difetti di manutenzione e conseguente responsabilità del conducente o del proprietario del veicolo. Un altro articolo pertinente risulta essere il 2049 c.c. che concerne la responsabilità di padroni e committenti per i danni arrecati da fatto illecito commesso dai loro sottoposti; per i nostri agenti potrebbe essere invocato nel caso di nolo di droni “a caldo” con operatore.

Spazio importante troverà poi l’art. 2051 c.c. relativo al danno cagionato da cose in custodia, norma che potrebbe essere destinata a neutralizzare i tentativi di de-responsabilizzazione del proprietario di robot, l’art. 2052 c.c. sui danni cagionati da animali che per il nostro ordinamento sono imputabili al proprietario sia che questi si trovino sotto la sua custodia, sia che siano smarriti o fuggiti, salvo provare il caso fortuito.

Infine autonomo rilievo avranno i dispositivi del c.c. che concernono le disposizioni generali in tema di solidarietà nel caso di imputabilità del fatto dannoso a più persone e dei criteri di quantificazione del danno.

In seconda battuta, risultano di estrema rilevanza le norme che disciplinano l’attività e la responsabilità dei soggetti che a vario titolo contribuiscono alla realizzazione e all’utilizzo dei robot e che possono pertanto fondare specifiche ipotesi di responsabilità aquiliana tipica. Particolarmente rilevante per la materia appare la responsabilità del venditore per i danni derivanti dai vizi della cosa venduta ex art. 1494 co. 2 c.c., che risulta applicabile anche nel caso di vizi delle singole componenti del robot, che possono quindi essere contestate direttamente dall’utente finale nei confronti del fornitore del componente a titolo di responsabilità contrattuale.

Sono poi certamente applicabili in numerosi casi le norme in materia di responsabilità per danno da prodotti difettosi e in materia di sicurezza dei prodotti, di cui agli artt. 102 e ss. del D. Lgs. n. 206/2005, codice del consumo, quindi la disciplina normativa e regolamentare applicabile ai robot, ai suoi componenti e ai relativi servizi accessori, quale ad esempio il D. Lgs. n. 17/2010 che attua la direttiva 2006/42 relativa alle macchine. Va poi considerato il regolamento ENAC del dicembre 2013 sui mezzi aerei a pilotaggio remoto: tale regolamento però disciplina soltanto una categoria di robot c.c. droni ovvero “SAPR” – Sistema Aeromobile a Pilotaggio Remoto o anche “UAV”, unhumaned aerial vehicles.

Queste clausole insieme ad altri elementi giuridici forniscono agli operatori un contesto specifico e dettagliato nell’ambito del quale, per queste specifiche categorie di robot, è possibile dettagliare in concreto la gestione delle responsabilità connesse alla produzione e all’uso dei droni; tale assetto è particolarmente interessante in quanto può essere utilizzato anche per implementare, nei limiti del possibile, soluzioni analoghe per la gestione di altre tipologie di robot.

Si rinvengono poi le normative tecniche standard come le ISO, che costituiscono parametri di riferimento qualitativi essenziali, per fondare o scongiurare l’imputazione del fatto dannoso.

Per quanto riguarda poi il tema della ripartizione delle responsabilità tra i soggetti agenti nel sistema di interazione, in qualunque senso intesa, con i robot, tali fatti costitutivi di responsabilità possono derivare:

- a) Da vizi “tradizionali” del prodotto (malfunzionamenti nell’hardware, errori del software, errori di progettazione);

- b) Dalla programmazione e dall'impostazione del funzionamento del robot, che possono essere di carattere generale o speciale (errate impostazioni della specifica missione);
- c) Dal pilotaggio;
- d) Dallo specifico contesto operativo dei robot (ambiente di funzionamento, condizione metereologiche, e così via. A tal fine rilevano in particolare i luoghi dove i robot operano e dove vengono prestati i relativi servizi: spazi chiusi privati, luoghi aperti al pubblico, spazi pubblici);
- e) Da fatti dei quali il soggetto imputabile non ha alcun controllo (in caso di responsabilità oggettiva o quasi oggettiva derivante ad esempio da cose in custodia, esercizio di attività pericolosa, responsabilità di padroni e committenti).

Per gli operatori del settore il punto critico sta nel fatto di identificare preventivamente il soggetto sul quale possono ricadere le conseguenze della responsabilità civile derivanti dall'uso dei robot e quindi la loro allocazione. Appare opportuno quindi specificare, sulla scorta di quelle già esistenti, clausole di esenzione di responsabilità, meccanismi di rivalsa nonché casi di responsabilità solidale per l'ambito della robotica.

L'aspetto che sottolineo essere cruciale nella nostra riflessione, e cioè l'allocazione più efficiente delle responsabilità delle azioni dannose commesse da agenti artificiali su soggetti, tali definiti dal diritto, che possono rispondere di tali azioni, appare il più difficoltoso⁸². Si tratta

⁸² Basti pensare a J. Gordley, *Foundations of Private Law: Property, Tort, Contract, Unjust Enrichment*, Oxford University Press, 2006; e, nell'ambito dell'analisi economica del diritto, R. Posner, *Economic Analysis of Law*, Little

tuttavia di un aspetto fondamentale dell'analisi, in quanto è generalmente esclusa nel nostro ordinamento la possibilità di limitare preventivamente la responsabilità extracontrattuale con apposite clausole di esonero, al contrario di quello che invece può avvenire in ambito contrattuale. Pertanto, uno dei modi con cui può essere gestito il rischio derivante dalla progettazione, costruzione ed utilizzo dei robot è quello di "contrattualizzare" per quanto più possibile i rapporti con i soggetti che a vario titolo si interfacciano con l'agente.

La stipula di contratti adeguati, che gestiscano in modo appropriato e per quanto più possibile dettagliato l'allocazione del rischio e delle relative responsabilità, è un aspetto dirimente per non disincentivare l'utilizzo di queste nuove tecnologie; ma ciò deve avvenire non solo sul piano strettamente giuridico, attraverso cioè la previsione di apposite e valide clausole, ma anche sul piano tecnico, attraverso un'approfondita ed adeguata analisi preventiva delle caratteristiche tecniche, delle prestazioni, dei livelli di servizio delle macchine e di tutto quanto sia suscettibile di essere oggetto di regolamentazione e disciplina, informazione che dovranno poi essere allegate e divenire parte del contratto. Sullo stesso piano, inoltre, si pone l'acquisizione di documenti comprovanti gli elevati standard qualitativi del prodotto. La stipula, quindi, di adeguati contratti appare allo stato lo strumento attraverso il quale possono essere gestiti la stragrande maggioranza dei rapporti tra i diversi attori che partecipano e si relazionano a vario titolo con il robot: in questo modo si riduce l'area dell'alea e della conseguente imprevedibilità dell'allocazione dei costi nel caso di danni.

Brown, Boston, 1973; nonché Id. *The Jurisprudence of Skepticism*, in "Michigan Law Review", 1988, 86(5): 827-891.

Nei casi in cui non possa essere concluso un contratto (si pensi ad esempio all'interazione della macchina con terzi casualmente presenti nel raggio d'azione, o all'uso di robot da parte di enti pubblici per offrire servizi alla cittadinanza), appare comunque consigliabile per quanto possibile prevedere e cercare di creare a tavolino una relazione tra l'attore, che nella scena del risarcimento del danno non potrà che essere la persona fisica (in quanto si è visto che non è ancora possibile citare a giudizio un robot, impugnandogli la dannosità dei suoi comportamenti) ed il soggetto destinatario delle interferenze causate dalla macchina. Qualche possibile accorgimento sarebbe quello di porre annunci con appositi cartelli nell'area di azione dei robot, così che il destinatario sappia che in quel raggio d'azione non è solo, oppure mediante un controllo degli accessi all'area e tracciatura del raggio d'azione del robot.

Tali iniziative, seppur non escludano a priori la responsabilità del soggetto, potranno valere quali limitazione "indiretta" di responsabilità, in quanto costituiranno certamente elementi a favore dell'agente nella valutazione del suo grado di diligenza, nel caso questi dovesse, ad esempio, provare di avere adottato tutte le misure idonee ad evitare il danno, al fine di evitare l'applicazione dell'art. 2050 c.c. che concerne l'imputazione per l'esercizio di attività pericolose, quali oggi sono considerate alcune utilizzazioni di robot.

Il tema dell'instaurazione di una relazione preventiva con il possibile soggetto passivo del danno va inoltre letto in un'ottica di concertazione tra i diversi soggetti agenti: la strada verso un'efficace gestione del rischio passa anche attraverso l'individuazione di azioni congiunte tra i vari attori, quali possono essere ad esempio il fornitore di servizi robotici ed il proprio cliente.

In una diversa prospettiva, un altro punto cruciale nella gestione della responsabilità è la tracciabilità delle operazioni compiute attraverso i robot. RegISTRAZIONI di sistema dei parametri rilevanti di funzionamento dei robot, riprese video dell'attività svolta (con camere a bordo oppure esterne), relazioni scritte sulle missioni compiute, rapporti relativi alla situazione ambientale in cui il robot opera (ad es. le condizioni meteo), sono tutti elementi che potranno essere utilizzati per dimostrare, ad esempio, che lo scontro del drone con i tralicci elettrici è dovuto ad un calo improvviso della batteria, piuttosto che ad una manovra errata o ad un colpo di vento improvviso.

Quale strumento di gestione del rischio vi è infine la stipula di apposite clausole assicurative che coprano il soggetto utilizzatore nel caso di comprovata responsabilità. Esse sono già in parte presenti e quindi previste dalla legge ma possono essere anche scelte dall'utente di volta in volta in base alla natura ed agli utilizzi che intenda fare con la macchina.

Sul punto, è peraltro auspicabile un rapido sviluppo sul mercato di proposte diversificate che con il tempo verranno a consolidarsi in una serie di polizze "standard".

Infine, terminiamo questo capitolo con un breve *excursus* circa la responsabilità penale nell'utilizzo di un'agente artificiale o semplicemente nell'interazione con lo stesso. Essa appare uno degli scenari più inquietanti che si possono presentare agli occhi di utenti o semplici terzi che entrano nel raggio di azione di macchine robotiche, le quali, come detto, possedendo i requisiti menzionati di interattività, autonomia ed adattatività possono far sorgere, conseguentemente alle proprie azioni, veri e propri reati

Tuttavia, quando si parla di responsabilità penale, sorge un'ulteriore problema giuridico di non poco conto. L'art. 27 della nostra Costituzione, infatti, si esprime in questi termini: "La responsabilità penale è personale"; esso sancisce così il principio inderogabile per il nostro ordinamento, cioè che la responsabilità da reato (fatte salve le ipotesi di responsabilità amministrativa degli enti di cui alla legge 231/2000) può essere riconosciuta esclusivamente nei confronti di una persona fisica. *Quid iuris* nell'ipotesi in cui un fatto di reato, ad esempio la lesione di terzo provocata dalla collisione di un agente artificiale con un oggetto, venga commesso da un robot che ha agito elaborando in autonomia gli input esterni che ha ricevuto?⁸³

Ed ancora, considerato che ogni reato è costituito da un elemento oggettivo, cioè dal fatto materiale, e da un elemento soggettivo rappresentato da coscienza e volontà di porre in essere un'azione o un'omissione prevista dalla legge come reato, come può essere configurato l'elemento soggettivo, cioè *l'animus nocendi* in capo ad un robot?

Delle tre caratteristiche dei robot più volte menzionate, appare sufficiente prenderne in considerazione anche solo una sola di esse: l'interattività. Essa consiste nel modo in cui l'agente artificiale risponde agli stimoli esterni dell'ambiente attraverso il mutamento degli stati interni o dei valori delle sue proprietà. Si può dire rilevare come droni o piccoli robot di servizio possano, nella inevitabile interazione tra la loro attività e la realtà fisica del mondo che li circonda, (interazione che, peraltro, rappresenta la ragione stessa della loro creazione), potenzialmente causare fatti che assumono rilevanza penale. Tale

⁸³ Per un'introduzione a questi temi si v. C. Blengino, *I reati informatici*, in M. Durante e U. Pagallo, *Manuale di informatica giuridica*, cit., pp. 219-245.

rilevanza, ovvero sia il rischio di cagionare danni a cosa o persone più o meno rilevanti, va declinato diversamente a seconda della capacità di movimento del robot (solo perché qui, per semplicità di discorso si fa riferimento ai robot umanoidi ma lo stesso può dirsi anche con altri tipologie di agenti artificiali, si pensi al giudice robot, per esempio, e alle varie forme di responsabilità in cui, seguendo l'argomento della sua autonomia, potrebbe incorrere).

In tal senso è quindi immediatamente evidente che un robot che svolge un'attività di telepresenza in un locale chiuso con limitata capacità di movimento è potenzialmente meno pericoloso nei confronti del mondo fisico che lo circonda rispetto ad un drone che svolge attività di sorveglianza su un centro abitato.

Dal punto di vista penalistico, è opportuno precisare che l'eventuale danneggiamento di beni materiali da parte di un robot, non integra alcun reato: il delitto di danneggiamento di cui all'art. 635 del codice penale punisce la condotta di chi dolosamente "distrugge, disperde, deteriora o rende, in tutto o in parte, inservibili cose mobili o immobili altrui". In altri termini, non è previsto nel nostro ordinamento un reato di danneggiamento colposo, ipotesi che ricorrerebbe nel caso di danni da oggetti materiali arrecati da un robot.

Diversa la situazione nel caso di danni arrecati a persone fisiche, condotta che viene qualificata come reato di lesioni personali punita sia a titolo di dolo che a titolo di colpa (di cui agli artt. Rispettivamente 589 e 590 c.p). Anche qui come nel caso della responsabilità civile, l'allocazione dei costi che da essa derivano per i danni fisici arrecati dall'interazione di un agente artificiale con uno o più esseri umani che entrano nell'alveo delle sue dinamiche motorie ed informatiche sarà determinata dall'analisi del singolo fatto concreto: a seconda della

causa d'azione del robot, escludendo quindi dialetticamente una sua capacità di agire in piena autonomia, la responsabilità potrebbe confluire in capo al produttore oppure al gestore della piattaforma operativa oppure ancora al pilota ad esempio nel caso di caduta a terra improvvisa di un robot, non dovuta a malfunzionamento interno.

Un secondo reato, che si fa palese soprattutto con l'utilizzo di determinate categorie di robot, concernente questa volta la tutela non dell'*habeas corpus* ma dell'*habeas data*, è quello concernente il trattamento illecito di dati personali disciplinato dall'art. 167 del D.lgs n. 101 /2018. Tale articolo, così come modificato, mantiene la medesima ossatura fondamentale rispetto a quelle del previgente Codice del 2003 ed è ugualmente rubricato. Ciò nonostante si rinvengono alcuni ritocchi. Gli elementi comuni ed innovativi che si reiterano in molte fattispecie oggi previste sono la clausola di esclusione ("Salvo che il fatto costituisca più grave reato"); il "chiunque" che amplia la portata applicativa, dal punto di vista soggettivo, precedentemente limitata al "titolare, al responsabile e alla persona designata ex art. 2 terdecies", convertendolo, così, in reato comune; e, soprattutto, l'introduzione del dolo specifico, che ricomprende non solo il fine lucrativo ("al fine di trarre profitto") ma anche il "nocimento dell'interessato".

Ci siamo ora soffermati brevemente sull'esegesi dell'articolo semplicemente con lo scopo di cogliere la ratio sottesa alle modificazioni avvenute a seguito dell'introduzione del Regolamento Europeo più volte menzionato e alla successiva introduzione del decreto summenzionato ed i relativi reati in esso contenuti.

Naturalmente altri e numerosi sono i reati relativi ad acquisizioni fraudolente di dati personali e loro utilizzi distorti, ne abbiamo citati

due in particolare per rendere chiaro il dato che gli agenti artificiali hanno un modo prioritario di ledere i nostri diritti personali, e tale modo è la violazione della normativa sulla privacy e sulla protezione dei dati personali. Quello che si vuole evidenziare in questa sede, è che i problemi che pongono, poniamo i cosiddetti giudici robot quando sentenziano (sia essa sentenza definitiva o atto endoprocessuale) trattano dati e spesso tanti, troppi dati per il fine che sono chiamati a svolgere, in tal modo la violazione dei dati personali, che è classificabile come reato autonomo spesso funge anche da quel tipo di reato che i penalisti definirebbero come reato-scopo, ossia la condotta dolosa che si realizza al fine di commettere un altro, talvolta più grave reato come per esempio i ben noti casi di discriminazione occorsi con i primi algoritmi predittivi (vedi caso Compass di cui si parlerà in seguito).

In conclusione, come abbiamo più volte fatto notare nel corso della trattazione sulla responsabilità, la creazione, la gestione e l'utilizzo dei robot può potenzialmente coinvolgere un numero indeterminato di persone: il progettista, il produttore materiale anche di un singolo componente, l'assemblatore, il gestore della piattaforma cloud che fornisce il servizio per la gestione dell'agente artificiale, fino all'utente finale, al pilota in casi di droni o a terzi che nulla, a rigore, centrerebbero con le fasi suddette di produzione ed utilizzo del robot.

La domanda che appare utile porci è pertanto come e per quali profili tali soggetti possono essere chiamati a rispondere penalmente?

Purtroppo, come sempre capita nei settori del diritto sottoposti a rapida evoluzione, come quello delle nuove tecnologie, non esiste e non esisterà mai una risposta univoca; infatti, saranno proprio le nuove tecniche ad "obbligare" il diritto a farle fronte e quindi a determinare la

giurisdizione, dal campo civile al ben più arduo penale, ma la continua evoluzione farà sempre sì che ciò che è stato fatto non basterà.

Per quel che in questo paragrafo interessa, possiamo dire, generalizzando, che i profili di responsabilità dovranno essere ricostruiti, caso per caso, a seconda delle circostanze di fatto, che hanno determinato, nella fattispecie in esame, la lesione di uno degli interessi protetti dal nostro ordinamento giuridico sopra delineati.

Ancora un volta un esempio potrà essere utile per concretizzare il discorso: ipotizzando che un drone cada rovinosamente al suolo colpendo un comune cittadino che passeggia sulla pubblica via cagionandogli lesioni gravi. Chi sarà penalmente responsabile dei danni subiti da tal soggetto?

La risposta, ancora una volta, dipende. Dipende cioè dai risultati di un'accurata analisi della causa che ne hanno determinato la caduta. Questa analisi, verosimilmente costituirà oggetto di una consulenza tecnica il cui esito consentirà di individuare le cause e di conseguenza i soggetti potenzialmente responsabili. Nel caso in cui, ad esempio, la causa della caduta venga ricondotta alla rottura di un componente meccanico dovrà essere valutata la posizione del progettista/produttore, mentre laddove la causa venga attribuita alla perdita della connessione internet con cui il drone comunica con la piattaforma di controllo a terra dovrà essere, invece, esaminata la posizione del gestore che ha fornito il servizio e del programmatore che non ha previsto l'ipotesi di perdita di connessione.

Importante notare che, l'individuazione della causa e del soggetto che risponde dello specifico settore coinvolto nel sinistro, ovviamente, non rappresenta ancora la "soluzione del caso". Trattandosi di responsabilità penale colposa dovrà, infatti, essere analizzato lo

specifico comportamenti per verificare se il soggetto aveva previsto adeguate misure di sicurezza atte ad evitare l'evento e, in caso affermativo, per quali motivi tali misure non abbiano nel concreto funzionato. Argomentazioni analoghe valgono, come si è detto, per il trattamento illecito di dati personali in cui l'allocazione della responsabilità va calibrata sulla scorta dell'attività posta in essere in concreto dai soggetti che, ai sensi del GDPR 679/2016 possono essere configurati come "titolare del trattamento", ovverosia coloro che decidono finalità e mezzi del singolo trattamento dei dati personali.

Infine, l'art. 41 del codice penale stabilisce il principio secondo cui il "concorso di cause preesistenti, simultanee o sopravvenute, anche se indipendenti dall'azione od omissione del colpevole, non esclude il rapporto di causalità fra l'azione od omissione e l'evento". Ciò in pratica significa che in ogni singola fattispecie concreta, sarà necessario analizzare l'eventuale sussistenza di concause con conseguente estensione dei profili di responsabilità penale in capo a più soggetti, i quali dovranno rispondere del reato per la parte di comportamento colposo loro attribuibile.

Si nota pertanto come il diritto, di fronte ai cambiamenti tecnologici in corso ed in generale di fronte ad una società che si fa di volta in volta più complessa, si trova in forte difficoltà a normare profili, soggetti e situazioni che però necessitano di essere "ricoperti" dal velo del diritto per evitare il far west o peggio disincentivare la produzione e l'utilizzo della tecnologia che non sempre, ma in molti aspetti sicuramente, è fonte di aiuto e valore aggiunto nella vita di ognuno di noi, nello svolgimento della professione ma anche quando ne siamo semplici destinatari passivi.

2.3 Una questione di coscienza

Quando parliamo di agenti artificiali, siano essi robot, robot umanoidi, droni, software ecc e quando descriviamo le loro azioni e ne tentiamo di definire i parametri di responsabilità diciamo in modo implicito che gli stessi, in un qualche modo agiscono e quindi pensano.

Questo sillogismo, che abbiamo già affrontato nei capitoli iniziali del presente lavoro, ci ha condotti a ragionare convenientemente in termini di scissione (lo abbiamo definito “divorzio”, dalla nota metafora utilizzata dal Prof. Luciano Floridi), in quanto, come lo studioso ci fa osservare, la capacità di agire è altro rispetto alla capacità di pensare e agire pensando, ossia capire perché agisco in quel modo e non in altro rispetto al mondo che mi circonda. Ebbene questa scissione, che appare evidente in molti aspetti ma tendiamo a dimenticarla in altri, ci porta al ragionamento ben noto tra gli studiosi dell’informatica giuridica e del diritto delle nuove tecnologie in generale, circa le proprietà ontologiche delle creature artificiali⁸⁴ e quindi, in primis, ai discorsi sulla loro coscienza, a partire dal tentare di comprendere cosa sia da intendersi per coscienza anche in un essere umano.

Osserva Santosuosso: “Resta, a questo punto, da affrontare il rilievo ontologico secondo il quale un robot, o un sistema di AI, non ha quei requisiti (come coscienza, Intenzioni, capacità di formulare piani o di darsi obiettivi) che fanno di una persona una persona nel senso generalmente accettato di “persona umana”, con la connessa attribuzione di diritti e doveri. Questa qualità ci riporta alla definizione

⁸⁴Riprendo qui la definizione che ne dà A. Santosuosso nel suo *“Intelligenza artificiale e diritto, perché le tecnologie di AI sono una grande opportunità per il diritto”*, Mondadori Università, 2020.

di AI ossia: “L’intelligenza artificiale è il campo dedicato alla costruzione di animali artificiali (o almeno, creature artificiali che i contesti appropriati appaiono essere animali) e, per molti, persone artificiali (o, almeno, creature artificiali che in contesti appropriati appaiono essere persone)”.

Continua Santosuosso, il punto di collegamento tra qualità ontologiche (e biologiche) umane a proposito di coscienza e la dimensione giuridica intesa in senso lato è la Dichiarazione Universale dei Diritti dell’Uomo del 1948⁸⁵ che stabilisce la piena simmetria tra l’entità che ha l’insieme dei diritti e delle libertà e l’essere umano. Tuttavia, le cose sono più complesse di quanto possa apparire ed è facile osservare che, da un lato, tale simmetria non è esistita prima della Dichiarazione del 1948 (tanto da renderla, appunto, necessaria) e, dall’altro, che è lo stesso art. 1 a riaprire immediatamente la questione quando, nella sua seconda parte, afferma: “Loro [tutti gli esseri umani] sono dotati di ragione e coscienza”.

L’affermazione che tutti gli esseri umani “sono dotati di ragione e coscienza”, presa alla lettera, sembra supporre che la ragione e la coscienza siano parte dell’essenza di tutti gli esseri umani e, quindi, costituiscano loro qualità intrinseche, senza le quali verrebbe meno il carattere dell’umanità. Se così fosse, la conseguenza sarebbe che un essere umano che manchi di tale qualità non sia un essere umano o non

⁸⁵ La Dichiarazione universale dei Diritti dell’Uomo fu adottata dalla Risoluzione dell’Assemblea Generale 271A nella sua terza sessione a Parigi il 10/12/1948: “Tutti gli esseri umani sono nati liberi e uguali in dignità e diritti” (articolo 1) e “A ogni individuo spettano tutti i diritti e le libertà enunciati nella presente Dichiarazione, senza distinzione alcuna per ragioni di razza, colore, sesso, lingua, religione, opinione politica o di altro genere, origine nazionale o sociale, ricchezza, nascita o altra condizione” (articolo 2).

sia un essere umano completo o non abbia il diritto di avere uguali dignità e diritti⁸⁶.

Su questa base può essere formulato un lungo elenco di ulteriori questioni su quale sia il significato di “ragione e coscienza”, sulla possibilità della loro presenza in artefatti non umani e sullo spazio da riservare alle entità non-umane che, in ipotesi, ne abbiano un certo grado. Si pensi, ma solo per fare un esempio, alla questione della società e dello status giuridico e a quanto essa sia stata storicamente la principale palestra della teoria del diritto a proposito di chi non veniva considerato umano e quindi titola di diritti e doveri.

Per esaminare il problema bisogna tenere conto di quattro aspetti essenziali. Il primo riguarda lo stato delle conoscenze in materia all’epoca della redazione della Dichiarazione Universale e gli obiettivi dei suoi estensori, il secondo, i tempi di sviluppo negli anni del dibattito sulla coscienza negli umani, il terzo riguarda le scoperte neuroscientifiche a proposito dell’attività mentale degli animali non umani, il quarto l’acquisizione che non esiste un solo tipo di coscienza i cui gradi possono essere misurati secondo un’unica scala, ma esistono diversi tipi di coscienza.

Sul primo punto, quello delle idee degli estensori della Dichiarazione, è interessante notare che vi fosse disaccordo circa l’idea di indicare ragione e coscienza tra i requisiti degli umani che per taluni essa fosse ridondante o fonte di fraintendimenti. Secondo la visione essenzialista, ad avviso di uno dei più importanti commentatori della dichiarazione, la ragione e la coscienza sono indicati come attributi dell’essere umano: averli qualifica per l’appartenenza alla famiglia umana, non averli rende

⁸⁶ Ma così non appare dall’analisi dei lavori preparatori. Sul tema e sui rilievi della ragione e coscienza negli artefatti non umani invio a Santosuosso 2015°.

questa appartenenza discutibile. Questa posizione è sostenuta da Charles Malik, secondo il quale “ The First article of the Declaration on Human Rights should state those characteristics of human beings which distinguish them from animals, that is reason and conscience”⁸⁷.

Tuttavia il testo fu alla fine approvato per due prevalenti ragioni: evitare un eccessivo riferimento alla cultura occidentale e porre il confine degli animali non umani, che costituivano il limite oltre il quale l’umanità cessava di esistere. Un punto di vista che oggi si definirebbe specista, ma che non sorprende che ci fosse all’epoca, perché la rivendicazione di diritti per gli animali risale solamente alcune decadi successive e la conoscenza scientifica sulla percezione e le forme di coscienza negli animali non umani arriverà, anch’essa, solo più tardi.

Che questa fosse la situazione del dibattito all’epoca della redazione della Dichiarazione non deve sorprendere, se si considerano i tempi di evoluzione del dibattito sulla coscienza negli umani. Roberto Van Gulik, nella sua rassegna filosofica sull’argomento sottolinea come “nonostante la rinnovata enfasi nello spiegare le capacità cognitive come la memoria, la percezione e la comprensione del linguaggio, la coscienza sia rimasta un argomento ampiamente trascurato per molti ulteriori decenni” ed è stato solo negli anni Ottanta e Novanta del XX secolo che “c’è stata una forte ripresa della ricerca scientifica e filosofica sulla natura e sulle basi della coscienza. Una volta che la coscienza è tornata in discussione, c’è stata una rapida proliferazione della ricerca con un’ondata di libri e articoli, nonché la creazione di riviste specializzate”. In termini di interpretazione giuridica ciò implica che l’attuale ampiezza semantica del lemma coscienza deve essere

⁸⁷ J. Morsink, *The Universal Declaration of Human Rights: Origins, Drafting, and Internet*, University of Pennsylvania Press, Philadelphia, 1999.

principalmente determinata facendo riferimento a concetti e idee di un'era successiva a quella degli estensori. Il problema (che di per sé è frequente nell'interpretazione delle leggi, che acquisiscono nel tempo concetti normativi nuovi) è reso più difficile dall'incertezza nel campo filosofico, dove ancora oggi le parole “cosciente” e “coscienza” sono usate con una varietà di significati e in una vasta gamma di modi senza che ve ne sia uno nettamente prevalente. Le parole “cosciente” e “coscienza” sembrano termini generali che coprono un'ampia varietà di fenomeni mentali e sono applicati sia a interi organismi- coscienza della creatura - sia a particolari stati e processi mentali – coscienza dello stato.

La coscienza negli animali non umani, che rappresenta il punto tre, è al centro dell'attenzione di un gruppo di eminenti studiosi e scienziati che nel 2012, presso il Churchill College, Università di Cambridge, ha preso una posizione netta contro l'idea che solamente gli umani posseggono il sostrato neurologico che genera la coscienza. In occasione della conferenza sulla coscienza negli umani e negli animali non umani alcuni di essi hanno scritto una “dichiarazione di Cambridge” che si chiude così:

evidenze convergenti indicano che gli animali non umani hanno i sostrati neuroanatomico, neurochimico e neurofisiologico insieme alla capacità di mostrare comportamenti intenzionali. Conseguentemente il peso delle prove indica che gli umani non sono gli unici a possedere il sostrato neurologico che genera la coscienza. Gli animali non umani, inclusi tutti i mammiferi e gli uccelli, e molte altre creature inclusi i polpi, posseggono anche questi strati neurologici.⁸⁸

⁸⁸ Dichiarazione di Cambridge, disponibile online.

Naturalmente non tutti gli scienziati sono d'accordo e ve ne sono alcuni eminenti che non condividono la radicalità o alcuni presupposti di questa posizione. Tuttavia, da un punto di vista etico e giuridico, non era immaginabile, solo pochi anni prima, una più forte confutazione dell'idea comune condivisa dagli estensori della dichiarazione universale. Essa, e comunque il dibattito intorno ad essa, è chiaramente connessa all'imponente sviluppo nelle ultime decadi degli studi neuroscientifici.

Ma se la coscienza, o almeno uno stato mentale che possa rientrare nell'idea di coscienza, non è confinabile ai soli esseri umani ci si può chiedere, ed è il nostro quarto punto, se esista un solo tipo di coscienza in natura.

Da un punto di vista evolutivo, la coscienza è un tratto che alcuni animali hanno. La questione è come esso sia distribuito e quali qualità abbia. Alcuni autori sostengono che la coscienza si può manifestar in un ampio spettro di gradi di ricchezza e di pienezza. In questo senso è rilevante la differenza tra le teorie della discontinuità: secondo le *discontinuity theories* vi è un punto particolare al quale la coscienza si è originata; le teorie della continuità (*continuity theories*) invece, concettualizzano l'evoluzione della coscienza in termini di “una transizione graduale nella coscienza dal non riconoscibile al riconoscibile”.⁸⁹ Se si assume un punto di vista da discontinuità (seguendo autori come John Searle, Daniel Dennet e Giulio Tononi, che rifiutano l'idea che la coscienza sia binaria: sì /no) la questione diventa

⁸⁹ C. Allen e M. Trestman, *Animal Consciousness*, in *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition).

quella di quali dimensioni e quali soglie qualitative possano consentire di distinguere differenti tipi di attività mentale.

Un recente studio sulle origini evoluzionistiche e genetiche della coscienza mette in luce le diverse traiettorie evolutive che possono essersi verificate. Ci sarebbe una comune base genetica per la coscienza, non condivisa tra tutti i differenti sensi, ma piuttosto all'interno dei sistemi sensori individuali attraverso i gruppi di vertebrati. La coscienza stessa sembra essere avanzata attraverso differenti percorsi nelle differenti linee dei vertebrati.

Tutto ciò alimenta le teorie dalla discontinuità e mette in discussione il concetto di una scala unica delle facoltà mentali degli animali. Quindi, alla domanda se la coscienza sia apparsa prima negli umani o in un precedente ascendente non umano, e chi/cosa fosse quest'ascendente, si può rispondere che “la coscienza può essere apparsa molteplici volte, come il volo alare, che ha avuto evoluzioni indipendenti in insetti, uccelli, pipistrelli e pterosauri”⁹⁰.

Alla luce di quanto esposto, può essere sostenuta l'idea che il diritto sia un artefatto (almeno nel senso di prodotto storico) e che anche le caratteristiche (ontologiche) richieste per il riconoscimento di diritti e libertà sono artefatti, nel senso di selezione sulla base della conoscenze scientifiche e delle scelte sociali. Dall' altro lato, quello giuridico, i diritti umani potrebbero essere meglio qualificati come “diritti di base”, che è un'espressione più neutra, in grado di comprendere anche entità (parzialmente o totalmente) non umane e di risolvere l'apparente ossimoro di diritti umani attribuiti a entità parzialmente o totalmente artificiali.

⁹⁰ C. Allen e M. Trestman, vedi nota supra.

Per quanto possa apparire paradossale, solo un diritto artificiale, nel senso detto e frutto di scelte consapevoli, può umanizzare le tecnologie di AI. In questo senso, l'AI obbliga a assumersi alcune responsabilità. Se dobbiamo immaginare una società popolata da molte entità non necessariamente totalmente umane, dobbiamo adottare un approccio inclusivo (anziché l'idea di usare il diritto come difesa della società esclusiva – umana – contro macchine e alieni). Esso offre i grandi vantaggi di conoscere, comprendere e dare una cornice giuridica e sociale ai rapporti tra *homo sapiens sapiens*, animali non umani e entità parzialmente o totalmente tecnologizzate.

Anche questo approccio è, ovviamente, discutibile. Stephen Hawking, per esempio, era tra coloro i quali pensavano che l'intelligenza artificiale potrebbe essere un vero pericolo in un futuro non troppo lontano, perché i robot potrebbero progettare miglioramenti per sé stessi e superare in astuzia tutti noi. Non credo sia una giusta previsione, ma, anche se lo fosse, essa costituirebbe un'ulteriore buona ragione per discutere dei diritti di base e delle responsabilità degli agenti artificiali.

Capitolo III

L'intelligenza artificiale applicata al diritto: il giudice robot

Nei capitoli precedenti ci siamo occupati degli sviluppi dell'intelligenza artificiale e di come essa si rapporta all'essere umano prima ed al diritto poi e di come spesso queste due relazioni siano rese complicate da un'evoluzione molto rapida della tecnica e da un conseguente lentezza del diritto nel normare una materia che necessita

di punti fermi per dare fiducia alla società che quindi sceglierà di avvalersene, ma nel contempo di non imbrigliarla in eccessivi formalismi legislativi attraverso i quali diventerebbe ardua qualsiasi evoluzione successiva.

In questo terzo capitolo, che è anche il centro focale del presente lavoro, ci occuperemo di un argomento tanto dibattuto tra gli addetti ai lavori quanto poco conosciuto a non giuristi o informatici, e cioè dell'uso degli algoritmi nella giustizia. Tale uso, lo evidenziamo fin dal principio deve essere inteso come ausilio agli operatori del diritto, siano essi giuristi, avvocati, giudici o notai, nel ricercare atti, norma, sentenza o dati importanti per il loro lavoro, sia intesa nel sua forma più evoluta di una vera e propria giustizia predittiva, vale a dire di algoritmi che non solo fungono da aiuto al giurista nello svolgere la propria professione ma tendono addirittura a sostituirlo, come il caso del giudice robot di cui ampiamente si darà.

Occorre sottolineare fin da subito che tale sostituzione è ben lungi dall'essere in atto ed anzi chi scrive la ritiene un passo se non impossibile, per lo meno allo stato attuale dell'avanzamento tecnologico difficile da realizzare per tutta una serie di questioni di cui in seguito si darà conto, problematiche che vanno da una forte difficoltà ad immettere in un algoritmo tanti dati quanti sono quelli che costituiscono la base per una ponderata decisione giudiziale, come leggi, regolamenti, atti di soft law, precedenti giurisprudenziali ma anche logica, capacità di sussunzione della norma dal caso concreto alla legge astratta e non da ultimo quella caratteristica a nostra avviso imprescindibile del giudicare, soprattutto nella materia penale che è l'empatia dell'organo giudicante, intesa essa non come sentimentalismo nei confronti del soggetto da giudicare ma come

intrinseca caratteristica umana del mettersi nei panni degli altri e immaginarne il pensiero, l'azione ed anche la motivazione che ha condotto all'azione eventualmente intrapresa.

Ma andiamo con ordine: innanzi tutto quanto parliamo di “giustizia digitale” o “giustizia predittiva” (come il lettore attento avrà notato le due definizioni non sono affatto sinonimi), intendiamo con la prima l'utilizzo della tecnologia, o meglio dell'intelligenza artificiale nel mondo della giustizia largamente inteso, ossia non solo come ausilio del giudice nello scrivere una sentenza, un'ordinanza o un decreto, ossia una dei tre atti finali in cui si esplica una decisione giudiziaria, oppure, in riferimento all'avvocato come prevedere il possibile esito di uno di questi tre atti del giudicante, bensì essa va intesa come ausilio per tutti coloro che operano nel mondo del diritto e della giustizia ma senza che vi sia necessariamente il carattere della previsione. Si pensi pertanto ad una banca dati giurisprudenziale che sulla base di input dati dal legale mi dia la giurisprudenza europea recente più seguita dalle corti d'appello italiane, oppure ancora, un software che mi scriva un contratto di acquisto. La seconda nozione invece, ossia quella che chiamiamo giustizia predittiva si riferisce invece all'utilizzo di quei software il cui fine ultimo è appunto quello di predire o meglio prevedere (sulla differenza dei due verbi mi soffermerò tra poco) la decisione dell'organo giudicante di una controversia allo scopo, per esempio, di evitare un causa giudiziale e privilegiare metodi di risoluzione della controversia stragiudiziale oppure ancora di evitare una seguente causa per lite temeraria. Occorre far notare sin da ora come tali strumenti, per quanto non necessitino di particolari abilità informatiche e nemmeno giuridiche, siano per ora utilizzati quasi

esclusivamente da avvocati o esperti del settore e non da comuni cittadini.

Ma in cosa consiste una decisione robotica e cosa si intende per giudice-robot?

Il verbo decidere deriva dall'unione del prefisso de = da e dal verbo latino caedere = tagliare, quindi letteralmente “decidere” significa tagliare da, eliminare, definire. Ed è proprio questa l'attività a cui è preposto il giudice, come suggerisce anche il primo comma dell'art. 527 del codice di procedura penale, individuare in via preliminare le possibili soluzioni delle questioni di fatto o di diritto riguardanti l'imputazione per poi procedere per eliminazione fino a quando non individua la soluzione dominante da sottoporre poi ad ulteriore controllo seguendo il così detto “criterio di falsificabilità”⁹¹ con cui si verifica se una certa ipotesi possa essere confutata o meno.

Quando si pensa di utilizzare i processi logici connessi alla robotica per emulare l'attività tipica del giudice, di decisione di una controversia, o dell'avvocato, nel sostenere una tesi difensiva in un caso specifico, l'informatica deve porsi ben oltre il semplice livello di immagazzinamento del dato. La complessità qui è proprio nel rapporto tra fattispecie e interpretazione della legge e conduce ad un prodotto che finisce per diventare frutto di un ragionamento logico.⁹²

⁹¹ L'espressione “principio di falsificabilità” identifica il criterio formulato dal filosofo Karl Popper per separare l'ambito delle teorie controllabili, che appartiene alla scienza, da quello delle teorie non controllabili, da Popper stesso identificato con la metafisica.

⁹² G. Mammone, *Considerazioni introduttive sulla decisione robotica*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019, p. 13.

La decisione giudiziaria diventa “robotica” quando si ricorre all’utilizzo di algoritmi sotto forma di software per svolgere quelle controversie di competenza della figura umana del giudice.⁹³

Questo accostamento di algoritmo matematico utilizzato nelle scienze giuridiche è molto meno recente di quanto si possa pensare. Nella Grecia antica, infatti, quando non vi era un problema matematico da risolvere, si effettuava un ‘offerta votiva agli dèi. Questo veniva definito in greco come *anathema*, parola che veniva accostata a quella di *thesis* che denotava invece l’azione di stabilire qualcosa anche nell’ambito delle scienze giuridiche.⁹⁴

Sembra veder realizzata la visione che ebbe Leibniz nel 1666 quando, nella sua opera *Dissertatio de arte combinatoria*⁹⁵, scrisse: “Le parti un giorno, di fronte ad una disputa, potranno sedersi e procedere a un calcolo”

Leibniz nello sviluppo iniziale del proprio pensiero filosofico ricerca una lingua comune come caratteristica universale del pensiero umano. Questa lingua potrebbe essere quella utilizzata da un giudice-robot in un procedimento di tipo logico-matematico che segue precisi algoritmi⁹⁶.

⁹³ M. De Felice, *Decisione robotica negoziale. Nuovi “punti di presa” sul futuro*. Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019, p. 179.

⁹⁴ P. Zellini, *La matematica degli dèi e gli algoritmi degli uomini*, Milano 2016, pp. 35-36.

⁹⁵ *La dissertatio de arte combinatoria* è un ‘opera giovanile di Gottfried Leibniz pubblicata nel 1666 a Lipsia. E’ una versione estesa della sua tesi di dottorato, scritta prima che l’autore intraprendesse seriamente lo studio della matematica.

⁹⁶ L. De Renzis, *Primi passi nel mondo della giustizia “high tech”: la decisione in un corpo a corpo virtuale tra tecnologia e umanità*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019, p. 139.

3.1 Giustizia, intelligenza artificiale, Unione Europea e Italia: La Carte etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari e il Programma Strategico di Intelligenza Artificiale 2022-2024 del governo italiano

Quando parliamo di giustizia robotica intendiamo quello sviluppo robotico che è avvenuto nell'attività giudiziaria, che prevede l'utilizzo di sistemi automatizzati di decisione giurisprudenziale, che usano algoritmi con un elevato grado di autonomia dell'intelligenza artificiale, che va ad emulare quella umana attraverso l'utilizzo di software sempre più sofisticati finalizzati alla risoluzione di problemi di vario ordine a seconda anche del loro grado di apprendimento.⁹⁷

Questo concetto di giustizia robotica viene oggi declinato come giustizia predittiva (predictive justice) che la Commissione Europea per l'efficienza della giustizia (CEPEJ) definisce in questo modo:

“Per giustizia predittiva si intende l'analisi di una grande quantità di decisioni giudiziarie mediante tecnologie di intelligenza artificiale al fine di formulare previsioni sull'esito di alcune tipologie di controversie specialistiche (per esempio, quelle relative alle indennità di licenziamento o agli assegni di mantenimento). Il termine “predittivo” utilizzato dalle società di legal tech è tratto dalle branche della scienza (principalmente la statistica) che consentono di predire risultati futuri grazie all'analisi induttiva. Le decisioni giudiziarie sono trattate al fine di scoprire correlazioni tra i dati in ingresso (criteri previsti dalla legge,

⁹⁷ M.R. Covelli, *Dall'informatizzazione della giustizia alla decisione robotica? Il giudice di merito*. Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019, p.131.

fatti oggetto della causa, motivazione) e i dati in uscita (decisione formale relativa, per esempio, all'importo del risarcimento). Le correlazioni che vengono giudicate pertinenti consentono di creare modelli che, qualora siano utilizzati con nuovi dati in ingresso (nuovi fatti o precisazioni introdotti sotto forma di parametri, quali la durata del rapporto contrattuale), producono secondo i loro sviluppatori una previsione della decisioni (per esempio, quale può essere le forbice risarcitoria).

Una volta inquadrato cosa si intende per giustizia robotica occorre una precisazione terminologica: se ci si sofferma un attimo sul termine giustizia predittiva si comprende come l'accento è posto sul termine "predire" ossia dal latino l'atto di preannunciare il futuro in termini divinatori o profetici; ebbene, e ne coglieremo meglio l'essenza anche dopo, avendo descritto come funzionano effettivamente gli algoritmi predittivi, sarebbe più giusto parlare di "previsione" anziché di "predizione" in quanto la macchine prevede, ossia enuncia una decisione a seguito di un'osservazione (su dati immessi). In sostanza in giudice robot non "vede" il futuro nell'emanare la decisione a guarda piuttosto al passato (per esempio al precedente giudiziario vincolante se opera in paesi di common law) e dà la parte per sentenziare.

Un'ulteriore domanda interessante è: cosa ci spinge a voler sostituire giudici umani con giudici robot?

Il Prof. Massimo Luciani⁹⁸, individua tre principali risposte a tale quesito: liberare dal lavoro gli esseri umani, erogare prestazioni più

⁹⁸ Massimo Luciani è professore ordinario di Istituzioni di diritto pubblico presso il Dipartimento di Giurisprudenza dell'Università degli Studi di Roma "La Sapienza". Già Presidente dell'Associazione Italiana dei Costituzionalisti ed Avvocato del Foro di Roma.

efficienti di quelle che sono fornite dall'uomo e infine garantire la certezza del diritto.

Per quanto riguarda l'ipotesi di liberare l'uomo dal lavoro, è necessario dare una divisione netta tra le tipologie di lavoro di cui vogliamo che il calcolatore ci liberi e quelle invece che forse permettono all'uomo una realizzazione di cui difficilmente vuole liberarsi; se infatti possiamo considerare positivamente il fatto che un robot ci liberi da tipologie di lavoro poco gratificanti e faticose, quando l'attività in cui intervengono i robot comincia ad essere un'attività che riveste un'importanza a livello sociale che garantisce soddisfazione e ci consente, in bene o in male, di esercitare del potere, le cose cambiano e diventa delicata e spesso tacciata di tecnocrazia l'introduzione e l'accettazione dell'intervento artificiale.

Questa seconda tipologia riprende sicuramente il concetto che aveva Kierkegaard di lavoro come quella di "lotta per il sostentamento". L'uomo aspira ad essere liberato da una tipologia di lavoro che nel mondo anglosassone viene identificato con le tre D⁹⁹: dull, dirty, dangerous e da un lavoro che invece di fa sentire realizzati.

Senza dubbio, quindi, un lavoratore di livello medio-alto cercherà di porsi in contrasto con la sua sostituzione da parte di un robot e questo anche se fosse dimostrato che la macchina è in grado di offrire prestazioni migliori delle sue. Questa resistenza peraltro è quasi sicuramente destinata al fallimento vuoi per un inevitabile processo di evoluzione della società, vuoi per un nuovo ruolo che l'uomo andrà a ricoprire di cui ancora al momento non è consapevole.

⁹⁹ P. Lin, K. Abney e G. Bekey, *Robot Ethics: Mapping the Issues for a Mechanized World*, in "Artificial Intelligence", California, 2011, p. 994.

Nel caso di attività umane meno complesse, ma sicuramente non necessariamente meno lesive di diritti la sostituzione dell'uomo con la macchina è già attuata. Un esempio sono stata le olimpiadi di Tokio 2020 dove per la prima volta è stato un robot a giudicare le prestazioni delle atlete di ginnastica ritmica. Per rimanere nell'ambito sportivo, molte sono ormai le discipline dove il giudizio dell'arbitro umano è subordinato a quanto deciso con l'utilizzo di supporti tecnologici come il Var nel calcio o l'occhio di falco nel tennis¹⁰⁰. La decisione della macchina è, in molti di questi casi, scevra da qualunque condizionamento di parte ed è pertanto una decisione indiscutibile. I supporti informatici agli arbitraggi hanno però contemporaneamente prodotto una rivoluzione nel mondo sportivo.

Questo ci riporta alla seconda ipotesi: l'utilizzo dei robot in quanto capaci di garantire prestazioni più efficienti di quelle fornite dall'essere umano. Una progressiva sostituzione dell'uomo con la macchina, soprattutto in ambiti dove riscontriamo condizioni di eccessivo carico di lavoro, pensiamo al campo sanitario o a quello della giustizia, sembra un intervallo di tipo necessario e provvidenziale.

L'efficacia in questo senso è intesa come economicità e rapidità nella risoluzione di problemi che questi sistemi riescono ad ottenere rispetto a quanto ottenuto con forza lavoro umana.

La lentezza del procedimento giudiziario in Italia è una problematica annosa e la richiesta di velocizzazione accompagna ogni ragionamento sulla giustizia.¹⁰¹

¹⁰⁰ L'Hawk-Eye (occhio di falco) è un sistema di moviola utilizzato comunemente nel tennis, cricket ed altri sport. Consiste nella riproduzione della traiettoria della palla e del percorso che ha seguito.

¹⁰¹ Si veda, ad esempio, C. Castelli e D. Piana, *Giustizia predittiva. La qualità della giustizia in due tempi*, *Questione Giustizia, Rivista di Magistratura Democratica*, 2019.

L'esigenza di una ragionevole durata del processo la troviamo esplicitamente espressa nell'art. 111 della Costituzione che la individua come uno dei principi necessari per avere quello che viene definito nello stesso articolo un "giusto processo". A fronte della velocità o come dice l'articolo della "ragionevole durata" troviamo poi esigenze apparentemente contrapposte o che almeno lo sono, purtroppo, nella pratica, quali il garantismo e l'effettività tra gli altri. Il giusto processo, si osserva, passa inevitabilmente da un corretto bilanciamento di tutti questi principi costituzionali.

Rendere più veloci i processi, significa nella pratica, come detto sopra, andare a discapito della qualità che si avrebbe invece in un tempo più lungo, tempi troppo brevi vanno a detrimento delle garanzie e del contraddittorio.

Avremo seri problemi di costituzionalità se, per tutelare una ragionevole durata del processo. Andassimo a discapito del rispetto del diritto di difesa. Una maggiore rapidità della decisione robotica potrebbe essere accettata solo se la prestazione fosse qualitativamente almeno uguale a quella umana.

Impossibile, però, non ammettere le potenzialità che un contributo tecnologico potrebbe apportare nel rendere migliore il nostro sistema giudiziario se questo contributo fosse, non sostitutivo, ma ausiliare della figura del giudice. L'eccessiva lunghezza dei processi italiani ci ha portato molte volte ad essere richiamati dalla Corte Europea dei Diritti dell'Uomo in numerose cause per aver violato l'art. 6 CEDU che statuisce la necessità di un termine ragionevole per i processi.

Questa problematica resta infatti una costante nel nostro paese e non solo a fronte della quale gli uffici giudiziari hanno cercato di correre ai ripari spesso sfociando però in fenomeno di cieco produttivismo a

discapito della qualità che conduce, tra le altre cose, oltre a rischiare di mandare innocenti in carcere o al sempre più frequente fenomeno della scadenza dei termini per le custodie cautelari che fa sì che molti soggetti poi giudicati innocenti restino per lungo tempo in carcere in attesa di giudizio crea anche l'avvio di numerose cause di risarcimento che, come noto, non sono a carico del magistrato che ha commesso l'errore o non ha svolto il suo dovere nei termini previsti dalla legge, bensì sono a carico dello Stato, ossia del contribuente.

La riduzione dei tempi, per tutti i servizi della pubblica amministrazione e non solo per la giustizia rimane una delle prime necessità e richieste della società. Se queste può essere controllata non risulta altrettanto semplice monitorarne la qualità. Una delle modalità che abbiamo è la percentuale di conferma o di smentita di una sentenza nei successivi gradi di giudizio ma naturalmente questo non basta. Quel che si evidenzia è una sempre maggiore perdita di fiducia del cittadino nella pubblica amministrazione in generale e nella giustizia in particolare.

Il cittadino infatti, sempre più spesso si vede un ribaltamento della decisione nella motivazione del giudice chiamato a sentenziare in secondo grado così da far perdere fiducia oltre che allontanare i cittadini dalle istituzioni, siano esse detentrici del potere politico o giudiziario. Qualche anno fa, venne redatto un questionario in cui si chiedeva al cittadino se preferiva essere giudicato da un giudice umano o da un giudice robot. La risposta della stragrande maggioranza dei cittadini sottoposti al test è stata la seguente: se sono innocente mi va bene essere giudicato da un giudice robot ma a fronte di una sentenza di colpevolezza preferisco che a emetterla sia un umano.

Questo test ovviamente ha dato inizio a molte discussioni circa quella che è, di fatto, la differenza tra un umano e una macchina e l'osservazione a parere di chi scrive più interessante, di cui si dirà meglio in seguito, riguarda il tema della totale assenza di empatia di una macchina, dove per empatia non si intende il sentimento di gradevolezza o di disgusto o di pena bensì quella capacità prettamente umana di “cogliere la risposta con la pancia e non con il cervello”, immedesimarsi in chi ho di fronte per capire se è colpevole o innocente e nel primo caso per cogliere le eventuali motivazioni che hanno condotto il soggetto a compiere quell'azione, ciò che, almeno nel diritto penale appare tutt'altro che marginale posto che tale comprensione può significare vedersi riconosciute attenuanti o non vedersi riconosciute aggravanti che peseranno non poco nel computo finale della pena.

Ma ora, facendo un passo indietro, occorre, prima di porci tutte le questioni circa i problemi o le potenzialità di una giustizia sempre più digitale, fare un'analisi delle principali fonti legislative europee e nazionali per capire come i nostri governi, ad oggi, intendono questo settore del diritto e quali strumenti economici e non intendono mettere in campo per migliorarlo; inoltre una breve ricognizione dello stato dell'arte italiano ed europeo sulla giustizia digitale appare necessaria.

Al congresso IAAIL del 2017, Richard Susskind cita un passo premonitore di Reed C. Lawlor del 1963:

Verrà un giorno in cui si sarà in grado di inserire un insieme di dati in una macchina che ha al suo interno precedenti, regole di diritto e regole di ragionamento in cui la macchina sarà capace di offrire, passo dopo passo, il ragionamento attraverso il quale si può essere in grado di arrivare a una decisione. Noi potremo studiarlo e decidere se la macchina ha proposto qualcosa di giusto o sbagliato. In alcuni casi la macchina non dirà quale

potrebbe essere la soluzione, ma vi è una probabilità che la risposta sia corretta, e questa probabilità è del 90%.¹⁰²

Questo capitolo, come il lettore avrà già avuto modo di notare è dedicato principalmente all'AI e al mondo della giustizia e, quindi, a cosa cambia o stia cambiando nel modo in cui i giudici decidono, in casi sia civili che penali. Le domande alle quali cercheremo di dare una risposta sono: i giudici utilizzano tecniche di AI? Se sì, quali? E i giudici italiani? In che modo tutto ciò incide sulle decisioni dei giudici? E sulla motivazione dei loro atti? I giudici sono parte delle istituzioni dello Stato: in che modo questo problemi si pongono nella P.A in generale? E negli apparati di polizia? A che punto sia con la previsione fatta da Lawlor? E' eticamente preferibile essere giudicati da un robot o da un umano?

Questi temi sono stati affrontati, per la prima volta in modo sistematico dalle istituzioni europee i giorni 3 e 4 dicembre 2018 a Strasburgo, dove la Commissione Europea per l'efficienza della giustizia¹⁰³ del Consiglio d'Europa (CEPEJ) approva il documento *European Ethical Character on the Use of Artificial Intelligence in Judicial System and their environment* e, così, accosta le parole “giudice” e “intelligenza artificiale”. D'intelligenza artificiale e giudici si può, quindi, parlare nei 47 Paesi del Consiglio d'Europa e anche in Italia, nonostante permangano diffuse resistenze.

Tra gli scopi dichiarati dal documento vi è quello di accrescere la prevedibilità del modo di applicazione della legge e la coerenza delle

¹⁰² Reed C. Lawlor, *What computer can do: analysis and prediction of judicial decisions*, in “American Bar Association Journal”, 49, 1963, pp. 337-344.

¹⁰³ La *European Commission for the Efficiency of Justice* (CEPEJ) è costituita dai rappresentanti di 47 Stati membri del Consiglio D'Europa.

decisioni giudiziarie. Proposito, questo, che conferma che la questione della predizione delle decisioni non è solo interesse degli avvocati, ma un aspetto importante del funzionamento dell'intero sistema giuridico e delle società. Soprattutto è un problema sociale unitario, che può essere guardato da più parti, avvocati, giudici, apparati amministrativi, singoli cittadini e via discorrendo. Secondo l'opinione risalente all'inizio degli anni Sessante del secolo scorso, messa in luce in questo capitolo, una predizione affidabile dell'attività dei giudici dipende dalla comprensione scientifica del modo in cui il diritto e i fatti incidono sui decisori rilevanti, cioè sui giudici.

Il documento CEPEJ si rivolge non solo ai giudici, ma anche a tutte le istituzioni incaricate di prendere decisioni nel campo legislativo e regolamentare sull'intelligenza artificiale.

La Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari (European Ethical Charter on the Use of Artificial Intelligence in Judicial System and their environment, 2018) dichiara apertamente che “nel 2018, l'uso di algoritmi di intelligenza artificiale nei sistemi giudiziari europei rimane principalmente un'iniziativa commerciale del settore privato rivolta a compagnie assicurative, dipartimenti legali, avvocati e singoli individui”. Per il momento i giudici negli Stati membri del Consiglio d'Europa non sembrano fare alcun uso pratico e quotidiano di software predittivi. Test locali e lavori accademici sono stati condotti per esplorare il potenziale di queste applicazioni, ma non sono ancora state applicate su vasta scala.

Questo è il punto di partenza realistico, che avrebbe forse consigliato un titolo meno altisonante, come Carta etica sugli usi solo possibili o futuri, dell'intelligenza artificiale. Il documento inoltre è basato su dati dichiaratamente scarsi e su una letteratura internazionale proveniente

prevalentemente dagli USA. Rimane, alla fine, l'importanza per i principi etici enunciati in via ipotetica e per il futuro:

- Principio di rispetto per i diritti fondamentali: assicurarsi che la progettazione e l'attuazione di strumenti e servizi d'intelligenza artificiale siano compatibili con i diritti fondamentali;
- Principio di non discriminazione: prevenire lo sviluppo o l'intensificazione di qualsiasi discriminazione tra individui o gruppi d'individui;
- Principi di qualità e sicurezza: nel trattamento delle decisioni giudiziarie e dei dati, utilizzare fonti certificate e dati immateriali con modelli elaborati in modo multidisciplinare, in un ambiente tecnologico sicuro;
- Principio di trasparenza, imparzialità ed equità: usare metodi di elaborazione dati accessibili e comprensibili;
- Principio "sotto controllo utente": preclude un approccio prescrittivo e garantisce che gli utenti siano attori informati e che abbiano il controllo delle scelte fatte.

Il documento distingue, quindi, i diversi usi possibili. Alcuni vanno incoraggiati, come le tecniche di *machine learning* per la ricerca della giurisprudenza e la visualizzazione dei risultati, l'utilizzo di *chatbots* per facilitare l'accesso alla giustizia, migliorarne l'efficienza rendendo possibile, ad esempio, effettuare valutazioni qualitative e quantitative e fare proiezioni circa le risorse umane e di bilancio necessarie. Altri usi richiedono considerevoli precauzioni metodologiche, come, ad esempio, il supporto alla ricerca di soluzioni transattive in cause risarcitorie sulla base di predizioni effettuate dalle compagnie

assicuratrici, le soluzioni di liti con sistemi online oppure l'uso di algoritmi nelle investigazioni criminali al fine di individuare dove un crimine sta per essere commesso (*predictive policing*). Altri usi richiedono ulteriori studi scientifici, come, ad esempio, la profilazione dei giudici o la previsione delle decisioni degli stessi o che vanno presi con estrema cautela, come l'utilizzo di sistemi di profilazione degli individui nei processi penali, che può portare a esiti discriminatori e a effetti deterministici e il rischio della cristallizzazione dei precedenti di cui parleremo meglio nel capitolo seguente dedicato alle problematiche tecniche e processuali del giudice robot.

Il documento esprime poi preoccupazione che possa essere posto in pericolo il primato della legge e la sovrana discrezionalità del giudice. Infine, il documento suggerisce i mezzi per monitorare questo fenomeno sotto forma di una Carta etica, sottolineando la necessità di un approccio cauto all'integrazione di questi strumenti nelle politiche pubbliche, con il coinvolgimento di tutte le parti interessate, siano essi professionisti legali, società di tecnologia legale o scienziati. Interessante la checklist finale, che consente di valutare agevolmente se i sistemi che sono introdotti rispettano i principi etici fondamentali. Per quanto concerne poi lo stato dell'arte della giustizia digitale europea, essa è in fervida evoluzione.

“Promuovere la giustizia digitale” è il titolo di uno dei progetti principali del Network europeo dei consigli superiori della magistratura nel triennio 2018-2021. Quest'organizzazione, fondata a Roma nel 2004, sostenuta dalla Commissione Europea e con un ufficio permanente a Bruxelles, è aperta a istituzioni come il nostro Consiglio

Superiore della Magistratura e analoghe appartenenti agli altri Stati membri dell'Unione Europea.¹⁰⁴

Il 4 maggio 2018, ad Amsterdam, il network discute di tecnologie informatiche, inclusi l'intelligenza artificiale e il *legal design thinking*. Dai resoconti emerge la consapevolezza che digitale è la nuova normalità e che le corti devono seguire il resto della società, che è già ampiamente avviata su questa strada. I temi concreti in discussione sono, in verità, quelli di base. Se si vuole, della digitalizzazione dell'uso delle tecnologie informatiche (automazione dell'ufficio, sistemi di flusso di lavoro/gestione dei casi, comunicazione digitale, siti web) e delle correlate difficoltà politiche, legislative (anche per gli effetti sulle norme procedurali), finanziarie e tecniche (es. la *governance* sui progetti IT e le scelte tecniche strategiche). Sono intraviste anche alcune opportunità in termini di miglioramento dell'accesso alla giustizia, e dell'imparzialità ed integrità, nonché l'eliminazione degli ostacoli fisici, economici e informativi.

Detto questo e apprezzata l'apertura a un mondo in cui le tecnologie informatiche sono la regola e non l'eccezione e, quindi, alla necessità di un cambio di mentalità nel guardare alle procedure, vi è veramente poco di intelligenza artificiale, nel senso proprio del termine, se non qualche riflessione, più che altro speculativa, su se l'utilizzo futuro di tecnologie di AI sia conforme al principio del giusto processo previsto all'art. 6 CEDU, (e dall'art. 111 della Costituzione italiana) che richiede che l'esito sia trasparente e che possa essere spiegato come si è raggiunto un certo risultato. Tale principio ribadito, seppur in altro

¹⁰⁴ A patto che siano indipendenti dal potere esecutivo e legislativo, o che siano autonome e che assicurino la responsabilità finale per il sostegno del potere giudiziario nell'impartire la giustizia in modo indipendente.

contesto anche dall'art. 22 del GDPR è stato anche oggetto di una recente sentenza del Consiglio di Stato chiamato a pronunciarsi sulla possibilità da parte del cittadino di essere destinatario di decisioni interamente automatizzate (ovviamente non solo per quanto concerne l'ambito della giustizia). In tale storica decisione il Supremo Consiglio ha statuito tra le altre cose che le decisioni prese dall'algoritmo (il caso posto alla sua attenzione riguardava una graduatoria insegnanti), pur avendo una funzione di supporto dell'attività umana non possono essere motivo di elusione dei principi alla base del nostro ordinamento e che, appunto, l'attività dell'algoritmo è solo ausiliaria dell'uomo, al quale spetta sempre "l'ultima parola".

Inoltre, la modalità digitale delle procedure e la comunicazione elettronica tra i soggetti coinvolti in procedimenti giudiziari sono per la Commissione Europea un elemento essenziale del funzionamento efficace del sistema giudiziario negli Stati membri. La giustizia elettronica europea è considerata un mezzo per migliorare l'accesso alla giustizia in un contesto paneuropeo e sviluppare tecnologie dell'informazione e della comunicazione. La giustizia elettronica dovrebbe contribuire allo sviluppo del mercato unico digitale, che è uno degli obiettivi dell'*eGovernment*¹⁰⁵.

L'obiettivo è migliorare l'accesso all'informazione nel settore della giustizia nell'Unione Europea (per esempio, EUR-Lex¹⁰⁶ è stato arricchito con nuove tipologie di documenti e strumenti di ricerca). Degno di nota è poi il richiamo alla dematerializzazione dei

¹⁰⁵ Piano d'azione dell'UE per l' egovernement 2016-2020 – Accelerare la trasformazione digitale della Pubblica Amministrazione.

¹⁰⁶ Si tratta di un sito gestito dall'Ufficio delle pubblicazioni dell'UE che dà accesso alla legislazione dell'UE e anche alla giurisprudenza della Corte di Giustizia dell'Unione Europea. Gli utenti vi possono accedere gratuitamente e hanno a disposizione vari metodi di ricerca in tutte le lingue ufficiali dell'Unione.

procedimenti giudiziari e stragiudiziali, che dovrebbe essere proseguita al fine di offrire un accesso più facile e più veloce ai tribunali e di agevolare il ricorso ai procedimenti stragiudiziali mediante l'uso di strumenti sicuri di comunicazione elettronica, segnatamente e-CODEX, in situazioni transfrontaliere.

Secondo il *Piano d'azione 2019-2023 in materia di giustizia elettronica*, l'intelligenza artificiale potrebbe svolgere un ruolo importante nel settore della giustizia, in quanto potrebbe sviluppare strumenti pratici per anonimizzare o pseudonimizzare automaticamente le decisioni giudiziarie, rendere reperibili in modo veloce e affidabile informazioni utili. I dati giuridici possono essere utilizzati in formato Open Data per aiutare i cittadini, le imprese e le autorità giudiziarie a studiare e raccogliere dati al fine di analizzarli e contribuire alle applicazioni che utilizzano tali dati, anche avvalendosi dell'intelligenza artificiale. Un vocabolario controllato e identificatori come ECLI (*European Case Law Identifier*, Identificatore europeo della giurisprudenza) potrebbe consentire di indicizzare il contenuto e agevolare il trattamento di dati e il reperimento di informazioni. L'intelligenza artificiale può rendere possibile l'analisi di Big Data in ambito giuridico. Il piano prevede lo sviluppo di uno strumento di AI per l'analisi delle decisioni giudiziarie e lo sviluppo di applicazioni destinate all'uso da parte dei magistrati nazionali. Il portale europeo della giustizia elettronica, dal quale è possibile accedere a tutti i servizi digitali, è stato rinnovato nella veste grafica e arricchito con pagine informative, strumenti di ricerca e moduli dinamici.

Quello delineato dall'Unione Europea è un sistema imponente e importante che punta a essere a disposizione di ogni operatore, cittadino, avvocato, giudice o altro. I sistemi di giustizia elettronica

possono essere utilizzati per facilitare il funzionamento di varie reti esistenti a livello europeo, quali la *rete giudiziaria europea in materia civile e commerciale* e la *rete giudiziaria europea in materia penale*. Vi è il limite, da non discutere qui, del carattere volontario dell'attuazione di questo orientamento da parte di ciascuno Stato membro. Tuttavia, nella misura in cui gli Stati si adeguano, la dematerializzazione dei procedimenti giudiziari e stragiudiziali e tutta la digitalizzazione, dalla pubblica amministrazione alla giustizia nei diversi settori, produce quella massa di dati che è il presupposto materiale per l'effettiva introduzione di tecnologie di AI; per ora una speranza, con una visuale attenta alla protezione dei dati e all'etica. E in Italia? Da noi, una parte notevole e crescente dell'attività giudiziaria viene svolta con modalità digitali. Ciò riguarda il processo civile telematico, che è l'espressione più avanzata, ma anche altri settori della P.A.

Il processo amministrativo (PAT) è diventato telematico dall'1 gennaio 2017 per tutti i nuovi ricorsi proposti davanti ai TAR e davanti al Consiglio di Stato. Dal 1 gennaio 2018 il PAT è applicato anche per i ricorsi che erano stati proposti prima del 1 gennaio 2017. I depositi di atti e documenti riguardanti tali ricorsi sono quindi eseguiti in forma digitale secondo le nuove regole del PAT.¹⁰⁷

Anche nella giustizia tributaria il processo tributario telematico (PTT) è attivo dal 15 luglio 2017 su tutto il territorio nazionale. La digitalizzazione delle fasi della notifica, del deposito del ricorso e dei

¹⁰⁷ Con il PAT l'utente ha l'opportunità, previa registrazione attraverso il portale dedicato, di accedere al sistema informatico della Giustizia Tributaria e di depositare gli atti e i documenti processuali già notificati alla controparte. Inoltre, i giudici tributari, i contribuenti, i professionisti e gli enti impositori, potranno consultare da casa o dai propri uffici il fascicolo processuale contenente tutti gli atti e documenti del contenzioso cui sono interessati.

documenti comporta vantaggi notevoli per tutti gli attori del processo tributario in termini di semplificazione, trasparenza degli adempimenti processuali e di durata del contenzioso.

Il processo civile telematico (PCT) è sicuramente la realtà più estesa e di più ampia portata e rappresenta la realtà più avanzata di digitalizzazione. E' applicato, sia pure in modo quantitativamente non omogeneo sul territorio nazionale, nei tribunali e nelle corti d'appello, mentre, per quanto riguarda la Corte di Cassazione è stata nel 2018 avviata la digitalizzazione dei sistemi di cancelleria e l'infrastruttura per avviare il processo civile telematico. Ancora non è attuato presso il giudice di pace, dove è stato esteso per ora solo l'impiego dei sistemi elettronici di gestione dei registri.

Per dare un'idea della realtà del PCT, basti considerare che, anche solo in termini di accessibilità dei sistemi da parte di utenti esterni agli uffici giudiziari, il sistema telematico, che rende disponibili i servizi telematici ai professionisti e agli enti, ha registrato un vertiginoso aumento di deposito di atti telematici negli ultimi 3 anni.

Vi è poi il processo penale telematico (PPT) che è in una posizione più arretrata per una serie di ragioni che non è questa la sede per approfondire. Tuttavia, vi sono importanti segni di una ripresa del processo di digitalizzazione a partire dal "Portale delle notizie di reato, la cui diffusione è in continuo aumento, accompagnato da interventi sempre più capillari volti alla totale integrazione, in attesa dello sviluppo del sistema unico, tra registri e gestore documentale", mentre sono state poste le basi in alcune realtà pilota per il flusso completo di informazioni dalle Procure ai Tribunali, alle Corti d'appello e alle Procure generali.

Quanto fin qui esposto mette in luce come negli ultimi anni, dal 2018 in poi, nel nostro Paese si stia cercando di correre ai ripari e di stare al passo con gli altri paesi europei e non tecnologicamente avanzati. Tuttavia molto c'è ancora da fare per due evidenti ragioni: la prima riguarda il giustizia digitale intesa qui nel suo significato più basilare ossia il fatto che seppur nella teoria, la gran parte dei tribunali e delle corti d'appello italiane abbiamo oramai banche dati a disposizione degli operatori del diritto, nella pratica tali sistemi sono ancora piuttosto arretrati e talvolta non migliorano in modo apprezzabile il lavoro del giurista, dall'altra occorre notare come spesso, parlando di giustizia digitale ci si riferisce ad una giustizia se non interamente tecnologizzata, sicuramente molto più avanzata di quella che utilizza una semplice banca dati come può essere Italgiure, utile ma lontana dalla giustizia digitale intesa in senso stretto.

Un fattore sicuramente determinante è l'atteggiamento mentale dei governi e non ultimo dei cittadini, operatori del diritto e non di fronte ai cambiamenti tecnologici.

Un raffronto tra la realtà della digitalizzazione dell'attività giudiziaria in Italia e negli altri Paesi non è tra gli scopi di questa tesi tuttavia vogliamo fare un breve riflessione.

Sinteticamente: Dory Reiling, considerata una leader della digitalizzazione a livello internazionale e chiamata a fare la relazione introduttiva alla meeting di Amsterdam del Network dei Consigli Superiori, intervistata da Anne Wallace, dice che in Olanda "fino a settembre 2018, oltre 2000 casi erano stati trattati in via telematica. Tuttavia, la procedura non è stata implementata come previsto negli altri nove tribunali di primo grado. Il motivo principale è che l'implementazione sarebbe troppo impegnativa e richiede troppe

risorse”. Quanto al Regno Unito il governo ha annunciato nel 2015 che avrebbe investito 700 milioni di sterline per digitalizzare integralmente le corti.

Come dicevamo poc’anzi il problema cruciale è quello della forma mentis, in primis dei giuridici se parliamo di giustizia. L’abito mentale richiesto per il lavoro giudiziario non aiuta l’innovazione. Il loro lavoro è rivolto a eventi passati e implica la decisione su chi ha commesso errori in qualcosa che non è andato per il verso giusto. Innovare, invece, significa sperimentare e sperimentare significa provare, e vi sarà sempre qualcosa che non va bene. Ma se questo significa che è possibile essere incolpati di ciò che non è andato bene, poche persone vorranno correre questo rischio, ovviamente. Alla fine il problema appare quello di riuscire a cambiare l’idea che le tecnologie servano a supportare le procedure cartacee, mentre esse in realtà servono a innovare radicalmente i processi. E questo si potrebbe scrivere un’altra tesi di dottorato pertanto ci fermiamo qui nel ragionamento.

All’esito di questo rapido excursus su cosa sta accadendo nel campo delle applicazioni tecnologiche alla giustizia in Europa e in Italia e su cosa si sta discutendo a livello europeo e italiano, si può dire che al momento, e fatta eccezione per le poche esperienze di cui si dirà più avanti e che riguardano solo il processo civile, non vi è alcuna reale applicazione di tecnologie di intelligenza artificiale. La discussione nei documenti esaminati è soltanto teorica, tratta talora da esperienze e pubblicazioni in altri Paesi (specie USA) e tesa a porre principi etici o giuridici per evitare la paventata rivoluzione dell’AI.

Non vi è necessariamente AI nei processi di digitalizzazione se, come spesso accade, essi sono meramente intesi a trasferire su supporto digitale le tradizionali attività cartacee o meramente intese alle

comunicazioni tra gli addetti ai lavori del mondo giustizia. E' bene chiarire che si può parlare di AI solo laddove la grande quantità di dati prodotti quotidianamente, in contesti e in sistemi digitalizzati, sia organizzata in un modo che consenta operazioni di *big data analytics* attraverso l'uso di tecniche di *machine learning*, al fine di estrarre informazioni da quei dati. Ciò richiederebbe struttura di raccolta e organizzazione di quei dati (con le loro tipiche caratteristiche di dati strutturati, non strutturati e semistrutturati, dove "strutturato" sta per "leggibile" e processabile da un computer, mentre esempio di dato non strutturato è un 'immagine), cioè di magazzini di dati.

Il Ministero della Giustizia italiano ha introdotto il *datawarehouse* della giustizia civile per scopi statistici e di analisi organizzativa e ora si propone di:

basare lo sviluppo dei nuovi sistemi sulla condivisione dei dati e la circolarità delle informazioni: la valorizzazione del dato e della sua aggregazione si tradurrà nella progettazione, realizzazione ed evoluzione di *datawarehouse* sempre più performanti. La gestione del dato, nella prospettiva futura e più aderente alle attuali tecnologie deve infatti superare la dicotomia di sistemi registro-centrici o documento-centrici. In particolare, gli sviluppi in corso tendono al superamento della integrazione delle informazioni da evento con le informazioni da atto ed alla costruzione di sistemi di rappresentazione cognitiva. L'obiettivo è ottenere una piattaforma comune di informazioni, collegate tra loro ed interdipendenti le une dalle altre in una unica catena del valore del processo.¹⁰⁸

¹⁰⁸ Relazione del Ministero della Giustizia resa in occasione dell'inaugurazione dell'anno giudiziario 2019.

Questo è il salto oggi possibile e alle porte, con la consapevolezza che importanti esperienze sono in corso a Singapore e in Corea del Sud e che, da questo punto di vista l'Italia è in una posizione buona, che potrebbe diventare anche ottima, se la costruzione di *datawarehouse* si dovesse accompagnare alla messa a disposizione dei giuridici di strumenti tecnici che consentano la produzione di documenti di buona qualità informatica.¹⁰⁹

In un paio di paragrafi precedenti abbiamo citato l'esperimento eseguito in Italia, all'audience di un corso intensivo europeo su *Neuroscience and Law*, circa la preferenza nell'essere giudicati da un giudice umano o da un robot, ora, partendo dalla premessa, più volte esplicitata relativa al fatto che, ad oggi, l'intelligenza artificiale sulla quale si basa anche la giustizia non è certo da considerarsi "forte" (nel senso della distinzione fatta dagli esperti di AI, ossia quella che replica gli stati cognitivi della mente umana) bensì è un AI che si limita a dare un risultato, sulla processo matematico reso possibile da una quantità elevata di dati che, una volta processati danno un risultato che rappresenta un correlazione e non una "causazione" dell'evento. Tale distinzione la esplicheremo dopo aver parlato del noto caso Loomis, divenuto celebre alle cronache anche col nome dell'algoritmo utilizzato, ossia COMPAS.

Tale precedente giurisprudenziale (perché tale è divenuto), è diventato l'emblema della difficoltà di chiarezza sulle applicazioni di AI in ambito giudiziario. Il signor Loomis, condannato per una sparatoria in cui era coinvolta anche la polizia, si è visto aumentare la pena per il

¹⁰⁹ A livello accademico si può citare il progetto LAILA, finanziato dal Ministero dell'Università (PRIN 2019), promosso dalle Università di Pavia, Torino, Napoli e Bologna.

rischio di recidiva in una misura che ha ritenuto eccessiva. Il difensore, intenzionato ad appellare, ha scoperto che il giudice aveva quantificato la pena facendo ricorso a un sistema automatico (artificiale) chiamato COMPAS, sistema usato nelle corti americane di alcuni Stati. Il caso è giunto fino alla Corte suprema del Wisconsin, che ha ritenuto non vi fosse lesione del diritto al *due process*, perché l'uso del sistema COMPAS non escludeva l'obbligo del giudice di spiegare quali fattori aggiuntivi avessero portato a quella specifica quantificazione. Il sistema COMPAS è un sistema privato basato sulla raccolta per molti anni di dati statistici sui condannati, sui fattori di rischio presenti nei vari casi (che sono per lo più alcolismo, droga, prostituzione, ma anche indirizzi di residenza, appartenenza a etnie extracomunitarie, precedenti penali del soggetto ma anche dei familiari soprattutto se con lui conviventi nonché livelli di reddito) e sulle recidive. Il giudice, inserendo i dati sui fattori di rischio del proprio imputato, riceve una risposta statistica che proietta sul caso in decisione i dati contenuti nel database.

La questione merita di essere messa a fuoco con cura. Sotto un primo punto di vista che è anche quello centrale, a parere di chi scrive, non si può addossare in capo ad una persona la possibilità statistica della commissione di altri reati, in altre parole, il giudice deve giudicare il caso concreto, pertanto non potrà mai ritenere una persona soggetta ad alto rischio di recidiva se altre persone, simili a lui per "profilo" hanno commesso altri crimini successivamente al primo per cui richiedono, per esempio, una misura alternativa alla detenzione (concessa ovviamente solo quando non si ha ragione di ritenere il soggetto recidivo). Vi sono poi una serie di altre questioni che vanno dalle questioni di proprietà dell'algoritmo alle base del software che emette la decisione ai problemi di *bias* del sistema, fino a quelli più legati alla

politica del diritto sulla perdita di responsabilità dell'organo giudicante qualora decida, nella maggior parte dei casi in decisione, di conformarsi alla decisione dell'algoritmo. Se il problema è la proprietà privata del sistema COMPAS (o simili), con i conseguenti diritti di privativa che escludono di conoscere nel caso specifico la logica e il modo di funzionamento del sistema, ci si può chiedere: se il Ministero della giustizia italiano, invece di limitarsi a fornire le apparecchiature digitali a giudici italiani, fornisse anche un sistema *open source*, ossia un sistema aperto e conoscibile nella sua logica e architettura, discusso e approvato da avvocati e giudici, sarebbe quindi accettabile? E se invece il sistema fosse interamente in mano pubblica, certificato e la sua architettura fosse conosciuta e condivisa dalla magistratura e dall'avvocatura, sarebbe invece da rifiutare?

Se il problema sono i possibili *bias* del sistema, essi potrebbero essere oggetto di discussione preventiva e di continuo monitoraggio applicativo, rimanendo fermo che il giudice può giustificare dopo averli portati a conoscenza, i fattori aggiuntivi e particolari che portano ad una determinata situazione.

Se invece il problema è quello di affidarsi a un aiuto tecnologico nella condanna ci si può chiedere: dovremmo fare più affidamento su un ragionamento giuridico esclusivamente umano, nel quale possono comunque annidarsi *bias* cognitivi del giudicante (coperti talora da motivazioni apparenti), piuttosto che su dati quantitativi, che possono essere ulteriormente discussi prima della sentenza finale?

Ciò detto è evidente che contano anche le differenze culturali: mentre in Italia alla domanda se si preferisce essere giudicati da un umano o da un computer, posta da esperti a esperti computazionali, la risposta, come visto, dipende dalla colpevolezza o innocenza del soggetto a cui

si rivolge la domanda, sperando, in quest'ultimo caso nella *pietas*, caratteristica prettamente umana o negli errori che oggettivamente un umano commette più di algoritmo ben addestrato, negli USA, in occasione di un dibattito sul futuro del diritto organizzato a New York dal centro CodeX-The Stanford Center for Legal Informatics, il *chairman* quando pose al pubblico la stessa domanda la risposta fu quasi univoca: l'uditorio rispose in stragrande maggioranza che avrebbe preferito un robot.

Allo stato, gli effetti della digitalizzazione nel mondo della giustizia non è banale. Per effetto di un sistema avanzato, infatti, il tradizionale atto giuridico (sia esso una sentenza, ordinanza, decreto, atto di citazione, comparsa piuttosto che legge, provvedimento amministrativo ecc.) una volta inserito in un ambiente informatico, subisce alcune trasformazioni importanti e correlate: in primo luogo, si delocalizza rispetto al suo autore/produttore (in quanto anche materialmente, viene a trovarsi in un server, che può essere in un luogo remoto, rispetto all'ufficio di provenienza). In secondo luogo, si trasforma da documento compatto in aggregato di informazioni, che possono essere disaggregate e che possono popolare ulteriori aggregati di informazioni, ai quali possono contribuire soggetti diversi e in posizione istituzionale, professionale e personale diversa. Gli elementi di questi aggregati possono perfino appartenere a fonti sovranazionali, comunitarie o di altri Paesi (posto che, da un punto di vista contenutistico, non si può escludere che un caso o una fonte extranazionale sia concettualmente più prossima al caso da decidere di quanto non lo sia una fonte o un precedente nazionale: sono queste le dinamiche tipiche del diritto transnazionale).

Questi i cambiamenti già in atto ai quali i giudici hanno dovuto conformarsi:

- I materiali di un fascicolo possono essere facilmente reperiti online durante lo studio della causa o la decisione, sia essa monocratica o collegiale, anche se l'ufficio di cancelleria è chiuso.
- Anche nella fase della decisione in Camera di consiglio è possibile accedere, tramite uno schermo di dimensioni appropriate, a materiali della causa, a banche dati, a fonti nazionali o internazionali.
- Il giudice può avere accesso ai materiali anche dalla propria abitazione o dall'estero.
- Il processo si dematerializza e l'uso della carta viene eliminato o, o almeno, si riduce fortemente.

Arrivati a questo punto, cosa accadrà quando su questa massa di dati, che ogni giorno di più vengono prodotti nel processo penale, si applicheranno le tecnologie di AI? Cioè quando su queste grandi moli di dati si comincerà a lavorare con strumenti di *big data analytics* e, dunque, di *machine learning*?

Questa realtà, ovviamente, non riguarda solo la giustizia. Alcuni elementi dell'attuale panorama tecnologico sono chiari e ampiamente descritti.

Siamo nell'era della quarta rivoluzione industriale, le cui caratteristiche principali sono la connettività, l'intelligenza distribuita, l'industrializzazione di ogni processo. Gli investimenti riguardano principalmente l'infrastruttura cloud/digitale, con data center di grandi capacità e comunicazione dati ad alta velocità. Le

diverse sfaccettature di tale realtà sono: a) *Internet di tutto*, che naturalmente include l'Internet delle cose (IoT), ovvero la connessione veloce attraverso il cloud (e 5G) di servizi, attività industriali, ospedali e tutti gli aspetti delle città intelligenti, e l'Internet delle persone (*Internet of people*, quando le entità connesse saranno esseri umani); b) l'enorme quantità di dati che producono tutte queste questioni; c) l'analisi dei big data come mezzo per governare tutti questi dati e sfruttarli attraverso l'uso di tecnologie di apprendimento automatico cioè l'intelligenza artificiale e la scienza dei dati.

L'analisi dei big data utilizza esclusivamente metodi e strumenti di *machine learning*, che rendono possibile esaminare grandi quantità di dati, scoprire correlazioni nascoste e dedurre altre informazioni. L'analisi dei big data è anche utile per progettare nuovi sistemi di ragionamento e decisione automatizzati e autonomi. Si tratta di un settore in continua evoluzione. Basto citare i Real-time Analytics, strumenti di analisi che realizzano risultati in tempi molto brevi, o il motore di ricerca di *Google Dataset Search*.

Naturalmente anche il mondo del diritto è coinvolto sotto due principali profili: il primo riguarda il diritto delle tecnologie, come diritto che regola le tecnologie, con riguardo alla proprietà dei dati, alla responsabilità, alla proprietà intellettuale, alla privacy, all'etica dei big data, all'etica della ricerca su Internet. Il secondo profilo riguarda il diritto come campo esso stesso di ricerca con sistemi di analisi di big data. In questo senso, il diritto è visto come un'entità tecnologizzata in cui è possibile applicare l'analisi dei big data coinvolgendo anche le professioni di avvocati e giudici.

Il processo decisionale basato sui dati come campo di ricerca sia nei settori della scienza-tecnologia sia in quello sociale, etico e giuridico è una settore in larghissima parte inesplorato. Quello che appare chiaro agli studiosi però è che “big data does not equal big knowledge”,¹¹⁰ almeno automaticamente.

Oggi possiamo dire che i processi decisionali sono presi sulla base di decisioni matematiche, ossia che si basano interamente o comunque parzialmente sui dati (per fare un esempio tra tutti, pensiamo alle misure presa durante la pandemia da Covid-19).

Tali decisioni basate sul ragionamento statistico fino ad arrivare all'apprendimento automatico e quindi all'AI vera e propria sono generate da processi decisionali basati sulla conoscenza tradizionale e sulla conoscenza basata sui dati, in altre parole le due coesistono in diverse combinazioni a seconda del campo di applicazione, delle situazioni e, in gran parte, della disponibilità di dati. Tuttavia, i due sistemi sono diversi per genere, ipotesi e ispirazione, tanto che non si può dire che questa convivenza sia semplicemente complementare. Sono allora necessari nuovi paradigmi etici e giuridici per affrontare questi aspetti?

Sulla linea che va dalle decisioni basate sulle regole verso quelle basate sui dati, sorgono i seguenti problemi che qui riassumiamo:

¹¹⁰ Gordon Webster, *Big Data does not equal Big Knowledge*, 19 maggio 2019. In tale articolo si rinviene il seguente interessante parallelo: “Applying the standard pantheon of data analytics and data visualization techniques to large biological datasets, and expecting to draw some meaningful biological insight from this approach, is like expecting to learn about the life of an Egyptian pharaoh by excavating his tomb with a bulldozer”.

- *Bias nel set di dati.* La qualità e le caratteristiche del set di dati utilizzato sono essenziali: da dove provengono i dati? Come vengono raccolti e selezionati?¹¹¹
- *Bias negli algoritmi applicati.* Anche per gli algoritmi applicati si pone il problema di qualità e caratteristiche. Come affrontiamo i pregiudizi umani immessi nella macchina? Può o dovrebbe questo processo di imitazione rimuovere i pregiudizi umani? Quali sono i pericoli di questo processo? Come può un sistema legale salvaguardare la sicurezza e la riservatezza dei dati personali necessari per addestrare tali algoritmi? Qual è l'attuale quadro giuridico? Il nuovo regolamento generale sulla protezione dei dati è uno strumento adeguato?
- *La spiegabilità dei risultati che l'AI ha prodotto.* Ammesso che il set di dati e di algoritmi applicati siano di qualità e caratteristiche adeguate, rimane il problema che questi processi non sempre sono spiegabili. (inteso nel senso di spiegabilità non di motivazione di una decisione).

La domanda che si può porre è la seguente: se i giudici decidono oggi seguendo le regole di diritto, con l'AI cambierà il loro modo di decidere, nel senso che le loro decisioni saranno non più prese secondo le regole, ma guidate dai dati forniti da sistemi di AI? Passeranno così da un sistema *rule-based* a uno *data-driven*? E' corretta l'opposizione *rule-based* v. *data-driven* ?

Quali regole segue invece il processo decisionale giuridico?

¹¹¹ Rashida Richardson, Jason M. Schultz, Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems and Justice*, in "New York University Law Review Online", Forthcoming.

Esso è il processo che più di altri è tradizionalmente basato su regole, ora con l'avvento della tecnologia esso è messo fortemente in discussione nella sua struttura portante in quanto subisce, così come gli altri processi decisionali testé descritti, l'influsso dei dati. Tuttavia, l'individuazione del vero problema in gioco richiede una comprensione dei termini e dei punti da cui si parte:

- “Processo decisionale basato sui dati” non è sinonimo di “decisione basata su prove”, come potrebbe apparire dal tenore letterale delle parole usate. *Data-driven decision making* non significa *evidence based* (ossia appunto “basata su prove”) in opposizione a decisioni arbitrarie: è un modo diverso e nuovo di ragionare e decidere che porta con sé varie problematiche di cui ci occuperemo nel prosieguo.
- “Processo decisionale basato su regole” (ossia *rule-based decision making*), inoltre, non è sinonimo di “decisione presa in modo giuridico”, poiché il processo decisionale basato su regole copre un campo meno ampio. Nel suo senso proprio e più rigoroso, è il modo di decidere secondo una regola scritta e chiaramente definita. Frederick Schauer, autore di uno studio fondamentale su *come giocare secondo le regole*, afferma molto chiaramente che “il processo decisionale governato da regole è un sottoinsieme del processo decisionale legale, piuttosto che essere congruente con esso¹¹². Secondo la sua decisione, molto esigente, se un tribunale legittimamente nominato prende una decisione secondo un criterio del “miglior interesse” del

¹¹² Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Clarendon Press, Oxford, 1991.

bambino o del paziente, o secondo il sistema di equità o determina l'entità di una condanna penale, quel tribunale prende sicuramente decisioni legali, che però non possono dirsi decisioni basate su regole, a causa della natura e della qualità intrinseca della norma che è stata applicata.

- Nel campo più ampio delle decisioni giuridiche (che include quelle basate sulle regole) si incontra un'importante divisione/opposizione, come quella tra formalismo e realismo giuridico, dove il punto cruciale è quello del ruolo che le regole svolgono nella decisione: se sono il criterio e la guida per distribuire diritti e torti o sono solo una giustificazione ex post di una decisione presa secondo altre ragioni, siano esse politiche, emotive, etiche o sociali. Il dibattito italiano sulla motivazione risente del fatto che nel nostro Paese non ha mai avuto un riconoscimento significativo la scuola del realismo giuridico, a differenza che in altri Paesi.¹¹³
- Un sistema di decisione basato su dati è un sistema in cui la decisione viene presa in base a ciò che emerge dall'applicazione degli algoritmi di apprendimento automatico. Da questo punto di vista è importante la distinzione “tra impostare un problema in avanti [*forward*] e un problema inverso [*inverse*], dove l'approccio *forward* che va dal modello ai dati osservabili, è quello usato tradizionalmente in approcci sperimentali o quasi sperimentali, mentre l'approccio inverso è il cuore

¹¹³ Edward H. Levi, *An Introduction to Legal Reasoning*, The University of Chicago Press, Chicago-London, 1945; Oliver W. Holmes, *The Path of Law*, in “Harvard Law Review”, 457, 10, 1897.

dell'apprendimento automatico, dove si usano i dati osservabili per costruire il modello piuttosto che usare il modello per assegnare peso causale ai dati osservabili". In altri termini il rapporto ipotesi-test, alla base del metodo scientifico moderno, è invertito ed è l'analisi di dati a far emergere un possibile modello da poi testare"¹¹⁴.

- In termini teorici, la questione della natura e dell'interpretazione dei risultati prodotti con l'uso di algoritmi è cruciale. Seguendo l'analisi della situazione proposta da Kevin Ashley "poiché un algoritmo di *machine learning* impara regole basate su regolarità statistiche, che possono sorprendere gli umani, le sue regole potrebbero non sembrare necessariamente ragionevoli per gli esseri umani. Le previsioni con *machine learning* sono basate sui dati. A volte i dati contengono caratteristiche che, per motivi spuri, come la coincidenza o la selezione parziale, sono associate ai risultati dei casi in una particolare raccolta. Sebbene le regole indotte dalla macchina possono portare a previsioni accurate, esse non si riferiscono all'esperienza umana e potrebbero non essere comprensibili per l'uomo come le regole costruite manualmente da un esperto. Poiché le regole che l'algoritmo di *machine learning* inferisce non riflettono necessariamente le conoscenze e le competenze giuridiche esplicite, potrebbero non corrispondere ai criteri di ragionevolezza di un esperto umano."¹¹⁵

¹¹⁴ Daniel Martin Katz, *Quantitative Legal Predictions or now I learned to stop worrying and start preparing for the Data- Driven Future of the Legal Services Industry*, in "Emory Law Journal", Vol. 62: 909, 2013, pp. 909-966.

¹¹⁵ Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: The New Tools for Law Practice in the Digital Age*, Cambridge University Press, Cambridge, 2017.

In conclusione, si può dire che i sistemi decisionali basati sui dati e quelli basati su regole hanno diversi vantaggi e svantaggi. I tipici problemi delle decisioni basate sui dati sono la quantità dei dati raccolti e i modi delle loro analisi e interpretazione, tanto che qualsiasi errore o pregiudizio in uno dei passaggi può influire pesantemente sulla decisione. Del tutto diversamente, il problema dei sistemi decisionali basati su regole è che basarsi su regole significa semplicemente che determinate regole sono state seguite, indipendentemente dalla qualità e/o dall'efficienza della decisione presa. Le combinazioni tra i due sistemi non sono agevoli e richiedono accortezze particolari.

La risposta alla domanda sopra posta, circa il cambiamento che l'utilizzo di sistemi decisionali *data-driven* induce nel lavoro dei giudici (come di qualsiasi decisore), è che certamente vi sarà un cambiamento importante nella logica della decisione. Ciò apre a un'ulteriore domanda: come possono coesistere la natura intrinseca degli schemi emergenti dell'analisi giuridica (e la loro limitata spiegabilità) e il diritto alla spiegazione delle decisioni politiche, che disposizioni costituzionali riconoscono?

A proposito di documenti normativi e buoni propositi per il futuro prossimo, nel mese di dicembre 2021, è stato emanato il "Programma Strategico Intelligenza Artificiale 2022-2024" dal governo italiano, in particolare dai Ministeri dell'Università e della Ricerca, dal Ministero dello Sviluppo Economico e dal Ministero per l'Innovazione Tecnologica e la Transizione Digitale di concerto tra loro. In tale documento, che per motivi di spazio non approfondiremo nel dettaglio, consta di una prima parte in cui si fa una ricognizione dello stato dell'arte italiano sul tema e poi una seconda in cui il governo si pone degli

obiettivi. In particolare vengono in rilievo una serie di dati utili come ad esempio si descrive che nel 2020 il mercato privato italiano dell'Intelligenza Artificiale ha raggiunto un valore di 300 milioni di euro, con un aumento del 15% rispetto al 2019, a circa solamente il 3% del mercato europeo, nettamente inferiore rispetto al peso italiano del PIL europeo che è circa al 12%. All'interno del mercato italiano, un controvalore di 230 milioni di euro è fornito ad aziende italiane, mentre i restanti 70 milioni di euro sono esportati verso aziende estere. In termini di applicazioni, le principali soluzioni AI presenti sul mercato rilevate da uno studio del Politecnico di Torino sono: 1) *Intelligence Data Processing*; 2) *Natural Language Processing*; 3) *Sistemi di suggerimento*; 4) *Automazione dei processi ripetitivi, Chatbot, Assistenti virtuali e Computer Vision*.

Per quanto riguarda il settore pubblico, il documento prosegue “le potenzialità dell'AI sono fortemente legate al processo di digitalizzazione del Paese. Secondo lo studio E-government Benchmark 2020 della Commissione Europea, l'Italia è caratterizzata da un basso livello di penetrazione e un livello medio di digitalizzazione. Permane quindi un divario tra l'offerta di servizi digitali e il loro effettivo utilizzo. Nonostante il livello di digitalizzazione nella PA sia in linea con quello di contesti europei simili, i servizi online non sono altrettanto utilizzati”.¹¹⁶

Nella sostanza il documento enuncia una serie di investimenti, anche con l'ausilio dei soldi che arriveranno grazie al Piano Nazionale di Ripresa e Resilienza, nei settori dell'innovazione tecnologica sia per favorire le imprese private emergenti nel settore come le start-up che

¹¹⁶ Citando letteralmente il testo del “Programma Strategico Intelligenza Artificiale 2022-2024 del Governo italiano.

per sostenere la crescita di spin-off innovativi, nonché una spinta per rafforzare le soluzioni di AI nella PA nell'ecosistema GovTech in Italia ed infine creare una sorta di canale comunicativo tra settori pubblico e privato impegnati nello sviluppo dell'AI con i centri di ricerca, nazionali ed europei per incanalare le competenze teoriche e renderle fruibili dal mondo industriale. Vengono anche rafforzati i programmi di dottorato nei settori dello studio di AI sia relativamente alle scienze dure che alle scienze umane connesse alla tecnologia.

Andando un po' più nel dettaglio del documento ciò che l'Italia punta a realizzare in questo biennio è: a) Rafforzare la ricerca dell'AI e i finanziamenti associati; b) promuovere misure per attrarre talenti; c) migliorare il processo di trasferimento tecnologico; d) aumentare l'adozione dell'AI tra le imprese e la PA e favorire la creazione di imprese innovative; e) rafforzare la leadership europea in ambito tecnologico; f) prepararci ai cambiamenti socio-economici che l'AI comporterà; g) stabilire un quadro giuridico appropriati basato su valori e principi etici.

Inoltre cinque sono i principi guida ossia: 1) L'AI italiana è un AI europea e quindi si auspica l'adozione di regole armonizzate con gli altri Paesi dell'UE; 2) Fare in modo che l'Italia diventi un polo globale di ricerca e innovazione dell'AI; 3) Rendere l'AI italiana sempre più antropocentrica, affidabile e sostenibile (dal punto di vista ambientale), che favorisca l'inclusione economica e sociale, responsabile e trasparente; 4) Far diventare le aziende italiane leader nella ricerca, nello sviluppo e nell'innovazione di AI, promuovendo programmi di AI per le imprese; e da ultimo 5) Fare in modo che le pubbliche amministrazioni italiane possano governare l'AI con l'AI.

A tal fine i tre Ministri sopracitati creeranno un gruppo di lavoro permanente sull'AI all'interno del Comitato Interministeriale per la Transizione Digitale per dirigere, monitorare e valutare l'attuazione di questa strategia, le sue successive iterazioni nonché coordinare tutte le azioni politiche sull'AI in futuro. Ciò comporta la possibilità di coinvolgere altri attori istituzionali, università e centri di ricerca, e ovviamente rappresentanti del settore privato.

All'interno di tutto ciò non vi alcuna specificazione del settore giustizia, la quale però viene fatta rientrare negli ingenti investimenti promessi per il settore della pubblica amministrazione.

3.2 Alle origini della motivazione delle decisioni pubbliche

La motivazione dei provvedimenti giurisdizionali e, in generale, delle decisioni pubbliche che riguardano specifici individui è un punto realmente critico nella applicazioni di tecnologie di AI al diritto.

L'utilizzo di tecnologie di AI è conforme al principio del giusto processo? Questa è la domanda che si pone il network europeo dei consigli superiori (ENCJ)¹¹⁷, richiamando l'art, 6 CEDU, secondo il quale l'esito del processo deve essere trasparente e deve poter essere spiegato come si è raggiunto un certo risultato. Anche la Commissione

¹¹⁷ La Rete europea dei Consigli di Giustizia è sorta con l'obiettivo di contribuire al consolidamento dello spazio europeo di libertà, sicurezza e giustizia, nel contesto di un processo di forte accelerazione delle politiche di cooperazione giudiziaria. Tale Rete ha come obiettivo quello di porsi come collegamento tra le istituzioni europee, le loro politiche e le varie magistrature nazionali, per favorire l'attenzione ai principi di autonomia e di indipendenza del potere giudiziario nell'elaborazione degli strumenti normativi di cooperazione.

europea per l'efficienza della giustizia (CEPEJ) si pone problemi analoghi. In particolare, si chiede se l'utilizzo di tecniche di AI possa avere effetti quasi prescrittivi, creando una nuova forma di normatività, che potrebbe integrare la legge limitando la discrezionalità sovrana del giudice e portando, in una prospettiva a lungo termine, a una standardizzazione delle decisioni giudiziarie, che verrebbero a essere non più basate sul ragionamento caso per caso dei tribunali, ma su un puro calcolo statistico. In definitiva, il problema è se “tali soluzioni siano compatibili con i diritti individuali sanciti dalla Convenzione europea dei diritti dell'uomo (CEDU)”, in particolare il diritto a un processo equo, il diritto ad un giudice naturale precostituito per legge, il diritto a un tribunale indipendente e imparziale e “l'uguaglianza nelle armi nei provvedimenti giudiziari). D'altra parte, anche l'art. 111 della Costituzione italiana prevede, sin dalla sua originaria formulazione, che “tutti i provvedimenti giurisdizionali devono essere motivati”, un aspetto importante del diritto a un giusto processo.

Insomma, la motivazione delle decisioni giurisdizionali si presenta come uno dei punti più delicati, forse il più delicato, del raccordo tra AI e principi giuridici. Al punto da far ritenere impossibile o preclusivo l'utilizzo di qualsiasi sistema di AI che impedisca la spiegabilità di ogni passaggio dalla specifica decisione giudiziaria come decisione dello specifico caso. Tali posizioni, e preoccupazioni, sono perfettamente comprensibili e meritano la massima attenzione. Tuttavia, mi pare sia necessario, quando si pongono in luce i limiti della decisione presa con l'ausilio di tecniche di AI, che si individui esattamente quale sia il termine di comparazione. In altri termini, bisogna spiegare cosa si presupponga a proposito di motivazione quando si esclude l'utilizzabilità di tecniche di AI nel corso di una decisione

giurisdizionale: si presuppone forse che, secondo quanto previsto dalle norme vigenti e quanto praticato nelle corti, ogni passaggio delle decisioni debba essere o sia effettivamente o potenzialmente spiegato o spiegabile?

La realtà presenta sfumature numerose e chiaroscuri che meritano attenzione. Il tema della motivazione è molto più complesso ed è stato trattato numerose volte in dottrina. Qui si seguito faccio solo cenno ad alcuni aspetti cruciali, a proposito di quale sia la natura della motivazione e cosa ad essa si richieda, con lo scopo di proporre qualche argomento che possa contribuire a ridimensionare alcune posizioni contrarie pregiudizialmente all'uso dell'AI nel decidere. Per esempio, si può cominciare ad accettare che gli umani hanno i loro pregiudizi e un loro modo unico e non conoscibile di prendere decisioni:

Non comprendiamo appieno il pensiero umano, ma accettiamo ancora i suoi errori. Se si commette un errore in un test di matematica, si potrebbe essere in grado di tornare indietro e capire dove si è verificato l'errore e come risolverlo per la prossima volta. Ma nessuno ha idea di come il cervello stesso sia arrivato internamente alle sue conclusioni! Le neuroscienze non sono ancora a questo punto e non abbiamo grossi problemi a cavarcela senza quella conoscenza. Se vogliamo affrontare alcune delle questioni etiche e giuridiche dell'intelligenza artificiale all'interno di settori sensibili agli errori, come l'assistenza sanitaria e la Finanza, non ha davvero senso lavorare sull'interpretabilità. Si tratta più di applicare l'AI nel modo giusto.¹¹⁸

¹¹⁸ George Seif, *AI doesn't need to be fully interpretable*, versione online visitato l'ultima volta il 22 novembre.

Da un punto di vista storico poi, è una conclusione quasi universalmente accettata tra gli studiosi che non esistesse un obbligo di motivazione sia nel diritto romano, sia nel diritto antico e medio-antico, in genere, mentre l'idea di fornire una spiegazione logico-razionale della decisione iniziò a far capolino nel periodo dell'Illuminismo e al tramonto della visione "tsistica" del giudice "bocca della legge".¹¹⁹

Per quanto riguarda l'Italia, l'obbligo di motivazione risale alle costituzioni giacobine di fine Settecento, con funzione principalmente interna al processo a tutela dei diritti delle parti, mentre la funzione di controllo estero del potere giudiziario era rimesso alla Corte di Cassazione, cui era devoluto il controllo sulla attuazione del principio di subordinazione del giudice alla legge.¹²⁰

Come si è visto, in Italia è l'art. 111 della Costituzione a prevedere per la prima volta in modo esplicito in un testo di legge non ordinaria¹²¹ l'obbligo di motivazione. Ma se si volesse trovare un ricco dibattito sull'origine di tale disposizione nei lavori preparatori, il risultato sarebbe deludente e a fatica si troverebbero indicazioni circa i contorni e il contenuto dell'obbligo di motivazione. Poco se n'è parlato nei lavori preparatori e, quando lo si è fatto, è costituito dal rinvio al principio di legalità, che, nello stabilire che "i giudici sono soggetti soltanto alla legge", pone, per un verso, un argine alle pretese di controllo del potere politico (anche se oggi sembrerebbe più il contrario) e, al tempo stesso, traccia un confine all'interno del quale i giudici sono autorizzati a

¹¹⁹ Barbara Biscotti, *Dispositivo e parte motiva nella sentenza: idee vecchie e nuove* in AA.VV., *Il giudice privato nel processo civile romano. Omaggio ad Alberto Burdese*, Vol. 1, Cedam, Padova, 2012, pp. 273-331.

¹²⁰ Vittorio Denti, *Sub art. 111 Cost.* in *La magistratura*, Vol. IV, in *Commentario alla Costituzione*, a cura di G. Branca, Zanichelli, Bologna, 1987, pp. 5 ss.

¹²¹ L'obbligo di motivazione era già prescritto nel codice di procedura civile del 1865 e in quello del 1940, nel testo vigente fino alla riforma del 2009, come "le ragioni di fatto e di diritto della decisione" (art. 132, co. 2, n. 4 c.p.c.).

esercitare il loro potere: l'operare nei limiti della legge fonda il potere dei giudici e trova dimostrazione e conferma nell'esplicitazione dei motivi che hanno condotto ad una certa decisione (art. 101, co. 2 Cost.). D'altra parte, anche se riconosciuto in tutti gli ordinamenti affini al nostro, non sempre l'obbligo di motivazione è esplicitamente previsto nelle costituzioni o nei *Bill of rights*. Per esempio, la costituzione tedesca, il *Grundgesetz*, non ha una norma corrispondente al nostro art. 111 Cost., tanto che "la giurisprudenza della Corte Costituzionale dovette far riferimento al 'principio dello stato di diritto' per affermare l'obbligo di motivazione, inteso come dovere generale dei pubblici poteri".

A livello delle carte internazionali né la Dichiarazione universale dei diritti dell'uomo né il Patto per di New York del 1966 né la Convenzione Europea dei Diritti dell'Uomo contengono l'obbligo di motivazione delle decisioni giudiziali, pur prevedendo altre cautele, come la pubblicità dei processi, l'imparzialità del giudice e il giusto processo.

In definitiva si può dire che in epoca relativamente recente sia emersa una sensibilità giuridica condivisa circa l'opportunità, e in una certa misura l'obbligatorietà, che chi detiene il potere giurisdizionale, come qualsiasi altro potere, debba rendere conto delle proprie decisioni argomentandole.¹²² Tuttavia, fermo restando il solido ancoraggio nei principi fondamentali degli ordinamenti occidentali (essenzialmente il

¹²² Per il potere amministrativo, oltre all'art. 41 CEDU, è degna di nota la Corte Costituzionale (sent. N. 310/2010), la quale statuisce: "l'obbligo di motivare i provvedimenti amministrativi è diretto a realizzare la conoscibilità e quindi la trasparenza dell'azione amministrativa. Esso è radicato negli artt. 97 e 113 Cost., in quanto da un lato, costituisce corollario dei principi di buon andamento ed imparzialità della pubblica amministrazione e, dall'altro, conente al destinatario del provvedimento, che ritenga lesa una propria situazione giuridica, di far valere la relativa tutela giurisdizionale".

giusto processo e il principio di legalità), l'obbligo di motivazione dei provvedimenti giurisdizionali non ha un unico e univoco modo di essere riconosciuto e, soprattutto, di essere inteso. Si può dire sinteticamente: una motivazione vi deve essere, ma che cosa essa sia è ampiamente da vedere.

A livello dottrinario molti profili sono stati posti in luce.

Secondo la formula di più largo impiego, la motivazione è "l'espressione dell'iter logico-giuridico attraverso il quale il giudice dice che è pervenuto alla decisione"¹²³. In termini tecnicamente forse più precisi, e ridotta al minimo, la motivazione può essere considerata come "l'insieme degli enunciati linguistici formulati dal giudice in funzione giustificativa delle proposizioni che adempiono a un ruolo prescrittivo (dispositivo)"¹²⁴. In epoca più recente la motivazione della sentenza è stata ritenuta, sia sotto il profilo della teoria del linguaggio, sia anche della teoria generale del diritto (debitrice, per questo aspetto, nei confronti della prima), un discorso giustificativo, ossia "un discorso atto a far risultare, mediante ragioni, che qualcosa è giusto".¹²⁵

Tuttavia, anche tali minime e comprensive definizioni possono essere seriamente criticate. In generale è diffusa l'idea che i motivi espressi nella sentenza siano incapaci di rispecchiare la ricchezza delle intuizioni e del lavoro logico che ha guidato il giudice alla formulazione del dispositivo. Di qui una degradazione del loro valore, ridotto alla stregua di apparato formale destinato a manifestarsi come punto di

¹²³ Vittorio Denti, op. cit.

¹²⁴ Ennio Amodio, *Motivazione della sentenza penale*, in *Enc. Dri.*, Vol. XXVII, Giuffrè, Milano, 1977, pp.199 ss.

¹²⁵ Patrizia Borsellino, *La motivazione della sentenza come contesto di argomentazione razionale*, in AA.VV., *La fabbrica delle interpretazioni*, Atti del Convegno annuale della Facoltà di Giurisprudenza (Università di Milano-Bicocca 19-20 novembre 2009), pp. 141 ss.

emergenza di una realtà dalle radici ben più profonde, anche se, per le parti, la motivazione è tutt'oggi una garanzia ineliminabile.

Vi è poi che fa notare che “la stessa natura giuridica della motivazione non sia identificabile al di fuori di un approccio globale del fenomeno”. La motivazione, quindi, può essere inadeguata e considerata sotto due profili, come segno linguistico e come fonte d'inizi. La motivazione “segno” riguarda l'esposizione dei motivi con i quali il giudice comunica le ragioni della decisione alle parti e al pubblico (e al giudice dell'impugnazione), mentre la motivazione come “fonte di indizi” allude alla sua attitudine a esprimere, oltre il suo significato intenzionale, altri fatti, quali lo stato psicologico, il livello culturale, le opinioni e lo stato sociale dell'autore.

Quali debbano essere i requisiti della motivazione è stato storicamente oggetto di indagine da parte della Corte di Cassazione, che, investita del controllo del rispetto dell'obbligo di motivazione da parte dei giudici, ha adottato un concetto di motivazione con significati assai diversi “dalla pura e semplice esistenza grafica dello scritto, alla presenza di proposizioni che assumono come punto di partenza certi contenuti, per giungere fino alla pretesa dell'analiticità dell'accentramento o della persuasività degli enunciati del giudice”.

Un crinale interessante è costituito dalla visione della motivazione come operazione logica (o persino psicologica) e come esposizione di tutti gli elementi probatori assunti nel processo, che si riferiscono a un fatto da accertare: in altri termini come il giudice seleziona ciò che è emerso nel processo ai fini della motivazione in fatto. Nel processo civile, può valere anche per i motivi di diritto che una parte ha fatto valere e che, sia se accolti sia se rigettati, richiedono una spiegazione e risposta.

Si profila così un'opposizione tra una visione della motivazione come oggetto/entità razionale in sé, i cui enunciati si giustificano per la loro reciproca coerenza interna e in relazione ad altri enunciati teorici, che possono provenire da precedenti giudiziari e/o della dottrina, e la visione della motivazione come elaborato logico giudico che contiene, e deve contenere, una completa analisi delle questioni, di fatto o tesi di diritto, emerse nel corso del processo (per esempio fatti e prove). La questione è rilevante in sede di analisi delle decisioni e la loro predizione.

Oggi ci si trova di fronte a tre tipi di difficoltà. Alle due principali tradizionali, quella incarnata dalla corrosione scettica tipica del realismo giuridico, e quella della motivazione intesa in senso meramente formale, si somma l'emergere dei problemi di spiegabilità di elaborazioni eventualmente prodotte dall'uso di sistemi di intelligenza artificiale, preoccupazione quest'ultima di cui si fanno carico i documenti internazionali.

Premesso che l'idea di fare semplicemente argine all'uso di AI non ha molto senso, perché, ammesso anche che sia possibile, comunque lascerebbe sul tappeto i primi due problemi, va ricordato che le diverse tesi tradizionali riguardano e mettono in luce funzioni e aspetti diversi della motivazione, che sono rilevanti e acquistano maggior peso quando si aggiungono i problemi di spiegabilità. Alcune tesi considerano la motivazione come mero sillogismo (dalla norma astratta alle sue applicazioni al caso specifico), altre come resoconto del percorso decisionale, altre come razionalizzazione a posteriori, altre come esposizione che parte dallo stato dell'arte.

Qui di seguito esaminiamo alcuni aspetti che assumono rilievo con l'emergere del problema della spiegabilità dei risultati raggiunti con

tecniche di AI. L'obiettivo è di capire se, cosa e come il problema dell'*explainability* aggiunga ai precedenti problemi e quali prospettive apra il possibile utilizzo di tecnologie di AI.

Un punto di partenza utile per la nostra indagine è costituito dalla convinzione che la motivazione non sia rappresentazione del processo decisionale. Che la motivazione non fosse o non potesse essere il “fedele resoconto” del processo attraverso il quale il giudice giunga alla decisione era chiaro già negli anni '50 del '900, quando un esimio giurista come Calamandrei sosteneva che essa fosse “l'apologia che il giudice elabora a posteriori della decisione stessa”.¹²⁶

Ciò autorizza il ricorso alla distinzione tra il momento dell'esplorazione e della ricerca della soluzione (giuridica) di un caso e quello della sua giustificazione. In termini pratici, un conto è il complesso percorso che il giudice compie per arrivare alla decisione della lite e “un altro è il discorso giustificativo che il medesimo costruisce, al fine di rendere conto della propria decisione”.

Una visione di questo genere è, alla fine dei conti, in linea con le previsioni normative.

Sono ben noti, e regolati dalla legge, i casi di manifesta scissione tra il momento della decisione e quello della motivazione. L'art. 544 del Codice di procedura penale prevede che “qualora non sia possibile procedere alla redazione immediata dei motivi in Camera di consiglio, vi si procede non oltre il quindicesimo giorno da quello della pronuncia” e che “quando la stesura della motivazione è particolarmente complessa, [...] il giudice [...] può indicare nel dispositivo un termine più lungo”. Le regole di procedura civile prevedono un'analogo possibile scissione tra momento decisorio e

¹²⁶ Piero Calamandrei, *Processo e democrazia*, Cedam, Padova, 1954.

momento di deposito della motivazione nelle cause in materia di lavoro (art. 420 c.p.c.), dove è prevista la lettura del dispositivo e il successivo deposito della motivazione, e in tutte le altre materia che applicano i riti del lavoro o riti a esso ispirati (come per esempio nella impugnazioni di sanzione CONSOB o della Banca d'Italia).

In tutti questi casi, penali e civili, il giudice decide di assolvere/condannare/dichiarare/constituire o altro, e poi ha un termine per motivare quella decisione presa, appunto, antecedentemente.

Si può quindi dire, a un primo livello di approssimazione, che, ben prima dell'avvento dell'AI, la motivazione non è resoconto di tutto quanto accaduto nel processo e nella fase decisoria, ma una sua giustificazione a posteriori. Questo significa che vi è uno iato, o un possibile disallineamento, tra motivazione e quello che un fedele resoconto potrebbe dire del percorso che ha portato a quella decisione.

In questo iato che riguarda sia il tempo della stesura sia la natura stessa dell'atto-motivazione, si possono inserire diversi elementi e considerazioni, di tipo sia teorico sia fattuale.

A livello teorico si considerino le applicazioni della *defeasible logic*, cioè di quella logica che supera i limiti della logica monotona (secondo la quale il sistema giuridico è una base assiomatica, alla quale si applica una logica monotona) e che, per evitare la derivazione di conclusioni incompatibili con nuove norme prevalenti, accetta che tale base assiomatica sia modificata corrispondentemente. Nota il Prof. Giovanni Sartor, a sostegno della necessità di adottare una logica *defeasible*:

Nel ragionamento giuridico hanno particolare importanza i precedenti. In generale, il fatto che un caso precedente P1 sia stato deciso in un certo modo

fornisce una ragione per decidere nello stesso modo un nuovo caso simile *c*, che condivida con P1, almeno in parte, gli aspetti o fattori che hanno motivato la decisione di P1. Si tratta di una ragione *defeasible*, che può essere superata da ragioni in contrario. In particolare, agli argomenti che richiamano il precedente P1, si possono opporre considerazioni basate su aspetti o fattori la cui presenza o mancanza distingue il caso C da P1, argomentando che tali differenze richiederebbero una decisione diversa per C. Tali considerazioni possono essere rafforzate dal richiamo ad altro precedente P2, nel quale si sia raggiunta una decisione diversa da quella di P1, e che condivida con C la presenza di fattori assenti in P1 o l'assenza di fattori presenti in P1.¹²⁷

Premesso che quanto detto a proposito del precedente vale, con solo lievi modifiche, anche per l'applicazione di norme di legge, chiunque abbia un minimo di dimestichezza con il modo in cui si giunge a una decisione giurisdizionale non farà fatica a riconoscere in questa esposizione teorica alcune dinamiche tipiche della formazione della decisione.

Lo studio preliminare della causa richiede, in una prima fase, la messa in ordine dei fatti esposti dalle parti e della norme/teorie giuridiche che esse hanno indicato a sostegno delle domande proposte al giudice. In questa fase il giudice ipotizza applicazioni di norma/precedenti e procede per tentativi e verifiche delle priorità logico-giuridiche delle

¹²⁷ Giovanni Sartor, *L'informatica giuridica e le tecnologie dell'informazione. Corso di informatica giuridica*, Giappichelli, Torino, III edizione, 2016. L'autore spiega così un ragionamento non monotono: "Supponiamo che Tizio abbia causato un danno a Caio, ma che Tizio fosse incapace di intendere e di volere, essendo in stato di ubriachezza. La prima delle nostre norme ci condurrebbe a ritenere che Tizio sia responsabile, ma la seconda norma, più specifica, fornisce un'eccezione alla prima, e quindi Tizio non sembrerebbe responsabile. Infine, la terza norma, escludendo l'applicazione della seconda (che avrebbe escluso la prima) ci conduce a ritenere che la prima si applichi e quindi che Tizio sia responsabile. Si tratta di un ragionamento non-monotono: se aggiungiamo nuove premesse a una teoria giuridica T vengono meno alcune conclusioni derivabili da T".

questioni e della tenuta delle ipotesi prospettate. Realizzato un sufficiente grado di chiarezza e convinzione, il giudice decide secondo un ragionamento che è nella sua mente, nei suoi appunti, in una bozza di sentenza, e poi, dopo aver formulato e letto il dispositivo, fornisce il quadro completo delle motivazioni in un testo (sentenza), che rende pubblico attraverso il dispositivo.

Ogni persona, non necessariamente un giudice, che abbia esperienza di scrivere, non dico un saggio teorico, ma anche una semplice lettera a chicchessia, sperimenta facilmente la delicatezza del passaggio dalla fase, magmatica e meditata, di formazione di un'idea di quello che si intende scrivere in una lettera e lo scrivere effettivamente una serie di enunciati che corrispondano a quell'idea, e sperimenta tutta la differenza che può intervenire tra pensare un enunciato e porlo in forma scritta. Questo accade quotidianamente ai giudici, che talora scoprono ulteriori buoni motivi della decisione presa sulla base dello studio preliminare e dell'esperienza e la difficoltà di porre una sequenza di parole, frasi e concetti che abbiano un fondamento logico e razionale.

Può capitare che si scopra che un fatto, al quale era stato dato un peso x , in realtà ne abbia uno nettamente maggiore o minore. A quel punto la decisione può cambiare nel tipo di giustificazione che del dispositivo si dà, che, a causa dell'emergere di elementi di fatto o di diritto che vadano in senso contrario, deve seguire una via diversa da quella immaginata inizialmente. Tutto questo, si badi bene, non è, e giustamente, oggetto di rendiconto nella motivazione.

Secondo un'idea corrente, la maggior parte dei sistemi di apprendimento automatico (*machine learning*), e soprattutto quelli di *deep learning*, sono essenzialmente scatole nere (*black boxes*), in cui

non si può davvero controllare come l'algoritmo raggiunga il risultato che raggiunge.

La *black box* è una metafora atecnica e suggestiva, al pari di altre, come per esempio essere un 'oracolo'.¹²⁸ In realtà, come si è visto per la motivazione in generale, anche la comprensione di cosa venga considerata una scatola nera richiede una chiarificazione su cosa si intenda per spiegazione. La scatola nera, infatti, si definisce proprio in opposizione a quello che è ritenuto spiegabile o spiegato, rispetto al quale la scatola nera, per la quale conosciamo solo gli stimoli in entrata (*input*) e le risposte in uscita (*output*), si presenta come un 'entità le cui operazioni interne non possono essere oggetto di indagine'.¹²⁹

E' stato fatto notare che la spiegazione può essere intesa in almeno due modi. Se si chiede a una persona perché ha fatto qualcosa, si parte dal presupposto che la persona abbia avuto una buona ragione per aver agito in quel modo e, fondamentalmente, ci si chiede quale sia stato il ragionamento attraverso il quale quella persona ha preso quella decisione, e ci si aspetta che abbia pesato i pro e i contro e abbia scelto una linea d'azione in previsione di un risultato atteso.

Se invece, ed è il secondo modo di intendere la spiegazione, ci si chiede perché qualcosa sia andato storto, si sta chiedendo una sorta di

¹²⁸ Che infatti non si trova in testi tecnici come quello di Russel e Norvig. La metafora dell'oracolo è usata, invece, nell'editoriale di "Nature" dopo la vittoria di Alpha GO: *Digital intuition*, "Nature", 28 gennaio 2016, Vol. 529, p. 437.

¹²⁹ La questione è al centro anche di sviluppi tecnologici, nel senso della chiarezza ai quali sono interessati anche nel mondo delle imprese. In effetti, *deep learning* ha la reputazione di essere un metodo "scatola nera". Diverse storie di successo per la costruzione di "scatole di vetro" hanno utilizzato modelli per migliorare la trasparenza e l'interpretazione. Alcuni progetti nel campo dell'intelligenza artificiale spiegabile sono riportati in Vishal Morde, *5 Key Lessons Learned from 2018 AI Summit*, NYC. Degno di nota anche l'interesse degli apparati militari al tema della spiegabilità: sul punto si veda Benjamin Powers, *What is Explainable AI and why does the military need it?* Versione online, visitato l'ultima volta il 22 novembre 2021.

spiegazione a posteriori di un fallimento. Ad esempio, dopo un incidente d'auto, si potrebbe volere una spiegazione di cosa ha causato l'incidente. L'autista era distratto? Un'altra macchina lo ha fatto sterzare? In questo caso, più che un ragionamento, si cerca, più o meno, l'evento critico che ha causato una reazione particolare al di fuori del normale comportamento.¹³⁰

La spiegazione che si vorrebbe ricevere o che ci si aspetterebbe, quando si pensa all'intelligenza artificiale, sembra essere quella del primo tipo (la ragione per la quale abbia scelto una linea d'azione sulla base di quale risultato atteso), anche se per lo più i sistemi sono simile al secondo tipo, ossia ricevono stimoli ai quali reagiscono. Di solito, quindi, bisogna prendere atto che è più difficile capire le ragioni per le quali è stata presa una decisione particolare.

In realtà, i sistemi di *machine learning* e di *deep learning* non sono scatole nere, nel senso proprio del termine, perché sono costituiti da un insieme di reazioni elementari, ciascuna delle quali sarebbe in sé spiegabile. Ma esse sono molto numerose, e quindi non facili da ricostruire da un umano, oppure, nel caso dei modelli di *deep learning* (o *neural networks*), le interazioni sono non-lineari, che significa non solo che non vi è un unico percorso tra input e output, ma che l'effetto di cambiare un input può dipendere dai valori di altri input. Questo rende molto difficili da concepire mentalmente cosa stia accadendo nel sistema, anche se i dettagli sono nondimeno trasparenti e del tutto disponibili per un 'eventuale ispezione.

La situazione si può anche rovesciare. Nelle ricerche sulla visione umana, per esempio, si sta tentando, con i sistemi di *machine learning*,

¹³⁰ Dallas Card, *The 'black box' metaphor in machine learning*, 2017, visitato l'ultima volta il 26 novembre 2021.

di imitare il comportamento umano usando solo gli input e gli output. In tal caso, se considerato dal punto di vista di un sistema di *machine learning*, è l'umano a costituire la scatola nera.

In sintesi, si può dire che difficoltà di conoscenza e di ricostruzione della decisioni sicuramente esistono, ma sono difficoltà materiali (come, per esempio, l'eccessiva onerosità e quindi non convenienza economica della ricostruzione) e non il risultato di chissà quale mistero del sistema.

D'altra parte, non sono questa difficoltà le uniche che si incontrano nella nostra vita sociale. Un segreto industriale, per esempio, è un ostacolo giuridico alla conoscenza che può avere, che si tratti della formula della Coca Cola o del software proprietario con il quale i giudici americani calcolano l'aumento di pena per il rischio di recidiva (vedi caso Loomis), effetti di blocco della conoscenza che non sono da meno. Gli esempi in ambito giuridico possono continuare. Un effetto *black box* è creato dal divieto, o dal non uso, della pubblicazione dell'opinione di minoranza, e quindi dissenziente, nella decisione di un collegio: l'effetto è di rendere inesplorabile il confronto di opinioni che si è sviluppato in un collegio.

Così come alcuni contributi scientifici nelle consulenze tecniche o nella perizie. Ricordo un caso in cui una parte, che agiva in giudizio contro l'autore di una contraffazione di un brevetto, si rifiutò di far accedere il consulente tecnico, nominato dal giudice, al proprio processo di fabbricazione, valutando più dannosa per sé la rivelazione delle specifiche tecniche di quel processo rispetto al rischio di non fornire la prova o di non consentire l'accesso alla prova. Ancora una volta un *black box* giuridico e tecnico. Vi sono poi alcune prassi decisionali basate su "inerzie" che somigliano a scatole nere, come quando un

giudice decide sulla base di una sentenza della Cassazione di cui esamina solo il dispositivo o la massima, senza prestare attenzione al caso concreto dal quale quella affermazione di diritto prende le mosse. In conclusione, le *black box*, nel diritto o nelle prassi, esistono, così come esistono nella medicina¹³¹ o nel pilotaggio degli aerei, come è emerso nei disastri dei Boeing 737 Max nel 2019. Si tratta di capire come gestirle e come integrare i processi decisionali. Non va, poi, sottovalutato il fatto che non tutto è negativo e che può essere importante, anche in campo giuridico, la scoperta di relazioni che prima non erano mai state considerate.¹³²

Sembra andare in questa direzione il progetto di ricerca su come decidere *About, By and Together with Algorithmic Decision Making Systems* (ADM), promosso dall'Hans-Bredow-Institute di Amburgo, che si propone di rispondere a domande come le seguenti:

In che modo gli umani prendono decisioni sugli altri umani rispetto a come i sistemi di ADM prendono le stesse decisioni sugli umani? In che modo gli esseri umani in congiunzione con i sistemi ADM prendono decisioni su altri umani? Quali sono i limiti in cui le macchine dovrebbero prendere decisioni sulle persone? E come possono gli Stati decidere se i sistemi ADM dovrebbero essere utilizzati all'interno dei sistemi di giustizia penale.¹³³

¹³¹ Willian Nicholson Price, *Big Data and black-box medical algorithms*, in "Science Transnational Medicine", Vol. 10, Issue 471, 2018.

¹³² Elizabeth A. Holmes, *In defence of the black box. Black box algorithms can be useful in science and engineering*, in "Science", Vol.364, issue 6435, 2019, pp. 26 ss.

¹³³ Il progetto è promosso dall'Hans-Bredow Institute (un istituto indipendente, non-profit che opera dal 1950 in collaborazione con l'Università di Amburgo).

E' utile, a questo punto, tornare a porsi una domanda, più profonda, se si vuole: di cosa è fatta una sentenza/motivazione, quali sono i suoi componenti, come è o deve essere organizzata al suo intento e cosa deve contenere rispetto alla totalità del materiale contenuto nel processo e nei suoi atti?

La risposta più convincente è, a mio avviso, che la motivazione, sia essa penale o civile, è l'insieme degli elementi che provengono dal processo e che hanno costituito oggetto di prospettazione da parte dei difensori delle parti (o dalle parti stesse) e che il giudice pone in un proprio ordine logico giuridico.

Non paia una definizione riduttiva. Essa è l'unica in grado di rispondere alla regola che vieta al giudice di decidere oltre le richieste delle parti o al di fuori di esse (*no ultra petita o extra petita*)¹³⁴ e al principio del contraddittorio, secondo il quale il giudice che ritenga “di porre a fondamento della decisione una questione rilevate d'ufficio” deve assegnare alle parti un termine per poter interloquire con il deposito delle memorie (art. 101 c.p.c.).

Dal che si deduce che il giudice deve pronunciare sulla questioni proposte dalle parte e non può porre a fondamento della decisione questioni sulle quali le parti¹³⁵ non abbiano potuto esercitare il diritto di difesa. Potrà certo, se lo ritiene, approfondire o anche sviluppare, ma a partire da una questione che sia stata oggetto d'interlocuzione.

¹³⁴ Codice di procedura civile, art. 112 rubricato: “corrispondenza tra il chiesto e il pronunciato” : “Il giudice deve pronunciare su tutta la domanda e non oltre i limiti di essa; e non può pronunciare d'ufficio su eccezioni, che possono essere proposte soltanto dalle parti”.

¹³⁵ Incluse le prove e il loro risultato, che non può essere ignorato in nome della pura consistenza razionale interna dell'elaborato sentenza. D'altra parte, il testo attuale dell' art. 360 c.p.c. prevede il ricorso per Cassazione “per omesso esame circa un fatto decisivo per il giudizio che è stato oggetto di discussione tra le parti”.

A stretto rigore si può dire che il giudice non aggiunge argomenti, ma organizza e gerarchizza quelli esistenti in modo funzionale alla decisione e che, quindi, una buona motivazione è una motivazione che rende esplicito il perché e il come, e secondo quale ordine, il giudice abbia deciso di aggregare quegli elementi giuridici: niente di più e niente di meno. Questo perché la decisione e la motivazione arrivano all'esistenza di un processo governato dalla regola fondamentale, non tanto e non soltanto della legalità, quanto del contraddittorio, espressione massima del diritto di difesa e del giusto processo.

Il passo successivo è chiedersi quale sia la natura di questi elementi/argomenti giuridici che vanno a popolare la motivazione della sentenza. Essi sono norme di legge, norme costituzionali o di trattati internazionali, norme provenienti da altri ordinamenti giuridici (caso sempre più frequente, soprattutto e non solo in ambito europeo), elaborazioni dottrinarie.

Tutti questi materiali, si noti, sono gli stessi e vengono tratti da fonti condivise con gli avvocati, gli studiosi dell'accademia, le amministrazioni in vista di loro decisioni o gli stessi legislatori. La differenza sta solo nel modo in cui ognuno di questi professionisti o di questi enti organizza le informazioni, modo che dipende dalle proprie finalità istituzionali e dai propri confini deontologici.

Quanto alla forma, queste informazioni sono espresse in linguaggio naturale, cioè nel linguaggio usato nella specifica comunità di vita. Nella sentenza possono concorrere, anche coesistere, diversi linguaggi naturali, come capita talora con fonti europee quando sono espresse solo nelle due lingue dell'Unione come l'inglese e il francese. Inoltre, in quanto contenuti giuridici inseriti in contesti informatici/digitali, essi

hanno natura di dati, che possono esser strutturati, semistrutturati o non strutturati.

Si può, quindi, precisare l'affermazione di cui sopra dicendo che il giudice non aggiunge dati (da leggersi come argomenti), ma li organizza e gerarchizza in modo funzionale alla decisione e che, quindi, una buona motivazione è una motivazione che esplicita il perché, il come e il modo in cui il giudice abbia deciso di aggregare quei dati (elementi giuridici).

Se gli elementi giuridici di cui è composta la sentenza sono dati bisogna trarne la conclusione che la sentenza (al pari con gli atti degli avvocati e delle fonti dalle quali quei dati provengono) è un aggregati di dati, che il giudice ha ricomposto e organizzato in un modo idoneo a giustificare razionalmente la sua decisione. Ecco perché la motivazione non aggiunge dati ma organizza dati in modo funzionale alla decisione, e una buona motivazione ha il compito di esplicitare il perché e il come del modo di aggregazione adottato. Tutto ciò è comune al campo civile e a quello penale, all'avvocato e al giudice, all'amministratore pubblico come al legislatore a qualsiasi livello. I loro atti sono tutti aggregati di dati scomponibili e ricomponibili in varia maniera.

Uno dei risultati pratici è che le sentenze depositate in sistemi informatici vanno a costituire un insieme di decisione che, se ben organizzato in un magazzino di dati, è interrogabile con ricerche per parola, come negli attuali database, o esplorabile con modalità tecniche più raffinate, come i più volte menzionati sistemi di *machine learning*. E' utile, però, che le sentenze (e qualsiasi altro atto processuale) siano organizzati in modo da andare a costituire un insieme di dati (*dataset*) di buona qualità, poiché migliore è la qualità iniziale del dato, migliore è l'esito delle operazioni di AI. Negli ultimi anni numerosi gruppi di

lavoro sono stati istituiti in ambito giudiziario e numerosi documenti sono stati dedicati alla cosiddetta sinteticità degli atti. Chi scrive sostiene che il problema non sia la sinteticità dell'atto al fine di essere catalogato e fungere da "precedente", bensì la sua struttura logica e l'organizzazione in parti adeguate, distinte e gerarchizzate. Un atto di tal genere, che alla fine sarà anche più sintetico, è candidato a essere un atto di buona qualità nell'ottica del *datawarehouse*.

Un *datawarehouse* ben organizzato e alimentato con dati di buona qualità può essere interrogato con algoritmi che possano aiutare il giudice rendendo visibili quelle informazioni che a occhio nudo non si riesce a vedere o, anche, proponendo una bozza delle sentenza da scrivere e dei precedenti e delle norme applicabili, una volta che al sistema siano state fornite le coordinate spazio-temporali e di materia della questione da decidere.

A quel punto al giudice resta il compito, più difficile e di più alto pregio intellettuale e professionale, di selezionare quel materiale e quella proposta di motivazione, di sfidarla, cambiando o precisando alcuni parametri, o alcuni elementi di fatto e di diritto che contraddicono e cambiano la consequenzialità della proposta del sistema.

Appare necessaria una visione della motivazione come creazione non isolata ma inserita in un flusso di informazioni organizzate come dati, dove essa rappresenta soltanto un passaggio, per quanto di grande importanza pratica e sociale.

Infine, e per chiudere il cerchio e spiegare il motivo per il quale ci siamo soffermati così a lungo sui parametri a nostro parere imprescindibili per una sentenza e per la sua motivazione, pensiamo che tali atti, tipici del lavoro del giudice, non paiono posti in pericolo con le nuove tecniche di AI, solo a condizione che con le nuove tecniche vengano rispettati

tuti quei parametri di cui detto sopra, in primis, la motivazione deve avere la spiegazione, quanto più dettagliata possibile, (ciò non significa necessariamente che debba essere lunghissima) dei passaggi logici e “matematici” sulla base dei quali si è arrivato a quella decisione; ciò pare necessario per giustificare il dispositivo e naturalmente per fondare eventuali ricorsi avverso il medesimo che, come noto, si basano proprio sulle “falle” nella motivazione.

3.3 La giustizia digitale tra modello induttivo e modello deduttivo

Como ormai noto, quando si parla di giustizia predittiva si intende la possibilità di prevedere l’esito di un giudizio tramite alcuni calcoli; non si tratta di predire tramite formule magiche, ma di prevedere la probabile sentenza, relativa ad uno specifico caso, attraverso l’ausilio di algoritmi. Il diritto può essere costruito come una scienza, che trova la sua principale ragione giustificativa nella misura in cui è garanzia di certezza: il diritto nasce per attribuire certezza alle relazioni umane, tramite una complessa attribuzione di diritti e doveri.¹³⁶

Lo scopo quindi è prevedere le decisioni giurisprudenziali mediante algoritmi “addestrati” all’analisi di *database* contenenti precedenti ed altre informazioni utili ad aumentare il grado di certezza del diritto e la qualità delle decisioni, nonché risolvere altri problemi della giustizia come i lunghi tempi processuali.

¹³⁶ Luigi Viola, *Interpretazione della legge con modelli matematici*, Processo, Ad, giustizia predittiva, Volume 1, Diritto avanzato, Milano, 2018 (II edizione).

La Francia ha messo a punto diversi progetti per il calcolo di probabilità delle sentenze. Ancora prima la Corte Suprema del Wisconsin ha legittimato l'uso di algoritmi nel caso Loomis (di cui si è detto) divenuto appunto celebre per aver evidenziato alcuni dei problemi dell'intelligenza artificiale applicata alla giustizia.

Le diverse start-up create negli anni hanno cercato di sfruttare i vantaggi dell'AI, che ci permette di analizzare quali benefici potrebbe portare la giustizia.

Lex Machina, ad esempio, creata nel 2008, è uno strumento con una varietà di funzioni per aiutare gli avvocati ad impostare la strategia. La sua articolazione, Timing Analytics, usa la tecnologia per stimare i tempi processuali, aiutare gli utenti a selezionare gli avvocati rispetto alla loro esperienza e prevedere le possibilità di successo davanti a un giudice o a un tribunale.

La funzione di Judge Dashboard della compagnia Ravel Law, invece, contiene casi, citazioni, schemi e decisioni di un giudice specifico per poter supportare gli avvocati a capire come potrà decidere su un caso. Orbene, da questo vediamo che gli algoritmi, se correttamente progettati e applicati ad un database ricco e trasparente, possono facilitare il lavoro dei professionisti e aumentare la qualità del sistema, Quest'ultima precisazione è fondamentale, perché introduce l'altro volto della giustizia predittiva.

In primo luogo, una banca dati di bassa qualità e poco trasparente può creare molti problemi.

In caso Loomis è un esempio, testimoniando le derive discriminatorie che può avere l'AI se usata su informazioni sbagliate, come la razza. Allo stesso modo, le applicazioni alle misure cautelari avvenute negli USA hanno riconfermato la penalizzazione delle minoranze etniche e

delle classi sociali povere in virtù di algoritmi addestrati più all'analisi meccanica che all'interpretazione del caso e banche dati con informazioni troppo focalizzate sulle caratteristiche dei soggetti.

E' noto come il potenziale di trasformazione dell'organizzazione e del funzionamento delle istituzioni pubbliche portato dalla introduzione delle tecnologie costituisce ad oggi uno dei temi portanti dell'agenda internazionale e nazionale indipendente dalla forma di stato o di governo presente nelle varie aree geografiche; anzi, spesso si nota una maggiore affezione e un maggior sviluppo della tecnologia (di sorveglianza ma anche di aiuto alle decisioni dei giudici) in paesi poco democratici rispetto a quelli in cui la democrazia è valore fondante.

Non vi sono dubbi poi che una più intensa diffusione delle tecnologie possa, se adeguatamente governata ed accompagnata da misure di carattere professionale e formativo coadiuvare le politiche di miglioramento della efficienza della *governance* giudiziaria. Tempi, costi e modalità di gestione dei fascicoli sono per fare un esempio, gli aspetti maggiormente considerati dalle organizzazioni di carattere intergovernativo o dai network giudiziari impegnati nel monitoraggio della qualità della giustizia.

Naturalmente la giustizia così come ogni altro apparato burocratico dello stato per funzionare bene deve porsi degli obiettivi che sono tutti a favore del cittadino, utente finale del servizio; tale soggetto appare centrale anche quando si parla di giustizia *hig-tech* o giustizia digitale o giustizia predittiva (sulla differenza fra i vari termini ci siamo soffermati poco sopra).

Secondo la definizione data da uno dei più importanti studiosi della giustizia predittiva essa "è ancora allo stato di progetto", ossia è in una fase embrionale in cui si inizia a verificare sul campo la possibilità di

tecnologie applicate a ambiti di tipo giuridico, giurisprudenziale o giudiziario. Non siamo infatti dinnanzi ad una applicazione di programmi software che lavorano su piattaforma su cui sono disponibili i fascicoli digitali e che hanno come funzione dominante quella di gestire passaggi, la scrittura, la revisione, la condivisione, la validazione del fascicolo. Si tratta invece di algoritmi che hanno come campo di applicazione contenuti decisionali, test di sentenze, decreti, atti del giudice, banche dati giurisprudenziali, appartenenti a sistemi anche molto diversi e distanti fra loro per sensibilità culturale e giuridica. Tali algoritmi sono strutturati come funzioni i cui argomenti sono appunto tali ambiti e i cui risultati sono “probabilità” di costo, orientamento decisionale, range di penalità, range di indennizzi ecc. La giustizia predittiva è dunque l’amministrazione della giustizia data con sofisticate applicazioni di tecnologia (algoritmi) con finalità di carattere analitico/induttivo (si scoprono pattern decisionali o pattern comportamentali analizzando e processando dati che riguardano casi e decisioni già avvenuti) sia con finalità prospettico/predittivo ossia si individuano “propensioni decisionali” e su queste basi vengono valutate le probabilità con le quali si può prevedere che la decisione del giudice, nel caso di soluzione giudiziale delle controversie, o del mediatore, nel caso di attivazione di meccanismi di *alternative dispute resolution*, converga su un punto che possiamo definire focale. In altri termini, come è ovvio, non si tratta di predire con esattezza puntuale il dispositivo di una sentenza, ma di individuare l’orientamento del ragionamento del giudice. Poiché tale ragionamento non ha mai la natura di un sillogismo lineare, ma si compone di passaggi analogici deduttivi induttivi, la predizione sarà focale e non puntuale.

Il termine giustizia predittiva è una locuzione oggi utilizzata non solo dagli operatori del diritto ma anche dal comune cittadino che auspica una giustizia più rapida, più trasparente, più “prevedibile” (per evitare processi persi in partenza per esempio) e, se vogliamo, più tecnologica. Su questo ultimo punto chi scrive desidera aprire una breve parentesi. Come il lettore avrà potuto apprezzare sin dalle prime pagine, il presente lavoro non mira né a demonizzare né a esaltare l’uso della tecnologia nel diritto e in particolare nella giustizia, ma sommessamente a porre in risalto e ragionare su quali possono essere i punti di debolezza, le criticità o proprio le impossibilità di utilizzo degli algoritmi in una materia da sempre regolata dall’uomo, ma anche evidenziare invece i punti di luce, vale a dire cosa l’uomo, sia esso giudice, avvocato o comune cittadino sotto processo possa legittimamente aspettarsi da una giustizia “tecnologizzata”.

Si tratta di cercare di capire se la tecnologia possa davvero contribuire a far sì che “le parti, un giorno, di fronte ad una disputa, potranno sedersi e procedere ad un calcolo”¹³⁷ operato da un software.

L’algoritmo è uno strumento logico-matematico funzionale alla risoluzione di problemi complessi, in maniera univoca e con un numero finito di passaggi¹³⁸; si tratta cioè di uno strumento di automazione¹³⁹. Gli algoritmi, codificati nel linguaggio dei codici sorgente, costituiscono la struttura dei software informatici e forniscono istruzioni predeterminate per elaborare dati (input) trasformandoli in decisioni (output).

¹³⁷ G.W. Von Leibniz, *Dissertatio de arte combinatoria*, 1666.

¹³⁸ M. G. Losano, *Corso di informatica giuridica*, Milano, Unicopli, 1983, Vol. 1, p. 321.

¹³⁹ L. Viola, *Attività amministrativa e intelligenza artificiale*, in *Cyberspazio e diritto*, n. 1-2 /2019, p.78.

Attraverso le informazioni fornite dagli algoritmi, ad esempio, si può impostare a sveglia ogni mattina sul proprio smartphone, si può lavorare al computer e si può attivare il navigatore GPS sulla propria auto; ad oggi è quasi impossibile non utilizzare quotidianamente sistemi attraverso strumenti elettronici di uso comune.

Allo stato dell'arte le capacità di elaborazione di alcuni sistemi informatici possono raggiungere prestazioni elevatissime che, solo poco tempo fa, sarebbero sembrate di esclusiva pertinenza dell'intelligenza umana. Per cui, superate le frontiere dell'informatica tradizionale, si tende a parlare di intelligenza artificiale.¹⁴⁰

Gli algoritmi complessi appartenenti al dominio dell'intelligenza artificiale possono consentire l'elaborazione di dati secondo una logica assimilabile al ragionamento inferenziale umano, che è il presupposto logico del funzionamento di ogni sistema giuridico.

In Italia, il principale modello di giustizia predittiva, proposto da autorevole dottrina, si basa sulla trascrizione in termini algoritmici del procedimento interpretativo previsto dall'art. 12 delle preleggi.

La dottrina precisa che si tratta di un modello che “non utilizza intelligenza artificiale, né può prescindere dall'essere umano e dalla sua creatività”.

Questa conclusione è condivisibile nella misura in cui l'art. 12 delle preleggi delinea solo il procedimento interpretativo della legge e quindi presuppone la selezione delle norme da interpretare.

¹⁴⁰ M. Somalvico, *L'intelligenza artificiale*, Milano, Rusconi Editore, 1987; S. Russel e P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2002; M. Wooldridge, *An Introduction to Multiagents System*, John Wiley & Sons, Chichester, Inghilterra, 2002; N. J. Nilsson, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann, San Mateo, CA, USA, 1998.

La selezione delle norma applicabili, quale operazione preliminare non regolata dall'algoritmo, non può in effetti prescindere dall'intervento creativo umano. Con dei semplici adattamenti, però, questo modello potrebbe evolvere in una forma di intelligenza artificiale operando anche la selezione automatica delle norme applicabili: in questi termini, potrebbe probabilmente identificarsi già con un primo embrione di intelligenza artificiale applicata al diritto.

Algoritmi applicati a processi decisionali complessi sono una novità solo in ambito giuridico mentre sono da tempo oramai una realtà in ambito economico.

E' stato stimato ad esempio, che circa l'80% delle decisioni prese dagli utenti di Netflix e il 30% delle scelte di acquisto operate dai clienti di Amazon sono influenzate da algoritmi che selezionano i contenuti da proporre in base alle precedenti visualizzazioni¹⁴¹. Il ragionamento qui che brevemente vorrei proporre è piuttosto semplice ma sicuramente molti studiosi della tecnologia la pensano in maniera differente. Vorrei in primis far notare la differenza che intercorre tra un algoritmo che mi consiglia cosa comprare sulla base delle mie scelte o delle mie visualizzazione precedenti da un algoritmo che mi giudica sulla base di mie scelte di vita, di condizioni oggettive della mia esistenza che non sono dipese da me, come per esempio il nascere in una famiglia di poco abbienti condizioni, o avere genitori con precedenti penali oppure ancora essere stata io stessa ad aver commesso reati in passato. E' ovvio che nel primo caso, per quanto l'algoritmo nel costringermi a

¹⁴¹ M. Morello, *Netflix, dietro le quinte del gigante dello streaming. Negli studios della tv globale*. Panorama 2018; per un'analisi scientifica si v. H. Wang; *Predicting the Incremental Benefits of Online Information Search for Heterogeneous Consumers*, *Decisions Science*, 2016. Infine, per un utile confront si v. C. Gomez-Urbe, *Un approccio globale per i consigli personalizzati*, Netflix Media Center.

visualizzare ciò che lui desidera influenza le mie scelte però nel fare ciò molto spesso, nel mondo della rete, in un universo cioè di conoscenza sconfinato, per certi versi mi aiuta a trovare più facilmente ciò che mi interessa e senza il quale magari spenderei molto più tempo. Gli algoritmi usati in campo giuridico e giudiziario invece sono suscettibili di ledere ben altri e più importanti diritti della persona verso la quale vengono utilizzati. Quello che qui si vuole sottolineare, come in verità si è cercato di fare durante tutta la stesura della presente tesi dottorale è quello di non demonizzare la tecnologia e soprattutto di porre in rilievo le differenze e i relativi problemi che questa pone a seconda degli ambiti di utilizzo.

Ancora, in campo finanziario gli algoritmi, supportati da potentissime capacità di calcolo, sono alla base della finanza quantitativa. Ad esempio, Bridgewater¹⁴² e Blackrock, rispettivamente uno dei più importanti gestori di Hedge funds e la più grande società di investimenti del mondo, utilizzano da tempo gli algoritmi per migliorare le decisioni di investimento e “sovraperformare” i rendimenti dei mercati finanziari (peraltro neutralizzando efficacemente anche i c.d. “bias cognitivi” umani).

In Italia, nella giustizia in particolare, l’uso degli algoritmi non ha visto una evoluzione a 360 gradi in ogni settore del diritto, ma ci sono stati settori nei quali l’automazione ha visto un avvio più deciso ed altri nei quali ancora oggi siamo molto indietro.

Nel processo amministrativo, per esempio, sono già da tempo utilizzate alcune procedure completamente automatizzate per il controllo della regolarità formale degli atti e documenti prodotti e queste operazioni vengono realizzate senza alcun intervento del magistrato.

¹⁴² R. Dalio, *Principles: Life and Work*, 2017.

In altre parole, le operazioni di acquisizione e registrazione di atti e documenti nel processo amministrativo sono già oggi subordinate all'esito positivo di un controllo affidato ad un software.¹⁴³

Benché non si tratti di una forma di applicazione di algoritmi "interpretativi" in senso stretto è già un primo passo che dimostra la non incompatibilità tra il mondo del diritto e la tecnologia digitale.

Il problema che con la tecnologia ed il diritto spesso si deve affrontare è il linguaggio. Esso è uno strumento di comunicazione che si compone di segni vocali o grafici, a seconda che sia utilizzato nella forma parlata o scritta.

Il linguaggio di programmazione, con cui sono costruiti i codici sorgente e quindi gli algoritmi, ha la stessa identica funzione del linguaggio scritto o parlato: veicolare informazioni. Si tratta semplicemente di una forma ulteriore di linguaggio, capace di essere interpretato e tradotto anche da macchine ovvero da operatori logico-informatici, quali i software. Storicamente il linguaggio informatico probabilmente rappresenta rispetto alla scrittura tradizionale un progresso paragonabile all'invenzione della scrittura rispetto al linguaggio verbale: un cambio di paradigma rivoluzionario.

Non ci sono quindi, almeno astrattamente, ostacoli ontologici per rappresentare ed interpretare i precetti tipici del dominio giuridico, di regola comandi e divieti, attraverso l'uso di questi codici comunicativi funzionalmente assimilabili a quelli tipici del linguaggio tradizionale.

La diffidenza e lo scetticismo verso queste (non tanto) nuove tecnologie deriva forse da una serie di fattori sociali ed emotivi, alimentati

¹⁴³ M. Ancona, *Decisioni automatizzate: qualcosa già esiste*, in *La nuova procedura civile*, n. 4 del 2019.

probabilmente anche dalla scarsa digitalizzazione della giustizia e dei suoi operatori.

Come anticipato, in ambito giuridico, alcuni temi radicalmente innovativi, come l'implementazione di algoritmi e intelligenza artificiale a fini interpretativi, sono sembrati a lungo una sorta di tabù. Il diritto, del resto, è un sistema di regole creato nel tempo e per sedimentazione sulla base dell'intelligenza sociale: si tratta di un prodotto della tradizione e non di una creazione innovativa. Il settore della giustizia di conseguenza non è avamposto della rivoluzione digitale ma per fortuna, quantomeno nel dibattito dottrinale, sembra emergere un crescente interesse per l'innovazione.

Da qualche anno oramai, si dibatte su algoritmi interpretativi e giustizia predittiva, prevalentemente a partire dal settore civilistico e processualcivilistico.

Di recente, queste tematiche sono state affrontate dalla dottrina e anche in relazione al summenzionato diritto amministrativo, tradizionalmente refrattario ad ogni approccio "predittivo" a causa soprattutto della frammentarietà e disorganicità delle fonti normative di riferimento (si pensi solo che fonti del diritto amministrativo oltre a leggi e regolamenti dello Stato ci sono le fonti regionali e poi tutta la produzione normativa europea) ed la conseguente aleatorietà intrinseca nella produzione giurisprudenziale. Questa nuova frontiera del dibattito offre forse nuovi spunti di riflessione sulla potenziale applicabilità dell'intelligenza artificiale in ogni branca del diritto.

La procedimentalizzazione delle operazioni interpretative, secondo la logica sequenziale di un algoritmo, si dimostra infatti ontologicamente compatibile con lo schema inferenziale giuridico, sia nella forma della

deduzione più rigorosa (norma-fatto-effetto) che in quella della deduzione più approssimativa (norma-potere-effetto).

L'inferenza della norma-potere-effetto, tipica espressione del diritto amministrativo, implica l'esistenza di spazi di discrezionalità, in quanto la produzione dell'effetto previsto dalla norma è subordinata al ricorrere di un elemento esterno alla fattispecie (la volontà funzionalizzata dell'Amministrazione), ma il procedimento interpretativo della norma rimane comunque inalterato (ex art. 12 preleggi). La peculiarità è il contenuto della norma, non il suo procedimento interpretativo. La norma, infatti, nel diritto amministrativo, può non descrivere una fattispecie autosufficiente, la cui verifica quindi non basta per attivare l'effetto giuridico previsto, ma delimitare, più o meno elasticamente, un ventaglio di opzioni tra cui l'Amministrazione ha il potere di scelta in funzione della tutela degli interessi pubblici concreti. Ne deriva che un algoritmo interpretativo potrebbe essere efficacemente applicato per ricavare quel ventaglio di opzioni descritte dalla norma, senza nessuna aprioristica preclusione applicativa.

Alla crescente attenzione teorica per le tematiche afferenti alla giustizia predittiva non corrisponde in realtà una di software interpretativi utilizzabili e potenzialmente capaci. In Italia manca una solida sperimentazione sul campo. La realtà probabilmente sarà più complessa di quella ipotizzata nella teoria, quando l'intelligenza artificiale diventerà una realtà operativa: l'esperienza aiuta ad apprendere pregi e difetti delle varie opzioni tecniche disponibili.

La dottrina infatti tende a specificare che “la giustizia predittiva è ancora allo stato di progetto, ossia è in una fase embrionale in cui si

inizia a verificare sul campo la possibilità di tecnologie applicate a ambiti di tipo giuridico, giurisprudenziale o giudiziario”.

Il ritardo sembra fisiologico: occorre coniugare due saperi molto distanti tra di loro, ossia quello degli informatici e quello dei giuristi ed è necessario che i due collaborino in ogni momento. Inoltre, e non ultimo, occorrono mezzi, umani ed economici e quindi la volontà politica di investire su tali progetti. Infine, ed è anche il punto focale troppo spesso sottovalutato, occorre la capacità di “vedere oltre”, ossia di avere in testa l’obiettivo e sapere cosa si vuole fare e dove si vuole arrivare.

Per esempio, sul piano della giustizia amministrativa, un possibile serbatoio di risorse potrebbe essere rappresentato da una società *in house* del Mef: Sogei S.p.A. Si tratta infatti della società fornitrice all’celebre App “Immuni”. E’ una società pubblica che già si occupa in teoria sia di “giustizia digitale” che di “modelli previsionali” sviluppando anche software di intelligence.

L’altra branca del diritto, forse quella che più si presta a sperimentazioni e utilizzo diffuso della giustizia predittiva è sicuramente il diritto civile (inteso nel suo significato generico che ricomprende la legislazione che va dal diritto di famiglia, al diritto condominiale, societario ecc).

In tali materie, infatti, spesso la decisione del giudice così come il lavoro dell’avvocato si basa su “modelli” spesso ripetitivi. Si usa dire tra i giuristi che il “diritto civile è oggettivo mentre quello penale è orientato sul soggetto” (ho posto volutamente la frase in un virgolettato in quanto per nulla tecnica). Con tale affermazione lo studioso delle scienze giuridiche intende dire che nel diritto civile c’è una oggettività nella causa, spesso il soggetto o i soggetti non vengono quasi in rilievo

e rileva il fatto oggetto della causa: immaginiamoci una causa per una lite condominiale sull'utilizzo degli spazi comuni di un immobile; qui per nulla rileva chi sono i due litiganti se non il fatto comune di essere soggetti che vivono in quel dato palazzo. Il giudice pertanto nel dirimere la causa si soffermerà sul dato oggettivo oggetto del litigio. Nel diritto penale, di cui si approfondirà meglio nella righe a seguire, il soggetto che commette un reato o la persona offesa non sono elementi secondari nel giudizio o nella loro difesa ma anzi loro in quanto soggetti rilevano, per esempio per quanto concerne le aggravanti o le attenuanti di cui ad esempio agli artt. 61, co. 5 e 62, co.2 oppure ancora si pensi all'art. 133 che statuisce che “il giudice deve tener conto, altresì, della capacità a delinquere del colpevole desunte: [...] dal carattere del reo, [...] dalla condotta precedente contemporanea al fatto e susseguente, dalla vita antecedente del reo, [...] ed infine dalle condizioni di vita individuale, familiare e sociale del reo.

Gli studiosi della giustizia predittiva sono quindi molto propensi al suo utilizzo nelle materie civilistiche anche per la ovvia ragione che le stesse spesso chiamano in causa diritti di proprietà o di credito ma più difficilmente diritti ritenuti supremi dall'ordinamento giuridico, quali il diritto alla libertà e via discorrendo.

Oltre alla terminologia “giustizia predittiva” ossia l'uso degli algoritmi per risolvere controversie o per coadiuvare il legale nello svolgere il suo compito difensivo, si parla sempre più frequentemente anche di “giurimetria”. Tale espressione sta ad indicare l'interpretazione della legge con modelli matematici, ossia una sorta di commistione tra le scienze giuridiche e le scienze dure quali la matematica. Tra i più esperti del settore sta avanzando anche l'ipotesi dell'implementazione di algoritmi per la “legislazione predittiva”; con tale terminologia si

intende il tentativo, già in fase di scrittura della legge, quindi algoritmi da far lavorare insieme al nostro legislatore per aiutarli a capire quali saranno le interpretazioni più probabili che quel dato testo normativo avrà tra i giuristi o anche tra i consociati una volta approvato e promulgato.

Ma tornando al problema dell'utilizzo della giustizia predittiva nel diritto civile, essa può essere d'ausilio non solo al fine di prevedere un possibile esito della causa sulla base dei precedenti giurisprudenziali e dell'orientamento di quel dato giudice o di quella specifica Corte, ma può svolgere una utile funzione deflattiva e quindi consequenzialmente anche di riduzione dei tempi di giudizio. Ciò in quanto se un avvocato conosce, prima di iniziare una causa giudiziale che l'esito sarà molto probabilmente a lui nefasto, è probabile che si orienti per una risoluzione stragiudiziale della controversia evitando in tal modo non solo di far pendere tempo e denaro al proprio cliente ma anche di ingolfare ulteriormente la macchina giudiziaria.

Nel diritto penale le cose sono molto diverse e più complicate (nel capitolo seguente analizzeremo più in dettaglio i problemi che sorgono), basti qui accennare che il diritto penale consta di prove irripetibili, di cross examination come fulcro del processo, tutti elementi che rendono difficoltoso inserire un algoritmo al loro interno, senza far poi riferimento al problema della discriminazioni che sorgerebbero con un algoritmo che fa riferimento a casi antecedenti simili nel cui ragionamento non vi sarebbe spazio alcuno per un'analisi della condotta dell'agente e delle sue qualità soggettive che, seppur "similare" ad un caso precedentemente analizzato, potranno sicuramente dare un esito giudiziale differente.

Ma nel parlare di giustizia predittiva non possiamo esimerci dal fare riferimento ai due principali “metodi” fino ad ora studiati e sperimentati, il metodo induttivo ed il mio metodo deduttivo.

Il metodo induttivo segue un’impostazione statistico giurisprudenziale, vale a dire che non prende come oggetto della sua analisi o comunque non ne fa il punto fondamentale il dato normativo, ma privilegia invece la casistica giurisprudenziale, vale a dire il precedente del giudice: si hanno precedenti di giurisprudenza ed in base a questi si prevedono le decisioni future.

A titolo esemplificativo: se dieci sentenze su cento precedenti dicono che nel caso x si applica y allora ci sarà il 10% di possibilità che in futuro il giudice a parità di fatto x si orienterà su y.

In questa direzione vanno alcuni progetti dei tribunali italiani, come per esempio le sopra citate Corti di Appello di Bari, Venezia e Brescia), oltre a realtà che vanno consolidandosi in altri Paesi come la Francia in primis; tra i più noti progetti non italiani, vi è, appunto, “Predictice”, società specializzata in *legal tech* che ha messo a disposizione uno strumento di carattere informatico che avvalendosi di algoritmi di calcolo permette di prevedere la probabilità di orientamento decisionale del giudice. La base sulla quale l’algoritmo viene applicato è costituita dalle decisioni delle Corti di Appello e dalle decisioni della Cassazione. Per il progetto italiano, bisognerebbe utilizzare ampiamente le banche dati condivise della giurisprudenza per avere trasparenza delle decisioni, conoscenza da parte del singolo giudice del livello di resistenza e accoglimento dei propri provvedimenti, riscontro da parte del presidente di sezione e del dirigente dell’ufficio e più in generale degli “operatori giudiziari” autori delle diverse tesi giurisprudenziali dell’ufficio e la probabilità di accoglimento della domanda avanzata in

una certa materia ovvero la possibilità che un reato venga ravvisato e la gamma di pene irrogate.

Ciò, a ben vedere e come anticipato in termini generici poco sopra, verrebbe a essere prezioso anche per contenere le domande giudiziali e avere complessivamente un effetto preventivo e deterrente oltre che creare un rapporto di trasparenza e di collaborazione attiva tra le parti in causa.

Non si tratta, come dicono i detrattori tecnologici di controllare la giurisprudenza, ma di acquisire consapevolezza della stessa e di aiutare in modo propositivo tutti, in primis i cittadino. Del resto l'art. 47 quater dell'ordinamento giudiziario non esalta il conformismo delle decisioni, ma tende ad evitare le divergenze inconsapevoli, cercando di superare l'inevitabile difformità che esiste nelle diverse giurisprudenze attraverso il confronto e l'acquisizione del valore della prevedibilità delle decisioni.

Gli studiosi della giustizia predittiva, in primis l'Avv. Luigi Viola, uno dei massimo esperti italiani in materia sostengono che la previsione su base statistico-giurisprudenziale, vale a dire il modello induttivo, non sia da privilegiare in quanto l'impostazione basta su meri calcoli statistici dei precedenti giurisprudenziali ha una portata limitata ai soli casi in cui ci siano numerosi precedenti, così da escludersi i casi più complessi relativi alle novità normative, non ancora oggetto di stratificati orientamenti giurisprudenziali; non è in linea con il nostro sistema di *civil law* che, per quanto sia sempre più orientato al porre attenzione ai precedenti giudiziari (si parla in tal senso di una sempre crescente "commonizzazione del *civil law*) nel contempo non è fatto obbligo al giudice di conformarsi all'orientamento prevalente e quindi ai precedenti dai quali invece può legittimamente discostarsi; vi è una

sorta di “stagnazione giurisprudenziale”, la quale non permette una corretta evoluzione del diritto e per di più, nel caso di errore in una sentenza, lo perpetra rendendolo così precedente e quindi lo applica anche in futuri; infine, ma sono solo alcuni, forse i più evidenti problemi che si pongono, vi è il problema cardine della predizione, ossia il fatto che essa, se correttamente intesa, non si basa sulla quantità delle sentenze emesse che vanno in una certa direzione ma si basa sulla correttezza de jure dell’atto giuridico, pertanto la Suprema Corte di Cassazione andrà appunto a dare rilievo e si conformerà alle decisioni della Corti di Appello che hanno deciso la causa con il criterio della correttezza degli argomenti di diritto e non sulla quantità di sentenze emesse dalle corti inferiori in un senso o nell’altro.

Nei sistemi di *civil law*, poi, il metodo interpretativo è deduttivo, diversamente da quelli di *common law* dove è prevalentemente induttivo.

L’Avv. Viola ci propone un esempio basato sul c.d. sillogismo aristotelico, per meglio comprenderne la ratio:

- a) Tutti gli uomini sono mortali (premessa maggiore)
- b) Socrate è un uomo (premessa minore)
- c) Socrate è mortale (conclusione).

La conclusione appena esposta si basa sul metodo deduttivo; è priva di vizi logici ed è in linea con il sistema italiano di *civil law* che consta di regola – fatto – effetto, cioè sentenza.

Ora, restando sullo stesso esempio invertiamo l’ordine:

- a) Socrate è mortale (conclusione)
- b) Socrate è un uomo (premessa minore)
- c) Tutti gli uomini sono mortali (premessa maggiore).

Quanto appena esposto si basa sul metodo induttivo; presenta almeno un vizio logico perché generalizza (premessa maggiore) partendo da una conclusione ed, infatti, che Socrate sia mortale e sia un uomo non implica necessariamente che tutti gli uomini siano mortali; tuttavia, è in linea con il sistema del common law: sentenza – fatto – regola.

Un altro esempio:

- a) La legge è uguale per tutti i cittadini ex art. 3 Cost. (premessa maggiore)
- b) Tizio e Caia sono cittadini (premessa minore)
- c) Tizio e Caia sono uguali di fronte alla legge (conclusione).

Anche in questo caso si è utilizzato il metodo deduttivo, che è privo di vizi logici: la conclusione è la diretta conseguenza della premesse.

Ora, di nuovo, cambiamo ordine, così come leggerebbe il metodo induttivo:

- a) Tizio e Caia sono uguali di fronte alla legge (conclusione)
- b) Tizio e Caia sono cittadini (premessa minore)
- c) La legge è uguale per tutti i cittadini (premessa maggiore).

Quando appena scritto è viziato perché generalizza singoli casi: che Tizio e Caia siano uguali di fronte alla legge non implica necessariamente che tutti siano uguali di fronte alla legge.

Come anticipato, si ritiene pertanto preferibile in Italia optare per il modello deduttivo, ossia per un'impostazione algoritmico-normativa tramite combinazione di dati,

Ciò che davvero rileva ai fini di un'interpretazione, non è il numero di precedenti giurisprudenziali ma il corretto utilizzo dell'art. 12 della preleggi, che è l'unica disposizione che si occupa *expressis verbis* di interpretazione della legge.

Dal suo dato testuale emerge che, per interpretare una legge, bisogna procedere nel modo che segue:

1. Attribuire il significato letterale, che è la c.d. interpretazione letterale;
2. Comporre l'interpretazione letterale con l'intenzione del legislatore, che è la c.d. interpretazione per *ratio*;
3. In assenza di precisa disposizione, nel senso di assenza di norma oppure di presenza di norma con significato equivoco, si può utilizzare la c.d. analogia *legis*, ovvero cercare nel medesimo complesso di leggi una situazione giuridica tipizzata analoga, con la stessa *ratio*;
4. Se il caso è ancora dubbio si farà riferimento ai principi generali dell'ordinamento (analogia *iuris*)

L'algoritmo è una sequenza predeterminata di operazioni per giungere ad un risultato ed è:

- Sequenziale, ossia contiene l'indicazione di una serie predeterminata di operazioni;
- Può essere condizionale, ovvero può contenere condizioni (se...allora);
- Può essere iterativo, ovvero si possono ripetere più volte le stesse istruzioni a seconda del verificarsi o del non verificarsi di una condizione.

L'art. 12 Preleggi è un algoritmo perché è una sequenza di operazioni per giungere ad un risultato.

Esso infatti è:

- Sequenziale, in quanto prevede un iter procedimentale che inizia con l'interpretazione letterale, teleologica o per *ratio*, per poi procedere a quella analogica e per principi generali.

- Condizionale, prevedendo alcune condizioni come l'assenza di precisa disposizione per legittimare l'interpretazione per analogia *legis*, nonché il caso dubbio per legittimare l'analogia *iuris*
- Iterativo, in quanto prevede un ciclo, causato da condizioni; se si attiva l'analogia *legis*, allora la disposizione analogica dovrà essere decodificata, ricominciando dall'interpretazione letterale.

L'impostazione che qui si suggerisce di utilizzare ha conseguenza sul piano pratico.

Ad esempio, nell'ipotesi di responsabilità dell'avvocato a cui si è fatto cenno, non basterà scrutinare solo la giurisprudenza precedente e verificare che probabilità di successo avrebbe avuto l'avvocato laddove avesse attivato correttamente l'azione, ma si dovrà vedere caso per caso quali argomenti avrebbe speso ed eventualmente anche andare esente da responsabilità per un'azione ritenuta temeraria *ex post*, quando gli argomenti sembravano *ex ante* utili.

In conclusione possiamo dire che il modello deduttivo, basato su un approccio scientifico al diritto, assicura razionalità al procedimento interpretativo ed aumenta la certezza del diritto e quindi in un certo senso anche la sua prevedibilità, riuscendo comunque a cogliere gli aspetti nuovi, frutto di una evoluzione giuridica senza dunque creare un diritto "dato" in un certo momento ed imm modificabile.

3.4 I profili ontologici della decisione robotica tra (im)prevedibilità e certezza del diritto

Un computer potrà davvero giudicare?

Rispondere a questa domanda non è semplice e ciò per una molteplicità di ragioni. In primis perché dobbiamo concordare preventivamente che cosa significa giudicare e, seconda questione preliminare capire se intendiamo il computer come “aiutante”, come mezzo che coadiuva l’essere umano giudice oppure se lo soggettivizziamo e lasciamo libero di “decidere”.

Riguardo al secondo dei quesiti preliminari, ad oggi, si direbbe altamente improbabile che gli strumenti oggi a disposizione, di natura per lo più statistica, possano essere in grado di adempiere adeguatamente a tale compito.

Tuttavia, come detto sopra, la ricerca si sta dichiaratamente spingendo in tale direzione, per lo meno sul piano delle intenzioni, e anche se queste non dovessero essere realizzate e i sistemi di giustizia predittiva non arrivassero a sostituire integralmente il giudice, è verosimile che possano comunque esercitare una significativa influenza sull’attività decisionale. Questo timore è stato espresso già nel 2016 dalla Commissione Europea per l’Efficienza della Giustizia (CEPEJ) del Consiglio d’Europa, che nella sue “Linee guida sulla cybergiustizia” ha evidenziato il pericolo che l’uso di software di analisi di precedenti giurisprudenziali da parte dei giudici possa minare la capacità discrezionale e soprattutto l’indipendenza.

E’ facile immaginare che per chi decide possa diventare difficile discostarsi dal risultato di analisi elaborata da sofisticati software in grado di processare migliaia di informazioni. Inoltre, il fatto che tali

strumenti possano non essere adeguati a decidere non significa che la sostituzione del giudice con uno strumento di *machine learning* non possa comunque avvenire, in un futuro.

Si rende dunque opportuna, a livello legislativo e governativo (perché a livello accademico già c'è da tempo) qualche riflessione su dove vogliamo settare l'asticella nell'uso della tecnologia nella giustizia, di quanti investimenti in tal senso si vogliono fare in termini economici e non solo e di quale sia la percezione del cittadino di questo particolare uso della tecnologia.

Dopo la crisi del formalismo giuridico, la prospettiva offerta dal ricorso all'intelligenza artificiale potrebbe essere quella di un ritorno all'antico: l'aspirazione alla decisione come "atto meccanico", scevro da componenti soggettive e irrazionali, realisticamente superata già dal secolo scorso, potrebbe tornare attuale grazie all'impiego degli algoritmi.

Il mito illuministico del giudice *bouche de la loi* potrebbe risolvere la diffidenza nei confronti dell'arbitrio del giudice che ben prima della nascita dell'intelligenza artificiale e della sua introduzione in ambito giuridico si immaginava di trasformare metaforicamente in un "automa".

Già Max Weber, tra gli altri, sosteneva che una delle condizioni necessarie allo sviluppo del moderno capitalismo fosse l'esistenza di un diritto "calcolabile in modo simile ad una macchina".

Egli credeva infatti che l'imprenditore dovesse essere posto nella condizione di prevedere, oltre ai prezzi delle merci, le spese ed i futuri guadagni, anche gli esiti delle decisioni giudiziali, e che ciò fosse possibile in presenza di un diritto "calcolabile", ossia razionale e "tecnicizzato".

Nella sfera del capitalismo ciò è auspicabile ed in parte possibile oggi, tuttavia l'aspirazione ad una decisione consistente in una meccanica applicazione di regole può essere in realtà considerata un "lascito della modernità", un prodotto dell'aspirazione illuministica ad un giudizio razionale, neutro, oggettivo e scevro da valutazioni personali etiche, politiche, religiose o filosofiche.

L'idea del giudice-automa nacque infatti con la concezione formale del diritto sviluppata da Montesquieu, il quale immaginava l'organo votato all'applicazione delle leggi come un essere inanimato, una entità spersonalizzata, con il compito di mero esecutore delle disposizioni redatte dall'assemblea legislativa eletta dal popolo.

Secondo il modello meccanicistico concepito dal filosofo e affermato dalla Rivoluzione francese, la decisione sarebbe dovuta consistere in un'operazione logica di sussunzione della fattispecie concreta a quella astratta prevista dalla norma, in un'applicazione "automatica" di semplici sillogismi deduttivi. La logica giuridica sarebbe quindi coincisa con la logica formale e il giudice, senza dispiegare discrezionalità alcuna e alcun potere, e prendendo le distanze dalle proprie opinioni, dalle proprie convinzioni etiche e dalle proprie emozioni, si sarebbe comportato proprio in modo simile ad una macchina.

Come scrisse Norberto Bobbio, però, "dall'epoca del cosiddetto feticismo legislativo molta acqua è passata sotto i ponti, e nessuno crede più seriamente al giudice come automa": ormai è universalmente accettato che l'attività della giurisprudenza è sempre inevitabilmente creativa, che l'ideale meccanicistico del giudice come "bocca della legge" è irrealizzabile.

C'è sempre, nel decidere, una quota insopprimibile di soggettività, almeno nel momento dell'interpretazione giuridica (essa infatti sicuramente non dovrebbe esserci nella fase di ricostruzione del fatto) e ciò può non essere necessariamente un male dal momento che la disposizione normativa è stata scritta per essere interpretata e la sua interpretazione può cambiare ed è giusto che cambi a seconda del momento storico in cui si applica, della sensibilità diffusa sul tema ecc; .l'interpretazione diventa un problema quando la stessa non rispecchia il "sentire dei consociati" quanto cioè si discosta troppo da quale interpretazione che non sono hanno dato altri giudice di quella medesima norma nello stesso contesto storico e socio-culturale ma anche da quella che darebbe un cittadino comune in quella circostanza.

Da un punto di vista pratico infatti l'interpretazione presuppone l'attribuzione di significato alle norme, e ciò non può prescindere, almeno in parte, da elementi extra-giuridici: per quanto chiare e precise possano essere le disposizioni, l'interprete dovrà sempre attribuire un senso ai vocaboli, risolvere antinomie o sopperire a inevitabili lacune, decidendo tra diverse opzioni interpretative. Dall'altro, anche la stessa prova di un fatto è sempre la "conclusione più o meno probabile di un procedimento induttivo, la cui accettazione è un atto che esprime un potere di scelta ad ipotesi esplicative alternative".

L'idea di un sillogismo giudiziario perfetto, dunque, è un'illusione. Si tratta di una consapevolezza maturata sin dagli ultimi anni dell'Ottocento a partire dal pragmatismo di Dewey e la *sociological jurisprudence* di Holmes, Pound e Cardozo, da cui si svilupparono in seguito le principali tesi del realismo giuridico americano che diedero

origine a quella che Morton White definì la “rivolta contro il formalismo”.

In generale, gli elementi che accomunano le teorie antiformalistiche erano la credenza dell'importanza della *law in action* (costituita dai comportamenti effettivi delle corti) rispetto alla *law in books*, l'affermazione della natura creativa dell'interpretazione e la concezione “predittiva” della giurisprudenza, secondo la quale il vero diritto non consisterebbe nell'insieme di norma vigenti, ma nella previsioni di come i giudici, i veri artefici del diritto, si comporterebbero.

Da queste premesse si svilupparono però posizioni molto eterogenee. Quelle senz'altro più radicali furono adottate da Jerome Frank, il quale, fortemente influenzato dalle teorie psicanalitiche di Freud, esasperò l'importanza dei fattori psicologici e soggettivi nella decisione del giudice, affermandone l'assoluta irrazionalità e imprevedibilità.

Accanto allo scetticismo delle norme tipico di molti realisti egli introdusse nella propria filosofia anche un radicale scetticismo dei fatti, secondo il quale l'accertamento giudiziale sarebbe totalmente dipendente dalla soggettività di chi partecipa al processo: testimoni, imputati, parti, periti e giudici.

Come per ogni essere umano, la loro visione è infatti influenzata da una grande quantità di fattori, come il carattere, le preferenze personali, le simpatie e i ricordi, i preconcetti e le esperienze passate, che viziano irrimediabilmente la loro percezione della realtà esterna e la ricostruzione dei fatti. Secondo Jerome Frank, dunque, la decisione sarebbe assolutamente imprevedibile, la verità processuale una pura finzione e la certezza del diritto un mito.

I più accreditati studi di psicologia cognitiva sembrano confermare le tesi del giurista americano. La ricerca ha infatti dimostrato che le euristiche (altrimenti dette scorciatoie del pensiero) naturalmente operate dalla mente umana, indispensabili per apprendere e “organizzare” rapidamente il mondo circostante, possono facilmente determinare distorsioni cognitive ed errori.

Secondo la teoria delle “due menti”, il pensiero sarebbe elaborato da due diversi “cervelli” o sistemi: il sistema 1, caratterizzato da processi mentali automatici e intuitivi, e il sistema 2, che opera invece attraverso meccanismi cognitivi “lenti”, coscienti e sottoposti al controllo razionale.

Il primo, il più antico e primitivo, funziona in via automatica e senza sforzo; il secondo, invece, più recente dal punto di vista evolutivo, elabora le operazioni più complesse che richiedono attenzione e coscienza.

Le due “menti” sono strettamente interconnesse, ma sembra che di fronte ad un problema sia il sistema 1 ad attivarsi per primo, formulando una rapida e intuitiva risposta, e solo in un secondo momento il sistema 2, che sottopone la soluzione proposta dal sistema 1 al vaglio della razionalità per convalidarla, correggerla o scartarla.

Questa teoria, che afferma, in sostanza, che ogni scelta umana è effettuata in maniera inconscia e in seguito rivestita di fondamento razionale, sembrerebbe proprio avvalorare la tesi secondo cui la decisione del giudice è, in fondo, intuitiva.

Inoltre, è noto che il sistema 1 non produce sempre intuizioni corrette e che il sistema 2 spesso non sia in grado di riconoscere gli errori da esso effettuati.

Le intuizioni del sistema 1 sono infatti estremamente “soggettive”, poiché dipendono direttamente da esperienze passate, emozioni, umore, stati d’animo e situazioni di ansia e stress e quindi grande peso hanno sul decidere le generalizzazioni, gli stereotipi e i pregiudizi che derivano dai meccanismi euristici di ragionamento.

Esistono poi importanti meccanismi psicologici inconsci che orientano verso la conferma delle prime impressioni, come ad esempio l’ancoraggio o il “bias della visione a tunnel”.

Di fronte ad un tale scenario, perché allora non accogliere di buon grado l’introduzione di strumenti di intelligenza artificiale nei sistemi giudiziari?

Se la decisione del giudice, come ogni decisione umana, è un prodotto di processi irrazionali, e addirittura può dipendere, come pare sia stato persino dimostrato, da ciò che ha mangiato un giudice a colazione (in realtà, è bene specificare che tale esperimento aveva un campione molto ristretto di giuridici, una ventina, un numero troppo esiguo per farne un campione significativo), non parrebbe forse più sensato affidarla ad una macchina, non influenzata da fattori emotivi, psicologici o chimici?

Che non ci si possa rassegnare al totale relativismo della funzione giudiziaria, del resto, parrebbe una conclusione implicita anche in alcune considerazioni dello stesso Jerome Franc, il quale avrebbe cercato di formulare alcuni rimedi volti a mitigare l’assoluta irrazionalità della decisione, come, per esempio, la formazione di una magistratura consapevole dei propri meccanismi inconsci, e così in grado (in teoria) di metterli a freno.

Ma anziché affidare la giustizia al *self restraint* dei giudici, fidandosi, paradossalmente, delle loro intuizioni e della loro valutazione

equitativa, rinunciando così definitivamente al “mito” della certezza del diritto, perché non aspirare invece a eliminare del tutto la componente soggettiva e irrazionale della decisione?

Non si può assolutamente rinunciare ad un diritto certo e ad una decisione, per quanto possibile, prevedibile e oggettiva: come scrisse il già citato filosofo torinese Norberto Bobbio proprio in critica alle tesi di Jerome Frank, la certezza del diritto è “un elemento intrinseco al diritto, e a togliere di mezzo la certezza non si libera l’umanità da una nociva illusione, ma la si priva letteralmente del sussidio e del rimedio del diritto”.

La volontà di creare un robot-giudice parrebbe proprio il frutto della grande sfiducia nei confronti della decisione umana, della cui discrezionalità, del resto, anche gli esponenti del formalismo giuridico, che speravano in un giudice *bouche de la loi* totalmente spersonalizzato, diffidavano.

Se però ci si è finalmente rassegnati al fatto che l’ideale del giudice-automa sia irrealizzabile, c’è chi non ha abbandonato l’antica ambizione ad un diritto dalla meccanica applicazione, e proprio per questo, si direbbe, alla luce dei grandi progressi della robotica e dell’intelligenza artificiale si è passati dall’aspirazione a un giudice-macchina, dall’impossibile concretizzazione, a quella di una macchina-giudice.

Dunque, alla luce delle grandi novità che stanno investendo il mondo giuridico, si potrebbe affermare che un robot-giudice possa finalmente concretizzare l’ideale meccanicistico dell’interpretazione della legge? Che un computer possa divenire, al posto del giudice, una *bouche de la loi*?

In realtà tutto ciò non sembra del tutto esatto.

E' importante infatti ricordare che gli odierni sistemi di *machine learning* operano tramite meccanismi totalmente diversi rispetto a quelli logico-formali che dovrebbero informare la decisione giudiziaria secondo il modello formalistico.

Se l'ideale meccanicistico si potesse realizzare tramite un software, lo strumento adatto sarebbe probabilmente un sistema esperto, ovvero uno strumento funzionante su base logica in grado, in teoria, di decidere riconducendo le fattispecie concrete a quelle astratte attraverso l'applicazione di sillogismi, e non un sistema di giustizia predittiva.

Infatti, strumenti come questi non operano applicando deduttivamente regole predeterminate, bensì elaborando grandissime quantità di dati costituiti da decisioni giudiziali rese in passato, individuando le connessioni che legano alcuni elementi contenuti nelle sentenze e il dispositivo, e proponendo infine le stesse soluzioni per i nuovi casi che presentano le stesse caratteristiche.

Questi software, dunque, non deciderebbero applicando la legge bensì le vecchie decisioni giudiziali.

Si creerebbe così uno scenario opposto a quello configurato da Montesquieu: l'automa-giudice non sarebbe una "bocca della legge", bensì una "bocca del precedente giurisprudenziale".

E' chiaro che anche quando Weber parlava di un diritto "calcolabile" non immaginava una situazione simile.

Egli, anzi, distingueva tra diritto continentale, o razionale-formale e quello inglese, di natura "empirica" basati sul fatto, che elaborava concetti orientati in base a fattispecie concrete, familiari all'esperienza quotidiana, ma che non conosce concetti generali elaborati mediante astrazione del concetto, mediante un'interpretazione logica del senso,

mediante generalizzazione e sussunzione, e quindi applicati sillogisticamente come norme.

Secondo il filosofo e sociologo, il capitalismo inglese non si sarebbe sviluppato grazie a , ma “nonostante” la struttura del suo diritto, assimilabile a un sistema razionale-materiale.

Il concetto di un diritto “calcolabile” corrisponde però sempre ad un modello razionale-formale, composto da norme generali ed astratte e applicate con gli strumenti della logica giuridica, e non a quello razionale-materiale fondato sul precedente.

E’ opportuno aggiungere che, comunque, ciò che deriverebbe dall’impiego di software di *machine learning* non è certo la realizzazione di un diritto razionale-formale, ma neppure, ovviamente, di un sistema razionale-materiale simile a quello inglese, se non per il semplice fatto che si utilizzerebbero i precedenti giurisprudenziali per giungere a una decisione. Come si vedrà, infatti, il loro impiego darebbe vita a un sistema “fondato sul precedente” ma caratterizzato, al contrario, da rigidità e incapacità di evoluzione.

Ma la domanda forse più interessante che dobbiamo porci è: come mai un algoritmo giudicante?

Il fatto che il modello applicato dagli strumenti di giustizia predittiva non corrisponda con quello formalistico non significa, di per sé, che l’introduzione di tali strumenti ai fini della decisione sia impossibile o inopportuna. Del resto, è comunque ormai universalmente accettato che neanche un giudice umano sia in grado di decidere nella maniera delineata dal’ideale meccanicistico, e non si può escludere a priori che una giusta decisione possa essere assunta con meccanismi diversi da quelli logici previsti dal modello razionale-formale.

E' dunque utile gettare uno sguardo più approfondito sui possibili benefici che la realizzazione di un robot-giudice di giustizia predittiva potrebbe apportare e sugli obiettivi che si prefiggono i sostenitori di tali progetti, riflettendo sugli aspetti che potrebbero rendere la decisione attuata da un sistema di *machine learning* migliore rispetto a quella umana, nel senso di più di “più conforme al sistema delle norme vigenti” e più giusta.

Innanzitutto, è impossibile non notare che, nonostante l'irrinunciabile principio di certezza giuridica stabilisca che ogni persona deve essere posta in condizione di valutare e prevedere le conseguenze giuridiche della propria condotta, si assiste oggi alla crisi della fattispecie e alla sempre maggiore creatività del giudice chiamato ad applicare la legge; che gli orientamenti giurisprudenziali sono spesso disomogenei, e che l'organo deputato alla nomofilachia incontra serie difficoltà nel perseguire la funzione che gli compete ex art. 65 ord. Pen.

Si potrebbe allora guardare l'intelligenza artificiale come allo strumento in grado di garantire una maggiore prevedibilità delle decisioni giudiziarie, rendendo così il diritto certo e, in un certo senso, “calcolabile”.

Poiché i software di giustizia predittiva operano elaborando una grande quantità di precedenti giurisprudenziali, una decisione assunta sulla base del risultato fornito da tali strumenti sarebbe infatti automaticamente conforme alle decisioni assunte in passato in casi simili: di conseguenza, illogici *revirements* giurisprudenziali sarebbero evitati, e sarebbero garantite stabilità e omogeneità interpretativa. Inoltre, ciò potrebbe rilevare non solo ai fini della maggiore prevedibilità della decisione, ma anche nell'ottica del rispetto del principio di eguaglianza.

Alcuni parlano in proposito di un principio di equità “orizzontale” o “fattuale”, secondo il quale circostanze fattuali simili dovrebbero ricevere lo stesso trattamento di fronte al giudice.

Un'altra importante ragione che spinge all'introduzione di strumenti di AI nel processo è l'obiettivo di risolvere il delicato problema dell'ingolfamento e lentezza della macchina giudiziaria: grazie infatti alla grande capacità di calcolo e alla potente capacità di elaborazione, i sistemi di *machine learning* possono fornire un sicuro apporto positivo in ambiti in cui è ingente il carico di lavoro, tra i quali rientra sicuramente quello della giustizia.

E' noto che la situazione di sovraccarico di lavoro negli uffici giudiziari è critica in molti Stati membri del Consiglio d'Europa, soprattutto in Italia, e l'accumulo di numerosi processi pendenti presenta conseguenze molto gravi, in particolare in relazione alla loro durata eccessiva.

L'esigenza di rapidità non è certamente da sottovalutare, e trovare una soluzione al problema del sovraccarico di lavoro nei tribunali e ai tempi eccessivamente lunghi della giustizia è urgente.

Il principio della ragionevole durata del processo, rientrante tra i principi del giusto processo, è sancito dall'11 della Costituzione, ma anche dalla Carta dei diritti dell'Unione Europea, all'art. 47 e dalla Convenzione Europea dei diritti dell'uomo e delle libertà fondamentali, che all'art. 6 recita: “Ogni persona ha diritto a che la sua causa sia esaminata equamente, pubblicamente ed entro un termine ragionevole da un tribunale indipendente e imparziale, costituito per legge”, rendendo il principio della ragionevole durata un diritto soggettivo direttamente azionabile.

E' plausibile che l'elaborazione automatizzata della decisione possa consentire una maggiore velocità nella definizione dei giudizi e lo smaltimento degli arretrati in tempo breve, risolvendo il grave problema dell'eccessiva durata dei procedimenti.

Tuttavia, è evidente che quello della ragionevole durata non è l'unico criterio di cui tenere conto in materia di giustizia, e che anche altri aspetti siano da considerare, tra cui il rispetto degli altri principi del giusto processo. Se una giustizia lenta può essere "ingiusta", ovviamente una giustizia rapida non è automaticamente "giusta": anche se consisterebbe in un metodo molto veloce, non si potrebbe certamente lanciare una moneta per decidere che debba risultare vincitore di una causa; tale ovvietà per sottolineare come "la maggior rapidità della decisione robotica potrebbe importare solo a condizione che la prestazione del giudice-robot e quella umana si equivalessero in termini qualitativi e soddisfacessero quel principio della tutela giurisdizionale effettiva che la giurisprudenza ha da tempo valorizzato.

Allo stato, c'è chi sostiene, ed io concordo con tale assunto, che una decisione algoritmica non soddisfi tali condizioni.

SE si mette da parte il forte carico di suggestione che l'idea di un robot-giudice porta inevitabilmente con sé e si guarda più concretamente al funzionamento dello strumento a disposizione, si comprende che affidare a un sistema di *machine learning* il compito di decidere non è sensato.

E' vero che la produzione automatizzata delle sentenza consentirebbe di ridurre i tempi dei processi, risolvendo il grave problema degli arretrati nella giustizia, inoltre sottrarrebbe alle valutazioni discrezionali dei giudici le loro decisioni rendendole quindi meno

variabili e disomogenee bensì più coerenti e prevedibili; tuttavia, è pacifico che solo su tale base non è possibile affidare la decisione giudiziaria a strumenti che sono sostanzialmente inadatti a tale scopo. Il problema principale non è il fatto, pur rilevante, che affidare la decisione a uno strumento che elabora quantità di decisioni di merito comporterebbe il passaggio a un sistema fondato sul precedente giurisprudenziale, anche se ciò determinerebbe sicuramente uno stravolgimento del sistema normativo in un ordinamento di *civil law* come quello italiano, in cui l'art. 101 dell' Costituzione stabilisce la soggezione del giudice soltanto alla legge.

Il fattore più critico non sarebbe, infatti, tanto l'uso del precedente quanto la maniera in cui è elaborato dai sistemi di giustizia predittiva. Se non fossero applicati dei correttivi, il precedente "sostituirebbe" ogni altra fonte normativa: anche se fossero promulgate nuove leggi, il software non potrebbe tenerne conto, poiché continuerebbe a processare le vecchie sentenze contenute nei suoi database, con ciò minando anche la semplice evoluzione giurisprudenziale che auspica, per la ovvia ragione del soddisfacimento del principio della certezza del diritto, che i giudici si conformino sì, ai precedenti ma non siano affatto costretti a farlo, possono infatti, se puntualmente motivato, discostarsene e con ragione.

Dunque, se ci si conformasse al risultato fornito dalla macchina, sempre fondato sulle stesse decisioni, nel deriverebbe la totale rigidità del sistema e la cristallizzazione del diritto.

Ciò avverrebbe anche e soprattutto perché un sistema di intelligenza artificiale non sarebbe in grado di elaborare un'interpretazione evolutiva dei precedenti.

Peraltro, anche quando un giudice “umano” è chiamato a decidere sulla base dei precedenti come avviene normalmente nei sistemi di *common law*, tra le altre cose, esaminare la situazione concreta, valutare gli elementi rilevanti del caso e individuare il precedente conforme tra le varie decisioni dei giudici superiore, astraendone la *ratio*.

Si tratta di operazioni che presuppongono un ‘intelligenza “forte” (nel senso del termine di cui si è fatto ampiamente riferimento nei primi capitoli di questa tesi), di cui gli odierni sistemi di AI non posseggono; non sono cioè in grado di comprendere le particolarità della situazione concreta e i motivi che hanno indotto i giudici a decidere in una maniera piuttosto che in un’altra.

Infine, ed è un punto normativo fondamentale, non solo l’apporto umano nelle decisioni giudiziali è ancora auspicato e fortemente richiesto per le motivazioni di cui sopra, ma è anche oggetto di obbligo di legge. Il più volte menzionato GDPR infatti, all’art 22 afferma: “L’interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona”. Sulla base di tale norma, almeno per quanto riguarda gli Stati appartenenti all’Unione Europea sembra potersi totalmente escludere la possibilità di un processo svolto interamente da un algoritmo.

Allo stato della tecnica, tuttavia, in Europa, ma lo stesso seppur con qualche lieve differenza si può dire anche per gli Stati Uniti, tutti i software di giustizia predittiva sono utilizzati come strumenti che coadiuvano il giudice in vari momenti del processo fino alla decisione finale ma mai in sua completa sostituzione; in altre parole non sono

oggi in uso (se si esclude l'uso a scopo di ricerca) algoritmi che sostituiscono interamente l'attività umana, anche e soprattutto perché lo stato della tecnica non è ancora maturo.

Tra i più diffusi programmi predittivi di “assistenza legale” possono essere citati, per esempio, il *Watson/Ross* dell'IBM, in grado di interagire con gli operatori giuridici, svolgere ricerche tra migliaia di documenti legali e monitorare mutamenti giurisprudenziali; il *Legal Analytics* di *Lex Machina*, disegnato per orientare l'azione e le strategie dei difensori effettuando previsioni sulle decisioni dei giudici e il famoso software francese, già nominato nel presente lavoro, *Predictice*, grazie al quale avvocati, consulenti e compagnie assicurative sono in grado, tra le altre cose, di conoscere le *chances* di successo di una causa.

Alcuni esponenti della dottrina ritengono peraltro che l'impiego del termine “giustizia predittiva” sia fuorviante. Dato che gli algoritmi di *machine learning* costruiscono modelli elaborando dati storici, sarebbe più corretto parlare di “previsione”, ossia di un'enunciazione compiuta in seguito ad un'osservazione e no di “predizione”, termine che indica l'atto di preannunciare il futuro su una base che non sarebbe, a rigore, statistica, bensì quasi divina.

3.5 La sentenza come sillogismo matematico: un giudice senza empatia

In seguito all'esperimento di Danziger¹⁴⁴ una banca d'investimento americana, la Bank of America Merrill Lynch, ha sottolineato come lo sviluppo e l'incremento dell'Intelligenza Artificiale ha tra i vantaggi quello di sopperire a mancanze e a contribuire all'eliminazione dei pregiudizi.

“Uno dei principali vantaggi dell'Intelligenza Artificiale è l'eliminazione dei pregiudizi umani. Il ricercatore Shai Danziger ha mostrato che la percentuale di sentenze favorevoli dei giudici israeliani salta al 65% dopo la pausa pranzo, e diminuisce gradualmente fino a 0 mano a mano che si avvicina la pausa successiva. Numerose attività umane, tra cui la medicina diagnostica e gli accertamenti antifrode per citare due esempi lontani tra di loro presuppongono un processo decisionale imparziali.

E' quasi lapalissiano l'affermare che coloro che svolgono queste attività, dove rientra anche la giustizia, dovrebbero essere persone quanto più distaccate dalle pressioni dell'opinione pubblica, da giudizi

¹⁴⁴ Si tratta del noto esperimento descritto in precedenza in cui si è cercato e trovato una sorta di correlazione tra ciò che i giudici hanno mangiato a colazione, o meglio tra una decisione presa verso ora di pranzo, momento in cui il giudice iniziava ad avvertire una certa fame ed una decisione invece data in momenti post-prandiali. L'esperimento ha, come noto, messo in luce che le sentenze emesse a seguito della consumazione di pasti da parte dei giudici hanno dato un esito proporzionalmente più favorevole all'imputato di quelle emesse in momenti pre-banchetto. E noto anche, però, che tale esperimento fu svolto su un campione di giudici eccessivamente basso, e come tale non può essere eretto a dato oggettivo, ma preso come esempio del fatto che i giudici sono esseri umani e come tali soggetti ad una serie di variabili umorali anche temporali e banali.

soggettivi e magari anche instancabili per velocizzare la macchina della giustizia.

Questa visione riprende il pensiero di Cesare Beccaria per il quale la tutela della dignità dell'imputato poteva essere assicurata solo da un giudice che avesse analizzato la causa a lui sottoposta con freddezza e asetticità idealmente assimilabili a quelle di una macchina.

A tal proposito bisogna ricordare come uno dei punti dal Libro Bianco dell'Agenda per l'Italia Digitale (AGID) è proprio promuovere l'utilizzo di software di Intelligenza Artificiale per poter arginare e combattere i pregiudizi e le disuguaglianze che affliggono le istituzioni e la nostra società.

Si ritiene che il giudice robot sia l'espressione massima di neutralità, ma come abbiamo potuto constatare nel corso del capitolo terzo, i programmi di giustizia predittiva si basano su dati presi da sentenze passate, elaborate da umani, in tal modo finiscono per incorporare tutti quei pregiudizi che sono tipici degli esseri umani e anzi finendo per ampliarli.

Un caso particolare e noto tra gli studiosi di questa affascinante materia è il caso Loomis: nel febbraio del 2013 Eric Loomis, cittadino statunitense, arrestato per ricettazione di un 'auto e resistenza a pubblico ufficiale, è stato condannato a sei anni di reclusione da un giudice che, per stabilire la pena, non si è basato solo sulla fedina penale dell'imputato, ma anche sul punteggio assegnato all'imputato da un algoritmo predittivo di valutazione del rischio di recidiva chiamato COMPAS.

Il già menzionato algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) esamina la propensione dell'imputato a ripetere il delitto sulla base di 137

domande cui deve rispondere o che vengono integrate da informazioni tratte dal casellario giudiziale. Si tratta di domande molto varie, si passa dal chiedere all'imputato se ha un telefono nella propria abitazione, se ha difficoltà a pagare le bollette, se ha precedenti penali o quale sia la sua storia e quella dei suoi familiari. Una volta esaurite le domande, l'algoritmo valuta su una scala da 1 a 10 quale sia il rischio di recidiva dell'imputato, e anche laddove sia sempre il giudice umano alla fine a dover determinare la pena e decidere se basarsi o meno su tale risultato, ormai nelle corti americane è obbligatorio utilizzare un sistema di questo tipo.

Questo algoritmo in particolare ha suscitato due tipologie di problematiche. In primo luogo, la società che lo ha sviluppato è una società privata che lo ha progettato e depositato il brevetto pertanto il codice sorgente non è aperto né pubblico ciò che crea un problema di trasparenza circa il procedimento che esegue l'algoritmo per giungere alla decisione finale e non da ultimo una sorta di mancanza di responsabilità dello Stato (detentore del potere di "punire" un proprio consociato ma conseguentemente responsabile di condotte sbagliate) in quanto esso può agevolmente andare esente da responsabilità lamentando che ha deciso l'algoritmo con un procedimento da lui (e da nessun altro) conoscibile e verificabile.

A tal proposito un gruppo di ricercatori del Bertelsmann Stiftung ha individuato una serie di regole per controllare l'uso dell'algoritmo che viene utilizzato per la decisione robotica:

- 1) E' necessario che i processi decisionali all'interno di un algoritmo siano sempre comprensibili.

- 2) Una persona fisica o giuridica deve essere sempre ritenuta responsabile degli effetti derivanti dall'uso di un sistema algoritmico.
- 3) Quando un sistema algoritmico viene utilizzato in maniera decisiva per risolvere una controversia o comunque una decisione che riguarda diritti di una persona, deve essere possibile richiedere una spiegazione di tale motivazione e poter presentare un reclamo.

L'altro aspetto problematico invece è stato evidenziato da un report di ProPublica del 2016 da cui risulta che alla popolazione afro-americana è attribuito un elevato tasso di recidiva nei due anni successivi alla condanna, in particolare due volte più alto rispetto alle altre popolazioni.

Infatti, l'algoritmo riteneva che altre etnie risultassero molto meno inclini alla recidiva, per cui mentre gli imputati afro-americani avevano maggior possibilità di vedersi assegnare alti punteggi di rischio di recidiva, la controparte bianca riceveva punteggi più bassi. Questo è solo un esempio di come i sistemi di Intelligenza Artificiale sfatino il mito del giudice imparziale, come recita la prima delle sei leggi di Kranzberg sulla tecnologia: "la tecnologia non è né buona né cattiva; non è neanche neutrale".

Sappiamo bene che le persone sono soggette a pregiudizi e che hanno la tendenza a discriminare e vorremmo che l'uso della tecnologia, essendo essa non dotata di pregiudizi in quanto non educata, non pensante ed incapace di provare emozioni, potesse risolvere il problema, ma dall'altra parte gli algoritmi sono costruiti su dati prodotti da persone e quindi finiscono con riprodurre proprio quegli stessi pregiudizi che abbiamo ed in taluni casi anche peggiorandoli,

ciò dovuto all'immissione di una enorme quantità di dati "pregiudizialmente viziati".

Non dobbiamo, però, cadere nemmeno nella trappola di considerare gli algoritmi il male perché come sottolineato poc'anzi essi assumono pregiudizi perché noi glieli immettiamo insieme ai dati che gli chiediamo di elaborare.

La vera sfida pertanto, è migliorare la conoscenza della tecnologia per poterla usare anche per combattere le disuguaglianze.

Un altro esempio interessante sempre in materia di pregiudizi (questa volta meno pregiudizievole per la categoria in questione ma perfetto per capire cosa intendiamo quanto parliamo di pregiudizi insiti nell'essere umano) è quello relativo ad uno studio effettuato dalle Università della Virginia e del Washington, relativo al sistema di riconoscimento di immagini. Lo studio ha dimostrato che se l'algoritmo vede una persona intenta a cucinare nella maggioranza dei casi tende a dire che è una donna anche se in realtà si tratta di un uomo ai fornelli.

Esempi come quelli citati hanno portato l'opinione pubblica a porsi in maniera ancora più critica e a diffidare ancor di più di questi sistemi, ma, se da una parte è certamente vero che gli algoritmi riflettono i pregiudizi dei loro creatori, dall'altra rappresentano, potenzialmente, un ottimo strumento per individuarli e rimuoverli.

In una ricerca effettuata per conto della McKinsey & Company, Jake Silberg e James Manyika hanno evidenziato che per combattere i pregiudizi è necessario che gli uomini sfruttino e lavorino insieme con gli algoritmi, comparando i risultati ricevuti e esaminando possibili spiegazioni per le differenze riscontrate nelle analisi. Se un umano riscontra che un algoritmo allenato su decisioni umane mostra

l'esistenza di pregiudizi, non dovrebbe semplicemente smettere di usare l'algoritmo, ma considerare come mai esso ha assorbito questi pregiudizi e correggerlo e poi, riflettere sul perché di queste storture che molto probabilmente si avrebbero anche se il giudice fosse un essere umano.

Un esempio di utilizzo dell'Intelligenza Artificiale in questo senso lo riscontriamo nella ricerca effettuata da Stamos T. Karamouzis e Dee Wood Harper per *l'International Journal of Law and Information Technology*. Oggetto delle ricerche erano le sentenze di condanna a morte nel sistema americano. L'obiettivo era dimostrare se queste fossero basate su dati arbitrari in conflitto con i principi di giusto processo e giustizia.

La pena capitale è un tipo di condanna con una storia molto antica, attualmente tra le democrazie occidentali, solo negli Stati Uniti è ancora vigente, ma anche qui è comminata assai di rado. Nonostante l'estrema pena, la percezione della *public opinion* era che la stessa fosse applicata in maniera arbitraria.

Per avere certezza della veridicità di questa opinione i ricercatori progettaronò un algoritmo basato su un sistema di apprendimento automatico, a più strati, che potesse verificare se il detenuto condannato a morte poi venisse effettivamente giustiziato, utilizzando variabili che erano indipendenti dal tipo di crimine commesso.

La rete neurale analizzò i profili di un migliaio di detenuti condannati di cui solo per circa la metà la pena era stata effettivamente eseguita.

La rete era stata impostata utilizzando 1000 profili, altri 66 erano stati utilizzati per una prima verifica, infine era stata testata sui rimanenti 200 casi.

Le variabili utilizzate per profilare ogni detenuto per formare la rete erano state:

- 1) Il numero di identificazione del detenuto;
- 2) Lo Stato di provenienza;
- 3) il sesso;
- 4) la razza;
- 5) se avesse o meno origine ispanica;
- 6) l'anno di nascita;
- 7) lo stato coniugale al momento in cui era stato arrestato per il delitto;
- 8) il livello di educazione conseguito in quel momento;
- 9) la fedina penale;
- 10) se era recidivo;
- 11) Altre informazioni relative ai tempi delle condanne

Dopo che la struttura della rete neurale era stata ottimizzata e testata, il passo successivo della ricerca era stato verificare la percentuale di capacità predittiva nella determinazione di esecuzione di un detenuto. La capacità in questione risultò essere pari al 90% dei casi, in pratica la rete era stata in grado di valutare correttamente 460 casi su 488 tra i detenuti non giustiziati e 448 casi su 512 di detenuti giustiziati. Questo esperimento ha permesso, non solo di trovare dei parametri utili ai fini della determinazione di un'effettiva esecuzione capitale o meno, ma, alla luce delle variabili utilizzate per descrivere i profili dei detenuti che prescindevano per la maggior parte delle sostanziali caratteristiche del crimine commesso, ha permesso agli esperti di evidenziare quanto sia giusto o ingiusto il processo di comminazione della pena di morte.

I *bias* amplificati dagli algoritmi, quanto sono analizzati consapevolmente, possono risultare strumenti utilissimi per la lotta alle disuguaglianze.. Il problema nasce quando i programmi hanno una funzione decisionale che andrà poi ad inficiare i destinatari, creando così disparità e distribuzione delle opportunità non omogenee. Per contrastare questo secondo fenomeno, è necessario che venga incentivata la ricerca e la marginalizzazione dei *bias*.

Lo studio effettuato con McKinsey illustra alcuni modi di minimizzare la discriminazione in questi algoritmi in modo da poter incrementare la fiducia delle persone in questi sistemi:

- 1) Pare necessario in primo luogo fare attenzione al contesto in cui l'Intelligenza Artificiale può aiutare a correggere i *bias* e verificare se possa esserci un alto rischio contrario, cioè che possa invece amplificarli.
Al momento dello sviluppo del sistema è importante anticipare una possibile formazione del pregiudizio all'interno del sistema, i programmatori devono sempre vigilare determinando con certezza i punti in cui i sistemi promuovono giustizia e quelli dove invece producono difficoltà.
- 2) Stabilire processi e pratiche per poter testare e mitigare i pregiudizi nei sistemi di AI. Ciò è possibile, per esempio, con algoritmi che siano il più trasparente possibile riguardo al processo e ai parametri utilizzato per giungere ai risultati.
- 3) Capire dove persistono i pregiudizi umani. L'AI riesce ad analizzare un grande numero di decisioni umane, quindi è

importante capire come queste decisioni siano effettivamente elaborate.

- 4) Investire di più nella ricerca e analisi dei *bias*: ciò potrebbe realizzarsi attraverso un supporto nello sviluppo di queste nuove tecnologie, nella messa a disposizione di più dati per i ricercatori che lavorano sulla questione pur rimanendo attenti a potenziali rischi di privacy.
- 5) Infine, è necessario promuovere iniziative in cui le macchine ed esseri umani lavorino insieme, valutando accuratamente i casi in cui l'AI possa prendere la maggior parte delle decisioni e quelle in cui, sia l'uomo a mantenere il controllo della macchina, o ad avere l'ultima parola. I programmatori dovrebbero orientarsi alla creazione di algoritmi che diano raccomandazioni o vari opzioni tra cui decidere e gli uomini abbiano una funzione di controllo o di decisione finale.

Visti gli enormi vantaggi che potrebbero seguire ad un funzionamento efficace di questi sistemi, una delle aree di ricerca più attive nel campo potrebbe essere proprio quella dello studio dei *bias* sia da un punto di vista statistico che sotto un profilo legale.

Come abbiamo avuto più volte modo di dire nel corso di questo lavoro, la tecnologia, così come l'Intelligenza Artificiale non è neutra, essa ha dei *bias* ontologici che però l'essere umano può certamente correggere, ma non è cattiva, dipende dall'utilizzo che se ne fa. In riferimento all'argomento del presente capitolo, l'AI può rivelarsi essenziale nella riduzione delle disuguaglianze sociali: in

particolare modo per quanto riguarda la conoscenza e garanzia dei diritti individuali.

Queste tecnologie permettono infatti di migliorare sensibilmente la qualità della vita degli individui, fino ad arrivare a perfezionare il giudizio umano riducendo i nostri pregiudizi, consci o inconsci che siano, facendoci avvicinare sempre più al concetto di giustizia ideale a cui mira il nostro ordinamento.

A questo punto, in conclusione di questo paragrafo è utile tornare al tema del titolo, ossia come le tecnologie modificheranno il ruolo del giudice e il sistema giudiziario in generale, anche se la vera domanda dovrebbe essere però non tanto *se* ma *quando* e *in che modo* ciò accadrà.

Già oggi possiamo vedere la tecnologia come parte integrante di alcuni tribunali e corti. Nei prossimi vent'anni si stima che l'AI giocherà un ruolo sempre più importante nell'occuparsi di dispute di modesto valore, come quelle riguardanti problemi assicurativi al di sotto di certe cifre, e avrà sempre di più un ruolo cardine nella formazione della decisione.

Si arriverà ad una sostituzione completa del giudice con un algoritmo? Esso ruberà il lavoro ai magistrati, avvocati, dottori commercialisti e giuristi? La risposta veloce è certamente no, quella più ponderata è dipende dal lavoro in oggetto.

Se si tratterà di lavoro ripetitivo, con un obiettivo ben definito ed una precisa lista di compiti da svolgere probabilmente potrà essere eseguito da un robot in modo rapido, efficace e soprattutto economico. Pensiamo ad attività come la redazione di un contratto di locazione o di una lettera di diffida, lavoro per cui non è indispensabile un essere umano in quanto si tratta di atti standardizzati.

In questi ambiti la sostituzione dell'uomo da parte della macchina può considerarsi molto probabile in futuro ed in modo quasi totale.

Per altri aspetti è chiaro quanto l'AI possa solo fungere da aiuto ma mai potrà sostituirsi ad attività che richiedono intelligenza nel senso umano, ossia capacità di agire in situazioni non determinate a priori.

Un esempio di tale attività tutta umana è la capacità di convincimento che un avvocato deve saper sfoderare durante un dibattimento e che varia da caso a caso, dipenda dal giudice che occorre convincere, dal grado di giudizio e via discorrendo.

Come può un algoritmo espletare con successo le tre forme di persuasione richiamate da Cicerone nel De Oratore, ossia:

“Dimostrare la veridicità della propria tesi, conciliarsi la simpatia degli ascoltatori e suscitare nei loro animi quei sentimenti che sono richiesti dalla causa”.

Per ora i sistemi esistenti di AI non sono in grado di fornire chiare interpretazioni a fattispecie concrete, molto spesso complesse, che richiedono un tipo di risoluzione creativa.

Come suggerisce Elliot Ash, nella sua conferenza per TEDxZurichSalon analizzando le problematiche della giustizia umana e di quella robotica, la peculiarità dei giudici umani è che hanno un concetto di giustizia che possono applicare a nuovi casi ed in risposta a nuove leggi che si evolvono insieme alla società.

I giudici robot, invece, per il modo in cui sono stati creati, questo non lo possono fare, possono solo replicare quello che i giudici umani hanno fatto.

Quindi, il ruolo ideale da un punto di vista tecnico ma anche etico è quello di coadiuvare i giudici umani a individuare le loro mancanze e i loro pregiudizi in modo da poterli superare e permettergli di

“ricordare” , per esempio, quanta più giurisprudenza possibile relativa ad un certo caso per facilitare il compito di reperimento di materiale da parte dell’umano che è mansione non solo faticosa ma anche dispendiosa di tempo.

Inoltre, a mio parere la componente emozionale è troppo importante per poter essere sostituita. E’ bene notare che per emozione si intende soprattutto l’empatia, cioè la capacità che deve o per lo meno dovrebbe avere un giudice quando stabilisce la colpevolezza o l’innocenza di una persona. Si tratta anche, forse, di quello che in gergo si definisce “sesto senso” cioè quella capacità che, seguendo ovviamente all’analisi oggettiva del fatto concretamente realizzatori e dopo aver applicato la fattispecie normativa astratta corretta aiuta in giudice ad “immaginare” “immedesimarsi” e decidere con la dovuta cognizione la causa.

La pronuncia del giudice non è unicamente una mera combinazione di motivazione e dispositivo, ma riassume in sé un processo che rievoca non solo la funzione del pensare ma anche, come detto, quella della intuizione.

Capitolo IV

La giustizia predittiva: problemi tecnici, procedurali ed epistemologici

Il capitolo finale del presente lavoro tratta, come si può evincere dal titolo la tematica dei rischi o se vogliamo dirla in modo più neutro delle differenze che tramutano spesso in difficoltà interpretative dell’uso degli algoritmi nella giustizia ed in particolare nel processo.

Non si è voluto fare la distinzione tra i differenti tipi di processo che, come ovvio, comportano rischi e responsabilità differenti nonché sono suscettibili di ledere interessi differenti ma si è voluto riassumere la principali caratteristiche di un processo svolto nei paesi di *civil law* e porre invece l'attenzione su tre principali ordini di problemi che sono appunto quelli che nel titolo definiamo "tecnici" ossia quelli più prettamente relativi alle fasi processuali e pre-processuali e analizzare quindi come la tecnologia ponga in essere cambiamenti radicali per esempio, nell'interpretazione della prova, o nella fase dibattimentale, problemi "procedurali", intesi essi non solo nel senso della procedura civile o penale ma proprio di metodo; qui si farà riferimento, per esempio, alla distinzione che intercorre tra causalità e correlazione e che la tecnologia talvolta tende a mettere sullo stesso piano, e per finire i problemi "epistemologici", altrimenti detti di politica del diritto l'affascinante tema dell'umanità del giudicare, dell'irresponsabilità di un giudice che si limita a motivare una decisione eterodiretta e presa da un soggetto non umano e di conseguenza il ruolo della Stato, inteso come unico detentore del potere i giudicare e di comminare sanzione e irrogare pene. Chi risponde di fronte ad una carcerazione preventiva subita da un soggetto poi dichiarato innocente?

4.1 La decisione presa da algoritmi: problemi e prospettive

La trasformazione digitale dei dati, anche di quelli giudiziari, si scompone in tre momenti distinti: la loro digitalizzazione, il loro

trattamento sotto forma di programma ed, eventualmente, la loro ulteriore correlazione sotto forma statistica.

Come indica infatti il termine, la “digitalizzazione” consiste nel codificare sotto forma di numeri eventi del mondo fisico, vale a dire nel convertire il fenomeno in un “segnale” che potrà essere trascritto sotto forma di una sequenza di numeri. La realtà esterna si trova quindi codificata sotto forma di dato, termine in genere utilizzato al plurale “dati”, poiché un dato isolato non ha alcun valore nella misura in cui non permette di operare su di esso alcun trattamento.

Tale passaggio si realizza innanzi tutto col trasferimento del dato dal fisico al semiotico, ovvero all’ambito dei segni. Si tratta prima di tutto di convertire la realtà fisica in segnale discreto per poi codificarlo tramite numeri. La codifica digitale rende, dunque completamente omogenei fenomeni molto diversi: un’immagine, un testo, la voce, le analisi del sangue, ecc. Qualsiasi ordine di fenomeni può ricevere una codifica digitale trattabile da un computer. Così, qualsiasi fenomeno, naturale o culturale che sia, può essere trascritto tramite una codifica digitale, ovvero in una scrittura di numeri.

Una simile forse di omogeneizzazione, infine, non può che realizzarsi mettendo da parte sia la percezione diretta, sia la lingua, che sono i vettori del significato: la trascrizione in un unico linguaggio viene infatti ottenuta respingendo tutto quello che potrebbe rimandare a un significato. Tale sospensione del senso viene prolungata da una seconda operazione: la programmazione.

Tale aspetto, prettamente nella mani del programmatore non lo approfondiamo: basti dire che la programmazione è l’organizzazione che avviene per i dati all’interno di programmi specifici per il loro

trattamento, una sorta di riordino per renderli comprensibili da una macchina e quindi fruibili.

Arriviamo poi al terzo aspetto, ossia a quello delle correlazione (di cui si dirà più approfonditamente nel paragrafo successivo).

L'aggregazione statistica che si crea con la programmazione viene a stabilire delle correlazioni finora passate inosservate, che consentono di tracciare delle tendenze nei fenomeni naturali e nelle interazioni sociali e, così facendo, di fare delle previsioni circa il futuro sulla base di eventi passati.

E' in questa fase che viene ad instaurarsi in modo più pregnante il cosiddetto "mito della delega alle macchine"¹⁴⁵ ossia si comincia a credere che non ci sia bisogno di interpretare i dati per poterli decifrare, ma che la delega ad appositi programmi dedicati sia sufficiente a tal fine.

Il diluvio di dati, secondo le parole di Chris Anderson , rende il metodo scientifico obsoleto.¹⁴⁶

Qualsiasi interpretazione, perfino scientifica, si trova quindi soppiantata da questa ideologia che porta a credere che il mondo parli da sé, perché ha avuto la buona idea di essere composto da atomi numerici e quindi calcolabili.

A prima vista, la comprensione statistica dei fenomeni naturali e delle interazioni sociali non è niente di nuovo siccome risale al XIX secolo e da allora l'uso della statistica per gestire gli Stati moderni non ha mai smesso di estendersi, senza il ricorso ai computer. Ma la raccolta dei

¹⁴⁵ A. Garapon, J. Lassegue., *La giustizia digitale, Determinismo, tecnologia e libertà*, Il Mulino, Saggi, 2021.

¹⁴⁶ C. Anderson., E' il titolo di un editoriale dell'ex direttore capo della rivista "Wired" del 23 giugno 2008: *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, in "Wired".

dati ha preso oggi una nuova piega, in virtù della sua automatizzazione informatica. Finora le tendenze statistiche in ambito sociale permettevano di definire quello che lo statistico Adolphe Quetelet nel XIX secolo aveva chiamato “l’uomo medio”, finzione matematica che consentiva di studiare i comportamenti collettivi e attuare politiche pubbliche per padroneggiarli¹⁴⁷. Oggi le rete consente non solo di conoscere le tendenze generali dei comportamenti e delle opinioni, ma anche di partecipare all’elaborazione di tali tendenze generali.

Inoltre tutto ciò che viene “informatizzato” diviene spesso altro dall’originale e talvolta assume significati anche differenti. Si pensi alla scrittura informatica, la quale è pacifico che non possiede tutte le proprietà del linguaggio.

Rappresentando le istruzioni in modo univoco, per evitare qualsiasi errore di elaborazione, i linguaggi di programmazione perdono una delle principali caratteristiche del linguaggio umano. Per illustrare tale caratteristica della lingua prendiamo un esempio tratto da Eugenia Grandet di Balzac:

A tanto problema gli uni rispondevano che Grandet non avrebbe concesso la figlia né all’uno né all’altro, e si aggiungeva che il vecchio bottegaio, roso dall’ambizione, cercava per genero qualche pari di Francia, al quale trecentomila lire di rendita avrebbero fatto accettare tutte le botti presenti, passate e future della famiglia.

La parola “botte” non indica qui certamente un “barile” ma rimanda piuttosto alla dimensione sociale del commercio del bottegaio: questa

¹⁴⁷ Come ha evidenziato J.Hacking, *The Taming of Chance*, Cambridge, Cambridge University Press, 1990.

sola parola racchiude tutto il racconto di Balzac. Il senso preciso della parola “botte”, che non è repertoriato in nessun dizionario, è tuttavia immediatamente riconosciuto come tale da qualsiasi lettore.¹⁴⁸

Questa straordinaria malleabilità del linguaggio umano l’accomuna a in tratto caratteristico dell’essere vivente, anch’egli capace di modificare la propria forma, nel corso della sua vita, senza estinguersi. Nel caso delle lingue informatiche, al contrario, esistono dei termini “primitivi” le cui regole d’uso ed il cui significato vengono stabiliti una volta per tutte. Sta proprio qui la profonda differenza rispetto al linguaggio umano nel quale ogni parola assume un significato valoriale differente a seconda del contesto, del tempo addirittura della modalità e del tono con cui viene pronunciata. L’enunciazione non è, pertanto, equiparabile a un ‘elaborazione statica come è l’attività informatica.

Un’altra differenza che è opportuno cogliere tra il linguaggio informatico e quello umano e stavolta con riferimento al diritto è la funzione espressiva e la funzione performativa del diritto.

Nel diritto parola può essere usata sia come potere espressivo per testimoniare, argomentare o deliberare, (tutte operazioni queste che mirano a stabilire un significato) sia come forza performativa (ovvero con la funzione di realizzare quello che proferisce), ad esempio per prestare un giuramento, acconsentire a un’ azione o pronunciare una sentenza. In questi ultimi casi la parola trae la sua forza non solo da quello che enuncia, ma soprattutto dal fatto che tale enunciazione è formulata in un preciso quadro giuridico-rituale, che è la condizione perché possa produrre i suoi effetti giuridici: attestare la verità, impegnarsi nel vincolo del matrimonio o condannare una persona.

¹⁴⁸ Cadiot, P., *La Metaphore, ou l’entrelacs des motifs et des themes*, in “Semen”, n. 15, 2022, pp. 41-59.

Prestare giuramento o dichiarare che una coppia è unita legalmente sono operazioni performative di cui sarebbe difficile fare a meno, perché è in ultima istanza su esse che si fonda l'istituzione del matrimonio e della testimonianza.

Del resto, gli atti più solenni come il matrimonio o il giuramento esigono una presenza reale: prestare giuramento o sposarsi implica che le parole vengano pronunciate forte e chiaro, solitamente in piedi, in un luogo ricco di simboli e in presenza di un pubblico. Per lo scritto il diritto richiede spesso la pratica rituale della firma per autenticare il soggetto, per formalizzare il suo impegno ed eventualmente rammentarglielo, il che presuppone sempre un rituale ed una forma.

Nel caso del diritto, il doppio carattere del linguaggio, espressivo da un lato e performativo dall'altro, è dovuto quindi in larga misura a un elemento non linguistico, di natura sociale: il dispositivo rituale o istituzionale che lo autorizza e che gli attribuisce la sua portata.

A differenza della scrittura alfabetica, in cui il raggruppamento di caratteri produce immediatamente il significato, la scrittura digitale è capace di una propria attività grafica, che ritarda l'avvento del significato. Essa non effettua più una registrazione del mondo ma un'elaborazione di dati. Questo è quello che nutre il "mito dell'autocontrollo della macchina" perché tale controllo sfugge necessariamente in una certa misura alla padronanza grafica da parte degli individui, a prescindere dal fatto che essi siano o meno degli informatici. Questo fatto, completamente nuovo, mette in primo piano l'aspetto performativo di questa scrittura meccanizzata.

Molti sono i problemi di fondo della digitalizzazione delle nostre vite e del diritto che per ragioni di spazio e anche di difficoltà di analisi non appare opportuno tratteggiare in questo lavoro.

Sicuramente tra le più evidenti però c'è quella che gli studiosi chiamano "giustizia sistemica". Per comprendere le trasformazioni all'opera nella giustizia digitale, occorre rapportarla a un modello di giustizia più ampio, che ritroviamo tanto nella governabilità neoliberale quanto nella giustizia globalizzata; essa è appunto la giustizia sistemica. Ci si orienta verso una giustizia di tipo sistemico quando il giurista non si basa più su una volontà politica, una legge, un'usanza o una tradizione, ma su un meccanismo tecnico che è neutro sotto il profilo morale. Il sistema tecnico prende il posto dell'operatore di giustizia sotto forma di una quarta parte.

Sarebbe tuttavia errato ritenere che si tratti di un semplice sistema di coordinamento paragonabile a un mercato o a un sistema tecnico, siccome la giustizia sistemica integra dei valori ma in un modo che le è proprio. I suoi principi sono ammissibili solo a condizione di essere integrati dalla tecnica (vedi gli ormai noti principi di justice by privacy o di privacy by design), o interiorizzati dagli individui. Ciò che manca rispetto all'ordinamento giuridico che conosciamo è la volontà comune che è all'origine del diritto, intesa essa come la consapevolezza di far parte di un gruppo di individui che si danno di regole vincolanti ed in quanto tali ognuno si sente pienamente parte del tutto solo a condizione di ottemperare a queste regole che egli stesso ha contribuito a darsi. Tale ordinamento si fonda su una Costituzione, una *Grundnorm* secondo Kelsen, legata a un ordinamento politico. Ora, l'idea di "sistema" si è emancipata, nel senso che il mercato e la governabilità neoliberale sono dei sistemi puri, indipendenti dall'ideologia. Si può così constatare un'inversione: non è più un sistema che si mette al servizio di una volontà politica collettiva ma, al contrario, una volontà politica che

deve integrarsi in sistema per sperare in una qualsivoglia efficacia delle proprie regole.

Da qui, il diritto non interviene più a partire da un atto di potenza pubblica, ma sposando la forma di un sistema, ovvero di automatismi che attingono la propria energia al di fuori della forza pubblica.

I due modelli di giustizia, ossia quella prima della rivoluzione tecnologia e quella post rivoluzione ossia sistemica, appaiono quasi contrapposti. Il primo è influenzato da eventi esterni che ne modificano in parte i presupposti e le regole normative mentre il modello sistemico e tecnico non è più integrato da riferimenti esterni ma esiste a prescindere dal resto, o meglio col resto si integra ma in una sorta di “individualismo virtuale”.

Un altro problema di tipo tecnico che viene a crearsi con l'utilizzo nella macchine nella giustizia riguarda la sempre maggiore “autonomia del soggetto” ossia perché diventa egli stesso espressione, tramite il proprio comportamento e stile di vita, della legge stessa.

Il soggetto è pertanto autonomo, ma non perché si fa da solo la propria legge, come nella filosofia classica di Rousseau e di Kant, ma perché appunto tramite le sue caratteristiche ed il suo comportamento “decide” la legge a lui applicabile. Tale legge, che non è stata decisa collettivamente da nessuna parte, è inserita nel suo comportamento individuale. Si tratta qui della situazione più opprimente:

Una norma che si adatta perfettamente al comportamento di ciascuno e che aderisce al corpo come una seconda pelle, non vi è nulla di peggio. Poiché non

si possono prendere le distanze da tale norma, in qualche modo la persona finisce per incarnarla¹⁴⁹.

Per il diritto antico, la generalità della legge era una condizione indispensabile alla libertà, che si definiva negativamente come ciò che non è proibito. Il diritto penale distingueva così molto nettamente l'atto repressibile, le motivazioni personali che possono spiegarlo ed il comportamento generale dell'individuo. Oggi non è più la personalità a poter eventualmente chiarire l'atto, al contrario è l'atto che chiarisce un comportamento, vero bersaglio, parrebbe, della giustizia digitale. Per via della sua limitazione all'individuo, il digitale corrisponde a un'estensione paradossale dell'ambito della norma.

La filosofia libertariana che è alla base del digitale e dell'ideologia anti-statale (intesa come superamento dello Stato nelle sue forme di territorialità, popolo e sovranità) della globalizzazione cospira contro qualsiasi istanza di regolazione che si incarni in istituzioni politiche perenni, esterne alle interazioni sociali da esse regolate. Le disposizioni astratte e generali della legge, così come la sua esteriorità, introducevano un'equidistanza tra i cittadini e la legge. Da qui la sua forza, di cui non ci si sbarazza così facilmente come credono gli entusiasti della Silicon Valley, in quanto la forma della legge e la sua triangolazione tramite la mediazione di un organo centrale come lo Stato territorializzato aveva la funzione di contenere la violenza, di orientare gli sguardi, di suscitare la fiducia degli individui.

¹⁴⁹ A tal proposito A. Garapon., *Bien juger. Essai sur le rituel judiciaire*. Nuova ed., Paris, Jacob, 1997; trad. it *Del giudicare. Saggio sul rituale giudiziario*, Milano, Cortina, 2006.

Contrariamente a quanto speravano i libertari, ciò che subentra all'indebolimento delle legge non è la scomparsa di qualsiasi controllo, ma il suo ri-orientamento.

Di qui l'argomentazione che si sta avendo una sorta di sopravvento dell'programmatore al giurista, o per meglio dire che colui che da sempre detiene il sapere giuridico e quindi il principale interprete dalla volontà del sovrano prima e dello Stato democratico poi stia di fatto perdendo il controllo di questa sue prerogativa durata secoli. In effetti nella giustizia predittiva, in un certo senso è proprio colui che scrive materialmente il codice di un algoritmo nato allo scopo di giudicare che, di fatto decide le modalità attraverso le quali la macchina realizzerà il suo compito. Non ultimo, non solo decide le modalità ma il *coder* è colui che fornisce al computer una sorta di *know-how* su come giudicare partendo per esempio dalla cernita dei dati che si decide di immettere all'interno della macchina affinché vengano da questa elaborati. Ciò naturalmente può dar luogo ad una miriade di problemi, alcuni dei quali evidenti di discriminazione di cui abbiamo avuto modo di parlare, per esempio, a proposito del noto caso COMPASS.

Non solo la previsione del comportamento può dare luogo, se estremizzata, alla mancanza di arbitrio, ma un' utilizzo smodato della macchine può avere conseguenze ancora più gravi e queste già in parte si possono constatare se si presta attenzione agli usi quotidiani che siamo costretti a fare della tecnologia. Il rischio è quello di passare da una delega legittima ed anzi auspicata alle macchine di alcuni lavori usuranti o ripetitivi alla paura di essere, da queste, relegati ad automi che non possono più esistere e svolgere le proprie attività quotidiane senza di loro. Tale sensazione che sicuramente in qualche particolare circostanza ognuno di noi avrà, almeno una volta nella vita, avvertito,

è resa ancor più evidente dagli strumenti tecnologici utilizzati durante la crisi pandemica. Un esempio fra tutti è rappresentato dall'uso dello strumento del green pass, il quale, prescindendo qui da considerazioni nel merito di tale strumento come argine alla diffusione del virus è divenuto certamente strumento necessario alla titolarità pratica della maggior parte dei diritti, anche fondamentali. Si pensi, per esempio, al soggetto boosterizzato al quale, per i motivi più vari non è stato aggiornato il green pass e lo stesso non solo, se superati i 50 anni non può neppure lavorare ma sicuramente non può svolgere la stragrande maggioranza delle attività umane. Tale paradossale situazione, peraltro, è risolvibile solo con la tecnica, ossia tramite quello strumento che ha, a monte, creato il problema, in altre parole non ci sono altre vie di uscita per correggere le storture di uno strumento informatico se non attraverso l'informatica medesima.

Da un punto di vista più filosofico, la possibilità di delegare alla macchina un certo numero di attività equivale quindi, sociologicamente, ad affidarsi a una nuova classe di scribi, gli informatici, per qualsiasi attività gestita per via informatica, ovvero per quasi ogni attività. Una simile delega alle macchine di tutto ciò che è meccanizzabile nelle interazioni sociali riorganizza qualsiasi divisione del lavoro e ogni gerarchia sociale. Tramite la robotizzazione, il controllo e la de-materializzazione delle informazioni, l'informatica ha cambiato la natura stessa del lavoro: qualsiasi lavoro d'ora in poi è legato alla costruzione, all'utilizzo e al mantenimento delle macchine, non solo in uno specifico luogo di produzione ma in qualsiasi tipo di interazione sociale.

Già dalla crisi del luddismo del XIX secolo, più i compiti erano ripetitivi, più colui che li eseguiva avvertiva il rischio di essere

rimpiazzato da una macchina. Il rischio si estende oggi a qualsivoglia forma di interazione sociale. Prende così forma una nuova norma che costituisce il fondamento della gerarchia sociale: più si fa sentire l'obbligo di passare attraverso un 'interazione meccanica e più l'individuo che vi si sottomette è relegato al gradino più basso della scala sociale. La relegazione passa ormai per l'obbligo di attenersi a interazioni meccanizzate nei rapporti sociali e di essere assimilato a una macchina per l'elaborazione di informazioni.

4.2 Causalità e correlazione

Una delle problematiche che pone la "previsione" ossia la capacità appunto di dedurre dal passato eventi che accadranno in futuro è la correlazione che spesso viene confusa con la causalità che, naturalmente, è altra cosa.

Un esempio per capirci: la più grande catena di distribuzione del mondo, Walmart, con un programma denominato Retail Link ha registrato i dati di tutti i prodotti così che potessero essere monitorati l'incidenza e il volume delle vendite e delle scorte. Nel 2004 decide di passare al setaccio i database delle transazioni pregresse, cioè gli articoli acquistati da ogni singolo cliente e il costo totale sostenuto, tutto ciò che aveva nel carrello, il giorno, l'ora e persino le condizioni atmosferiche al momento dell'acquisto.

Cosa ha scoperto? Ha scoperto che in previsione di un uragano aumentavano non solo le vendite, come è ovvio, di torce elettriche, ma anche quelle di merendine dolci.

Naturalmente Walmart di questa scoperta ne fece in motivo di business in quanto ogni volta in cui venivano annunciate delle tempeste, il colosso commerciale impilava scatole di merendine all'ingresso dei suoi punti vendita, accanto alle attrezzature anti-uragano, per facilitare il compito ai clienti che entravano e uscivano di corsa, facendo così lievitare il fatturato.

Tuttavia questo è esempio, seppure ci spiega una evidente e provata correlazione tra la previsione dell'evento uragano ed il consumo di prodotti dolciari, in realtà nulla ci dice circa la causa, ossia la motivazione alla base della voglia dei cittadini americani di consumare più dolci durante eventi atmosferici estremi. Probabilmente la ragione sta nel fatto che di solito questi prodotti hanno una lunga conservazione e possono essere tenuti anche fuori frigo (nel caso di mancanza di corrente) ma come detto la correlazione è una cosa mentre la causazione dell'evento un'altra.

La "causalità" è un concetto di origine mitica che è stato rielaborato dalle scienze naturali. Mostra che un evento spazio-temporale porta immancabilmente a un altro. La correlazione, dal canto suo, si basa su una affinità statistica fra numeri che non sono necessariamente intesi come avvenimenti nello spazio-tempo. Ad esempio, le banche hanno notato una correlazione fra l'aumento dell'indebitamento nei conti bancari intestati a entrambi i coniugi e il tasso di divorzio. Questa correlazione non fornisce la causa di tale rapporto ma è sufficiente per dedurre una regola di uso interno delle banche che, ad esempio, rivedranno di conseguenza la loro politica dei prestiti.

Nei due esempi fatti si evince quindi una mancanza di spiegazione del fenomeno ma una regola dalla quale trarre conseguenze. In fondo, se ci pensiamo, cosa interessa a Walmart sapere la motivazione per la quale

si mangiano più merendine dolci durante le tempeste e per le banche la causa del divorzio dovuto a indebitamento? A loro basta la correlazione da un punto di vista di pratiche commerciali e di soluzioni da adottare a loro vantaggio.

L'aggregazione di dati potrebbe lasciar pensare che le correlazioni fra due fenomeni nel passato siano sufficienti per prevedere il loro rapporto nel futuro, senza che ci si debba interrogare sulla ragione che permette di porre in relazione i due fenomeni. In verità, l'uso dei numeri nei due casi non è affatto lo stesso. Nel caso fisico-matematico, i numeri servono a descrivere in modo assolutamente rigoroso l'evoluzione di una realtà spazio-temporale tramite un'equazione che ha la forza di una legge; il significato dei numeri dipende da una realtà che essi non descrivono dall'esterno ma che incarnano internamente. Nel caso statistico, i numeri sono dei segni univoci, che rimangono "agnostici" rispetto alla ragione del rapporto che stabiliscono e si limitano a una descrizione esterna che si manifesta nell'espressione di una regola.

Il significato dei numeri nelle correlazioni statistiche dipende allora dalla dimensione della serie studiata, in quanto la sua estensione gioca un ruolo analogo a quello dell'equazione nel caso fisico-matematico: è là che si concentrano il peso ontologico e la pertinenza delle correlazioni da scoprire. Ora, un database troppo ristretto non permette di correlare alcunché, né tanto meno di prevedere, mentre un database troppo esteso permette di dedurre qualsiasi correlazione e quindi nessuna valida previsione.¹⁵⁰

Pertanto, la dimensione del database non può essere lasciata a discrezione dei programmi, poiché questi si applicano in modo

¹⁵⁰ C.S Calude e G. Longo., *The Deluge of Spurious Correlations in Big Data*, in "Foundations of Science", Vol. 22, n. 3, 2017, pp. 595- 612.

indifferenziato a qualsiasi database, indipendentemente dalla sua consistenza. La dimensione che il database deve avere affinché i fenomeni si rivelino appropriati è quindi intrinsecamente esterna ai programmi stessi.

E quindi cosa si potrebbe rispondere alla domanda, ma il futuro può essere dedotto dal passato?

L'idea di "previsione statistica" è problematica in quanto dipende da due condizioni: da un lato, come visto, da un giudizio sulla dimensione dell'insieme studiato e, dall'altro, da una supposizione secondo la quale ci sarebbe un'omogeneità tra futuro e passato. Ora, si tratta qui di un presupposto, dato che l'approccio causale ha mostrato al contrario che futuro e passato non sono omogenei: il futuro dipende da condizioni tali per cui un cambiamento, se pure impercettibile, può produrre delle evoluzioni certamente prevedibili una volta iniziate, ma imprevedibili prima che si verifichino, in quanto sotto la soglia di misurazione.

Si vede allora come la convinzione comunemente diffusa secondo cui i programmi per mettere in correlazione sarebbero di per sé "intelligenti", e sarebbe quindi possibile delegare interamente la lettura dei dati, che parlano da soli, è illusoria.

Bisogna dunque avere ben chiara la distinzione tra legge e regola: la legge è ontologicamente appesantita dal suo rapporto con lo spazio-tempo, mentre la regola è una corrispondenza tra numeri.

E' sbagliato credere che basta che dei programmi tengano automaticamente conto di un gran numero di casi per assimilare le regole a delle leggi: esse non sono come le prime e non dipendono dalle stesse condizioni di validità.

A questi punto è utile chiedersi quali problemi pone l'applicazione della previsione alle decisioni giudiziarie. Che ne è del calcolo della previsione nell'assetto delle sentenze giudiziarie?

La risposta è in tre punti. Innanzitutto, per quanto detto sopra, la prevedibilità così come è stata pensata nell'epoca del determinismo è oggi superata, nelle scienze esatte (matematica e logica) come nelle scienze naturali (fisica, chimica e biologia), e non c'è alcun motivo di pensare che l'uso di programmi nell'ambito del diritto non sia soggetto agli stessi limiti.

Inoltre, non c'è motivo di credere, vieppiù, che l'uso di tali programmi, che si tratti della classificazione della giurisprudenza, della valutazione e correzione delle storture cognitive dei giudici o degli avvocati, o ancora della valutazione della loro reputazione, possa smettere di espandersi, poiché non esiste alcun limite che si possa opporre alla evoluzione tecnologica e purtroppo a questa, talvolta malata, volontà di rendere tutto digitale, veloce, semplice e a portata di clic.

Tuttavia, penso che la persistenza e la preservazione di un indeterminato non calcolabile siano necessari per costruire un senso giuridico in continuo divenire. Il diritto non è riconducibile alla somma delle prescrizioni positive, poiché offre anche la capacità di trasformarle, pur mantenendo il suo potere prescrittivo. Il calcolo predittivo priva il diritto della sua capacità di trasformazione e evoluzione.

La distinzione epistemologica ormai ammessa dalle scienze esatte fra determinismo e prevedibilità si ritrova nel diritto come distinzione tra prescrizione e prevedibilità. Lo dimostra il fatto che la norma di diritto si evolve in modo imprevedibile nel corso del tempo, ma non per questo mantiene meno il suo potere prescrittivo, al massimo muta l'oggetto

della prescrizione in maniera più repentina se si ha un “cambio” attuato dall’organo legislativo e più graduale se ha ad oggetto, per esempio, una norma non scritta, che fa riferimento a degli usi.

Ma quali sono i principali rischi sistemici nell’utilizzo dei programmi predittivi nella giustizia e quindi di un uso smodato delle correlazioni? Occorre sin da subito evidenziare quanto i programmi predittivi sono ancora troppo recenti perché ci si possa soffermare compiutamente sui loro pericoli accertati; è possibile tuttavia individuare alcuni possibili rischi.

Il primo è senza dubbio la rarefazione dei giudizi. Ciò per due principali ordini di questioni; il primo in quanto se un algoritmo ha quelle date informazioni e sulle stesse base le sue decisioni, ciò significa che le sue future decisioni andranno sempre nella stessa direzione ed anzi amplificheranno la valenza del precedente diventando ancora più “cogenti”; il secondo motivo attiene ad un minore ricorso alla giustizia da parte del cittadino che in tal modo sapendo già come la sua causa andrà, nel 90% dei casi a finire, sarà dissuaso dall’agire in giudizio. Questo secondo aspetto ha un elemento positivo senza dubbio, che è quello di un minore ingolfamento della macchina giudiziaria, tuttavia si scontra col dato, che ogni giudizio dovrebbe basarsi su elementi soggettivi e quindi, estremizzando si potrebbe arrivare a dire che due casi simili possono essere risolti in modalità opposte, ed anzi talvolta deve proprio essere così, altrimenti tanto vale a qual punto avere un macchina che decide perché l’attività del giudicare diventerebbe, a tutti gli effetti, ripetitiva.

Un altro elemento negativo dato da un uso sregolato degli algoritmi predittivi nella giustizia e collegato al problema appena trattato è quello relativo al rafforzamento delle tendenze maggioritarie e ciò lo si nota

soprattutto nei paesi si common law, come sussiste la regola del precedente giudiziario vincolante. Qui infatti se da un certo punto in poi la gran parte dei giudici deciderà certe controversie in un modo, nel futuro e con un meccanismo che incrementa se stesso, sempre più giudici saranno portati a sentenziare in quel modo, uccidendo così il fondamentale principio dell'evoluzione giurisprudenziale che nei paesi di common law è anche l'unico modo (o quasi) attraverso il quale si attua l'evoluzione del diritto.

Un ultimo rischio, infine, ma non per importanza, è di vedere la giustizia predittiva trasformarsi in una sorte di normatività secondaria, che si sostituirebbe alla regola di diritto stessa. La norma realizza ciò che enuncia, e che nella specie spinge verso un'uniformazione delle pratiche.

Un giudice poco coraggioso è tentato di considerare che, se la maggioranza dei suoi colleghi decide in un certo senso, è meno rischioso per lui seguire la loro decisione. E' l'effetto gregge della giustizia predittiva: essa spinge verso il conformismo. Naturalmente questo si ripercuote anche sulla professione dell'avvocato che perorano soluzioni alternative o che vogliono far cambiare una giurisprudenza. Infine e in netta correlazione con quanto sopra si avrà in tal modo una sorta di "pietrificazione del tempo"¹⁵¹.

Sostiene il giudice francese esperto di giustizia digitale Antoine Garapon che "la giustizia predittiva sconvolge così la funzione ordinatrice del tempo in quanto introduce *ab initio*, se non la soluzione per lo meno una soluzione molto probabile. Il tempo della procedura non è quindi più del tutto lo stesso avendo perso il suo punto di fuga, che è la decisione finale. Con gli strumenti predittivi, la posta in gioco

¹⁵¹ A. Garapon., *La giustizia digitale*, Il Mulino, Saggi, 2021, p. 185.

non è più la decisione giuridica, ma la risoluzione sociale del caso attraverso le pieghe di una negoziazione. I diversi attori della giustizia devono integrare un parametro aggiuntivo nelle loro scelte e un nuovo tempo, che non è né il futuro, poiché non si verificherà mai, né il condizionale, poiché non è in un mondo ipotetico; si tratta di un tempo virtuale che non esiste propriamente e che non è quindi né presente né futuro ma ciò nonostante li determina entrambi. La giustizia predittiva accresce il presente. Il ragionamento, d'altronde più strategico che giuridico, è accresciuto dalla conoscenza di una proiezione virtuale di tale decisione. Occorre considerare tale forma digitale di partecipazione alla decisione come un livello supplementare delle informazioni, che arricchirà il fascicolo: un aiuto, fra gli altri, nel prendere una decisione. C'è da scommettere che fra qualche anno, in ogni causa, o per lo meno in quelle "importanti" per valore economico o oggetto della medesima, saranno fornite da ciascuna parte delle valutazioni predittive, forse contraddittorie fra loro; esse compariranno nel fascicolo come informazione nuova che coadiuva il lavoro di ciascun attore giuridico, senza sostituirlo.

La possibilità di una previsione perfetta implica un'abolizione del tempo, poiché il futuro diventa presumibilmente tanto conosciuto come il passato. Tale crollo del tempo permette quindi di considerare il futuro come già presente, in una forma generalizzata di prolessi: di conseguenza, sapendo come si svolgerà il futuro, potremmo considerare che esista già qui e ora, trasformando così allo stesso tempo la predizione in una ingiunzione.

La giustizia predittiva, se non vuole passare per una giustizia divinatoria, misteriosa e intimidatoria come gli antichi oracoli, deve primariamente rendere pubblici i suoi algoritmi e non nascondersi

dietro al segreto industriale e gli attori del processo battersi affinché la centralità del contraddittorio comunque non venga meno così come gli altri grandi principi della giustizia e del diritto.

4.3 Discontinuità tra il fatto e il procedimento e una inevitabile de-ritualizzazione del processo

Secondo Yuval Noah Harari, nel futuro:

Scanner cerebrali potranno rivelare menzogne e inganni e premendo un bottone [...]. D'altra parte, mentire implica l'attivazione di aree del cervello differenti da quelle attivate quando si dice la verità. Non siamo ancora arrivati a questi punto, ma possiamo immaginare che in un futuro non troppo lontano gli scanner a risonanza magnetica potrebbero funzionare come infallibili macchine della verità.¹⁵²

Questo mito della delega del giudizio a delle macchine non è forse all'opera in quella folle speranza di risolvere il problema della verità tramite la diagnostica per immagini? O tramite un'altra macchina, quella della verità appunto che sarebbe infallibile perché in grado di leggere ed interpretare sulla base del numero di battiti cardiaci se una persona sta mentendo?

In tali condizioni, la verità non è più una sfida per la comprensione, non è più una questione di intima convinzione, ma il frutto di una messa in correlazione di dati oggettivi che sono oggetto di elaborazione piuttosto che di interpretazione.

¹⁵² Y.N. Harari., *Homo Deus: A Brief History of Tomorrow*, London, Vintage, 2017.

A cosa serve giudicare, se la verità emerge da sola tramite il semplice intervento della tecnologia che è infallibile? Il cervello, considerato l'interfaccia digitale tra l'uomo e la natura, non passa più attraverso la coscienza, ma direttamente tramite dei numeri sottoposti a un calcolo. La scrittura digitale converte in numeri la parte più intima dell'uomo: la sue intenzioni, e perfino la "sua" verità, ma una verità che gli sfugge;¹⁵³ da qui la somiglianza dell'uomo tecnologico con l'uomo delle società tradizionali che si rimette all'ordalia come forma di verità superiore alla propria percezione e di questa si fida ciecamente seppur non ne conosce le modalità di azione.

La tecnica tende a far dimenticare le condizioni fisiche e sociali che hanno determinato l'elaborazione di una prova. E questo vale per tutta la tecnica: nessuno pensa al flusso di elettroni che percorre il cavo di alimentazione quando premiamo l'interruttore per accendere una lampada. Lo sfruttamento digitale dell'immagine, poi, ha trasformato profondamente il suo ruolo sociale: non solo gli usi dell'immagine si sono divulgati (moltiplicazione della presenza sugli schermi e riduzione del costo), ma l'autorità dell'immagine si è modificata sotto il profilo qualitativo. L'immagine scientifica, dall'astronomia al microscopio elettronico, passando per la diagnostica per immagini, è diventata un elemento costitutivo di un'incontestabile oggettività.

L'immagine è un dato oggettivo, inconfutabile nella sua dimensione di immagine. Le riconosciamo la potenza di riprodurre una realtà talvolta più vera della realtà stessa.

¹⁵³ Mentre i suicidi e gli atti di auto-mutilazione sono in crescita negli USA, Facebook ha creato uno strumento che permette di rilevare automaticamente i comportamenti e i pensieri suicidi a partire da alcune parole chiave dei messaggi postati sul social network.

Nel processo, tutto questo ha importanti ripercussioni. Sebbene i per sé non significhi subito qualcosa, l'immagine viene presentata direttamente ai giudici, senza alcuna mediazione orale o scritta. Quando un 'immagine, o meglio, una realtà ricostruita da avvocati negli USA, viene proiettata a una giuria, l'effetto sulle menti non è lo stesso di quando si ascolta un esperto o un testimone che racconta ciò che ha visto. Un giurato non reagisce nello stesso modo alla lettura di un rapporto o all'ascolto di un testimone rispetto alla visione di un video. Quest'ultima gli propone un'altra percezione dei fatti rispetto a una prova, che parla tanto con quello che mostra, quanto con lo sforzo di immaginazione che sollecita. La sentenza diventa quindi una questione di vedere e non vedere. Il giudice deve vedere tutte le prove, tutti i testimoni ma nient'altro. E' la condizione perché continui a fingersi il testimone dei testimoni. Se così non fosse, e vedesse i fatti come "se ci fosse", rischierebbe di diventare un testimone fra altri e di perdere la sua veste di terzo.

Delegando la costituzione e la diffusione delle immagini a dei trattamenti informatici, si rafforza la funzione di "copia del reale" di alcune immagini. La tecnica digitale la nomina implicitamente foriere di una nuova forma di oggettività: ciò che è stato fotografato o firmato sembra avere un significato immediatamente oggettivo, senza che ci sia bisogno di far convalidare quelle immagini da procedure collettive, come il processo o il dibattito scientifico. Più esattamente, questa deliberazione collettiva si limita a una tacita acquiescenza verso la delega alle macchine.

Rispetto alla prova tradizionale, ricondotta in sostanza a un testo scritto o a una parola, le nuove modalità di riproduzione della realtà, presente

e passata, pretendono di liberarsi di qualsiasi procedura ed interpretazione.

La nostra epoca è affascinata dall'intelligenza artificiale, che, ad esempio, annuncia di poter ormai ricostruire il viso di un individuo a partire da una goccia di sangue. Se tale scoperta fosse confermata, si aggiungerebbe all'efficacia probatoria del DNA, che consente di identificare un individuo a partire da un capello. La tecnica spinge quindi alla credenza collettiva che il corpo parli da solo, e che sia sufficiente a smascherare il criminale. Lo stesso potrebbe dirsi per le intercettazioni telefoniche, che diventano in qualche modo l'equivalente tecnologico della confessione: una confessione che non ha più bisogno di soggetto, una prova che si conquista oltre la volontà dei soggetti, analogamente ai loro corpi che li tradiscono.

Un'altra difficoltà poi che si incontra nel processo tecnologico sorge sul tempo di sfruttamento delle prove conservate in forma digitale, che può essere infinito, come mostra l'esempio della virtopsia (ossia l'autopsia virtuale). Si tratta di una scansione estremamente completa di un cadavere, associata a un congelamento dei liquidi corporei, che permette di rifare l'autopsia tutte le volte che lo si desidera, senza più toccare il corpo.

Una volta veniva insegnato nelle scuole di criminologia che un 'autopsia poteva essere praticata una sola volta siccome l'esame devastava a tal punto i tessuti che non era più possibile considerare attendibili dei nuovo tentativi. Questo limite è oggi superato: con la virtopsia si può perfino immaginare di risolvere dei crimini secoli dopo il verificarsi dell'evento.

Per quanto concerne poi il momento del processo, anche qui si ha un cambio, o forse potremmo dire un impoverimento rispetto al vecchio

processo tradizionale scandito da ritualità, momenti ben precisi che si susseguono ordinati e via scorrendo.

La possibilità di districare il groviglio del processo classico per mezzo delle tecniche del digitale, infatti, suscita molte speranze: accelera le procedure e lotta contro il male endemico della lentezza della giustizia, ne abbassa i costi, può favorire l'accesso al diritto, riduce lo stress delle vittime e degli imputati chiamati a comparire. Abbiamo visto però che tutto ciò non è comunque senza danni, poiché in taluni casi è rafforzato l'impatto dei sentimenti. I benefici sembrano quindi importanti per l'amministrazione della giustizia, ma queste innovazioni non influenzano il significato del processo tanto sul piano soggettivo quanto su quello sociale?

La regola delle tre unità stabilisce un certo rapporto con lo spazio e con il tempo. Le tecniche di digitali influenzano in profondità questi due aspetti; pur mantenendo l'apparenza della medesima struttura spazio-temporale, esse minano una componente fondamentale, sebbene invisibile, del processo, vale a dire la sua efficacia simbolica.

Si può notare poi un secondo fattore di discontinuità ossia quella tra la presenza e il luogo dell'azione.

Quando da una prigione in Sicilia un detenuto testimonia in un processo che si tiene a Chicago, non è veramente lì; certamente, non subisce la pressione del processo, il peso della compresenza di altri detenuti, ma subisce di contro la pressione della prigione; ad esempio, potrebbe forse avere una catena ai piedi, il che è peraltro anche contrario al principio secondo cui un imputato deve sempre poter testimoniare senza impedimenti. La tecnica introduce un'asimmetria nell'esperienza simbolica; limitandoci solo a questo esempio ma se ne potrebbero fare a centinaia, i giudici si trovano in una maestosa aula di tribunale, mentre

il detenuto probabilmente in una lugubre cella: essi non vivono lo stesso evento.

Bisogna diffidare dei preconcetti: viene spesso avanzato l'argomento secondo cui l'interrogatorio di un testimone su internet, via Skype, sua meno stressante di uno condotto durante l'udienza. Ciò dipende anche da numerosi altri fattori: dalle condizioni di ascolto, dall'età dei testimoni, dalle relazioni che si sono strette durante le udienze e il processo, tutti aspetti imponderabili che nessuno conosce e può valutare per il prossimo. Ma il problema va al di là dei rispettivi meriti della videoconferenza o dell'udienza dal vivo. Nel *trial* come nel processo, la performatività proviene dall'incontro fisico. Il giuramento è l'incarnazione di un uso performativo della parola. L'efficacia di questo enunciato performativo dipende sia dalla formula sia dalle condizioni nelle quali viene pronunciata. Si può dunque comprendere come la voce, portatrice dell'individualità dei soggetti, sia alla base del processo di *common law*, e che la sua digitalizzazione e tutte le tecniche digitali che si fondano sulla registrazione e sulla ripetizione minacciano la sua potenza di verità.

Vi è poi la discontinuità tra la comunicazione e l'informazione. Le interazioni sullo schermo sono ridotte a uno scambio di informazioni; siamo a metà strada fra lo scambio fisico e la comunicazione via email. Il video priva la relazione giuridica del linguaggio del corpo e di tutto quello che suscita la compresenza; la rende funzionale, riducendola alla sua funzione procedurale. La presenza fisica è anche una condizione dell'intuizione, pensiamo a due persone durante il dibattito e alla loro gestualità, ai tremolii della voce ecc.

Ora, l'intuizione, come abbiamo avuto modo di discutere sopra nel capitolo dedicato al "giudice senza empatia" è inesistente in una

macchina. Che si tratti dell'intuizione del medico che palpa il corpo del paziente, o di quella di un giurato americano che valuta la credibilità di un testimone, o di quella dell'avvocato che percepisce l'attenzione della giuria a cui si rivolge: ogni volta, l'intuizione consente il gioco, in tutti i sensi. Per convincere e anche per ingannare. L'intuizione consente di poter modificare l'interpretazione dei dati, senza cambiare i dati stessi: ad esempio, la procedura di ricorso in appello si basa sulla possibilità di trarre conclusioni diverse dagli stessi fatti, cosa che una macchina non potrebbe fare, non potendo che ripetere in modo rigorosamente identico gli stessi risultati a partire dagli stessi dati analizzati.

Si ha poi la discontinuità tra il gesto e la parola. In un processo bisogna accompagnare la propria parola a una postura fisica e a una concentrazione adeguate. La parola acquisisce una forza di persuasione grazie a una particolare gestualità: è nel modo, e non nel contenuto che si gioca l'efficacia simbolica, perché è qui che risiede la performatività. Si pensi alla differenza tra un consenso espresso tramite una parola ufficiale davanti a un terzo simbolico (come un sindaco o un prete) e l'atto di premere un tasto del computer.¹⁵⁴

Il secondo gesto è estremamente ambiguo siccome si può cliccare per accedere a una pagina e ottenere un servizio, senza consapevolezza dei diritti a cui si rinuncia (vedi la privacy).

Vi è poi un aspetto tra il reale ed il virtuale di un'udienza che ci sta particolarmente a cuore e si tratta della discontinuità tra vedere, udire e perfino percepire gli odori di un luogo quale può essere l'aula di un tribunale.

¹⁵⁴ D. Cohen, *Ampleur et qualites du consentement donne par clic de souris*, in "Les Petites affiches", n. 164, 2015.

Sullo schermo di un pc è possibile infatti vedere e udire ma non percepire gli odori, o anzi non percepire l'odore in senso stretto e figurato (nel senso di "capire come butta").

Eppure, l'esperienza fisica con altre persone passa attraverso una corporeità condivisa e quindi anche attraverso una percezione olfattiva. Una cella di un carcere ha un odore, così come l'aula bunker di un tribunale, spesso posta nei sotterranei e chiusa per la maggior parte del tempo. La visione d'insieme offerta dall'aula di tribunale cambia il modo di ascoltare. Un discorso, che sia una testimonianza o un'arringa si ascolta e si percepisce anche attraverso i suoi silenzi o dall'intensità del silenzio che provoca quando termina. Gli avvocati si ascoltano ma si guardano anche per vedere l'effetto che sortiscono sul giudice o sull'uditorio in generale. La credibilità di ciascun attore del processo si deduce da tutte queste esperienze nel loro insieme. La tecnica impoverisce l'esperienza della giustizia, che diventa un coordinamento di propositi tenuti separati. Non vi è più un'esperienza comune, un vissuto condiviso. L'atto di giudicare si priva della propria semiologia, ovvero di una massa di informazioni e di interpretazioni che perlopiù non sono consci. Lo schermo del pc isola una parte della comunicazione e ne sottolinea un'altra.

Infine, ma solo volendo riassumere si ha la discontinuità tra azione e "tempi morti". La tecnica, infatti, cambia anche la percezione dell'evento di giustizia; il coordinamento digitale evita qualsiasi attesa nell'aula di tribunale. L'attesa prima di una udienza non è una perdita di tempo, essa consente di vedere la vita di persone che non avremo mai occasione di incrociare altrove, di comprendere meglio il modo in cui

l'istituzione giudiziaria rende giustizia. L'attesa è come il margine della relazione, l'affianco che permette di prepararsi.¹⁵⁵

Quale differenza intercorre tra il seguire un'udienza su Skype o assistervi fisicamente? In un caso, è necessario per lo meno prendere mezza giornata di permesso da lavoro, recarsi in tribunale, aspettare, avere paura del confronto con un universo estraneo e per lo più percepito come ostile, nell'altro, non avremo abbandonato il quadro quotidiano (forse nemmeno i pantaloni del pigiama!) e saremo in un ambiente familiare e rassicurante.

Il secondo ed ultimo aspetto che si vuole trattare nel presente paragrafo di tesi riguarda la cosiddetta de-ritualizzazione.

Di fronte a un disordine, a un conflitto o a una violenza, la tecnica e il rituale giudiziario adottano attitudini molto diverse. L'iterazione informatica mira a una ripetizione sempre uguale di procedure del tutto definite in anticipo, mentre la ripetizione rituale mira a riportare ciò che è inquietante a ciò che è conosciuto, l'estraneo al familiare. Il rituale genera senso grazie ai simboli ivi contenuti. Si caratterizza troppo spesso il rituale come ripetizione, poiché l'uso obbligatorio di testi canonici, formule, gesti e posture ne è la componente più immediatamente percepibile.

Il rituale giudiziario, anch'esso per certi versi ripetitivo è fatto per introdurre un vuoto di senso, uno svuotamento temporaneo, un distacco dalla realtà, in vista della ricostruzione dei fatti.

La tecnica digitale, invece, mira alla previsione, a partire dalla ripetizione dello stesso elemento, in un quadro di pensiero che tende a supporre che più l'iterazione di un processo ha funzionato in passato, più funzionerà nel futuro. Ora, nel trattamento giudiziario, per

¹⁵⁵ A. Garapon., *La Giustizia digitale*, Il Mulino, Saggi, Milano, 2021.

definizione, non si conosce l'esito di un processo e ci si accorda all'interno di un quadro rituale da cui ci si attende che possa trasformare il conflitto agli occhi delle parti. Il processo è una tappa intermedia in cui tutto si rigioca. Un quadro simbolico come quello dell'aula di tribunale è indispensabile perchè la realtà assuma un nuovo significato. Tuttavia, la riproduzione tecnica non prende in considerazione la componente sociale della giustizia. Concentrandosi sull'efficacia e rovesciando la soluzione dei conflitti sul lato privato, manca della necessaria partecipazione collettiva alla risoluzione di conflitti e disordini sociali. Perde anche la possibilità data al popolo di rimettere in gioco la legittimità delle istituzioni ed eventualmente l'occasione per esse di guadagnarsela ancora.

Riprendendo l'esempio del giuramento e del consenso, che abbiamo detto essere tra gli esempi di uso performativo del diritto, possiamo dire che è la congiunzione tra gesti e voce che rende possibile la performatività. L'efficacia simbolica del linguaggio non può dipendere solo dal linguaggio medesimo ma è frutto anche di gesti fisici.

Il rituale spesso è noioso, opaco, inspiegabile ma è proprio per questo che risulta attrattivo ed emozionante ed il trasposto che suscita, insieme alla emozioni appare fondamentale per il suo corretto funzionamento. Il rituale sarà ancora più potente perché non può essere riprodotto, contrariamente alla tecnica digitale che si fonda su una reiterazione infinita.

Nella maggior parte dei casi i giuristi si infastidiscono per essere considerati dei mediatori della loro parola, digitale o simbolica che sia. Essi sono spesso intimiditi davanti ai rituali, così come davanti alle macchine, per via della loro comune "opacità", sebbene di natura molto diversa. L'opacità grafica non è dello stesso ordine del mistero del

rituale. L'elaborazione informatica fornisce sempre la stessa qualità, la stessa affidabilità, del resto è quello che da essa ci si aspetta, ma non trasforma chi fa ricorso, diversamente dall'ordine simbolico. Possiamo quindi parlare del formalismo del linguaggio informatico e del formalismo del rituale, ma si tratta di due realtà molto diverse.¹⁵⁶

4.4 Verso l'ultra-soluzione: problemi di politica del diritto e il rischio di un giudice (ir)responsabile

Come diceva l'antropologo americano Gregory Bateson mezzo secolo fa: “La cibernetica rappresenta il boccone più grosso che l'uomo abbia strappato dal frutto dell'Albero della Conoscenza negli ultimi duemila anni”.¹⁵⁷

Diversamente dalle precedenti grandi rivoluzioni giuridiche, ad esempio quella del codice civile o del costituzionalismo, il cambiamento digitale non va ascritto né ai giuristi, né al legislatore, che sembrano anzi subire questa mutazione, né tantomeno ai decisori pubblici. No, questa rivoluzione è portata avanti da giovani uomini d'affari, freschi di scuole di commercio, da matematici e *geeks*, i quali hanno visto nella giustizia un campo inesplorato, l'ultimo forse, che hanno deciso di coltivare con i propri modelli e la propria logica ben lontana dal diritto. Le start-up che si lanciano nella *legaltech* (tecnologie informatiche applicate al diritto) sono create da ingegneri e imprenditori in una consapevole logica di profitto.

¹⁵⁶ Op. cit. A. Garapon, *La giustizia digitale*, Il Mulino, Saggi, Milano, 2021.

¹⁵⁷ G. Bateson, *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution and Epistemology*, London, Intertext, 1972, p. 481.

Da qui un primo shock culturale: la giustizia predittiva e la *legaltech* sono nate dalla penetrazione di una logica capitalista nella professione di avvocato. Tale logica conduce alla subalternità del giurista allo strumento tecnologico e, ovviamente, al suo inventore. Tali soggetti sono i veri detentori del potere non solo giudiziario perché ne stabiliscono modalità e mezzi ma anche politico. La forza e la legittimità hanno cambiato schieramento: non stanno più nelle istituzioni ma in società private, per lo più californiane che siedono ormai nei più importanti tavoli mondiali.

IL digitale non solo apporta nuovi concetti, ma conferisce anche un nuovo statuto al concetto che ha valore solo se operativo ed effettivo. Si tratta quindi di una rivoluzione cognitiva che invaderà e già ha invaso sia l'ambito politico quanto quello delle scienze umane e che proviene dalle scienze dure:

Adottare il punto di vista operativo implica molto più che una semplice restrizione del senso in cui noi intendiamo i termine "concetto": implica un cambiamento decisivo in tutto il nostro modo di pensare, in quanto non ci permetteremo più di usare, come strumenti, concetti dei quali non possiamo renderci adeguatamente conto in termini di operazioni.¹⁵⁸

Il digitale, in quanto forma simbolica, ovvero pensato non solo come tecnica, ma come un nuovo modi di costituire e di istruire un certo tipo di fenomeni e di rapportarci a essi per agire, riconosce ormai solo un approccio operativo, pratico. Né è testimone la spettacolare promozione dell'effettività, soprattutto nel campo del diritto: mentre la politica

¹⁵⁸ P.W. Bridgman., *The Logic of Modern Physics*, New York, Macmillan, 1927, p. 31.

promette diritti che non si concretizzano mai del tutto, internet pretende di realizzarli immediatamente grazie alle sue piattaforme.

Questo approccio è chiaramente un approccio basato su “diritto e matematica” riprendendo l’espressione inglese *law and...* che sottolinea sia la continuità storica che la discontinuità epistemologica con gli altri movimenti che hanno messo sotto pressione il diritto nel corso degli ultimi due secoli. Dopo che il diritto ha cessato di fondarsi su una trascendenza religiosa o sul diritto naturale, ha dovuto cercare altrove i propri punti di riferimento, sia per fondare le proprie regole, sia per preservare la propria legittimità. Esso ha bisogno di cercare un sostegno esterno per essere meglio compreso, ma anche per accrescere la propria autorità. Questo polmone esterno di legittimità è stato da prima ricercato nella storia, nei costumi, nella società stessa e nei suoi rapporti di forza (*law and society*), poi nell’economia intesa come ragionamento per gestire la scarsità e spiegare le scelte razionali (*law and economics*), e perfino nella letteratura (*law and literature*). Oggi lo si ricerca nel paradigma determinista della prevedibilità, che deriva dalle scienze informatiche e statistiche.

La novità dell’approccio “diritto e matematica” si manifesta su più livelli. Innanzitutto, perché il diritto cerca il proprio punto di riferimento in una risorsa completamente astratta, che non ha più niente di umano. L’altra novità di tale approccio è di essere attivo e non solo esplicativo. La giustizia digitale agisce direttamente, moltiplicando gli automatismi, le previsioni e le correlazioni, che si impongono come vere regole di diritto. L’inestricabile collusione tra la forza espressiva e la performatività di tale approccio è imputabile alle proprietà stesse della scrittura digitale. Vi è un’omologia fra la rivoluzione grafica e i tipi di pratiche e di conoscenze che essa genera.

Il titolo di questo paragrafo parla di “ultra-soluzione”, cosa intendiamo con tale termine? Intendiamo il mito della delega alle macchine di tutte quella attività anzi soprattutto di quelle che presuppongono delle responsabilità in capo all’individuo che le pone in essere.

Tuttavia da più parti ci insegnano che non bisogna idolatrare la tecnica. Innanzi tutto perché non esiste una tecnica “pura”, dal momento che possiamo assimilarla solo grazie a un’attività simbolica già all’opera. Una tecnica non è mai priva di una qualche forma di precomprensione,¹⁵⁹ né di un’ordina da cui trae la propria efficacia. Noi percepiamo il digitale come una tecnica che sconvolge tutto, in quanto fredda e meccanica, mentre in realtà essa attiva al contempo dei nuovi miti e dei miti pure antichi ma senza sostenerli apertamente.¹⁶⁰ L’uso di queste tecnologie è più veloce della loro “messa in cultura”, poiché gli usi e la cultura non avanzano allo stesso passo. Quali sono i miti risvegliati dal digitale? Per esempio la credenza di un mondo preordinato dal calcolo prima di ogni intervento umano; la credenza in un prolungamento del corpo tramite la tecnica; la credenza in un superamento dell’umano; infine, la credenza di una possibile sostituzione del piano teleologico-politico, per auto-fondare la società. Il digitale si integra innanzi tutto con una riflessione sul gesto tecnico come prolungamento del corpo stesso, (tutti pensiamo al nostro cellulare come una sorta di prolungamento della nostra mano, peraltro

¹⁵⁹ Il concetto di “precomprensione” nell’ermeneutica indica la situazione in cui si trova l’interprete che si accinge a comprendere un testo, costituita da uno sfondo di giudizi antecedenti prodotti sia dalla memoria culturale della tradizione in cui è inserito, sia dal suo personale orizzonte.

¹⁶⁰ J. G. Ganascia., *Le Mythe de la Singularité. Faut-il craindre l’intelligence artificielle?* Paris, Seuil, 2017.

anche il design medesimo è fatto a tal fine), che trae la propria origine dal pensiero tedesco sulla tecnica fin dal XIX secolo.¹⁶¹

Questa rappresentazione di qualcosa che è nostro ma esterno al corpo, postula che le funzioni della mente siano esteriorizzabili in una materialità dotata di un modo di essere propriamente tecnico, ovvero operativo e funzionale. Essa si fonda su un presupposto dualistico, che ipotizza un trasferimento fra memoria individuale interna e memoria culturale esterna, e questo, indipendentemente dal supporto, che sia organico o silicio.

Tuttavia, alcuni studiosi, sostengono una tesi interessante. Essi pensano che ciò che noi classifichiamo come intelligenza artificiale o macchina in grado di giudicare, cioè qualcosa che alla fine è un nostro “prolungamento” ma comunque altro da noi, in realtà si tratta di una prospettiva falsata. Ciò che la rivoluzione tecnologica, e quindi l’uomo oggi desidera non è un dualismo uomo – macchina, bensì una sorta di super-uomo, cioè un uomo integrato da una macchina. Questa prospettiva rompe con l’interpretazione dualistica più sviluppata nella rappresentazione del digitale come una “cosa esterna ed artificiale”. Delegare o meno l’attività di giudizio a un algoritmo realizzabile da una macchina è quindi un’ importante posta in gioco sociale.

L’immaginario del prolungamento, se non addirittura del superamento dell’umano, tramite l’azione di dati che non avrebbero più bisogno di essere interpretati poiché degli algoritmi se ne farebbero carico una volta per tutte, in realtà integra a parere di chi scrive, una sorta di nuovo

¹⁶¹ E. Kapp, *Grundlinien einer Philosophie der Technik. Zur Entstehungsgeschichte der Kultur aus neuen Gesichtspunkten*, Braunschweig, Westermann, 1877, p. 41: “Le parti della mano, il suo palo, il suo pollice e le sue dita, l’apertura, il cavo, la distanza fra le dita, il fatto di girare, afferrare, stringere la mano sono, sia indipendentemente sia insieme con l’avambraccio teso o piegato, la matrice comune dello strumento che prende il nome di questa attività”.

mito, ossia quello della delega alle macchine come l'unica via di salvezza per l'umanità intera. Tale mito, e qui torniamo alle origini, non solo del digitale ma anche del nostro lavoro di tesi, comincia con Turing ossia con questa visione che ricorda la *science-fiction*, ossia una sorta di film di fantascienza che nel bene o nel male, spiega che le macchine prenderanno il sopravvento sull'essere umano sia che questi lo voglia o meno.

Se pensiamo, in effetti, tutti i discorsi sulla rivoluzione digitale fanno costantemente ricorso alla figura retorica della prolessi: si chiede oggi di credere a una realtà della delega alle macchine che però si realizzerà non prima di domani. Tale retorica incoraggia a creder sin da oggi in ciò che sarà effettivo solo domani, poiché il domani è meccanicamente ineluttabile. Si può osservare il carattere circolare del ragionamento: si tratta di credere davvero nella delega alle macchine, perché il carattere ineluttabile di tale delega è inscritto nella stessa delega alle macchine. La prolessi trasforma così il discorso sul digitale in oggetto di fede, sostituendo all'analisi dei presupposti una temporalità immaginaria, agli antipodi della rigorosa analisi scientifica.

In fondo le nostre società sono state pensate tramite una fondazione teologico-politica del potere. Dalla religione si è passati a forme più democratiche provenienti dalla politica e dalle istituzioni secolarizzate. La rivoluzione digitale permette di perseguire il processo di auto-produzione della società tramite questo nuovo modo di vincolare e di influenzare i comportamenti individuali, al pari della religione, del sacro, del potere politico e quindi del diritto,

In questo mito della delega alle macchine, il digitale è dotato di una dimensione ontologica: le correlazioni rivelano uno stato del mondo che occorre accettare così com'è con fiducia; o meglio, che si impone

come norma per valutare il mondo vissuto. Il digitale si configura così come un nuovo diritto naturale interno al mondo. Si pone come un nuovo referente in funzione del quale valutare tutte le istituzioni giuridiche, una misura comune, come il dizionario per la lingua. Questo nuovo diritto naturale mette in tensione tutte le istituzioni giuridiche. La sfida della giustizia digitale non è solo tecnica, ossia come “addomesticare” un nuovo strumento e come regolarlo, ma è soprattutto normativa, nella misura in cui il digitale introduce un criterio per valutare il diritto alla stessa stregua del vecchio diritto naturale, o del positivismo o addirittura del noto approccio economico al diritto.

La giustizia predittiva spesso viene visto come il sogno profetico di liberare i giudizi da tutte quelle storture e quei pregiudizi immessi dall'uomo. Al contrario quindi il giudizio degli uomini viene visto come un ripiego storico di cui ci si è dovuti accontentare, nell'attesa di poter utilizzare un referente veramente scientifico ed impeccabile. Questi algoritmi e i loro utilizzo rivelano dunque sia una de-simbolizzazione sia una re-simbolizzazione. Una de-simbolizzazione, perché il sapere algoritmico fa emergere la fragilità umana dei giudici, i limiti della giustizia “giuridica” e l'artificialità delle finzioni, tuttavia necessarie per la vita del diritto; e una re-simbolizzazione, cercando di sostituire all'imperfetta giustizia degli uomini una certezza scientifica assoluta, che non avrebbe più bisogno degli individui.

Il capo di una start-up si domanda in cosa consiste, in fondo, il lavoro di un giudice:

Migliaia di pagine di procedura da sintetizzare, rapporti con esperti da leggere, testi di legge e giurisprudenza da consultare, per emettere finalmente una sentenza. Un lavoro titanico che nessun cervello normale è in grado di portare

a termine perfettamente, a maggior ragione se posto in un contesto di sovraccarico cronico. Le sentenze giudiziarie vengono così emesse in un contesto di razionalità limitata. Questo lavoro diventerà gradualmente alla portata di un'intelligenza artificiale, instancabile ed in grado di evitare tutte le storture cognitive causate dal nostro cervello [...]. Se una macchina è in grado domani di “digerire” in pochi secondi l'insieme dei casi simili e dei giudizi passati, adattando in maniera estremamente accurata la propria decisione al caso specifico, non potremmo risparmiare pesanti procedure, ed eliminare anni di ritardo? [...]. In realtà, un'intelligenza digitale renderebbe giustizia meglio rispetto a una persona influenzata dalla propria ideologia e dal contesto del processo [...]. Nella scienza, il criterio cardine è la riproducibilità di un'esperienza; il criterio della giustizia non dovrebbe più precisamente essere quello della riproducibilità di un giudizio? Quest'ultimo non dovrebbe dipendere né dalla persona che giudica né dalle circostanze.¹⁶²

Ora anche ad ammettere che tale signore dica il vero, e non ne sono così sicura, tale tipo di giustizia per essere tale, deve avere una qualche forma di legittimazione, ossia essere voluta e riconosciuta dai consociati come una giustizia giusta.

Quindi possiamo dire che il più volte menzionato mito della delega alle macchine offre un' “ultra-soluzione” alle difficoltà della giustizia, ovvero soluzioni che non risolvono le difficoltà in senso stretto, ma cercano piuttosto di eliminarle. Il termine è stato forgiato dallo psicologo Paul Warzlawick della Scuola di Palo Alto (California), in riferimento al fallimento e ai conflitti familiari. L'ultra-soluzione della giustizia consiste nel porre rimedio ai mali costitutivi della giustizia in modo radicale, eliminando il fattore umano, Così l'uomo crea problemi

¹⁶² L. Alexander eo. Babeau., *Confions la justice a l'intelligence artificielle*, in “Les Echos”, 21 settembre 2016.

per via della sue passione, della sua dismisura e del desiderio di potenza? Affidiamo l'organizzazione della coesistenza umana a della macchine. La giustizia è sempre un po' azzardata in quanto basata su prove incomplete e argomentazioni sempre inclini all'inganno? Rimettiamoci al calcolo. Il potere è irrazionale e probabilmente tirannico? Facciamolo scomparire. La deliberazione è infinita e non si arriva mai a un calcolo? Automatizziamo la decisione. Le scelte degli uomini sono infinitamente diverse? Rimettiamoci alla capacità di calcolo che è un atto logico.

Ovviamente tali domande e risposte avevano lo scopo di rendere sarcastica la decisione di devolvere interamente la giustizia in mano a della macchine i quanto l'illusione che loro siano migliori di noi, sebbene sia molti a crederlo veramente, è per ora, solo un' illusione.

Volevo poi concludere il capitolo con un esempio che spero chiarisca al lettore la volontà, a parer mi ancora piuttosto inconsapevole, del legislatore, nazionale, comunitario e internazionale di rendere sempre più il diritto "*machine friendly*". Un esempio fra tutti è il cosiddetto *legal design* ossia il "disegnare" il diritto per far si che sso possa essere compreso da una macchina e anche per renderlo in un certo senso binario, che è poi il codice attraverso il quale un computer "parla" con un umano.

Un esempio è il concetto di *privacy by design* (esso è forse il più eclatante ma ve ne sono anche altri); con tale norma (sulla quale non mi dilungo a spiegare a cosa si riferisce perché abbiamo avuto modo di farlo nei primi capitolo del presente lavoro di tesi), il legislatore europeo ha voluto chiaramente creare una norma a partire non da cosa sia giusto o meno ma da come un computer, oggi, possa interpretare e rendere pratico il concetto. Infatti, dire che "la privacy deve essere

immessa nel software che si occupa dell'attività X sin dalla progettazione di fatto significa delegare a alla macchina (al suo programmatore) il livello di privacy che intende dargli, avuto riguardo a come quel software intendere il concetto.

Conclusioni

Nel corso della pagine precedenti abbiamo analizzato come la tecnologia stia cambiando quasi ogni aspetto della nostra vita e questo indipendentemente dalla nostra volontà. Come abbiamo avuto modo di descrivere in più occasioni ognuno di noi durante la giornata si imbatte in agenti artificiali o algoritmi più o meno sofisticati. Si pensi a quando decidiamo di comprare un biglietto aereo piuttosto che del treno ci interfacciamo con algoritmi che ci dicono orari, prezzi, se ci sono ancora posti disponibili nella carrozza che preferiamo e via discorrendo. Ciò fino ad arrivare al più “temuto” e sicuramente delicato utilizzo degli algoritmi nel campo della giustizia, sia esso come algoritmo di *predictive policing* o di *decision making* o *predictive justice*.

La domanda che vogliamo porci e alla quale cerchiamo qui nelle nostre conclusioni di dare una risposta è la seguente: l'algoritmo è il nuovo alchimista?

Finora un informatico ha sempre scritto del codice di programmazione, sotto forma di regole, per risolvere un problema, ma in un mondo complesso non si possono definire tutte le regole in anticipo perché sarebbero troppe ma anche perché sarebbe impossibile scrivere regole per scenari non ancora dati. Purtroppo, o anzi per

fortuna, il mondo non è una scacchiera. Pensiamo, ad esempio, alla guida autonoma e alla complessità della viabilità delle città: nessun set di regole predeterminate può far fronte ai continui cambiamenti dovuti alle interazioni con altri soggetti complessi, come gli automobilisti, i pedoni, gli animali ecc.

Gli algoritmi analizzano le relazioni nei dati, non i valori o il significato che rappresentano, e questo è un punto chiave fondamentale. Perciò l'algoritmo non "predice" e non "pensa", ma si limita a costruire modelli seguendo le nostre orme. In altri termini, l'algoritmo è un meccanismo produttivo che usa i nostri dati come materia prima: scova le correlazioni ed astrae le regole.

L'AI è quindi una creatrice di regole, seguendo le quali costruisce una sua rappresentazione del mondo. Ma lo fa in modo irresponsabile.

Tutto il lavoro di apprendimento culmina in un risultato che ha del misterioso, proprio come quello dell'alchimista. Gli alchimisti pensavano di poter trasformare gli elementi chimici a loro piacimento, senza rivelare il metodo, e per questo erano considerati alla stregua di maghi. Poi, nonostante l'avvento del metodo scientifico, è sopravvissuta l'idea esoterica che fossero in grado di raggiungere risultati incredibili e inspiegabili. Come l'algoritmo alla base del *deep learning* che, per ogni dato che analizza, modificando contemporaneamente milioni di parametri dà un risultato inspiegabile.

La logica delle poche righe di codice è molto semplice, ma quello che si trova nell'insieme dei dati e il modo in cui viene interpretato dalle reti neurali risulta inspiegabile. Avremmo dunque bisogno di un bugiardo per gli algoritmi, così come li abbiamo per i farmaci, al fine di conoscere posologia e controindicazioni.

Come abbiamo scritto sopra e ribadiremo in breve ora, l'algoritmo ha un suo modo di calcolare e ci sono rischi che non prende in considerazione.

L'AI può impiegare infatti alcune tecniche che rendono opaco il suo funzionamento: per questo si parla di "scatola nera", perché accanto ai risultati non sono fornite spiegazioni. Quando leggiamo di una particolare implementazione dell'AI chiamata *deep learning*, si intende, come detto, una rete di migliaia di neuroni simulati (nel senso che sono ispirati al funzionamento di un neurone biologico, ma ne costituiscono un'estrema semplificazione), disposti in strati interconnessi. I neuroni nel primo strato ricevono ciascuno un *input* ed eseguono un calcolo prima di emettere un nuovo segnale. Questi *output* alimentano, in una rete complessa, i neuroni del livello successivo e così via, fino a quando non viene prodotto un *output* complessivo.

Senza inoltrarci nuovamente nei dettagli di questo metodo di apprendimento, basti dire che da un punto di vista matematico, il *deep learning* è del tutto trasparente: consiste in una serie di moltiplicazioni di matrici a cui si applicano funzioni non lineari. Tuttavia, il suo risultato finale è probabilistico, nel senso che non si può essere certi che il risultato sia corretto.

Oltre alle ragioni tecniche, dietro la *black box* ci sono anche ragioni e problematiche politiche. La scatola può infatti risultare nera non solo per l'inumana complessità dei calcoli, ma anche perché talvolta viene appositamente opacizzata, così da proteggere il segreto e dunque il valore commerciale. Il pubblico spesso è tenuto all'oscuro perché altrimenti si svelerebbero le scelte controverse, o la semplicità di

calcolo dietro ciò che viene venduto come un “sistema di Intelligenza Artificiale”.

Anche se lo strumento lavora in modo oscuro, noi abbiamo il dovere di scegliere il fine, che deve essere sempre trasparente.

L’interazione ciclica tra la macchina e l’intelligenza umana determina ciò su cui ci concentriamo, le storie e le immagini che vediamo, i suggerimenti e gli annunci mostrati accanto a qualsiasi contenuto ci venga sottoposto sui *social*.

Come farà l’uomo per non essere disumanizzato dalla macchina, per dominarla, per renderla moralmente arma di progresso?, si chiedeva Giuseppe Ungaretti nel 1953. Abbiamo bisogno di una nuova etologia per studiare il comportamento delle AI, e di una nuova filosofia per approfondire la nostra e la loro evoluzione.

L’età moderna non sarebbe nata senza le spiegazioni scientifiche dei fatti e senza la loro diffusione per mezzo della stampa. Da allora in poi, il metodo scientifico e il confronto tra esperti hanno abbondantemente rimpiazzato la fede come principale criterio della conoscenza umana. L’illuminismo, che poneva la ragione umana alla base di ogni conoscenza, ha dato vita al processo di secolarizzazione e sono queste le fondamenta su cui abbiamo vissuto finora.

Ora però in rete si diffondono fatti non spiegabili scientificamente, rilanciati da macchine autonome (bot) basate su algoritmi che prevedono il comportamento umano. Definiamo *fake news* queste notizie prive di alcuna valenza scientifica, come facilmente possiamo verificare provando a risalire alle loro fonti. Ma come faremo a riconoscere le fonti, se a generarle fossero macchine autonome? Questa è un’altra ragione per cui, se non possiamo più fidarci della conoscenza empirica su quello che accade alla macchina, non è

corretto parlare di Intelligenza Artificiale, perché si tratta solo di un poderoso calcolo numerico.

Ora che l'AI è integrata nei servizi pubblici essenziali e in altri processi ad alto rischio, come, appunto, giustizia e sanità, dobbiamo prestare particolare attenzione ai suoi risultati e alle sue conseguenze. Abbiamo accennato alla scatola nera dentro la quale avvengono calcoli non spiegabili, essa viene definita dall'Information commissioner's office (Ico) del Regno Unito e dall'Alan Turing institute come "qualsiasi sistema di intelligenza artificiale i cui meccanismi interni e logica sono protetti perché proprietari e/o inaccessibili alla comprensione umana. Ci sono organizzazioni senza scopo di lucro, come l'Ong tedesca AlgorithmWatch, che si pongono l'obiettivo di far luce sui processi decisionali algoritmici che hanno rilevanza sociale. AlgorithmWatch analizza gli effetti di questi processi sul comportamento umano al fine di massimizzare i benefici per la società. La no profit aiuta a sviluppare idee e strategie per raggiungere l'intellegibilità degli algoritmi grazie a un mix di tecnologie, regolamentazione e istituzioni di supervisione.

Se i dati prodotti dalle macchine si moltiplicano geometricamente, noi abbiamo bisogno, di nuovo, di altre macchine per estrarre il senso, così da ottenere conoscenza di quello che sta accadendo ed eventualmente predire quello che potrebbe accadere. Noi tutti, con l'aiuto delle macchine, sappiamo che la strada per il futuro non è prestabilita deterministicamente.

A tal proposito Alessandro Vespignani evidenzia che:

Un altro elemento importante da considerare è che i metodi predittivi basati su apprendimento automatico e Intelligenza Artificiale

assumono implicitamente che il nostro futuro è il passato. Ovvero, dato che le predizioni sono basate sullo studio di associazione tra dati del passato, il futuro deve essere già scritto in qualche modo nel passato [...]

Ma nel caso delle predizioni influenzali quali benefici posso avere dal dare in pasto al computer i dati influenzali di venti o trenta anni fa? Questi dati si riferiscono a popolazioni diverse, sistemi di rilevamento che sono cambiati, a protocolli ospedalieri e sistemi di salute pubblica completamente diversi da quelli attuali. Il rischio è di peggiorare le predizioni, perché il passato che viene analizzato non ha niente a che vedere con il presente e il futuro.

E' un po' come guidare guardando nello specchietto retrovisore. L'AI è una disciplina utile e pericolosa perché se non si segue la prassi del *continuous training and testing*, un giorno potrebbe non funzionare più, o meglio fornirci previsioni errate, e non saremo nemmeno così addestrati per rendercene conto in tempo. Se c'è un errore negli algoritmi procedurali la colpa è del programmatore, ed è una colpa che bisogna assumere come dato probabile posto che il programmatore è un essere umano. Invece, se c'è un errore nel *deep learning* questo è riferibile alla probabilità statistica ed è normale che succeda.

A ben vedere, ciò che accade dentro una scatola nera è un paradosso perché le previsioni sono verificabili ma non è possibile comprendere come siano state ottenute.

Pongo qui al lettore un ultimo schema redatto dalla Royal Society che descrive quali debbono essere, a parer loro, le caratteristiche per avere un'AI che sia la migliore possibile:

- *Interoperabile*, il che implica una qualche comprensione di come funziona la tecnologia;
- *Spiegabile*, così che un'ampia gamma di utenti possa capire perché o come la conclusione è stata raggiunta;
- *Trasparente*, ossia che abbia un certo livello di accessibilità ai dati e all'algoritmo;
- *Giustificabile*, nel senso che esiste una comprensione del caso a sostegno di un certo risultato;
- *Contestabile*, che presupponga cioè il possesso da parte degli utenti delle informazioni necessarie ad opporsi a una decisione o a una classificazione

Un esempio può riguardare proprio la giustizia: perché si è deciso per un periodo di detenzione più lungo? Perché è stata negata la condizionale?

Può essere utile aggiungere qualche considerazione riguardo alle previsioni delle macchine.

L'*interpretabilità* riguarda la misura in cui si è in grado di prevedere cosa accadrà, dato un cambiamento nei parametri di *input*. In questo modo possiamo capire quali fattori o dati siano stato decisivi nella produzione del risultato che ha avuto un certo impatto su un individuo.

La *spiegabilità* è la misura in cui l'architettura interna di una macchina può essere spiegata in termini umani. Con i fenomeni fisici e naturali, siamo abituati prima a usare gli strumenti e poi a comprenderne le leggi di funzionamento. Oggi persino gli esperti che impiegano i *deep learning* fanno fatica a comprenderne il

funzionamento. L'AI presenta alcune caratteristiche che non riusciamo a vedere o percepire. Pensiamo all'olfatto dei cani: tanto il loro fiuto è sensibile agli odori, e per questo ci risulta utile, tanto è "sensibile" l'AI quando si tratta di scovare correlazioni.

In termini di *trasparenza* poi, se usiamo un sistema non trasparente andiamo incontro a un cambiamento di natura filosofica: sappiamo cosa volevamo, ma non siamo certi di cosa abbiamo ottenuto. Questi perché tende a prevalere il modello per cui è meglio progettare reti di apprendimento e utilizzare enormi quantità di potenza del computer, invece di provare a definire il modo di aumentare la conoscenza umana del processo di decisioni algoritmica. In altri termini, essendo faticoso descrivere le regole e le relazioni tra gli oggetti e le macchine, spesso si prende la strada dell'apprendimento automatico, purtroppo senza la necessaria riflessione.

La riflessione è necessariamente filosofica, considerando che i due ambiti, tecnico e filosofico, si intersecano sempre di più quando si tratta di AI.

Possiamo riassumere i rischi in tre macrocategorie: l'abuso di sorveglianza, e pensiamo al riconoscimento facciale o alla localizzazione; il rafforzamento di pregiudizi, con previsioni o classificazioni errate; la creazione di una realtà alternativa basata su informazioni o immagini false.

Già il mondo è in parte inspiegabile, giacché non conosciamo tutte le leggi naturali ma ora lo stiamo rendendo ancora più difficile da spiegare. E ciò avviene per mano nostra, una mano artificiale. Eppure, a volte lasciamo che si creda che l'AI abbia un potere salvifico, e che porti necessariamente a grandi sviluppi. Chi non approfondisce il tema

pensa che si possa usare per tutto. Con un certa sicumera, che va smontata.

Come sottolineato numerose volte in precedenza con l'AI si rischia "l'allucinazione statistica".

L'AI si concentra infatti sulla correlazione tra dati e sull'automazione della previsione, non sul rapporto causa-effetto. In proposito c'è un libro di Judea Pearl, uno degli scienziati più importanti nel campo dell'AI e Dana Mackenzie, in cui si esprime una critica alla direzione intrapresa dalla maggioranza degli scienziati, che trascurano il rapporto causa-effetto concentrandosi solo sul miglioramento della correlazione.

Per semplificare, nei suoi studi Pearl introduce il concetto di inferenza causale e postula che oltre ad affrontare le incertezze, i metodi di apprendimento automatico dovrebbero anche, implicitamente o esplicitamente, utilizzare il ragionamento causale. Ma un sistema simile non è ancora disponibile.

Se e quando svilupperemo algoritmi basati sul rapporto causa-effetto, allora non conteremo più solo sui numeri ma anche sulla logica e l'eliminazione di parte dei *bias*.

Per quanto concerne poi gli algoritmi che prendono decisioni talvolta al posto degli umani, la questione si complica. La decisione è un'arte compresa tra essere e divenire, tra coscienza e realtà fisica. Per questo passiamo dal senno del poi alla lungimiranza anticipata. Anche se talvolta decidiamo di non decidere, la bontà di una decisione è quasi sempre associata al senno del poi.

Per le decisioni critiche, quelle che toccano la sfera della libertà e della salute, finora ci siamo sempre affidati a esperti umani per tre ragioni fondamentali: l'accuratezza, perché abbiamo la capacità di

ridurre al minimo gli errori gravi; la fiducia, perché dobbiamo fidarci del ragionamento che porta alla decisione; la responsabilità morale, perché siamo moralmente responsabili delle nostre decisioni, o se vogliamo vederlo dall'altro lato, perché chi decide si assume anche la responsabilità delle decisioni prese.

Tuttavia oggi ad essere rilevanti non sono più le risposte bensì le domande. E' evidente che le macchine non hanno l'abilità necessaria a rispondere e rispondono sulla base delle domande che noi poniamo loro.

Bertand Russel, scrisse che: “ Tutta la conoscenza umana è incerta, inesatta e parziale”.

L'AI non è consapevole di questi nostri limiti e ovviamente nemmeno dei suoi. Stiamo essenzialmente delegando la nostra fiducia a un'entità che, come abbiamo capito, non è assolutamente in grado di comprendere quello che comporta ogni sua azione e che non può essere responsabile. Stiamo cercando, attraverso le macchine, di deresponsabilizzarci, quando invece dovremmo essere più responsabili per evitare rischi sistemici e imprevedibili. Alcune cause di questo comportamento, a parer mio, sono storiche: le persone connesse in rete sono lontane nello spazio, mentre i tempi di comunicazione si accorciano; altre sono tecniche ed economiche: la facilità di ottenere risposte su tutto in tempo reale; altre ancora sociali: non ci fidiamo più di nessuno e nemmeno di noi stessi.

Se questo è vero, allora stiamo cambiando il mondo alla cieca: siamo ciechi davanti alla mole di dati, ciechi di fronte ai pregiudizi contenuti in dati e modelli, ciechi perché non sappiamo dove stiamo andando. Eppure, ci stiamo affidando alle macchine più che agli esperti umani dei vari settori e pensiamo che l'AI sia agnostica. Il problema poi è

che fidandoci ci comportiamo come se il mondo fosse un dominio lineare, mentre non è così. Un dominio lineare è prevedibile e ha un basso grado di interazione tra gli elementi che contiene; in un dominio complesso, invece, i legami causali sono difficilmente visibili e la prevedibilità estremamente bassa. In ingegneria e in fisica, alcuni fenomeni si possono spiegare e prevedere nel lungo termine con semplici leggi lineari, mentre altri fenomeni sono intrinsecamente caotici e non si possono prevedere nel lungo termine sulla base di pochi elementi. In maniera simile, in psicologia e economia ci sono fattori e cause che semplicemente non capitano, o che hanno così alti livelli di interdipendenza che le previsioni ex ante sono inaccessibili dal punto di vista computazionale e quindi le previsioni diventano disponibili solo ex post.

Dovremmo prendere le decisioni anche raccogliendo le risposte dalle macchine, non con dati provati e non con la fede, perché la fede non chiede prove. Ecco che cambia il modo in cui noi prendiamo le decisioni.

La perdita di conoscenza e di autonomia fanno parte di un processo iniziato nel XX secolo, nel corso del quale stiamo invertendo il rapporto gerarchico tra noi e le macchine. Oggi siamo più portati a mettere in dubbio la risposta a una domanda dataci da una persona oppure quella di un assistente virtuale? A parte i pochi esperti che conoscono tecnicamente come funziona un assistente virtuale e qualche ipocrita risponderemo tutti che ci fidiamo di più di Alexa. Molti esperti sostengono che sarebbe certamente meglio far risolvere all'AI i problemi deterministici e quelli probabilistici e lasciare a noi tutto il resto, ossia decidere in processi nuovi o incerti oppure quando servono astrazione, intuizione e creatività.

Allo stato della tecnica concordo con lor, non so se lo farò in futuro.

EPILOGO

Mentre mi accingevo a scrivere le conclusioni della mia tesi dottorale e proprio mentre definivo limitata l'AI e soprattutto limitata in quei settori in cui prevale il sentimento, l'intuizione, la coscienza, mi sono imbattuta, in modo piuttosto casuale, in un articolo, pubblicato sulla rivista di informatica giuridica della collana del CIRSFID, facente parte di un più ampio lavoro dal titolo "Nuove questioni di informatica forense" che titolava: "*Deep learning* per l'analisi del *sentiment*" a cura di Corrado Federici.

L'analisi del sentiment (AS) è una specialità dell'elaborazione del linguaggio naturale orientata a valutare lo stato d'animo dell'autore di un testo, spesso scritto in risposta ad una precedente pubblicazione. I casi tipici che vengono presi ad esempio sono i commenti in calce a un articolo apparso su un blog o a un post presente nella bacheca di un profilo Facebook.

L'autore spiega che i campi di applicazione di AS sono praticamente illimitati, potendo spaziare dal marketing per stabilire il posizionamento di un marchio, alla politica per valutare l'orientamento dell'elettorato su specifici argomenti, alla business intelligence per suggerire i prodotti più aderenti alle esigenze del cliente, fino alle scienze mediche per permettere alle strutture ospedaliere di fornire servizi migliori e non da ultimo l'utilizzo più smodato, allo stato si ha nel settore della pubblica sicurezza.

L'intento di questo epilogo non è chiaramente quello di contraddirmi né di esaltare gli usi dell'AI e del *machine learning* ma semplicemente quello di confermare ancora una volta una visione, che credo il lettore di questo mio lavoro abbia inteso, che sia critica verso la tecnologia ma non demonizzante e tanto meno sicura di quello che avverrà in un futuro prossimo.

E' stata troppo repentina l'evoluzione della tecnologia negli ultimi 20 anni per provare anche solo a prevedere o mettere paletti ideologici circa quello che sarà magari tra un anno.

L'unica cosa che penso invece possa e debba essere prevista è la finalità del suo utilizzo.

Bibliografia

AA. VV., *Nuove questioni di informatica forense*, Informatica Giuridica, Collana del CIRSFID, 2022.

AA. VV., *Le nuove tecnologie. Fattori criminogeni e risposte sociali*. Milano, 1997.

AGID (Agenzia per l'Italia digitale), *Libro bianco "IA al servizi del cittadino"*.

AGRAWAL, A. K., GOLDFARB, A., *Prediction Machines*, Harward Business School Press, Brighton, 2018.

ALBERTINI, L., *I contratti di accesso ad Internet*, in Giust. Civ., 1997, n. 2, p. 96.

ALBERTINI, L., *Osservazioni sulla conclusione del contratto tramite computer e sull'accettazione di un'offerta in Internet*, in Giust. Civ., 1997, n. 2, p. 21.

ALLEN, G., VERNER, G., e ZINSER, J., *Prolegomena to Any Future Artificial Moral Agent*, in "Journal of Experimental and Theoretical Artificial Intelligence", 2000, 12, pp. 251-261

AJANI, G., EBERS, M., *Uniform Terminology for European Contract Law*, Nomos, Baden-Baden 2005.

ALPA, G., *Computers e responsabilità civile.*, Milano, 1985.

ALPA, G., *Contratti di massa, profili generali*, voce in Enc. Dir., vol. I, Milano, 1997.

ALPA, G., *Cyber law. Problemi giuridici connessi allo sviluppo di Internet*, in Nuova giur. civ. comm., 1998, n. 2, pp. 385 ss.

ALPA, G., *Diritto della responsabilità civile*, Roma-Bari, 2003.

ALPA, G., *I diritti degli utenti nei contratti di assicurazione conclusi mediante Internet*, in Contratti, 2000, n. 12, pp. 1168 ss.

ALPA, G., *Istituzioni di diritto privato*, Torino, 1997.

- ALPA, G., *La Responsabilità civile*, in Trattato di dir. civ., Milano, 1999, n. 4, pp. 447 ss.
- AMOR, D., *E-Business. Vivere e lavorare in un mondo interconnesso*, Milano, 2000.
- ANTOLISEI, F., *Manuale di diritto penale. Parte generale*, Milano, 2003.
- ANTOLISEI, F., *Manuale di diritto penale. Parte speciale. Reati contro la persona*, Milano, 2003.
- ANTONUCCI, A., *E-Commerce. La direttiva 2000/31/CE e il quadro normativo della Rete*, Milano, 2001.
- ARENA, G., *La tutela della riservatezza nella società dell'informazione*, in Scritti onore di Pietro Virga, vol. I, 1994, N.1, pp. 63 ss.
- ASIMOV, I., *Io, robot*, Mondadori, Cles, 1975.
- ATELLI, M., *Il diritto alla tranquillità individuale, Dalla Rete Internet al "door to door"*, Napoli, 2001.
- BALDWIN, R., *The Great Convergence: Information Technology and the New Globalization*, The Belknap Press of Harvard University Press, Cambridge, 2016.
- BALLARINO, T., *Internet nel mondo della legge*, Padova, 1988.
- BARABASI, A.L., *Link. La scienza delle reti*, Einaudi, Milano, 2004.
- BARBARO, C., *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?*, *Questione Giustizia*, Fascicolo 4/2018.
- BARCA, A., *Il commercio elettronico e la conclusione del contratto on-line: il d.lgs. n. 70/2003. Aspetti giuridici e risvolti pratici della contrattazione digitale a seguito del d.lgs. n. 70/2003*, in *Diritto e diritti*, 2010.

- BARICCO, A., *The Game*, Einaudi Editore, 2018.
- BATTISTA, P., *La vita privata non esiste più*, in “Corriere della sera”, 22 luglio 2019.
- BATTISTON, R., *L’impatto dell’AI sulla società*, BioLaw Journal, n. I, 2020, pp.473-475.
- BAUMAN, Z., *La società dell’incertezza*, Bologna, 1999.
- BAUMAN, Z., *Modernità liquida*, Roma-Bari, Laterza, 2002.
- BEITZ, C., *The ideo of Human Rights*, Oxford University Press, Oxford, 2019.
- BEKEY, G.A., *Autonomous Robots: From Biological Inspiration to Implementation and Control*. The Mit Press, Cambridge, Mass., Londra, 2005.
- BERLIN, I., *Quattro saggi sulla libertà*, (1969), tr. It. Feltrinelli, Milano, 1989.
- BENEDETTI, G., *Tutela del consumatore ed autonomia contrattuale*, in Riv. Dir. Civ., 1998, n. 1, p. 17.
- BENTIVEGNA, S., *A colpi d tweet*, Il Mulino, Bologna, 2015.
- BENTIVEGNA, S., *Campagne elettorali in rete*, Laterza, Roma-Bari, 2006.
- BENTIVEGNA, S., *Politica e nuove tecnologie della comunicazione*, Laterza, Roma-Bari, 2002.
- BERNERS-LEE, T., *L’architettura del nuovo web. Dall’inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, tr. It. Milano, 2001.
- BIANCA, C.M., in *Diritto civile*, Vol. 5, La responsabilità, Giuffrè, Milano, 2012.

- BIANCA, C.M. – BUSINELLI, F.D., (a cura), *Tutela della privacy. Commentario alla l. 31 dicembre 1996 n. 675*, in *Le nuove leggi civili commentate*, 1999, p. 228.
- BISI, S., *Identificabilità delle persone sulla rete Internet*, in *Nexa Polito*, 2011.
- BLENGINO, C., *I reati informatici*, in M. Durante e U. Pagallo, *Manuale di informatica giuridica*, pp. 219-245, Giuffrè, 2011.
- BRIDGMAN, P.W., *The Logic of Modern Physics*, New York, Macmillan, 1927, p. 31.
- BOBBIO, N., *L'età dei diritti*, Torino, 1997.
- BOBBIO, N., *La certezza del diritto è un mito*, in “*Rivista internazionale di Filosofia del diritto*”, 28, 1951, pp.146-152.
- BOCCHINI, F. e QUADRE, E., *Diritto privato*, quarta edizione, Giappichelli, Torino, 2011.
- BOLOGNINI, L. – PAGANINI., *Libertà di Internet e reati: sì all'anonimato protetto*, in *Corriere.it*.
- BOLOGNINI, L.-FULCO, D.- PAGANINI, P., (a cura), *Next privacy. Il futuro dei nostri dati nell'era digitale*, Milano, 2010.
- BONCINELLI, E., *Capire la mente*, Castelvechi, Roma, 2017.
- BRADFORD, A., *The Brussels Effect*, in “*Northwestern University Law Review*”, 2012, Vol.107(1), pp. 1-68.
- BRANCA, G., *Sulla responsabilità oggettiva per danni cagionati da animali*, in *Riv. Trim. Dir. Proc. Civ.* 1950, p. 225.
- BREUKER, J., (a cura di), *Law, Ontologies and the Semantic Web. Channeling the Legal Information Flood*, IOS Press, Amsterdam, 2009.
- BUSCEMA, M., *L'arte della previsione. Intervista sull'intelligenza artificiale*”, *letture.org*.

- BUSNELLI, F.D., voce *Illecito civile*, in Enc. Giur. Treccani, vol. XV, Roma, 1989.
- BUTLER, S., *Darwin Among the Machines*, The Press, 13 giugno 1863.
- BUTTARELLI, G., *Banche dati e tutela della riservatezza. La privacy nella società dell'informazione*, Milano, 1997.
- CALAMANDREI, P., *Discorso sulla costituzione agli studenti milanesi*, 26 gennaio 1955.
- CALAMANDREI, P., *Processo e democrazia*, Cedam, Padova, 1954.
- CARDON, D., *Che cosa sognano gli algoritmi*, Mondadori Università, Milano, 2016.
- CARLEO, A., *Decisione robotica*, Il Mulino, Bologna, 2019.
- CARTABIA, M., *The multilevel protection of fundamental rights in Europe: the European Pluralism and the need for a judicial dialogue*, in *The protection of fundamental rights in Europe: lesson from Canada*, Casonato, C., (a cura di) in *Quaderni del Dipartimento di Scienze Giuridiche*, Università degli Studi di Trento, Trento, 2003.
- CASA, F., *Dalle scienze cognitive alle applicazioni giuridiche dell'intelligenza artificiale*, in U. Pagallo (a cura di), *Prolegomeni d'informatica giuridica*, Cedam, Padova, 2003, pp. 63-131.
- CASANOVA, P. PAGALLO U. PALMIRANI, M. SARTOR, G., (a cura di), *AI Approaches to the Complexity of Legal System. Law, Social Intelligence, mMAS and the Semantic Web*, Springer, Berlin-Heidelberg, 2014.
- CASSESE, S., *La crisi dello stato*, Laterza, Roma-Bari, 2002.
- CASTELLI, C., Presidente della Corte d'Appello di Brescia in un'intervista fatta da Claudia Morelli, in *Altalex.com*, 2017.
- CASTELLI, C., e PIANA, D., *Giusto processo e intelligenza artificiale*, Maggioli Editore, Santarcangelo di Romagna, 2019.

- CANZIO, G., *Nomofilachia e diritto giurisprudenziale*, in *Contratto e impresa*, n. 2, 2017.
- CASTELLI, C. – PIANA, D., *Giustizia predittiva. La qualità della giustizia in due tempi* in *Questione Giustizia*, maggio 2018, consultabile sul sito www.questionegiustizia.it.
- CASTELNUOVO, D., *Clausole generali e prevedibilità delle norme penali*, *Questione Giustizia*, Fascicolo 4/2018.
- CAVOUKIAN, A., *Privacy by Design: The Definitive Workshop*, in “Identity in the Information Society”, 2010, Vol. 3(2), pp. 247-251.
- CHAITIN, G., *Teoria algoritmica della complessità*, tr. It. Giappichelli, Torino, 2006.
- CHIRIATTI, M., *Incoscienza artificiale, come fanno le macchine a prevedere per noi*, Luiss University Press, 2021.
- CHIRIATTI, M., *#Humanless. L’algoritmo egoista*, Hoepli, Milano, 2019.
- CHIUSI, F., *Vivere con gli algoritmi*, Valigia Blu, 11 novembre 2020.
- CIACCI, G., BUONUOMO, G., *Profili di informatica giuridica*, Cedam, Padova, 2018.
- CIVININI, M.G., *Il valore del precedente nella giurisprudenza della Corte Europea dei diritti dell’uomo*, *Questione Giustizia*, Fascicolo 4/2018.
- CRISCI, S., *Intelligenza artificiale ed etica dell’algoritmo*, *Foro Amministrativo (II)*, fasc. 10, 1 ottobre 2018.
- CORASANITI, G., *Esperienza giuridica e sicurezza informatica*, Giuffrè, Milano, 2003.
- COSTA, C., *Intelligenza artificiale e giustizia: tempi ancora prematuri per l’applicazione di algoritmi predittivi nei tribunali*, in “AI 4 Business” .

- COSTANZI, C., *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, *Questione Giustizia*, Fascicolo 4/2018.
- CUCCHIARA, R., *Giustizia: l'Intelligenza Artificiale è un'opportunità, basta fare sistema*, Intervista di Claudia Morelli, in “Altaex”, 12 novembre 2018.
- CURZIO, P., *Il giudice e il precedente*, *Questione Giustizia*, Fascicolo 4/2018.
- DAL LAGO, A., *Populismo digitale*, Raffaello Cortina, Milano, 2017.
- DAMASIO, A., *L'errore di Cartesio*, Adelphi, Milano, 1995.
- DANFORTH ZERONDA, N., *Street shooting: covert photography and ublic privacy*, in “Vanderbilt Law Review”2010, vol. 63, pp. 1131-1159.
- DANZIGER, S., LEVAV, J., AVNAIM-PESS, L., *Extraneous factors in judicial decisions*, in PNAS, vol. 108, n. 17, 2011.
- DE BIASE, L., *Il lavoro del futuro*, Editore Codice, Torino, 2018.
- DE FELICE, M., *Decisione robotica negoziale. Nuovi “punti di presa” sul futuro*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019.
- DE RENZIS, L., *Primi passi nel mondo della giustizia “high tech”: la decisione in un corpo a corpo virtuale tra tecnologia e umanità*, contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019.
- DEBORD, G., *La società dello spettacolo*, Massari Editore, 2002.
- DELLA MORTE, G., *Big data e protezione internazionale dei diritti umani*, Editoriale Scientifica, Napoli, 2018.
- DENNET, D., *Kings of Minds: Towards an Understanding of Consciousness*, Basic Book, New York, 1996.
- DENNET, D., *The illusion of consciousness*, TED Conferences, 2003

DI GIORGI, R.M., *L'intelligenza artificiale: teoria e applicazioni del diritto*, in R. Borruso, Di Giorgi, R.M. Mattioli, L. Ragona, M., *L'informatica del diritto*, Giuffrè, Milano, 2004.

DI GIOVINE, O., *Salti mentali. Analogia e interpretazione del diritto penale*, *Questione Giustizia*, Fascicolo 4/2018.

DI STASIO, G., *Machine learning e reti neurali nel diritto civile, applicazione del machine learning a casi di diritto condominiale*, *Rivista online I-LEX*.

DOMINGOS, P., *The master algorithm: how the quest for the ultimate machine learning will remake our world*. New York, Basic Books, 2015.

DOMINGOS, P., *L'algoritmo definitivo*, Bollati Boringhieri, Torino, 2016.

DONINI, M., *Fattispecie e case law? La "prevedibilità del diritto" e i limiti alla dissoluzione della legge penale nella giurisprudenza*, *Questione Giustizia*, Fascicolo 4/2018.

D'ORAZIO, R. – ZENO ZENCOVICH, V., *Profili di responsabilità contrattuale e aquiliana nella fornitura di servizi telematici*, in *Dir. Inf. e infor.*, 1990, n. 1, pp. 421 ss.

DREEKE, R., *Il codice del profiler*, Sperling & Kupfer, 2021.

DUHIGG, C., *O poder do habito*, Sao Paulo, objetiva, 2012.

DURANTE, M., *Il futuro del web: etica, diritto, decentramento*, Torino, 2007.

DURANTE, M., *What is the Model of Trust for Multi-agent Systems? Whether or not E-Trust Applies to Autonomous Agents*, *Knowledge, Technology & Policy*, 2010

DURANTE, M., *Normativity, Constructinism and Constraining Affordances*, in "Ethics and Politics", 2011, XIII (2): 180-200.

DURANTE, M., *L'intelligenza artificiale*, in Durante, M., Pagallo, U., (a cura di), *Manuale d'informatica giuridica e diritto delle nuove tecnologie*, UTET, Torino, 2012.

DURANTE, M. *Deadling with Legal Conflicts in the Information Society. An Informational Understanding of Balancing Competing Interests*, in "Philosophy & Technology", 2013, 26(4): 437-457.

DWORKING, R., *L'impero del diritto*, tr. It. Milano, Il Saggiatore, 1986.

EASTERBROOK, E., *Cyberspace and the Law of the Horse*, in "University of Chicago Legal Forum", 1996.

ETZIONI, A., *How Patriotic is the Patriot Act? Freedom versus Security in the Age of Terrorism*, Routledge, New York-Londra, 2004.

FABRIS, A., *Etica per le tecnologie dell'informazione e della comunicazione*, Tr. It. Carocci, Roma, 2018.

FAGGIN, F., *Sarà possibile fare un computer consapevole?*, Mondo digitale, n. 61, dicembre 2015.

FERRARI, I. – BECKER, D., *Algoritmo e preconetto*. Jota, disponibile su www.jota.info.br.

FERRARIS, M., *Postverità e altri enigmi*, il Mulino, Bologna, 2017.

FLANAGAN, M., HOWE D. C., HISSENBAUM, H., *Emboding values in technology: theory and practice*, in J. van den Hoven and J. Weckert (a cura di) *Information Technology and moral philosophy*, pp. 322-353, Cambridge University Press, New York, 2008.

FLORIDI, L., *On the Intrinsic Value of Information Objects and the Infosphere*, in "Ethics and Information Technology", 2003, 4 (4), pp. 287-304.

FLORIDI, L., *Four Challenges for a Theory of Informational Privacy*, in "Ethics and Information Technology", 2006, 8.3, 109-119.

- FLORIDI, L., *The Method of Levels of Abstraction*, in “Minds and Machines”, 2008, vol.18 (3), pp. 303-329.
- FLORIDI, L., *La rivoluzione dell'informazione*, tr. It. Codice, Torino, 2012.
- FLORIDI, L., *Infosfera. Etica e filosofia nell'età dell'informazione*, Giappichelli, Torino, 2009.
- FLORIDI, L., *Philosophy of computing and information. 5 Questions*, Automatic Press/VIP, 2008.
- FLORIDI, L., e TADDEO, M., *The Ethics of Information Warfare*, Springer, Dordrecht, 2014.
- FLORIDI, L., CABITZA, F., *Intelligenza artificiale, l'uso delle nuove macchine*, Bompiani Editore, 2021.
- FLORIDI, L., *Etica dell'intelligenza artificiale, sviluppi, opportunità, sfide*, Raffaello Cortina Editore, 2022.
- FIDELBO, G., *Il precedente nel rapporto tre sezione unite e sezioni semplici: l'esperienza della Cassazione penale*, Questione Giustizia, Fascicolo 4/2018.
- FINN, E., *Che cosa vogliono gli algoritmi. L'immaginazione nell'era dei computer*, Einaudi, Torino, 2017.
- FORZA, A., MENEGON, G., RUMIATI, R., *Il giudice emotivo. La decisione tra ragione ed emozione*, Mulino, Bologna, 2017.
- FRANCESCHELLI, V., *Computer e diritto: convenzioni internazionali, legislazione, giurisprudenza e modelli contrattuali*, Maggioli, Rimini, 1989.
- FRIDMAN, J., *The problem with prediction*, aeon.co, 25 gennaio 2021.
- GABORIAU, S., *Libertà e umanità del giudice: due valori cardinali della giustizia. La giustizia digitale permetterà la fedeltà a questi*

- valori? *Qualche riflessione a partire dall'esperienza francese*, *Questione Giustizia*, Fascicolo 4/2018.
- GAGGI, M., *L'immoralità degli algoritmi*, in “*la Lettura*”, in “*Il Corriere della sera*”, 7 aprile 2019.
- GALGANO, F., *Diritto civile e commerciali*, vol. 2, Cedam, Padova, 2004.
- GAMBARO, A., e SACCO, R., *Sistemi giuridici comparati*, in “*Trattato di diritto comparato*” diretto da Sacco, R., Utet Giuridica, seconda ristampa 2021.
- GARAPON. A., *Ti faresti giudicare da un algoritmo*, *Questione Giustizia*, Fascicolo 4/2018, (intervista a cura di E. Fronza-C. Caruso).
- GARAPON, A., - LASSEGUE, J., *La giustizia digitale*, il Mulino, 2021.
- GARFINKEL, F., e SPAFFORD, G., *Web, Security & Commerce*, O'Reilly, Sebastopol, 1997.
- GARTHOFF, J., *Legitimacy is not authority*, in “*Law and Philosophy*”, 2010, vol. 29(6), pp. 669-694.
- GATT, L., *Intervento al III Congresso Nazionale di Giustizia Predittiva*, *Radio Radicale*, 3 luglio 2020.
- GATT, L., *Per un' intelligenza artificiale antropocentrica*, *Rivista Diritto, Mercato e Tecnologia*, 2021.
- GIALUZ, M., *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in “*Diritto penale contemporaneo*”, 29 maggio 2019.
- GIUSTI, A., *Il precedente nel rapporto tra sezioni unite e sezioni semplici: l'esperienza della Cassazione civile*, *Questione Giustizia*, Fascicolo 4/2018.
- GOODENOUGH O. R., *Gettong to Computational Jurisprudence 3.0*, in *The Challenge of Innovation in Law: the Impact of Technology and*

- Science on Legal Studies and Practice*, a cura di Santosuosso, A., Goodenough O. R., e Tomasi, M., Pavia University Press, 2015
- GRAMSCI, A., *Oltre il sovranismo*, Cento Autori, 2020.
- GRAZIANO, M., *Ripensare la coscienza*, Edizione Codice, Torino, 2020.
- GREENFIELD, A., *Tecnologie radicali*, Einaudi Editore, 2017.
- GROSSI, P., *Storicità versus prevedibilità: sui caratteri di un diritto post-moderno*, Rivista Questione Giustizia, 2016.
- GROSSI, P., *Prima lezione di diritto*, Laterza, Bari-Roma, 2003.
- Gruppo di Esperti MiSE sull'Intelligenza Artificiale, "Proposte per una strategia italiana per l'Intelligenza Artificiale, mise.gov.it, luglio 2019.
- HAO, K., STRAY, J., *Can you make AI fairer than a judge? Play our courtroom algorithm game*" MIT Technology Review, 17 Ottobre 2019.
- HARARI, Y. N., *Sapiens. Da animali a dei. Breve storia dell'umanità*, Bompiani, 2017.
- HARARI, Y. N., *Homo Deus: Breve storia sul futuro*, Milano, Bompini, 2017.
- HART, H. L. A., *Il concetto di diritto*, tr. It., Torino, Einaudi, 1991.
- Hill, j. l., *The law is what the judge ate for breakfast, The Political Centrist*, Vanderbilt University Press, 2009.
- HOLMES Jr. O. W., *The Common Law*, Little Brown and Co., Boston 1881.
- HOLEMS Jr. O. W., *The Parh of Law*, in "Harvard Law Review" 457, 10, 1897.

- IACONESI, S., *Ubiquitous Commons: A Browser Plugin to Enable User-generated, Peer-to Peer-enacted Licencing Models for Online Data*, 2014.
- IRTI, N., *Un diritto incalcolabile*, Giappichelli Editore, Torino, 2016.
- IRTI, N., *Il tessitore di Goethe*, Contenuto in A.Carleo, *Decisione robotica*, Bologna, 2019.
- JAYNES, J., *The Origin of Consiousness in the Breakdown of the Bicameral Mind*, Mariner Book, Boston, 2020
- JESSUP, P. C., *Transnational Law*, Yale University Press, New Haven, 1956, p. 2.
- JOY, B., *Why the Future Doesn, t Need Us*, Wired Magazine, aprile 2020.
- KAMMINGA, M., e SCHEININ, M., *The impact of human rights law on general international law*, Oxford University Press, Oxford, 2009.
- KAPLAN, J., *Intelligenza artificiale, guida al prossimo futuro*, Luiss University Press, Roma, 2017
- KARNOW, C., *Liability for Distibuted Artificial Intelligence*, in “Berkeley Technology and Law Journal”, 1996, 11, pp. 147-183.
- KASPAROV, G., *Deep Thinking*, Fandango Editore, Frosinone, 2019.
- KATZ, D. M., *Quantitative Legal Prediction or How I Learned To Stop Worrying and Start Preparing for The Data- Driven Future of The Legal Services Industry*, in “Emory Law Journal”, vol. 62: 909, 2013, pp. 909-966.
- KATZ, D. M., e BOMMARITO, BLACKMAN, *A general approach for predicting the behavior of the Supreme Court of the United States*, in “PLoS ONE”12(4), 2017.
- KELSEN, H., *Dottrina pura del diritto*, tr. It Einaudi, Torino, 1952.
- KURZWEIL, R., *La singolarità è vicina*, tr. It., Apogeo, Milano, 2008.

- LAWLOR, R.C., *What Computers can do: Analysis and Prediction of Judicial Decisions*, in “American Bar Association Journal” 49, 1963.
- LEENES, R., e FEDERICA, L., *Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design*, in “Law, Innovation and Technology”, 2014, vol.6(2): 193-220.
- LESSIG, L., *Code and other Laws of Cyberspace*, Basic Book, 1999.
- LIN, P., ABNEY, K. BEKEY, G., *Robot Ethics: Mapping the Issues for a Mechanized World*, in “Artificial Intelligence”, California, 2011.
- LUCIANI, M., *La decisione giudiziaria robotica*, Contenuti in A. Carleo, *Decisione robotica*, Bologna, 2011.
- LUKAS, G., *La distruzione della ragione*, III ed., Torino, 1974.
- MAMMONE, G., *Considerazioni introduttive sulla decisione robotica*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019.
- MANZOTTI, R., TAGLIASCO, V., *Si può parlare di coscienza artificiale?*, Sistemi intelligente, n. 1, 2002.
- MARMO, R., *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo, dati e machine learning, neural network, deep learning*, Hoepli, Milano, 2020.
- MATTEI, U., *Beni comuni. Un manifesto*, Bari, 2011.
- MATTEI, U. - QUARTA, A., *Punto di svolta. Ecologia, tecnologia e diritto privato. Dal capitale ai beni comuni*, Aboca, 2018.
- MAUGERI, M., *I robot e la possibile “prognosi” delle decisioni giudiziali*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019.
- MCCORDUCK, P., *Storia dell'intelligenza artificiale*, Muzzio, Padova, 1987.
- MCKENDRICK, J., *Artificial Intelligence May Amplify Bias, but also can help eliminate it*, Rivista Forbes, giugno 2019.

- MONATERI, P., *Dominus mundi: Political Sublime and the World Order*, Oxford Hart Publishing, 2018.
- MORIN, E., *Etica e identità umana*, Raffaello Cortina Editore, Milano, 2015.
- MOROZOV, E., *I signori del silicio*, Codice Edizioni, 2016.
- MOROZOV, E., *L'ingenuità della rete*, Tr. It. Edizioni Codice, Torino, 2011
- MULLAINATHAN, S., *Biases, human and algorithms*, Conferenza “Legal Challenges of the Data Economy”, Parigi, 22 marzo 2019.
- MURRAY, A.D., *The Regulation of Cyberspace: Control in the Online Environment*, Routledge-Cavendish, New York 2007.
- NIDA-RUMELIN, J., WEIDENFELD, N., *Umanesimo digitale*, Franco Angeli, Milano, 2019.
- NIEVA- FENOLL, J., *Intelligenza artificiale e processo*, Giappichelli, Torino, 2019.
- NORVIG, P., e RUSSEL, S., *Intelligenza Artificiale. Un approccio moderno*, tr. It. Pearson, Milano, 2005.
- NUMERICO, T., *Big data e algoritmi. Prospettive critiche*. Letture.org
- O'NEIL, C., *Weapons of math destruction: how big data increase inequality and threatens democracy*. Crow, New York, 2016.
- Osservatorio Digital Innovation del Politecnico di Milano, *Alla scoperta del Deep learning: significato, esempi e applicazioni*, 19 aprile 2019.
- ORWELL, G., *1984*, Mondadori Edizioni, 2016.
- OROFINO, M., *L'inquadramento costituzionale della rete internet: da nuovo mezzo per la libertà di espressione a presupposto per l'agire dell'individuo nella società dell'informazione*, in Rete italiana dei

CDE, UE: diritti e cittadinanza, Ufficio delle pubblicazioni dell'Unione Europea, Lussemburgo, 2014.

OROFINO, M., *Un caso concreto di governance multilevel: la normativa europea in materia di comunicazioni elettroniche*, in *Research Papers*, 2005 Centro Studi sul Federalismo, 2007

OROFINO, M., *La libertà di espressione tra Costituzione e Carte europee dei diritti. Il dinamismo dei diritti in una società in continua trasformazione*, Giappichelli, Torino, 2014.

OROFINO, M., *Profili costituzionali delle comunicazioni elettroniche nell'ordinamento multilivello*, Giuffrè, Milano, 2008.

OROFINO, M., *L'ordinamento della comunicazione" tra direttive comunitarie riforma del Titolo V della Parte II della Costituzione: considerazioni introduttive allo studio di un ordinamento settoriale in tensione tra "uniformazione" comunitaria e "differenziazione" regionale*, Giappichelli, Torino, 2003.

PAGALLO, U., *Teoria giuridica della complessità*, Giappichelli, Torino, 2006

PAGALLO, U., *Robotica* in Durante, M., e Pagallo, U., (a cura di), *Manuale d'informatica giuridica e diritto delle nuove tecnologie*, Giappichelli, Torino, 2012.

PAGALLO, U., *Sul principio di responsabilità giuridica in rete*, in "Il diritto dell'informazione e dell'informatica", XXV, 4-5, 2009, PP. 705-734.

PAGALLO, U., *Privacy e design*, in "Informatica e diritto", 2009, 1, pp. 123-134.

PAGALLO, U., *The Laws of Robots: Contracts, Crimes and Torts*, Dordrecht, Springer, 2013.

- PAGALLO, U., *Il diritto nell'età dell'informazione*, Giappichelli, Torino, 2014.
- PARENT, W. A., *Privacy, Morality and the Law*, Little Brown, Boston 1973.
- PARISI, F., *Il contratto concluso mediante computer*, Cedam, Padova, 1997.
- PARODI, C., SELLAROLI, V., *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in “Diritto penale contemporaneo”, Fascicolo 6/2019, pp. 47-71.
- PASCUZZI, G., *Il diritto dell'era digitale*, il Mulino, Bologna, 2010.
- PASQUALE, F., *The black box society: the secret algorithms that control money and information*, Harvard University Press, Cambridge (Mass), 2015.
- PEEK, N., HOLMES, J.H., SUN, J., *Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure and Analytics*, in “Yearb Med Inform”, 9 (1), 2014, pp. 42-47.
- PITRUZZELLA, G., *Internet, non è più tempo di irresponsabilità*, in “Corriere della sera”, 21 luglio 2019.
- POPPER, K., *La società aperta e i suoi nemici*, Armando Editore, Milano, 2004.
- POSNER, R., *Orwell versus Huxley: Economics, Technology, Privacy, and Satire*, 24 Philosophy and Literature I, 2000.
- POSNER, R., *Privacy, Surveillance and Law*, Chicago, 2007.
- PROSSER, W., *Privacy*, in “California Law Review”, 1960, vol. 48(3), pp. 383-423.
- PUNZI, A., *Judge in the machine. E se fossero le machine a restituirci l'umanità del giudicare*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019.

- RAFARACI, T., (A CURA DI), *L'area di libertà, sicurezza e giustizia: alla ricerca di un equilibrio fra priorità repressive ed esigenze di garanzia*, Giuffrè, Milano 2007.
- REEED, C., *Maching Law for Cyberspace*, Oxford University Press, Oxford, 2012.
- RIFKIN, R., *The Age of Access*, New York, 2000, trad. It. *L'era dell'accesso. La rivoluzione della new economy*, Milano, 2000.
- ROBOLAW, *Guidelines on Regulating Robotics. EU Project on Regulating Emerging Robotic Technology in Europe: Robotics facing Law and Ethics*, 22/09/2021.
- RODOTA', S., *Elaboratori elettronici e controllo sociale*, Bologna, 1973.
- RODOTA', S., *Il problema della responsabilità civile*, Milano, 1964.
- RODOTA', S., *Intervista su privacy e libertà*, Roma-Bari, 2005.
- RODOTA', S., *La privacy tra individuo e collettività*, in *Pol. Dir.*, 1974, pp. 557.
- RODOTA', S., *La vita e le regole. Tra diritto e non diritto*, Bologna, 2006.
- RODOTA', S., *Una Costituzione per Internet*, in *Politica*, 2006, n. 82, pp177 ss.
- RODOTA', S., *Vivere la democrazia*, Editore Laterza, 2018.
- SADIN, E., *La siliconizzazione del mondo*, Tr. It. Einaudi, Torino, 2018.
- SAMUEL, A. L., *Some studies in machine learning using the game of checkers*, in *IBM Journal of research and development*, 1959.
- SANTOSUOSSO, A., *Intelligenza artificiale e diritto*, Mondadori Università, 2020.

SANTOSUOSSO, A., *Diritto, scienza, nuove tecnologie*, Wolters Kluwer, Milano, 2011.

SANTOSUOSSO, A., *Se l'agente non è necessariamente un essere umano. Alcune riflessioni giuridiche*, in "Notizie di Politeia", vol XXX, 2014, pp. 28-48.

SANTOSUOSSO, A., *The human rights of nonhuman artificial entities: an oxymoron? Yearbook of Science and Ehtics*" in "Jahrbuch Fur Wissenschaft und Ethik", De Gruyter Publisher, 2015, pp. 203-237.

SANTOSUOSSO, A., *Autonomous Systems and the Law: Why Intelligence Matters*, in Eric Hilgendorf, Uwe Seidel (eds), *Robotics, Automatics and the Law: Leagl issues arising from the Automomics for Industry 4.0 Technology Programme of the German Federal Ministry for Economic Affairs and Energy*, January 2017.

SARTOR, G., *Intelligenza artificiale e diritto: un'introduzione*, Giuffrè, Milano, 1996.

SARTOR, G., *Cognitive Automata and the Law: Eletronic Contracting and the Intentionality of Software Agents*, in "Artificial Intelligence and Law", 2009, 17, 4, pp. 253-290.

SARTOR, G., *L'informatica giuridica e le tecnologie dell'informazione. Corso di informatica giuridica*, Giappichelli, Torino, III edizione, 2016.

SARTOR, G., *Logica e diritto*, in *Le direzioni della ricerca logica in Italia*, a cura di Hykel Hosni, Gabriele Lolli e Carlo Toffalori, Edizioni della Normale, Pisa, 2015, pp. 283-314.

SCHAUER, F., *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and the Life*, Claredon Press, Oxford, 1991.

- SCHERMER, BART, V., *The limits of privacy in automated profiling and data mining*. Computer Law and Security Review, 27(1), pp. 45-52, 2011.
- SCHILLING, T., *The Autonomy of the Community Legal Order: An Analysis of Possible Foundations*, in “Harvard International Law Review”, 1996, 37, pp. 389-410.
- SCHONBERGER, V.M. – KUKIER, K., *Big Data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, 2015.
- SCODITTI, E., *Giurisdizione per principi e certezza del diritto*, Questione Giustizia, 2012.
- SLAUGHTER, A. M., *Judicial Globalization*, in “Virginia Journal of International Law”, vol. 40, 2000.
- SINI, C., *L'uomo, la macchina, l'automa. Lavoro e conoscenza tra futuro prossimo e passato remoto*, Bollati Boringhieri, Torino, 2009.
- SOMEK, A., *The Legal relation. Legal Theory After Legal Positivism*, Cambridge University Press, Cambridge (UK)- New York, 2017.
- SORO, A., *Persone in rete*, Fazi Editore, Roma, 2018
- SOTIS, C., *Ragionevoli prevedibilità e giurisprudenza della Corte Edu*, Questione Giustizia, Fascicolo 4/2018.
- SOURDIN, T., *Judge vs. robot? Artificial intelligence and judicial decision making*, UNSW Law journal, vol. 41.
- SPANTIGATI, F., *La complementarietà tra tecnologia e diritto*, in Jus, 1995, pp. 53 ss.
- STANLEY, J., *How Propaganda Works*, Princeton University Press, Princeton, 2005.
- STEINER, C., *Automate This*, Penguin, New York, 2012

- STOLKER, C., *Rethinking the Law Schools: Education, Research, Outreach and Governance*, Cambridge University Press, Cambridge, 2014.
- SWARTZ, A., *Guerrilla Open Access Manifesto*, Archive.org, 2018.
- TEGMARK, M., *Vita 3.0 Essere umani nell'era dell'intelligenza artificiale*, Cortina, Milano, 2018.
- TEUBNER, G., (a cura), *Global law without State*, Dartmouth, 1998.
- TISCORNIA, D., *Intelligenza artificiale e diritto*, in Nannucci, R., (a cura di) *Lineamenti d'informatica giuridica*, ESI, Napoli, 2002.
- TSIOURAS, I., *La sicurezza dell'informazione*, tr.it. Franco Angeli, Milano, 2004.
- TURING, A. M., *Computing Machinery and Intelligence*, Mind, New Science, v. 59,n. 236, pp. 433-460, 1950.
- TURING, A. M., *On Computable Numbers, with an Application to the Entscheidungsproblem*, in "Proceedings of the London Mathematical Society", Series 2, 42, 1936.
- TURING, A. M., *Lecture on the Automatic Computing Engine, 1947*.
- TURTLE, H., *Text Retrieval in the Legal World*, in "Artificial Intelligence and Law", 1995, 3, pp. 5-54.
- VACCARO, S., *Gli algoritmi della politica*, Eleuthera, 2020.
- VACIAGO, S., *Internet e responsabilità giuridiche*, Piacenza, 2002.
- VERUGGIO, G., *Euron Roboethics Roadmap*, in "Proceedings Euron Roboethics Atelier", Genova, 2006.
- VESPIGNANI, A., *L'algoritmo e l'oracolo*, Il Saggiatore, Milano, 2019.
- VIGANO', F., *Il principio di prevedibilità della decisione in materia penale*, in *Diritto Penale Contemporaneo*, 2016.

- VINCENTI, E., *Massimizzazione e conoscenza della giurisprudenza nell'era digitale*, *Questione Giustizia*, Fascicolo 4/2018.
- VINCENTI, E., *Il problema del giudice-robot*, Contenuto in A. Carleo, *Decisione robotica*, Bologna, 2019.
- VIOLA, L., voce *Giustizia predittiva*, in *Enc. Giur. Treccani* (treccani.it 2020).
- VIOLA, L., *Interpretazione della legge con modelli matematici*. Processo, Adr, giustizia predittiva. Volume 1, *Diritto avanzato*, Milano, 2018 (II edizione).
- WALLACH, W., MARCHANT, G., *Toward the Agile and Comprehensive International Governance of AI and Robotics*, in “*Proceedings of the IEEE*”, Vol. 107, No. 3, 2019.
- WARWICK, K., *Intelligenza Artificiale. Le basi*, Dario Flaccovio Editore, Palermo, 2015.
- WARREN, S.D. – BRANDEIS, L.D., *The Right To Privacy*, 5 *Harvard Law Rev.* 193, 1890.
- WESTIN, A. F., *Privacy and Freedom*, Atheneum Press, New York, 1967.
- WIENER, N., *L'uso umano degli esseri umani*, Bollati Boringhieri, Torino, 1970.
- YEUNG, K., *Towards an understanding of regulation by design*, in “*Regulating Technologies: legal futurs, regulatory frames and technological fixes*, pp. 79-108, (a cura di) Brownsword, R., e Yeung, K., Hart, Londra, 2007.
- ZELLINI, P., *La matematica degli dei e gli algoritmi degli uomini*, Adelphi, Milano, 2016.
- ZELLINI, P., *La dittatura del calcolo*, Adelphi, Milano, 2018.
- ZICCARDI, G., *Hacker. Il richiamo delle libertà*, Venezia, 2011.

ZICCARDI, G., *Il controllo del codice informatico tra diritti, etica e business*, in *Politica*, 2006, n. 82, pp. 183 ss.

ZICCARDI, G., *Tecnologie per il potere*, Raffaello Cortina Editore, 2019.

ZITTRAIN, J., *Perfect Enforcement on Tomorrow's Internet*, in "Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes, a cura di Brownsword, R., e Yeung, K., Hart, Londra, 2007, pp.125-156.

ZUBOFF, S., *Il capitalismo della sorveglianza*, Luiss University Press, 2019.

ZUMBANSEN, P., "Transational Law" , in Jan Smits (ed.), *Encyclopedia of Comparative Law*, Edward Elgar Publishing, Cheltenham, 2006.