

Accorgimenti dell'intelligenza artificiale dei (per i) giuristi

Federico Casa

Abstract: This paper aims to analyze some of the issues affecting the study of artificial intelligence applied to the work of jurists, seeking to highlight how the results obtained today, although less satisfactory than those achieved in other disciplines, depend mainly on the implementation of learning algorithms. At the same time, the essay tries to “take stock” of the latest landfalls of artificial intelligence applied to law, noting how the main goal today has become predictability rather than understanding of legal experience. The contribution offers some opening paragraphs, in which the author clarifies the meaning of the terminology adopted; the concluding ones are about the relationship between legal methodology and the technology produced by artificial intelligence.

1. Logica e algoritmi

L'utilizzo dell'intelligenza artificiale (IA) è oggi considerata una scienza, anche se sempre più spesso si fa rilevare che, piuttosto che una disciplina scientifica, si tratta di una tecnologia dotata di una propria autonomia¹. Tale dibattito non interessa in questa sede; ci basti affermare che pacificamente essa intende riprodurre, a determinate condizioni, l'intelligenza umana; si discute in questo senso di àmbiti della scienza computazionale².

La discussione oggi riguarda non tanto che cosa occorra intendersi per "intelligenza umana", quanto quale sia quella componente dell'intelligenza umana (o del comportamento umano), che l'IA intenda riprodurre. Genericamente è possibile affermare che essa è definibile in relazione alla fedeltà di questa rispetto alla prestazione umana; altri orientamenti preferiscono una definizione maggiormente formale d'intelligenza artificiale, intesa come razionalità, nel senso che essa consente di precisare il ragionamento e il conseguente comportamento che deve essere tenuto in relazione a una determinata situazione. Secondo altre classificazioni ancora l'intelligenza è ora una qualità del pensiero e del ragionamento; ora una caratteristica dell'agire umano³.

Ne derivano quattro possibili definizioni, che hanno evidentemente anche diversi riferimenti filosofici e antropologici. In ogni caso, per quanto interessante, si tratta di un tema questo che, a tutto concedere, è di filosofia della scienza, e in particolare riguarda le assunzioni metafisiche che costituiscono il presupposto delle teorie scientifiche⁴. D'altro canto, le evoluzioni stesse dell'epistemologia della scienza dipendono da una diversa visione filosofica del significato e degli obiettivi del sapere scientifico (idea del progresso, rapporto con la verità, assunzioni metafisiche, relazione con la metodologia scientifica, caratteristiche della genesi di una teoria scientifica)⁵.

Nonostante il presente contributo non sia un manuale, anzi, visto che esso intende proporre una visione critica dell'IA applicata all'esperienza giuridica, rilevando alcuni accorgimenti cui essa ricorre, in ogni caso, alcune precisazioni e chiarimenti terminologici si rendono necessari. Si tratta di alcune nozioni di particolare complessità anche tecnica, che esprimeremo in modo molto semplicistico, ma che risultano necessari ai fini della comprensione delle questioni successive⁶.

¹ Fabris (2021), pp. 20-33.

² Durante (2019), pp. 23-45.

³ Russel, Norvig (2021), p. 4.

⁴ Bartley III (1983), pp. 7-36.

⁵ Buzzoni (2008), pp. 9-14.

⁶ Sartor (2021), pp. 271-300.

Come noto, un calcolatore è un esecutore generale di algoritmi; esso è in grado eseguire qualsiasi manipolazione simbolica, vale a dire qualsiasi trasformazione che conduca da certi simboli d'*input* ad altri simboli di *output*, una volta utilizzato un algoritmo che specifichi come compiere tale trasformazione⁷.

In prima approssimazione, è possibile affermare che un algoritmo è una procedura chiara e univoca, tanto precisa da poter essere attuata da un esecutore automatico, senza che a questo debbano essere fornite altre indicazioni. Si tratta, per utilizzare una metafora, di una ricetta per preparare una torta; un *software* svolge una funzione analoga a quella di una ricetta da cucina. Entrambi sono destinati a un esecutore: la ricetta richiede l'intervento del cuoco, il programma deve essere messo in atto dal computer, al quale è indicata la procedura di calcolo da seguire per realizzare un certo risultato a partire da determinati dati di *input*. Tanto il programma quanto la ricetta prescrivono una serie di procedure per risolvere un determinato problema, caratterizzato nei termini di una certa situazione di partenza (dati di *input*) e di un certo risultato (*output*) che occorre conseguire. Non vi è dubbio che, se nella ricetta le attività da svolgere sono materiali, nell'algoritmo si tratta di elaborazioni simboliche, che, con riferimento al linguaggio di programmazione e ai rapporti di questo con il linguaggio naturale, impongono la soluzione di problemi tecnici molto complessi; non è in ogni caso questa la sede per affrontare tali questioni⁸.

Sono infine definiti di *machine learning*, detti anche basati sull'apprendimento, quegli algoritmi che capovolgono il funzionamento degli algoritmi tradizionali; essi partono dai dati e dal risultato desiderato, al fine di proporre l'algoritmo che consente di passare dagli uni agli altri; si può fondatamente affermare che quelli di apprendimento sono algoritmi che scrivono altri algoritmi⁹.

Anticipando alcune nozioni che saranno meglio affrontate in seguito, ai fini che qui interessano, può essere utile proporre una classificazione degli algoritmi di *machine learning* oggi utilizzati dall'IA in relazione ai procedimenti che i loro programmatori adottano per affrontare i problemi che essi si propongono di risolvere.

Una prima griglia concettuale è relativa ai teorici della programmazione c.d. "simbolisti", secondo i quali è necessario fornire al programma alcune conoscenze iniziali, affinché esso proceda al completamento dell'inferenza¹⁰.

⁷ Crafa (2021), pp. 63-78.

⁸ Harel (2002), p. 7.

⁹ Russel, Norvig (2022), pp. 7-66.

¹⁰ Domingos (2016), pp. 77-82.

Come noto, una dimostrazione può dirsi deduttiva, allorché le sue premesse, se vere, forniscono ragioni conclusive per la verità della sua conclusione; da questo punto di vista, l'argomento o è valido o non lo è; se le sue premesse sono vere, la conclusione è sicuramente vera. L'argomento deduttivo può dirsi perlopiù decisivo nel senso che non occorre aggiungere altro a conferma della sua validità: "tutti gli uomini sono mortali/ Socrate è un uomo// Socrate è mortale" (dove / è il simbolo che indica il passaggio ad una premessa e // il passaggio alla conclusione). Nella deduzione corretta, non è possibile che le premesse siano vere e la conclusione falsa; invece, il fatto che un argomento sia valido non garantisce la verità della sua conclusione; d'altronde, la verità o la falsità della conclusione di un argomento non determina in se stessa la validità o l'invalidità di quell'argomento. La premessa maggiore contiene il predicato della conclusione (eroe nell'esempio che segue); la premessa minore il soggetto della conclusione (soldati); la conclusione indica soggetto e predicato; nelle due premesse compare il termine medio (codardi), che non è mai presente nella conclusione: "nessun eroe è codardo/ alcuni soldati sono codardi// alcuni soldati non sono eroi" Quelli che si sono maggiormente affermati nella letteratura giuridica sono i sillogismi aristotelici c.d. "categorici", in cui la conclusione è un enunciato individuale (del tipo "modo Barbara II", si dice nel gergo dei logici, per distinguerli da quelli che contengono solo enunciati universali, "modo Barbara I")¹¹.

In taluni casi il sillogismo può contenere enunciati ipotetici; questi ultimi interessano molto l'IA e il giurista. Negli anni Ottanta del secolo scorso, si era persino diffusa l'idea che attraverso questi fosse possibile sostituire l'attività del giudice, perché l'utilizzo di tale logica nella programmazione di algoritmi avrebbe consentito di riprodurre il sillogismo giudiziale¹². Certo l'idea di fondo era che l'esperienza giuridica fosse costituita solo da norme e il ragionamento del giurista fosse essenzialmente quello sillogistico¹³; oggi da allora pare trascorso un secolo. In ogni caso si abbia riguardo ad un sillogismo ipotetico-misto, in cui la premessa categorica (premesse minore) afferma l'antecedente della premessa condizionale (premesse maggiore); la conclusione afferma la sua conseguenza: "se c'è A allora B/ c'è A// allora B", questo è lo schema, che consente una dimostrazione sulla base del *modus ponens*. Questo è il modello di sillogismo giuridico proposto da un importante giurista come Robert Alexy, che adotta la simbologia della logica dei predicati¹⁴.

¹¹ Copi, Cohen (1999), pp. 257-264.

¹² Cossutta (2003), pp. 163-190.

¹³ Borruso (1997), pp. 654 ss.

¹⁴ Alexy (1998), pp. 177-182.

Il ragionamento induttivo costituisce invece una generalizzazione a partire dall'osservazione; le premesse non danno ragioni decisive per la conclusione, ma consentono di formulare un giudizio che possiede un determinato grado di probabilità, in relazione all'attendibilità della generalizzazione, la cui validità dipende spesso anche dal verificarsi di altre circostanze: "il cigno A è bianco/ il cigno B è bianco// tutti i cigni sono bianchi"; in realtà, è possibile che ci sia un cigno nero. È noto a tale proposito l'esempio del tacchino induttivista di Bertrand Russell, il quale, ogniqualvolta vedeva il proprio padrone, riteneva (con inferenza induttiva) che era giunto il momento di mangiare; e ciò fino al giorno di Natale, quando vide il proprio padrone, che però era venuto in vista del pranzo di Natale¹⁵.

Quello che i simbolisti chiedono al sillogismo di attuare è una inferenza induttiva che, secondo questa prospettiva, è denominata deduzione inversa. In altri termini, a partire da un sillogismo mancante della premessa maggiore, l'algoritmo dovrebbe individuarla e applicarla ad altri casi non previsti; questo tipo di algoritmi consente di generalizzare i dati a disposizione ai casi non previsti nell'inferenza logica; la deduzione inversa trova importanti applicazioni nell'identificazione di possibili effetti collaterali di nuovi farmaci¹⁶.

Secondo i c.d. "conessionisti", invece, l'algoritmo di apprendimento è generato sulla base dell'idea che, siccome l'organo dell'apprendimento è il cervello, l'intento è confrontare il risultato ottenuto con quello desiderato; quale risultato di questa disamina, è la possibilità modificare le relazioni tra le reti neurali; si tratta di quell'operazione che è denominata "retropropagazione"¹⁷.

I c.d. "evoluzionisti" ritengono che la madre di ogni forma di apprendimento sia la selezione naturale, e conseguentemente mettono in relazione i programmi con il modo in cui la natura fa evolvere gli organismi. Di contro, i c.d. "baynesiani" calcolano la probabilità di ogni ipotesi condizionandola ai dati osservati e su tale base formulano delle predizioni. L'idea di fondo è che, rispetto ad un accadimento, non vi siano ipotesi più vere di altre, ma solo il calcolo *a posteriori* di ogni possibile ipotesi che tenga conto di tutti i valori ottenuti nel momento in cui si fanno delle previsioni; si pensi agli studi sulla interazione dei geni rispetto a particolari malattie. La capacità di apprendere la struttura di una rete baynesiana è strettamente connessa al problema dell'estrazione d'informazioni causali dai dati. Ci si è sempre domandati se sia possibile apprendere reti baynesiane in modo tale che la rete ricostruita mostri influenze causali reali. Per molti anni gli statistici hanno evidente-

¹⁵ Larson (2022), pp. 156-186.

¹⁶ Domingos (2016), pp. 83-120.

¹⁷ Charniak (2019), pp. 7 ss.

mente preferito evitare la questione, ritenendo che i dati ottenuti mediante osservazioni, contrapposti a quelli generati con *test* sperimentali, potessero fornire solo informazioni a livello di correlazione. Nei primi anni di questo secolo si è cominciato a sostenere il contrario, dimostrando che in molti casi è effettivamente possibile accertare la causalità, in modo tale che l'algoritmo consenta di esprimere oltre alla consueta probabilità condizionata, le cause e gli effetti di un'azione¹⁸.

Da ultimo, per i c.d. "analogisti" è necessario individuare la regola di una situazione, al fine di poterla applicare a contesti somiglianti. Il ragionamento analogico ha lo scopo d'integrare la conoscenza di un oggetto meno conosciuto ricorrendo ad uno o diversi oggetti più conosciuti con cui il primo ha in comune delle proprietà rilevanti¹⁹. L'analogia, come noto, anche in ambito giuridico, consta di un'inferenza abduttiva, che consiste nella determinazione della proprietà rilevante; di una componente induttiva, che è rappresentata dall'operazione di generalizzazione a séguito dell'osservazione; infine, di una componente deduttiva, secondo la quale l'oggetto meno conosciuto avrà la proprietà dell'oggetto conosciuto²⁰.

In letteratura è stato rilevato che anche gli algoritmi di apprendimento faticano a svolgere il ragionamento abduttivo, il quale mira a formulare un'ipotesi che spieghi quanto riscontrato nella realtà; si tratta di una congettura sulla causa di un accadimento, che non può mai essere proposta attraverso un argomento deduttivo. Un tipico argomento abduttivo è l'inferenza alla migliore spiegazione, che si caratterizza per l'idea che, se un'ipotesi fornisce una buona spiegazione di un insieme di fenomeni, e se nessuna ipotesi fornisce una spiegazione altrettanto buona, allora è probabile che l'ipotesi sia vera²¹. A differenza dell'induzione che generalizza delle osservazioni, l'abduzione è un'inferenza che formula delle ipotesi esplicative e che trova il suo ambito di applicazione, per quanto riguarda il giurista, soprattutto nella determinazioni delle questioni di fatto e probatorie (la premessa minore del sillogismo giudiziale): "Tizio ha subito un'intossicazione alimentare/ ieri sera ha mangiato del pesce crudo nel ristorante di Caio// ieri Caio nel suo ristorante ha venduto del pesce avariato". Più generalmente: "è probabile che se c'è F (fatto), c'è I (indizio)/ c'è I// è probabile che ci sia F"²².

Tali notazioni consentono anche di chiarire la differenza dell'abduzione dal c.d. "sillogismo statistico", il quale mira ad ottenere una conclusione del ragionamento solo probabile. Infatti, è vero che una determinata conclusione

¹⁸ Pearl (2010), pp. 34 ss.

¹⁹ Ng (2019).

²⁰ Tuzet (2020), pp. 21-34.

²¹ Iacona (2010), pp. 67-70.

²² Carcaterra (2012), pp. 166- 167.

può essere solo probabile, ma ciò accadere per due ragioni ben distinte: o perché è probabile la premessa (si pensi all'applicazione di una massima di esperienza), oppure perché si tratta di un argomento induttivo²³.

Se l'algoritmo apprende da un insegnante oppure dall'ambiente stesso, si tratta di un "apprendimento supervisionato"²⁴; in ambito giuridico, esso intende suggerire una soluzione per un caso futuro, avendo appreso informazioni dagli archivi giurisprudenziali, e facendo riferimento, nel suggerire la soluzione, al contenuto della sentenza cui ha improntato la decisione. Nell'apprendimento per "rinforzo", il programma è in grado di valutare il risultato della propria azione ed eventualmente somministrarsi una ricompensa o una punizione; questi sistemi non sono utilizzati in ambito giuridico²⁵. Infine, il c.d. "deep learning", anche detto "apprendimento non supervisionato", il quale è utilizzato per riunire insieme oggetti che presentano somiglianze e/o connessioni rilevanti, anche non immediatamente evidenti. L'idea di fondo è che, data l'enorme quantità di testi disponibili (anche dal punto di vista giuridico) e leggibili da parte delle macchine, approcci basati sull'apprendimento automatico possono dare risultati migliori rispetto a quelli fondati su analisi semantiche, le quali presuppongono risolti importanti questioni correlate al linguaggio. D'altro canto, con il *deep learning* il compito di risolvere la comprensione di nuove parole o frasi è affrontato attraverso la determinazione di spunti base forniti da parole con significati analoghi che appaiono in contesti simili²⁶.

Torneremo in séguito sulla questione, e ciò con particolare riferimento all'esperienza giuridica. Allo stato ci basti rilevare che il comune denominatore degli algoritmi di apprendimento (compresi quelli giuridici) consiste nella considerazione che si è modificato il parametro stesso del suo "successo". Fino agli anni Ottanta del secolo scorso il comportamento intelligente doveva emergere da inferenze logiche eseguite sulla base di conoscenze dichiarative; si pensi ad un algoritmo per eseguire diagnosi mediche partendo da "principi primi" in un sottocampo specialistico. Oggi è ormai abbandonato l'approccio, secondo il quale occorre identificare un meccanismo nascosto, quello che ha generato i dati stessi che occorre analizzare, e ciò a favore del compito più semplice di predire gli avvenimenti futuri. Si tratta di generare un comportamento che, a partire da semplici *pattern* statistici (le relazioni che troviamo nei dati), sia "probabilmente approssimativamente corretto". Un po' come decidere di agire ignorando quello che si proponeva

²³ Canale, Tuzet, (2021), pp. 48-49.

²⁴ Russel, Norvig, (2022), pp. 9-13.

²⁵ L'aveva proposto già Turing (1950), pp. 433-460.

²⁶ Russel, Norvig (2022), pp. 213-235.

Albert Einstein, il quale, prima di agire, voleva “conoscere i pensieri di Dio”; l’intento, infatti, è conseguire i propri obiettivi anche senza avere appreso le regole che governano il mondo²⁷.

2. Definizioni e riferimenti filosofici

Possono darsi quattro possibili definizioni d’intelligenza artificiale, a seconda di quello che la scienza computazionale ritenga di dover riprodurre: (i) l’agire umano; (ii) il pensare razionale; (iii) il pensare in modo umano; (iv) l’agire in modo razionale. Alcune brevi notazioni su questi diversi modi di ricostruire l’IA aiutano a capire non solo i suoi mutati obiettivi e pertanto il suo funzionamento, ma anche il rapporto di questa con l’esperienza giuridica²⁸.

Ciascuna di queste definizioni è affermata e seguita in letteratura; ognuna di queste potrebbe essere quindi accettata, anche se esse danno conto in misura diversa della natura, dei fini e dei procedimenti dell’IA, accentuandone alcuni profili rispetto ad altri.

Con riferimento alla nozione *sub* (i), sia solo consentito osservare che il noto test di Turing intendeva verificare se il computer fosse un grado di superare un test, in cui un esaminatore umano, dopo aver posto alcune domande in forma scritta, fosse in grado di capire se le risposte provenissero da una persona. Il test consisteva in una comunicazione tra un essere umano e un computer in una stanza, ed un altro essere umano in una stanza separata con il compito d’interrogarli, per il solo tramite di domande e risposte scritte. Qualora l’interrogante non fosse riuscito a distinguere la macchina dall’uomo, allora questa poteva essere ritenuta intelligente²⁹.

La definizione *sub* (ii) mette in risalto la possibilità di riprodurre in un programma le “leggi del pensiero”; d’altro canto, tutta la tradizione logica della filosofia occidentale ha sempre inteso formulare enunciati riguardanti gli oggetti del mondo e le relazioni tra questi. Infatti, già a metà degli anni Sessanta esistevano programmi che ritenevano di essere in grado di risolvere, in linea di principio, qualsiasi problema che potesse essere descritto in termini logici. Lungo questa direzione, per fare qualche esempio, i problemi della politica e della guerra, le quali escludono la completa conoscenza delle regole che le governano, hanno visto il ricorso alla teoria della probabilità, che permette un ragionamento rigoroso pur in presenza d’informazioni incerte (sono deduzioni a partire da leggi di esperienza). Dal punto di vista della storia

²⁷ Cristianini (2023), pp. 37-52.

²⁸ Mitchell (2022), pp. 27-54.

²⁹ Nilsson (2005), 87-103.

della filosofia, si tratta dell'idea di riprodurre quella concezione del pensiero razionale, secondo la quale occorre accettare come sensate solo le descrizioni della conoscenza basate su regole d'inferenza logiche appartenenti a un sistema formale non auto-contraddittorio. A ben vedere, l'idea di ridurre l'inferenza logica a un processo puramente meccanico applicato a un linguaggio formale era già stata proposta da Wilhelm Leibniz alla fine del Seicento³⁰. Come altrettanto noto, George Boole nel 1847 aveva introdotto il primo sistema completo e funzionale di logica formale nel suo libro *The Mathematical Analysis of Logic*. La logica di Boole era modellata sulla classica algebra dei numeri reali e usava come metodo principale d'inferenza la sostituzione di espressioni logicamente equivalenti; per quanto avanzata, essa non era però riuscita a esprimere tutta la potenza del calcolo proposizionale³¹.

Da una prospettiva più strettamente filosofica, sin dai primi anni del Novecento lo snodo cruciale è divenuto la designazione del rapporto linguaggio-significato come determinazione del metodo di verifica o falsificazione di un'affermazione; come noto, oggi si discute di verità/falsità di un enunciato (dal punto di vista linguistico), di una proposizione (dal punto di vista logico), di una credenza (dal punto di vista semantico). Si tratta della lezione non prescindibile dell'empirismo logico della Scuola Analitica soprattutto del Circolo di Vienna. Per Gottlob Frege, dal senso (*Sinn*) e dalla denotazione (*Bedeutung*) di un'espressione va tenuta distinta la rappresentazione (*Vorstellung*), traducibile anche con il termine "significato", poiché essa è un'immagine soggettiva, mentre solo il senso è oggettivo. Sarà in questo modo affermata la pretesa di poter costruire un sistema formale non auto-contraddittorio e completo, in cui i significati sono introdotti come affermazioni di verità/falsità riferite a oggetti determinati. Risalendo a questi presupposti, la prima esposizione completa della logica proposizionale moderna è rinvenibile nell'opera *Begriffsschrift* del 1897 dello stesso Frege; oggi è soprattutto utilizzata la logica matematica del primo ordine, che estende l'ambito di applicazione della logica proposizionale mediante l'introduzione di quantificatori esistenziali e universali, predicati, funzioni variabili e costanti. In Carnap la tripartizione di Frege diventava la bipartizione estensione/intensione; e ciò sulla base della tesi che gli oggetti abbiano una loro realtà indipendente dalle classificazioni umane e che tra oggetto e simboli linguistici esista una corrispondenza bidirezionale; l'intento era evidentemente superare la tripartizione di Frege. Il presupposto teorico accettato anche da Wittgenstein era, infatti, che il raggiungere le unità teoriche elementari potesse permettere di affermare l'essere vero/non vero di ogni

³⁰ Pagallo (2005), pp. 19-57.

³¹ Russel, Norvig (2021), p. 255.

cosa, basandosi sulla diretta percezione dell'oggetto (prova empirica). Questo progetto filosofico-scientifico mirava a ricostruire tutti i passaggi che dalla sensazione portano all'elaborazione di un concetto e alle rappresentazioni complesse. Il sistema logico-formale permetteva allora di giungere a ulteriori risultati anche in assenza della prova empirica e quindi di supporre l'esistenza di realtà non percepite o non percepibili. In questa prospettiva, la logica, e non la matematica come voleva Boole, è proposta come capace di rappresentare il sistema formale di tutti i "sistemi formali", in grado quindi di ricomprendere all'interno della sua formalizzazione anche la matematica stessa. L'idea di fondo è che, a partire da pochi referenti "primitivi" nel mondo della natura, sia possibile, ragionando all'interno di un sistema formale, giungere alla scoperta del "vero" riguardo al mondo³².

Ai nostri fini, interessa osservare come, secondo questa nozione dell'IA, è soprattutto evidenziato il ruolo della logica proposizionale (e della logica del primo ordine) nella costruzione di algoritmi e programmi informatici. Il grado di verità rispetto alla realtà di una formula, risultato di un algoritmo, può essere inoltre oggetto di un vero e proprio impegno epistemologico: rispetto a ciascun fatto, è possibile esprimere la verità della proposizione che lo rappresenta rispetto a un mondo possibile³³. In questi termini, invece, le teorie basate su algoritmi probabilistici consentono di esprimere qualsiasi grado di credenza (da zero a uno).

L'IA che si è sviluppata nel corso degli anni Sessanta e Settanta del secolo scorso mirava allora a progettare programmi e costruire algoritmi che, una volta identificati i problemi, raccogliessero la conoscenza rilevante, la definissero attraverso un vocabolario di oggetti, predicati e connettivi logici, per poi, attraverso le regole d'inferenza, ottenere una proposizione finale (conclusione) che consentisse di fornire una risposta al quesito iniziale. Una caratteristica applicazione dei programmi fondati sulla conoscenza (e non sull'apprendimento) sono quelli i cui algoritmi sono basati sulla "concatenazione in avanti"; essi non sono se non un'applicazione informatica della dimostrazione sulla base del *modus ponens*, già proposto nel 1897 da Frege, il quale aveva basato il suo sistema d'inferenza (logica del prim'ordine) su una collezione di schemi logicamente validi, a cui aggiungere una singola regola d'inferenza (appunto il *modus ponens*). Dato un insieme di fatti noti, il concatenamento in avanti esamina se le condizioni di qualche regola ("se tutti gli uomini sono mortali") possano dirsi soddisfatte per altri fatti (non solo "se Socrate è un uomo" ma anche "se Socrate è un filosofo"). In caso positivo si aggiunge la conclusione della regola ai fatti noti; si passa poi a

³² Romeo (2012), pp. 87-108.

³³ Davidson (2006), pp. 107-145.

considerare se, grazie alle nuove conclusioni, sia possibile trarre, allo stesso modo, le conclusioni di altre regole (aggiungendo anch'esse ai fatti noti), e così via. Il processo termina quando si è conseguita la conclusione desiderata o si sia stabilita l'impossibilità di raggiungerla. La concatenazione in avanti sulla base della dimostrazione del *modus ponens* è sicura, perché essa consente d'ignorare le proposizioni irrilevanti. Invece, gli algoritmi basati sul concatenamento all'indietro procedono nel senso opposto; si tratta di una inferenza logica non deduttiva, poiché, data una conclusione ("allora Socrate è mortale"), essa ricerca una regola ("se tutti gli uomini sono mortali"), il cui conseguente ("Socrate è un uomo") coincida con la conclusione stessa³⁴.

Con riferimento invece alla definizione *sub* (iii) ("pensare umanamente"), l'IA deve perlopiù occuparsi d'introspezione (catturare i pensieri mentre scorrono), di sperimentazione psicologica (osservare una persona in azione) e del c.d. "*imaging* cerebrale" (osservare il cervello in azione)³⁵. Se il comportamento del *software* corrisponde, per quanto riguarda il suo *input/output*, a quello di una persona, potrebbe essere provato che alcuni dei meccanismi del programma operano anche negli esseri umani. Si tratta di riprodurre il funzionamento della mente umana; in questa prospettiva, come noto, non vi è ormai alcuna opposizione tra funzionamento della mente e del cervello, poiché il rapporto biologico è ben precisato. Appartiene, infatti, agli anni Ottanta del secolo scorso la discussione che opponeva il paradigma simbolico cognitivista "*rules and representations*", in cui il patrimonio di conoscenze e relazioni è inserito al momento della programmazione, e gli algoritmi che invece simulano l'attività cerebrale (neuroscienze)³⁶.

Allen Newell e Herbert Simon, che nel 1961 avevano sviluppato il *General Problem Solver*, non si erano infatti accontentati di scrivere un programma che risolvesse problemi, ma avevano anche inteso confrontare tale programma con la sequenza e la temporizzazione di quello che essi avevano osservato essere il ragionamento umano. È il campo interdisciplinare delle scienze cognitive, che è in grado di unire modelli computazionali sviluppati dall'IA e tecniche di sperimentazione psicologica, e ciò nell'intento di costruire teorie precise e verificabili sul funzionamento della mente umana. Anzi, oggi è possibile affermare che si è ormai precisata la differenza tra un algoritmo che esegue efficacemente un'azione umana, e quindi è un buon modello dell'esecuzione umana, da un algoritmo che funziona bene, a prescindere dal suo riprodurre il pensiero e l'azione umana³⁷.

³⁴ Russel, Norvig (2021), pp. 289-317.

³⁵ Cimmino (2012), pp. 105-143, con riferimento al metodo di analisi delle scienze della mente.

³⁶ Domingos (2016), pp. 121-149.

³⁷ Russel, Norvig (2021), p. 5.

Le scienze cognitive si distaccano però dall'analisi del linguaggio, adottando un metodo scientifico, che non è solo logico-dimostrativo e fondato sui giudizi analitici a priori, così come aveva fatto la Scuola Analitica, ma anche sintetico a priori, come nel caso dell'induzione completa aritmetica, e a posteriori, come nelle scienze della natura; è stato giustamente rilevato che le scienze cognitive hanno ripreso l'ipotesi convenzionalista e relativista di Henri Poincaré. L'intendimento di un'IA così costruita non è tanto allora proporre proposizioni corrette, ma riprodurre il sistema cognitivo umano; il senso (*Sinn*) non è oggettivo, poiché è costruito all'interno di un sistema cognitivo, in cui la percezione e l'aspetto irrazionale hanno la precedenza nelle determinazioni sia pratiche sia epistemiche dell'uomo, mentre la razionalità ha solo una funzione di controllo. La realtà non solo non è separabile dall'osservatore, ma le classificazioni del mondo sono persino teleologicamente orientate; esse non definiscono la realtà risalendo a parametri direttamente derivanti da essa, ma commisurandola alle caratteristiche anche biologiche dell'individuo che la percepisce e classifica³⁸.

Come si diceva, l'idea di riprodurre il pensiero umano è patrimonio acquisito della modellazione cognitiva, secondo la quale occorre comprendere il pensiero umano a partire dalle caratteristiche naturali dell'uomo, dal cervello, dalle sue capacità cognitive, linguaggio compreso; in questo senso l'approccio delle neuro-scienze all'IA è perfettamente compatibile con la modellazione cognitivista³⁹.

3. Agenti intelligenti e sistemi giuridici.

In letteratura è però largamente prevalsa la definizione *sub* (iv), al punto tale che la teoria degli agenti razionali è divenuta il modello *standard* dell'IA; secondo questa prospettiva, l'IA costituisce la scienza che progetta e sviluppa agenti razionali, i quali sono in grado di operare autonomamente, percepire l'ambiente, persistere in un'attività per un lungo arco di tempo, capaci ad ottenere il migliore risultato oppure, in situazioni d'incertezza, il miglior risultato atteso. In questa prospettiva, detto banalmente, la razionalità è intesa matematicamente ed è misurabile in relazione ai risultati conseguiti in un determinato ambiente (utilità); infatti, un agente razionale si comporta in modo tale da massimizzare il valore atteso della misura della sua prestazione. Tale approccio è condiviso anche dalla tradizione della psicologia cognitiva nell'area della risoluzione dei problemi; gli obiettivi sono suddivisi in “desi-

³⁸ Romeo (2012), pp. 100-115.

³⁹ Seung (2013).

deri” (generali) e “intenzioni” (perseguite in un dato momento)⁴⁰. L’etica del comportamento degli agenti razionali è evidentemente consequenzialista, poiché l’azione preferibile è quella che consegue la massima utilità rispetto all’obiettivo prefissato⁴¹.

Questo approccio è particolarmente significativo, perché esso libera l’IA non solo da ogni mediazione e condizionamento filosofico, ma non intende nemmeno più attribuire un senso alla ricerca informatica come corollario di una determinazione teorica, né ritrovare il meccanismo “nascosto” di cui i dati sono una semplice emergenza⁴². Non siamo in grado di dire con sicurezza se ciò corrisponda ad una puntuale opzione epistemologica; probabilmente no⁴³. Quello che è sicuro è che essa ha un risvolto metodologico importante. Il comportamento dell’agente prescinde da ogni possibile comprensione della realtà; ciò che caratterizza il suo agire non è l’effetto di una determinata concezione del mondo. In questi termini l’obiettivo non è più tanto quello di regolare il proprio agire con riferimento ad una precisa teoria scientifica o filosofica e così conseguire gli scopi affermando qualcosa di vero o di relativamente vero rispetto alla realtà, quanto proporre previsioni. Non si tratta di scoprire i rapporti di causa che regolano il mondo, ma di suggerire delle probabili future regolarità. Il comportamento dell’agente intelligente è misurato in termini di conseguimento di risultati; queste sono le migliori previsioni possibili, anche a costo di non sapere comprendere come queste siano state conseguite e quale altra possibile interpretazione possa essere loro attribuita. È stato persino rilevato che, se visto dalla prospettiva della filosofia della scienza, tutto ciò costituisce un ulteriore cambio di paradigma, così come lo era stata in fisica la teoria dei quanti rispetto alla meccanica di Newton⁴⁴.

Ciò chiarito, è possibile affrontare il nodo cruciale di questo contributo, e così domandarsi se i sistemi di IA aiutino il giurista a risolvere uno dei suoi problemi cruciali: la soluzione “giusta” del caso concreto⁴⁵.

Il problema è notissimo; alla sua soluzione sin dagli anni Ottanta del secolo scorso erano dedicati i sistemi esperti, i quali dovevano essere ben distinti da quelli che erano utilizzati dall’informatica giuridica documentaristica, che si occupa dei programmi informatici che agevolano la ricerca nelle banche dati giuridiche. I sistemi esperti contengono una base di conoscenza costituita da regole, che sono elaborate da algoritmi, i quali applicano i procedimenti della logica (perlopiù deduttiva) al fine di risolvere casi concreti; naturalmente

⁴⁰ Bratman (1987), pp. 47 ss.

⁴¹ Palazzani (2020), pp. 52-80.

⁴² All’opposto di ciò che sta accadendo nella fisica teorica, cfr. H. Päs (2024), pp. 20-45.

⁴³ Larson (2022), pp. 238-250.

⁴⁴ Cristianini (2023), pp. 42-43.

⁴⁵ Lo Giudice (2023).

tutta la conoscenza è fornita all’algoritmo al momento della programmazione⁴⁶. L’idea di fondo, per quanto spesso opportunamente sottaciuta, è che l’esperienza giuridica si riduca ad una serie di regole e che il ragionamento giuridico sia quello proprio del sillogismo giuridico⁴⁷.

Con l’eccezione di alcuni sistemi diffusi soprattutto nella pubblica amministrazione (australiana, inglese ed olandese) e a scopi ben limitati, non sono oggi noti sistemi giuridici basati sulla conoscenza che possano risolvere casi controversi. È vero che il sistema “HYPO”, che si appoggia su una conoscenza giuridica costituita da precedenti, soprattutto nell’ambito della violazione dei segreti industriali, ha dato prova di una qualche utilità, ma a fini molto limitati e con riferimento a questione controverse particolarmente semplificate. Esso dovrebbe essere in grado di determinare, a partire dalle fattispecie concrete oggetto dei casi già decisi, ricorrendo all’analogia dei “fattori” (argomenti a favore dell’attore o del convenuto), come possa concludersi una controversia avente ad oggetto un caso, con riferimento al quale non vi sia stata ancora alcuna controversia decisa in sede giudiziale⁴⁸.

Diversamente si pone la questione con riferimento ai sistemi programmati a partire da algoritmi di apprendimento. Tra quelli che si stanno affermando in ambito giuridico, un buon esempio è rappresentato dal programma “Claudette”, attraverso il quale è possibile riconoscere, attraverso l’apprendimento automatico supervisionato, le clausole potenzialmente abusive dei contratti conclusi tra i consumatori e le piattaforme *on line*. L’operazione consiste nell’addestrare alcuni algoritmi, affinché insegnino al sistema ad effettuare determinazioni e valutazioni analoghe a quelle specificate e implementate durante l’attività di addestramento. Tali algoritmi (analogisti) generano un “modello”, che fornisce al sistema criteri per la determinazione della vessatorietà di alcune clausole non note. Appare possibile affermare che il sistema, sulla base degli esempi forniti dai giuristi, mira a costruire un concetto di vessatorietà delle clausole, applicabile anche a pattuizioni diverse da quelle che i giuristi, che hanno implementato il programma, hanno ritenuto abusive. Fissata la valutazione dei suoi risultati sulla base dei criteri del richiamo, precisione e della media armonica tra precisione e richiamo, le risposte per l’80% dei casi coincidono con quelle fornite da giuristi esperti⁴⁹. Il risultato appare molto soddisfacente; occorre però proseguire nell’analisi degli algoritmi di apprendimento, e ciò per meglio comprendere che cosa stia accadendo in ambito giuridico.

⁴⁶ Taddei Elmi (2007), pp. 87-125.

⁴⁷ McCarty (1990); pp. 189-200.

⁴⁸ Sartor (2022), pp. 331-334, 343-346.

⁴⁹ Sartor (2022), pp. 347-356.

4. Accorgimenti.

Anche nell'ambito dell'attività degli studi legali si sono ormai affermati programmi che consentono la stesura di statuti e contratti commerciali, anche complessi, in relazione ai quali l'utente (l'avvocato) può limitarsi a specificare le clausole che egli ritiene opportuno inserire ai fini della regolamentazione di quel particolare rapporto societario o commerciale (un patto successorio valido in caso di morte di un socio; la clausola che stabilisce un accordo di esclusiva a carico di un agente commerciale). Contemporaneamente, vanno anche raffinandosi programmi che consentono di determinare, rispetto ad una determinata questione, pur proposta in modo astratto, quali siano le decisioni giurisprudenziali di legittimità e di merito che siano andate in un senso e quelle che sono state di senso contrario. Ciò avviene attraverso la c.d. "indicizzazione semantica", che organizza i termini da utilizzare nella ricerca specificando i rapporti tra gli stessi, non tanto attraverso parole chiave e/o connettori logici, ma tramite la ricerca semantica, secondo la quale il risultato che offre il programma appare il frutto della comprensione del significato contestuale dei termini oggetto di ricerca.

Il tema è di grande interesse ed ha in parte a che vedere con la predizione dei casi giudiziari e/o di quelle circostanze che possano influire sulla decisione o determinarla nel suo contenuto; si pensi ai sistemi che determinano in capo al condannato il rischio della recidiva (che comportano un inasprimento della pena) e possibili violazioni della libertà vigilata. Occorre però segnalare che un conto è la previsione del giudice medio, un conto è la decisione corretta, che in questa prospettiva potrebbe essere predicabile solo in casi molto limitati; si pensi alla verifica *a posteriori* dei comportamenti tenuti da imputati che hanno goduto della libertà vigilata, la quale (sola) effettivamente potrebbe consentire decisioni più "giuste"⁵⁰.

Proviamo allora a rispondere alla domanda iniziale: in che modo gli algoritmi di apprendimento possono aiutare il giurista a conseguire la soluzione giusta? Oggi il giurista può conoscere come sia già stata decisa la questione giuridica che gli è sottoposta. Lo poteva già fare con l'informatica giuridica documentaristica, anche se la ricerca semantica rende certamente più accurati i risultati ottenuti. E ciò non tanto però perché egli sia in grado di riprodurre modelli ontologici dell'esperienza giuridica o il ragionamento del giurista, quanto perché egli ha a disposizione un numero di dati sull'esperienza giuridica, che gli consentono di utilizzare algoritmi di apprendimento,

⁵⁰ Zaccaria (2022), pp. 119-144.

i quali evidenziano connessioni tra gli stessi e la determinazione di correlazioni probabilistiche inattese⁵¹.

In altri termini, attraverso gli algoritmi di apprendimento, che utilizzano l'analogia e le reti baynesiane, il giurista oggi, rispetto alla sua controversia, ha più elementi di un tempo; ma ciò non tanto perché egli sia in grado di meglio comprendere le ragioni dei contendenti⁵², oppure perché abbia imparato a formalizzare più accuratamente le c.d. "ontologie" giuridiche⁵³, quanto perché i dati dell'esperienza giuridica non sono difficilmente rappresentabili come semplici informazioni (le leggi in vigore, le sentenze di merito e di legittimità, le teorie della dottrina, gli atti giudiziari di un procedimento).

Certo, alcune decisioni della Corte Suprema statunitense sono state previste in relazione all'oggetto, alla procedura che ha condotto alla decisione, all'orientamento politico dei giudici; lo stesso dicasi per l'esito di alcune controversie in tema di brevetti, e ciò con riferimento alle caratteristiche delle parti, degli avvocati e dei giudici. Eppure, il risultato è stato conseguito, perché oggi è possibile elaborare un numero di dati persino difficilmente misurabile. I risultati conseguiti e/o conseguibili in ambito giuridico dipendono dall'utilizzo degli stessi metodi che consentono di estrarre modelli predittivi da un grande insieme di dati (i *big data*). Ciò che rende grande una "massa" di dati è una caratteristica funzionale: la possibilità di scoprire correlazioni e fare previsioni; come noto, gli algoritmi che operano su questi dati sono in grado di annotarsi e ricordare le loro previsioni e sulla base di queste fare altre previsioni ancora⁵⁴.

D'altro canto, chi opera con l'IA ha abbandonato l'idea che le macchine intelligenti possano identificare i "veri" meccanismi dietro ai dati che analizzano; ci si accontenta dell'obiettivo di fare previsioni utili, anche se queste non insegnano nulla del modello che riproducono⁵⁵; in questo senso l'IA ha rinunciato alla realtà in nome della prevedibilità.

Si tratta di un accorgimento, una scorciatoia è stato detto, un ridimensionare le ambizioni sulla base dell'idea che in ogni caso ciò comporta un progresso. Certo, da un altro punto di vista, gli agenti dell'IA per natura sono funzionali ad ogni genere di operazioni, anche di politica legislativa⁵⁶; rimane il problema che in questo modo il giurista corre il rischio di sapere di avere probabilmente delle ragioni da proporre, ma di non ritenere necessario conoscerle.

⁵¹ Barberis (2022).

⁵² Moro; Bettarello, Venezia (2015), pp. 1-46.

⁵³ Sartor (2022), pp. 178-181.

⁵⁴ Sarra (2017), pp. 47-53.

⁵⁵ Cristianini (2023), pp. 65-66.

⁵⁶ Santosuosso, Sartor (2022), pp. 1760-1782.

Riferimenti

- Alexy (1998) = Alexy R., *Teoria dell'argomentazione giuridica. La teoria del discorso razionale come teoria della motivazione giuridica* (1978), tr. it. Mazzoni C.M, Varano V., Giuffrè, Milano 1998.
- Barberis (2022) = Barberis M., *Giustizia predittiva: ausiliare e sostitutiva. Un approccio evolutivo*, in *Milan Law Review*, 2/2022.
- Bartley III 1983 = Bartley III W.W., *Come demarcare la scienza dalla metafisica*, tr. it. Prodi E., Borla, Roma, 1983.
- Borruso (1997) = Borruso R., voce *Informatica giuridica*, in “App. Enciclopedia del Diritto”, Milano, Giuffrè 1997, pp. 654 ss.
- Bratman (1999), = Bratman M., *Intentions, Plans and Practical Reason*, CSLI Publications, Stanford 1999.
- Buzzoni (2008) = Buzzoni M., *Filosofia della scienza*, Morcelliana, Brescia 2021.
- Canale, Tuzet (2021) = Canale D., Tuzet G., *La giustificazione della decisione giudiziale*, Giappichelli, Torino (2021).
- Carcattera (2012) = Carcattera G., *Presupposti e strumenti della scienza giuridica*, Giappichelli, Torino 2012.
- Charniak (2019) = Charniak E., *Introduction to Deep Learning*, Mit Press, Boston 2019.
- Cimmino 2012 = Cimmino L., *Introduzione all'epistemologia della mente*, Rubbettino, Soveria Mannelli, 2012.
- Copi, Cohen (1999) = Copi I.M, Cohen C., *Introduzione alla logica*, tr.it. Lupacchini R., il Mulino, Bologna 1999.
- Cossutta (2003) = Cossutta M., *Questioni sull'informatica giuridica*, Giappichelli, Torino 2003.
- Crafa (2021) = Crafa S., *Dalle competenze alla consapevolezza digitale: capire la complessità e la non neutralità del software*, in Moro P. (a cura di), *Etica, Diritto e Tecnologia. Percorsi dell'informatica giuridica contemporanea*, FrancoAngeli, Milano 2021, pp. 63-78.
- Cristianini (2023) = Cristianini N., *La scorciatoia. Come la macchine sono diventate intelligenti senza pensare in modo umano*, il Mulino, Bologna, 2023.
- Davidson (2006) = Davidson R., *Sulla verità* (2005), tr. it. Levi s., Editori Laterza, Roma-Bari 2006.

- Domingos (2016) = P. Domingos, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, tr. it. Migliori A., Bollati Boringhieri, Torino 2016,
- Durante (2019) = Durante M., *Pensiero computazionale. L'impatto delle ICT su diritto, società, sapere*, Meltemi, Milano 2019.
- Fabris (2021) = Fabris A., *Etica delle nuove tecnologie*, Zanichelli, Bologna 2021.
- Iacona (2010) = Iacona A., *L'argomentazione*, Einaudi, Torino 2010, pp. 67-70.
- Harel (2002) = Harel D., *Computer a responsabilità limitata. Dove le macchine non riescono ad arrivare*, tr. it. Civalleri L., Einaudi, Torino 2002.
- Larson (2022) = Larson E.J., *Il mito dell'intelligenza artificiale. Perché i computer non possono pensare come noi* (2021), tr. it. Micalizzi P., FrancoAngeli, Milano 2022.
- Lo Giudice (2023) = Lo Giudice A., *Il dramma del giudizio*, Mimesis, Milano 2023.
- McCarty (1990) = McCarty L.T., *Artificial Intelligence and Law: How to Get Them From Here*, in *Ratio Juris*, vol. III (1990), pp. 189-200.
- Mitchell (2022) = Mitchell M., *Intelligenza artificiale. Una guida per essere umani pensanti*, tr. it. Ferraresi S., Einaudi, Torino 2022.
- Moro (2015) = Moro P. (a cura di), *CollectiUS. Una piattaforma giuridica interattiva*, Giappichelli, Torino 2022.
- Ng (2024) = Ng A., *Machine Learning Yearning*, DeepLearning AI, 2024.
- Nilsson (2005) = Nilsson N.J., *Intelligenza artificiale*, Apogeo, Milano, 2005.
- Pagallo (2005) = Pagallo U., *Introduzione alla filosofia digitale*, Giappichelli, Torino 2005.
- Palazzani (2020) = Palazzani L. *Tecnologia dell'informazione e intelligenza artificiale. Sfide etiche al diritto*, Studium, Roma 2020.
- Päs (2024) = Päs H., *L'Uno. L'idea antica che contiene il futuro della fisica* (2023), tr. it. Frediani S., Bollati Boringhieri, Torino 2024.
- Pearl (2010) = Pearl J., *An Introduction To Causal Inference*, Berkeley Electronic Press, 2010.
- Romeo (2012) = Romeo F., *Lezioni di logica ed informatica giuridica*, Giappichelli, Torino 2012.
- Russel, Norvig (2021) = Russel S., Norvig P., *Intelligenza artificiale. Un approccio moderno*, vol. 1, tr.it. Amigoni F., Pearson Italia, Milano-Torino, 2021.

- Russel, Norvig (2022) = Russel S., Norvig P., *Intelligenza artificiale. Un approccio moderno*, vol. 2, tr.it. Amigoni F., Pearson Italia, Milano-Torino, 2022.
- Santosuosso, Sartor (2022) = Santosuosso A., Sartor G., *La giustizia predittiva: una visione realistica*, in *La giustizia predittiva*, *Giur. it.*, 2022, pp. 1760–1782.
- Sarra (2017) = Sarra C., *Business Intelligence ed esigenze di tutela: criticità del c.d. Data Mining*, in Id., Moro P. (a cura di), *Tecnodiritto. Temi e problemi di informatica giuridica e robotica*, FrancoAngeli, Milano 2017, pp. 41-63.
- Sartor (2022) = Sartor G., *L'informatica giuridica e le tecnologie dell'informazione*, Giappichelli, Torino 2022.
- Seung (2013) = Seung S., *Connettoma. La nuova geografia della mente*, tr.it., Codice, Torino 2013.
- Taddei Elmi (2007) = Taddei Elmi G., *Corso di Informatica Giuridica*, Edizioni giuridiche Simone, Napoli 2007.
- Turing (1950) = Turing A.M., *Computer Machinery and Intelligence*, in *Mind*, LIX (1950), pp. 433-460.
- Tuzet (1920) = Tuzet G., *Analogia e ragionamento giuridico*, Carocci, Roma 2020.
- Zaccaria (2022) = Zaccaria G., *Postdiritto. Nuove fonti, nuove categorie*, il Mulino, Bologna 2022.

