

Amedeo Santosuosso*

*Intelligenza artificiale, conoscenze neuroscientifiche
e decisioni giuridiche*

Abstract: The paper addresses some topics at the intersection of neuroscience, artificial intelligence (AI) and law.

The problem of AI is not only the development of more powerful machines, but the cognitive architecture of the intelligence model that is assumed as a reference. But until we better understand how human cognition works, we will not even make decisive steps on 'general' artificial intelligence.

Law is a rich test case and an important field for the development of logic-based artificial intelligence, particularly with regard to logical models of legal argumentation.

The question of the relationship between law and logic (and which logic) and between law and AI intertwines with legal traditions (mainly rationalist or historicist ones) and with the orientations present in the IA debate, and gives rise to complex and open questions. In particular, the relationship between logicist (or rationalist) and non-logicist (or historical) orientations in AI and law is considered.

Keywords: Artificial intelligence and law; Logicist approach; Neuroscience and artificial intelligence; Non-logicist approach.

Indice: 1. Un nodo problematico all'incrocio tra diverse discipline – 2. Intelligenza e capacità di calcolo – 3. Architetture cognitive e conoscenze neuroscientifiche – 4. Intelligenza artificiale, logica e diritto – 5. Processi decisionali nell'era delle decisioni basate sui dati – 6. Giudici, motivazioni e scatole nere (nuove e vecchie) – 6.1. Black boxes (nuove) – 6.2. Black boxes (vecchie)

1. Un nodo problematico all'incrocio tra diverse discipline

Questo contributo intende mettere a fuoco un nodo problematico all'incrocio tra diverse discipline. Il nodo problematico è quello delle decisioni giuridiche in un'epoca in cui gli apporti conoscitivi alla decisione non vengono più soltanto dall'esperienza dell'operatore, dai suoi studi passati o dal suo ricorso a fonti cartacee (libri, riviste e simili), ma sempre più spesso giungono da raccolte digitalizzate (database) o che usano tecniche di intelligenza artificiale -IA- (*machine learning*

* Docente di *Diritto, Scienza, Nuove tecnologie* presso il Dipartimento di Giurisprudenza dell'Università di Pavia; docente di *Diritto e tecnologie dell'informazione e comunicazione* presso l'Istituto universitario di Studi superiori (IUSS) di Pavia. amedeo.santosuosso@unipv.it

-ML- e *deep learning* -DL-). Le discipline coinvolte sono il diritto, l'intelligenza artificiale e le neuroscienze cognitive.

L'intelligenza artificiale fa affluire al nodo problematico due ulteriori (relativamente indipendenti) profili, di cui il primo è dato dalla logica che impronta i modi in cui i sistemi di ML e DL estraggono conoscenza da grandi insiemi di dati giuridici, e il secondo è dato dal rapporto tra l'intelligenza delle macchine (nella misura in cui di intelligenza si possa parlare) e intelligenza umana, come studiata dalle neuroscienze cognitive (che è il tema del rapporto tra IA e neuroscienze cognitive).

I rapporti tra le discipline coinvolte (diritto, IA e neuroscienze cognitive) sono resi viepiù complessi, avendo le neuroscienze connessioni dirette sia con l'IA, sin dalle sue origini, sia con i processi mentali che portano alla decisione umana, e quindi anche giuridica, indipendentemente dall'IA.

Dunque, le questioni che confluiscono nella definizione dell'incrocio problematico di cui ci si occupa in questo scritto appartengono a discipline diverse, tanto che un compiuto sviluppo di questa ricerca richiederebbe una pluralità di approcci disciplinari, come quello di esperti di IA (in particolare di procesamento del linguaggio naturale -NLP- e modellazione dell'argomentazione), esperti di neuroscienze cognitive e esperti di diritto e teoria del diritto.

Naturalmente l'autore non compendia in sé tutte queste competenze. È, e rimane, un giurista, pur nello scrupolo della documentazione in materia e pur avendo una lunga storia di lavoro interdisciplinare con il mondo scientifico. I lettori appartenenti ai diversi campi che convergono nell'incrocio potranno trovare insufficienti alcuni riferimenti di dominio o individuare aspetti e elementi conoscitivi che avrebbero potuto essere sviluppati in modo più approfondito. Propongo ai lettori di adottare uno spirito che ho trovato utile nei lavori interdisciplinari, e, quindi, di provare a guardare criticamente il nodo problematico che si tenta di descrivere, piuttosto che le inevitabili insufficienze nelle conoscenze di settore.

Nel paragrafo 2, si pone il tema del rapporto tra intelligenza e mera capacità di calcolo (e della necessità di distinguere i due aspetti); nel paragrafo 3, il rapporto tra architetture cognitive e conoscenze neuroscientifiche viene introdotto, precisando le differenze tra IA e *general purpose AI*, che è l'ambizione irrisolta, sin dall'inizio della storia dell'IA, di riprodurre i processi mentali umani in macchine; nel paragrafo 4, viene discusso il tema, certo non nuovo, dei rapporti tra logica e diritto, che però assume oggi nuova rilevanza e nuovi profili in relazione all'intelligenza artificiale; nel paragrafo 5, vengono discussi i diversi modi dei processi decisionali nell'era delle decisioni basate sui dati e alcune loro implicazioni di metodo; nel paragrafo 6, ci si immerge, con un approccio improntato a un realismo moderato, nelle molteplici dinamiche che caratterizzano le decisioni giudiziarie e, in particolare, viene trattato il punto più delicato, quello del raccordo tra IA e principi giuridici, che fa persino dubitare della possibilità di utilizzare qualsiasi sistema di IA che non renda possibile la spiegabilità di ogni passaggio della specifica decisione giudiziaria come decisione dello specifico caso (vecchie e nuove scatole nere). Nel paragrafo finale (il 7) si delinea un percorso di possibile ricerca

multidisciplinare, rispetto al quale questo scritto può essere considerato solo una prima esplorazione.

2. Intelligenza e capacità di calcolo

L'intelligenza artificiale ha fatto negli ultimi dieci anni enormi progressi nello sviluppo di alcuni algoritmi che hanno permesso l'analisi automatica delle immagini per riconoscere i volti e la visione artificiale per la guida di auto e camion¹. Dopo cinquant'anni di ricerche, negli ultimi due lustri, l'intelligenza artificiale ha conosciuto una rivoluzione grazie agli sviluppi delle reti neurali profonde ed è stato possibile costruire un algoritmo capace di battere il campione del mondo nel gioco del Go, un gioco nettamente più difficile degli scacchi².

Tuttavia, non si può dire che queste macchine siano intelligenti, poiché esse non sono capaci di costruire una rappresentazione del mondo o di dare vita a processi creativi, sostiene Mark Mézard, un fisico studioso delle reti neurali e direttore delle *École Normale Supérieure* di Parigi³. È la comprensione di quello che esse stesse riescono a fare, che appare un limite a oggi invalicabile. Roger Penrose, matematico illustre, emerito dell'Università di Oxford e Premio Nobel per la fisica nel 2020, è molto netto. In un'intervista rilasciata in occasione dell'evento *AI for Good* (2018) tesse le lodi delle incredibili capacità di calcolo dei sistemi di IA, della loro utilità ai fini dei calcoli per la progettazione, ma è molto chiaro nel porre la distinzione tra *agire sulla base dell'esperienza*, che è quella che può portare una macchina a vincere nel Go contro l'umano più esperto (come già accaduto per gli scacchi con Kasparov), e la *comprensione di quello che si sta facendo*. Così le macchine sviluppano abilità senza avere la conoscenza teorica di quale sia il fondamento di quelle azioni⁴. Melanie Mitchell, professore di computer science, affonda la sua critica verso coloro i quali pensano che *Superintelligent AI might be the last event in human history*⁵ e mette in luce la complessità dell'intelligenza umana, le cui limitazioni sono anche alla base dei suoi punti di forza:

The problem with such forecasts is that they underestimate the complexity of general, human-level intelligence. Human intelligence is a strongly integrated system, one whose many attributes – including emotions, desires, and a strong sense of selfhood and autonomy – can't easily be separated. [...] the notion of superintelligence without humanlike limitations may be a myth⁶.

1 In questo contributo mi avvalgo di mie precedenti ricerche pubblicate in Santosuosso [2020], in particolare alle pp. 8-13, 21-25, 96-120. A tali fonti rinvio per un'argomentazione più ampia e completa e per esaustivi riferimenti bibliografici.

2 *Digital intuition* 2016.

3 Ordine 2019.

4 Kan 2018.

5 Hawking S., Stuart R., Tegmark M., Wilczek F. 2014: la pagina attualmente non è disponibile, ma è citata da uno dei coautori in Russell. 2017. Si veda anche Russell 2019.

6 Mitchell 2019.

Alla fine, si può forse convenire con chi, come la scienziata Fei-Fei Li, sostiene che non vi è niente di artificiale nell'intelligenza artificiale: «è ispirata da persone, è creata da persone e, cosa più importante, incide sulle persone. È uno strumento potente che si sta solo ora cominciando a capire come funzioni, e questa è una profonda responsabilità»⁷. Fare intelligenza artificiale implica, però, una visione sugli umani, su cosa deve essere l'umanità che vogliamo per il futuro. Di qui la necessità di un'importante riflessione etica.

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, provenienti da diverse discipline e da background diversi per genere, età, etnia e cultura, concorrano tra loro. Questa ispirazione è al centro di un istituto fondato nel 2019 dall'Università di Stanford (USA), l'istituto per un'intelligenza artificiale centrata sull'uomo (*Stanford Institute for Human-Centered Artificial Intelligence*)⁸. 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.

Quello di Stanford è un gruppo di ricercatori di avanguardia, 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. È interessante, poi, che l'aver una visione dell'umanità (*need to represent humanity*) venga sollecitato 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, lo sviluppo della consapevolezza che la più completa lista di cose (che possano essere riconosciute da una macchina) non ci dice ancora nulla sulle relazioni tra quelle cose. È necessario, a tal fine, sviluppare parimenti un insieme di dati sugli attributi delle cose e sulle relazioni tra le stesse (il latte versato nella tazza o fuori la tazza sono due scene che coinvolgono gli stessi oggetti, ma le relazioni sono molto diverse!!!) e che tale passaggio richieda la collaborazione di neuroscienziati ed esperti di processi cognitivi umani, che dicano come ciò accada esattamente negli uomini.

Dalla visione delle cose, alle loro relazioni, a come ciò accada negli umani, alla centralità di questo passaggio nello sviluppo attuale dell'intelligenza artificiale, alla necessità di collaborare con scienziati appartenenti a diverse discipline... è come se da tutto ciò scaturisse anche la necessità di una visione sull'uomo!

7 Fei-Fei Li: "There's nothing artificial about AI." "It's inspired by people, it's created by people, and—most importantly—it impacts people. It is a powerful tool we are only just beginning to understand, and that is a profound responsibility". <https://www.wired.com/story/fei-fei-li-artificial-intelligence-humanity/> (visitato il 28 marzo 2019).

8 *Stanford Institute for Human-Centered Artificial Intelligence*, <https://hai.stanford.edu/news/stanford-s-new-institute-will-unite-humanities-and-computer-science-study-guide-and-develop> dove si trovano le dichiarazioni dei due co-direttori, Fei-Fei Li e Etchemendy. In particolare Fei-Fei Li: "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."

Vista in questa luce risulta comprensibile l'affermazione di Fei-Fei Li quando dice che non vi è niente di artificiale nell'intelligenza artificiale. 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'IA sia ispirato dalla nostra comprensione dell'intelligenza umana; la tecnologia deve essere guidata dalla nostra comprensione di come 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'IA così come non vi è niente di artificiale nell'energia elettrica, nel motore a vapore, nella stampa a caratteri mobili ecc., nel senso che nessuno direbbe mai 'stampa artificiale', 'elettricità artificiale' o simili. E allora bisognerebbe dire intelligenza delle macchine, o, più esattamente, capacità computazionale delle macchine, piuttosto che IA.

La congiunzione "e" del titolo di questo paragrafo va, dunque, intesa in funzione disgiuntiva, e cioè *intelligenza* versus (mera) *capacità di calcolo*.

3. Architetture cognitive e conoscenze neuroscientifiche

È come se solo oggi si fosse usciti dalla prima fase sperimentale della computazione digitale, iniziata con Turing e poi definita nel 1956, e per la prima volta si cominciasse a fare sul serio. Vi è una sorta di indicazione filosofica, per cui fare intelligenza artificiale significa avere una visione sugli umani, su cosa deve essere l'umanità che vogliamo per il futuro. Si tratta di una visione fondata scientificamente, che parte dal mondo della stessa intelligenza artificiale e da un settore come quello della visione, che è cruciale. Parte da scienziati che sollecitano una riflessione etica, una sollecitazione che tende a colmare un deficit sentito diffusamente. La visione di Fei-Fei Li si collega con le riflessioni del neuroscienziato Joshua Greene.

Il problema dell'IA non è solo lo sviluppo di macchine più potenti, ma quello dell'architettura cognitiva del modello di intelligenza che si assume come riferimento. Ma finché non capiremo meglio come funziona la cognizione umana e finché non avremo fatto dei passi decisivi su questo piano, non faremo neanche dei passi decisivi sull'intelligenza artificiale 'generale', come ci ricorda con forza Joshua Greene. La questione è duplice. Il cervello umano è certamente lungi dall'essere compreso, mentre l'intelligenza artificiale generale, che si sta cercando di replicare nelle macchine, esiste solo negli esseri umani, sui quali non è etico fare sperimentazione. È probabile che lo sviluppo della ricerca su organoidi cerebrali (sviluppati con materiale umano ma fuori dal corpo umano) possa aprire nuove strade di ricerca⁹.

9 Sui rapporti tra IA e neuroscienze il riferimento di base è a quello che è considerato il testo più ampio e autorevole nel campo dell'intelligenza artificiale: Russell S. – Norvig P. 2010 (il primo capitolo in particolare e passim). Inoltre si vedano, fra i molti, Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. 2017; Greely H. T. 2017.

Il problema, dunque, non è solo lo sviluppo di macchine più potenti, ma quello dell'architettura cognitiva. Se l'idea alla base dell'IA è quella di agente intelligente è importante il modello di intelligenza che si assume come riferimento.

Ci si può chiedere 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 in realtà). Cercare di capire come gli uomini facciano ciò è un passaggio essenziale per tentare di costruire un'intelligenza artificiale generale (*artificial general intelligence*)¹⁰, che è quella altrimenti chiamata *human-level AI* oppure *strong AI*.

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 tecnici 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 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.

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 esso dica qualcosa come *cane che insegue gatto*. Greene prosegue 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 che si colloca 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.

10 Perry 2019. Qui di seguito sono riportati alcuni passi salienti da me tradotti.

4. Intelligenza artificiale, logica e diritto

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 si fa giustamente 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 destinatario di eventuali restrizioni e la sua applicazione deve essere spiegata e giustificata. Sono, quindi, importanti la chiarezza dei significati e la coerenza dei ragionamenti, e quindi l'uso appropriato della logica¹¹.

Naturalmente questo 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 perché, precisano Prakken e Sartor, 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 a essere applicata in circostanze non previste al momento della sua approvazione, senza contare le formule astratte come 'buona fede', 'correttezza', 'abuso di segreti industriali' o del diritto in generale, o 'ragionevolezza', tutte formule che richiedono di essere tradotte in norme aventi un contenuto concreto.

Quindi, concludono gli autori, 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, e la natura conflittuale (*adversarial*) delle procedure (dove ogni parte punta a convincere il giudice della giustezza della sua tesi), tutto ciò fa sì che il ragionamento giuridico vada oltre il significato letterale delle norme di legge. È, quindi, necessario uno studio dell'argomentazione giuridica (*formal and computational study of argumentation*), un settore nel quale vi sono non solo da applicare le tecniche di IA, ma anche da sviluppare esperimenti e ricerche che possano rappresentare un avanzamento proprio di quelle tecniche.

Evidentemente il tema è complesso e comprende numerosi profili sia teorici, sia giuridici 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 mille linguaggi naturali. E si può, anche, ricordare che il fatto che la logica possa essere un utile strumento in ambito giuridico non significa sposare necessariamente una logica monotona, secondo la quale «se si accettano le premesse di un'inferenza conclusiva, se ne debbono necessariamente accettare anche le conclusioni», mentre sono ormai prevalenti, specie in ambiti ad alta incertezza, le co-

11 Prakken, Sartor 2015.

siddette logiche non monotoniche, come per esempio la *defeasible logic*. Secondo tale logica, posta una certa premessa, dalla quale discende una certa conseguenza (secondo la logica monotonica), si può accettare di dover cambiare la conclusione se una nuova informazione rivela l'erroneità dell'inferenza prima ottenuta. L'esempio divenuto classico nel campo dell'IA è il seguente: se io dico che Tweety è un uccello, l'ascoltatore deduce che Tweety possa volare, ma se io poi lo informo che Tweety è un pinguino, l'inferenza evapora di conseguenza, com'è giusto che sia¹².

Va, poi, tenuto presente che la questione dei rapporti tra diritto e logica (e quale logica) e tra diritto e IA si intreccia con le tradizioni giuridiche (principalmente quelle razionaliste o storiciste) e con gli orientamenti presenti nel dibattito sull'IA, e da luogo a complesse e aperte questioni. Tra le altre si considerino le seguenti.

Nell'ambito della ricerca sull'IA sono presenti tre principali orientamenti: uno logicista e uno non-logicista. Il secondo ha al suo interno un approccio simbolico, ma non logicista, e altri approcci connessionisti e neurocomputazionali.

L'aspetto interessante è che questi diversi orientamenti, per quanto presenti sin dalle origini della moderna IA 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 efficiente 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 IA di livello umano (*human-level AI*).

Quanto alle tradizioni giuridiche sia sufficiente ricordare i due principali orientamenti, di tipo razionalistico e di tipo pragmatico o storico, i cui due campioni del passato 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 a fine Ottocento il secondo, sono diventati l'emblema dei due principali orientamenti circa il ruolo della razionalità nel diritto: orientamenti che segnano un discrimine che, per certi versi, è più importante di quello tra *common law* e *civil law*¹³.

A questo punto sorge una domanda che coinvolge aspetti teorici e aspetti applicativi: una combinazione di diversi approcci può essere usata anche nel diritto? Si può adempiere all'obbligo di motivazione se si usano i *neural networks*? L'IA può organizzare anche attività che non appaiano 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 IA.

12 Sartor 2010.

13 Coke 1628, bk. 2: ch. 6, sect. 138. Holmes W.O. Jr. 1881, Lecture I.

Si può dire, in conclusione del paragrafo, e sulla base dei pochi punti teorici richiamati, che l'intreccio intelligenza artificiale-logica-diritto somigli molto a un intricato incrocio di diverse e ramificate strade¹⁴.

5. Processi decisionali nell'era delle decisioni basate sui dati

Oggi le decisioni che vengono prese nei più diversi campi sono distribuite lungo una scala che va dal processo decisionale basato su regole al ragionamento statistico fino all'apprendimento automatico e all'IA vera e propria. Di fatto, il processo decisionale basato sui dati e il processo decisionale basato sulla conoscenza tradizionale (come i sistemi basati su regole) coesistono in diverse combinazioni a seconda dei campi 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?

A prima vista il processo decisionale giuridico, tradizionalmente basato su regole (quale che ne sia la loro natura), è messo in discussione da sistemi decisionali basati su analisi dei dati. Tuttavia, l'individuazione del vero problema in gioco richiede una comprensione un po' più approfondita.

È utile partire dal chiarimento di alcuni punti:

– Processo decisionale basato su dati non è sinonimo di decisione basata su prove, come potrebbe apparire dal tenore letterale delle parole usate. *Data-driven decision making* non significa decisione *evidence based* (basata su prove) in opposizione a decisioni arbitrarie: è un modo diverso di ragionare e decidere.

– Processo decisionale basato su regole (*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 esso».¹⁵ Secondo la sua definizione, molto esigente, se un tribunale legittimamente nominato prende una decisione secondo il criterio (pur previsto da una legge) del «miglior interesse» del bambino o del paziente, o secondo il sistema di equità o determina l'entità di una condanna penale (in cui un numero illimitato di fattori può svolgere un ruolo), 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 (che non è chiaramente definita e che lascia ampi margini di ulteriore definizione a opera proprio del giudice).

14 Santosuosso, Pinotti 2020.

15 Schauer 1991: 11 [traduzione mia].

– Nel campo più ampio delle decisioni giuridiche (che include, quindi, 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 (personali, politiche, sociali, emotive e più) ragioni. Il dibattito italiano sulla motivazione risente del fatto che nel nostro paese non ha mai avuto un riconoscimento significativo la scuola del realismo giuridico, molto più presente in altri paesi¹⁶.

– 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 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». ¹⁷ In altri termini, il rapporto ipotesi-test, alla base del metodo scientifico moderno, è invertito ed è l'analisi, con tecnologie *machine learning*, di un appropriato insieme di dati a far emergere un possibile modello da poi testare.

– 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 (2017), “poiché un algoritmo di *machine learning* (ML) impara regole basate su regolarità statistiche, che possono sorprendere gli umani, le sue regole potrebbero non sembrare necessariamente ragionevoli per gli stessi umani. Le previsioni ML sono basate sui dati. A volte i dati contengono caratteristiche che, per motivi spuri, come la coincidenza o la selezione parziale, sono associati ai risultati dei casi in una particolare raccolta. Sebbene le regole indotte dalla macchina possano portare a previsioni accurate, esse non si riferiscono all'esperienza umana e potrebbero non essere così comprensibili per l'uomo come le regole costruite manualmente da un esperto. Poiché le regole che l'algoritmo ML estrae non riflettono necessariamente le conoscenze o le competenze giuridiche esplicite, potrebbero non corrispondere ai criteri di ragionevolezza di un esperto umano”¹⁸.

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 qualità dei dati raccolti e i modi delle loro analisi e interpretazioni, tanto che qualsiasi errore sistematico (*bias*) 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

16 Levi 1949; Holmes 1897.

17 Katz 2013: 909.

18 Ashley 2017: 111 [traduzione mia].

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 di qualsiasi decisore (e quindi anche dei giudici), 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 dall'analisi giuridica (e la loro limitata spiegabilità) e il diritto alla spiegazione delle decisioni pubbliche, che disposizioni costituzionali fondamentali riconoscono?

6. Giudici, motivazioni e scatole nere (nuove e vecchie)

La motivazione delle decisioni giurisdizionali si presenta come uno dei punti più delicati, forse il più delicato, del raccordo tra IA e principi giuridici. Al punto da far ritenere impossibile o preclusivo l'utilizzo di qualsiasi sistema di IA che impedisca la spiegabilità di ogni passaggio della 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 IA, che s'individuino 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 IA 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 complesso ed è stato trattato numerose volte in dottrina. Qui di seguito faccio solo cenni ad alcuni aspetti cruciali e propongo qualche argomento che possa contribuire a ridimensionare l'ansia da spiegabilità della decisione, che sembra pervadere alcune posizioni contrarie pregiudizialmente all'uso dell'IA nel decidere. Per esempio, si può cominciare ad accettare che gli umani, assumendo che i giudici tali siano¹⁹, 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

19 L'affermazione non è provocatoria, solo che si pensi all'avanzare di modalità, che si possono chiamare *human free*, di prendere decisioni.

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'IA nel modo giusto²⁰.

Le diverse tesi tradizionali riguardano e mettono in luce funzioni e aspetti diversi della motivazione, che sono rilevanti e acquistano maggior peso laddove è in gioco la difficoltà di spiegabilità. Alcune tesi considerano la motivazione come mero sillogismo (dalla norma astratta alla sua applicazione al caso specifico), altre come resoconto del percorso decisorio, altre come razionalizzazione a posteriori, altre come esposizione che parte dallo stato dell'arte (scopo/statement/dichiarazione di scienza conoscitivo).

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 giunge alla decisione era chiaro già negli anni Cinquanta del Novecento, quando Calamandrei sosteneva che essa fosse «l'apologia che il giudice elabora a posteriori della decisione stessa»²¹. Ciò autorizza il ricorso alla distinzione tra il momento dell'esplorazione e della ricerca della soluzione (giuridica) di un caso (*context of discovery*) e quello della sua giustificazione (*context of justification/explanation*). In termini pratici, un conto è il complesso percorso che il giudicante 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»²².

Una visione di questo genere è, alla fine dei conti, in linea con le previsioni normative²³. In tutti questi casi, penali e civili, il giudice decide di assolvere/condannare/ dichiarare/constituire o altro, e poi ha un termine per motivare quell'*antecedente* decisione. Si può quindi dire, a un primo livello di approssimazione, che, ben prima dell'avvento dell'IA, la motivazione non era resoconto di tutto quanto accaduto nel processo e nella fase decisoria, ma 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.

20 Seif 2018.

21 Calamandrei 1954: 101, citato in Biscotti 2012: 10.

22 Biscotti 2012: 11.

23 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 provvede 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'analoga possibile scissione tra momento decisorio e momento di deposito della motivazione nelle cause in materia di lavoro (art. 420 cpc), dove è prevista la lettura del dispositivo e il successivo deposito della motivazione, e in tutte le altre materie che applicano il rito del lavoro o riti a esso ispirati (come nelle impugnazioni di sanzioni CONSOB o della Banca d'Italia).

In questo iato, che riguarda sia il *tempo* della stesura sia la *natura* stessa dell'atto-motivazione (resoconto \neq giustificazione), 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. Per esempio, 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.

La possibilità di raggiungere una soluzione giuridica di un caso utilizzando anche un sistema di *machine learning* o *deep learning*, che produca un risultato che i giudici fanno proprio e che, dopo averlo fatto proprio, riescono a giustificare in termini giuridici tradizionali, fa scandalo solo se si assuma l'idea che la motivazione debba essere una specie di verbale di quanto accaduto nel corso del processo decisionale, che ha inizio con il primo studio della causa e che si conclude con la camera di consiglio (e anche dopo, fino al deposito della motivazione). Ma, lo si è visto sopra, la motivazione processo/rendiconto del percorso e dell'andirivieni, che ogni causa un minimo complessa richiede, non è teorizzata o richiesta da nessuno.

Quei risultati di sistemi di IA possono essere l'*invisibile* che viene reso *visibile*, cioè quello che è presente nell'esperienza giuridica che ha preso la forma di quelle raccolte di dati e che noi non vediamo o perché i dati sono tanto numerosi da essere per noi defaticante, o di fatto impossibile, un'esplorazione o perché la nostra esplorazione è affetta da uno dei bias cognitivi umani che ci impediscono di vedere qualcosa che pure sarebbe sotto i nostri occhi e che, invece, un sistema automatico (che avrà altri bias, ma non i nostri) riesce a correlare. Tutto sommato, e al netto della distanza di tempo, non siamo lontani dalla visione di Norberto Bobbio:

non diversamente da un artigiano che sia insieme anche un artista, questi strumenti il giurista non se li trova già pronti per l'uso davanti a sé: per usarli deve prima di tutto andarseli a cercare (e qualche volta sono lontani e nascosti), poi forgiarseli secondo lo scopo proposto, e talora anche fabbricarsi con le proprie mani.²⁴

Quello che è importante è che il coacervo, talora disomogeneo e apparentemente disordinato della fase di *discovery*, trovi il modo di prendere la forma e di metamorfizzarsi in un testo ispirato a quella razionalità che Patrizia Borsellino dice di "tipo logico o 'quasi-logico'" e che sia degno del nome motivazione.

6.1. Black boxes (nuove)

Questa idea, improntata a un realismo moderato, non è condivisa da chi sostiene che la decisione/motivazione resa sulla base delle risposte date da un sistema di IA sia irrimediabilmente inficiata dalla natura di *black box* di quel sistema. 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 es. essere un 'oracolo'²⁵. 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'unità le cui operazioni interne non possono essere oggetto di indagine²⁶.

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 potrebbe essere 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 difficile 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*, 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 IA, è l'umano a costituire la scatola nera.

25 "Digital intuition" 2016.

26 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 i seguenti: "Reversed Time Attention Model (RETAIN): *The RETAIN model was developed to help doctors understand why a model predicts patients at risk of heart failure.* Local Interpretable Model-Agnostic Explanations (LIME): *A post-hoc model that provides an explanation of a decision after it has been made.* Layer-wise Relevance Propagation (LRP): *This is particularly useful for computer vision for identifying important pixels by running a backward pass in the neural network*" (si veda Vishal Morde, 5 Key Lessons Learned from 2018 AI Summit, NYC Hype, Reality, Ethics, Disruption and More..., Dec 10, 2018, <https://towardsdatascience.com/https-medium-com-vishalmorde-lessons-learned-from-ai-summit-d7c071cb4c06> visitato il 19 agosto 2019.

In sintesi, si può dire che difficoltà di conoscenza e di ricostruzione delle decisioni sicuramente esistono, ma sono difficoltà materiali (come, per esempio, l'eccessiva onerosità e, quindi, la non convenienza economica della ricostruzione) e non il risultato di chissà quale mistero del sistema.

6.2. Black boxes (vecchie)

D'altra parte, non sono queste 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 rischio di recidiva, effetti di blocco della conoscenza che non sono da meno. Gli esempi in ambito giuridico possono continuare. Un effetto black box è creato, per esempio, dal divieto di, o dal non uso, della pubblicazione dell'opinione di minoranza, e quindi dissenziente, nella decisione di un collegio giudicante: 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 e nelle perizie. Vi sono poi alcune prassi decisionali basate banalmente 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 boxes*, nel diritto o nelle prassi, esistono, così come esistono nella medicina o nel pilotaggio degli aerei, come è emerso nei disastri dei Boeing 737 Max nel 2019. Si tratta di capire come integrare i processi decisionali. Non va, poi, sottovalutato 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. Sembra andare in questa direzione il progetto di ricerca su come decidere *About, By and Together with Algorithmic Decision Making Systems* (ADM), promosso da *The Hans-Bredow-Institut* 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 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 penale²⁷.

27 Il progetto è promosso da *The Hans-Bredow-Institut* (un istituto indipendente e non-profit che opera dal 1950 in collaborazione con l'università di Amburgo): <https://hans-bredow-institut.de/en/projects/opportunities-and-limitations-of-adm-systems> (visitato il 23 maggio 2019).

7. La complessità dei percorsi del giudicare

È 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.

Qui il cerchio si chiude. La motivazione, e il suo alto valore costituzionale, non sono posti in pericolo dall'uso di tecniche di IA. L'importante è avere la giusta visione del lavoro del giudice e della funzione del diritto nelle nostre società²⁸. La riflessione sui processi decisionali in ambito giudiziario, proprio per il suo rilievo etico e costituzionale, può costituire un esercizio utile per una migliore teorizzazione e pratica della decisione anche in ambito amministrativo e delle pubbliche istituzioni in generale. Aspetti come quello della trasparenza, del contraddittorio e della spiegabilità sono ampiamente comuni, fatte salve specifiche differenze di ordinamento. Anche le questioni metodologiche che le decisioni basate su big data pongono possono dare spunti utili negli ambiti più diversi, compreso quello della ricerca sia giuridica sia scientifica. E si potrebbe continuare.

In generale, se si prende atto che i sistemi avanzati di elaborazione delle informazioni costituiscono un nuovo orizzonte (da alcuni chiamato *intelligent automation continuum*) nel quale è l'interazione tra diverse discipline e tra componente umana e tecnologica la vera sfida nella nuova "era dell'insieme" (*the Age of With*)²⁹, ne discende che il nodo che si è, a un primo livello, messo a fuoco in questo scritto merita uno sviluppo e una ricerca più profonda e aperta agli apporti interdisciplinari.

Bibliografia

"Digital intuition", *Nature*, 28 January 2016, vol. 529: 437.

Ashley K.D., *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge: Cambridge University Press.

Biscotti B. 2012, "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, Padova: Cedam.

Bobbio N. 1970, "Essere e dover essere nella scienza giuridica", in T. Greco (ed.) 1970, *Studi per una teoria generale del diritto*, Torino: Giappichelli.

Calamandrei P. 1954, *Processo e democrazia*, Padova: Cedam.

²⁸ Per un'ordinata esposizione di questa visione del diritto, v. Santosuosso 2020.

²⁹ Per l'*intelligent automation continuum* si veda Kapoor S. 2020; per l'*Age of With* si veda Deloitte 2019.

- Coke E. 1628, *The First Part of the Institutes of the Laws of England, or, A Commentary on Littleton*, London, bk. 2.
- Deloitte 2019, Automation with intelligence, <https://www2.deloitte.com/content/dam/Deloitte/tw/Documents/strategy/tw-Automation-with-intelligence.pdf> (visitato il 20 novembre 2020)
- Greely H. T. 2017, Neuroscience, Artificial Intelligence, CRISPR-and Dogs and Cats. *UCDL Rev.*, 51, 2303
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. 2017, Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245-258
- Hawking S., Stuart R., Tegmark M., Wilczek F. 2014, “Transcending Complacency on Superintelligent Machines”, *The Huffington Post*, available at https://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html.
- Holmes O.W. Jr. 1881, *The Common Law*, Boston: Little Brown and Co.
- Holmes O.W. Jr. 1897, “The Path of Law”, *Harvard Law Review* 457, 10.
- Kan C. 2018, “Intervista a Roger Penrose”, available at <https://www.youtube.com/watch?v=dpSpwzyO0vU>.
- Kapoor S. 2020, IDC Perspective: Six Capabilities from Leading RPA Service Providers That Advance Financial Services Institutions Toward Intelligent Automation in: <https://www.capgemini.com/wp-content/uploads/2020/04/An-IDC-perspective-on-Intelligent-Process-Automation.pdf> (visitato il 20 novembre 2020)
- Katz D.M. 2013, “Quantitative Legal Prediction Or How I Learned To Stop Worrying And Start Preparing For The Data-Driven Future Of The Legal Services Industry”, *Emory Law Journal*, Vol. 62: 909-966.
- Levi E.H.1949, *An Introduction to Legal Reasoning*, Chicago-London: The University of Chicago Press.
- Mitchell M. 2019, “We Shouldn’t be Scared by ‘Superintelligent A.I.’”, *The New York Times*, Oct. 31.
- Ordine N. 2019, “Intervista a Mark Mézard”, *La Lettura*, 27 gennaio 2019.
- Perry L. 2019, “AI Alignment Podcast: Human Cognition and the Nature of Intelligence with Joshua Greene”, available at <https://futureoflife.org/2019/02/21/human-cognition-and-the-nature-of-intelligence-with-joshua-greene/>, visitato il 24 marzo 2019.
- Prakken H., Sartor G. 2015, “Law and logic: A review from an argumentation perspective”, *Artificial Intelligence* 227.
- Santosuosso A. 2020, *Intelligenza artificiale e diritto*, Milano: Mondadori Università.

- Santosuosso A., Pinotti G. 2020, “Bottleneck or Crossroad? Problems of Legal Sources Annotation and Some Theoretical Thoughts”, *Stats* 2020, 3(3), 376-395; available at <https://doi.org/10.3390/stats3030024> – 09 Sep 2020.
- Sartor G. 2010, *L'informatica giuridica e le tecnologie dell'informazione: Corso di informatica giuridica*, Torino: Giappichelli.
- Schauer F. 1991, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Oxford: Clarendon Press.
- Seif G. 2018 (December 18), “AI doesn’t need to be fully interpretable”, Available at <https://towardsdatascience.com/ai-doesnt-need-to-be-fully-interpretable-4767b110dc1e>, accessed 19 agosto 2019.
- Russell S., Artificial intelligence: The future is superintelligent. *Nature* 548, 520–521 (2017) doi:10.1038/548520a (visitato il 30 dicembre 2019)
- Russell S. 2019, *Human Compatible*, New York: Viking Press.
- Russell S.– Norvig P., *Intelligenza Artificiale. Un approccio moderno*, a cura di Francesco Amigoni, Pearson, Prentice Hall, 2010 (I edizione inglese 1994).
- Vishal Morde, 5 Key Lessons Learned from 2018 AI Summit, NYC Hype, Reality, Ethics, Disruption and More..., Dec 10, 2018, <https://towardsdatascience.com/https-medium-com-vishalmorde-lessons-learned-from-ai-summit-d7c071cb4c06> visitato il 19 agosto 2019.