

Giovanni Sartor

Introduzione

(doi: 10.4477/97020)

Rivista di filosofia del diritto (ISSN 2280-482X)

Fascicolo 1, giugno 2020

Ente di afferenza:

Università degli Studi Mediterranea di Reggio Calabria (Unirc)

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

L'articolo è messo a disposizione dell'utente in licenza per uso esclusivamente privato e personale, senza scopo di lucro e senza fini direttamente o indirettamente commerciali. Salvo quanto espressamente previsto dalla licenza d'uso Rivisteweb, è fatto divieto di riprodurre, trasmettere, distribuire o altrimenti utilizzare l'articolo, per qualsiasi scopo o fine. Tutti i diritti sono riservati.

L'INTELLIGENZA ARTIFICIALE E IL DIRITTO

a cura di Giovanni Sartor

Introduzione

1. Lo sviluppo dell'intelligenza artificiale

L'intelligenza artificiale ha attraversato ricorrenti momenti di euforia e depressione, fin dai suoi inizi negli anni Cinquanta: a previsioni eccessivamente ottimistiche sono seguite profonde disillusioni, i cosiddetti inverni dell'intelligenza artificiale (Nilsson 2010). Tuttavia, è fuor di dubbio che l'Intelligenza Artificiale abbia ottenuto grandi successi negli ultimi anni. Da un lato, essa si è data una solida base scientifico-culturale: il nucleo originario costituito da informatica, ingegneria, matematica e logica è stato esteso con modelli e concetti provenienti da altre discipline come la statistica, l'economia, le neuroscienze, la filosofia e il diritto. Dall'altro lato, si è realizzata un'ampia gamma di applicazioni di successo, che sono entrate nella quotidianità dei singoli individui e della società, oltre che nelle attività economiche ed amministrative: la ricerca di informazioni (i motori di ricerca); il riconoscimento di voci, immagini e volti; la diagnostica medica; la traduzione automatica; l'analisi e la classificazione di documenti; la risposta a quesiti in linguaggio naturale (*question-answering*); i giochi, dai videogiochi agli scacchi e al go; le transazioni commerciali ad alta velocità (*high-speed trading*); i robot industriali e domestici; i veicoli autonomi in cielo, terra, e mare; l'ottimizzazione dei processi industriali e commerciali; gli assistenti virtuali; ecc.

L'enorme successo dell'intelligenza artificiale degli ultimi anni è anche legato al cambiamento del paradigma che ne ha guidato la ricerca e lo sviluppo. Fino a poco tempo fa, si credeva che per sviluppare un sistema intelligente fosse necessario fornire alla macchina una completa rappresentazione formale della conoscenza, espressa mediante regole e concetti, e accompagnata

da algoritmi capaci di compiere inferenze e ragionamenti. Diversi formalismi logici (logiche classiche, modali, descrittive, argomentazione formale, ecc.) e modelli di ragionamento (deduttivo, *defeasible*, induttivo, probabilistico, revisione delle credenze, ragionamento basato sui casi, ecc.) sono stati elaborati e impiegati a tal fine (Van Harmelen, Lifschitz e Porter 2008), ma alla ricchezza dei risultati teorici spesso non hanno fatto seguito corrispondenti realizzazioni operative. Per esempio, i sistemi esperti – consistenti di un’ampia base di conoscenze specialistiche e di un motore inferenziale che usa tale base di conoscenza per rispondere a quesiti – hanno spesso deluso le aspettative. Da un lato, le risposte di tali sistemi sono talvolta incomplete e spesso incapaci di considerare le particolarità dei casi concreti, mentre dall’altro, l’ampliamento e l’aggiornamento della conoscenza richiede sforzi gravosi, rappresentando un grave ostacolo alla crescita e all’adattamento dei sistemi intelligenti (il cosiddetto “collo di bottiglia” della rappresentazione della conoscenza – *knowledge representation bottleneck*). Non sono mancate talune applicazioni di successo, pur limitate ad ambiti settoriali. Nel diritto, le ricerche di intelligenza artificiale “tradizionale” (la cosiddetta Good Old Fashioned AI) hanno condotto allo sviluppo di modelli logici di grande interesse teorico – per esempio, nell’analisi dell’argomentazione (Prakken e Sartor 2015) e del ragionamento con i precedenti (Ashley 2017), e alla realizzazione di alcuni risultati operativi, ma non hanno trasformato la pratica del diritto. Tra le applicazioni di successo, possiamo ricordare, il sistema Oracle Policy Automation (evoluzione dei precedenti sistemi Softlaw and Ruleburst, vedi Dayal e Johnson 2000), usato per numerose applicazioni giuridiche dalle pubbliche amministrazioni di diversi paesi, tra cui Australia, Regno Unito e Stati Uniti.

I sistemi di intelligenza artificiale hanno fatto un vero salto di qualità quando l’attenzione si è spostata dalla rappresentazione logica della conoscenza alla possibilità di applicare metodi di apprendimento automatico (*machine learning*) a grandi masse di dati. Nel modello dell’apprendimento automatico, anziché fornire tutta la conoscenza alla macchina, l’uomo le fornisce un metodo di apprendimento, da applicare ai dati cui la macchina ha accesso, per estrarre automaticamente da quei dati le indicazioni su come svolgere il compito affidatole. Pertanto, come osservava Turing ([1951] 1996) già negli anni Cinquanta, “una macchina capace di apprendere” realizza gli obiettivi che le sono assegnati, senza che l’uomo abbia indicato alla macchina come procedere, e anzi senza che egli abbia consapevolezza di ciò che accade all’interno della macchina.

Tuttavia, se è vero che il grande successo dell’apprendimento automatico ha messo in ombra le tecniche di rappresentazione della conoscenza esplicita e formalizzata, quest’ultima rimane altamente significativa. In molti settori, infatti, la modellazione logica esplicita della conoscenza e del ragionamento

può essere complementare all'apprendimento automatico. I modelli logici possono spiegare il funzionamento dei sistemi di apprendimento automatico, controllare e governare il loro comportamento secondo standard normativi (compresi i principi etici e le norme giuridiche), convalidare i loro risultati e sviluppare le implicazioni logiche di tali risultati in base a conoscenze concettuali e teorie scientifiche. Nella comunità dell'intelligenza artificiale, e in particolare dell'intelligenza artificiale per il diritto, è generalmente riconosciuta la necessità di combinare modellazione logica e apprendimento automatico, sebbene esistano opinioni diverse su come raggiungere questo obiettivo e sugli aspetti che devono essere coperti dai due approcci (per una discussione sui limiti dell'apprendimento automatico, vedi di recente Marcus e Davis 2019).

2. L' intelligenza artificiale e il Diritto

Nel corso di oltre cinquant'anni di storia, fin dai primi contributi in materia (Frosini 1968, Losano 1969) la disciplina dell'Intelligenza Artificiale e il Diritto hanno perseguito due direzioni principali e interconnesse, focalizzandosi da un lato sull'elaborazione di modelli computabili del diritto e dall'altro sulle questioni etico-giuridiche e sociali legate all'intelligenza artificiale.

Sotto il primo profilo, la ricerca si è focalizzata su due prospettive convergenti, da un lato l'uso delle tecniche di intelligenza artificiale per cogliere aspetti del diritto, dall'altro lo sviluppo di nuovi modelli per l'intelligenza artificiale, basati sulle peculiarità dell'intelligenza giuridica, vale a dire delle forme del ragionamento e della conoscenza proprie del diritto. Ciò ha portato alla creazione di nuove tecnologie di intelligenza artificiale per il diritto, utilizzate nel corso degli anni dagli studi legali, dalla magistratura, dalla pubblica amministrazione e da altre organizzazioni pubbliche e private.

La ricerca si è dedicata inizialmente alla rappresentazione di norme, concetti e casi, e alla loro elaborazione formale, secondo metodi di inferenza specificamente giuridici, e più recentemente all'applicazione di tecniche di *machine learning* a contenuti giuridicamente rilevanti (per una rassegna degli studi di IA e diritto, si veda Ashley 2017).

I primi sistemi di intelligenza artificiale in ambito giuridico possono farsi risalire agli anni Settanta, quando si realizzarono le prime applicazioni giuridiche della nascente tecnologia dei sistemi esperti (Waterman e Peterson 1980). L'idea di base è quella di creare basi di conoscenza giuridica computabili, che formalizzano contenuti normativi estratti da regole, casi e dottrine, cui applicare metodi di inferenza automatica. Le inferenze automatiche si sono ispirate alla logica, il che ha condotto ad una stretta correlazione tra intelligenza artificiale e logica giuridica.

Gli studi logici nell'intelligenza artificiale applicata al diritto non si sono peraltro limitati all'applicazione di logiche preesistenti, ma si sono estesi allo sviluppo di nuove logiche, in particolare per il ragionamento *defeasible*, il ragionamento basato sui casi, l'argomentazione (McCarty 1982, Ashley 1990, Prakken 1993, Gordon 1995, Hage 1997), la teoria dei giochi (Roth et al. 2007), la dinamica dei sistemi normativi (Governatori e Rotolo 2010). La formalizzazione si è estesa ai concetti giuridici (McCarty 2002, Sartor 2006), fino a giungere a principi e valori, e ai metodi per il loro bilanciamento (Sartor 2013, Horty 2019). Le ricerche sulla modellizzazione delle norme e dei ragionamenti hanno per lo più avuto carattere teorico conducendo soprattutto a realizzazioni prototipali. Peraltro, alcuni sistemi esperti giuridici hanno avuto ampie applicazioni, in particolare nell'ambito della pubblica amministrazione.

Nelle ultime due decadi, un ruolo via via più importante è stato svolto dalle tecnologie di apprendimento automatico (*machine learning*).

La disponibilità di ampie raccolte di dati giuridici – sentenze, decisioni amministrative, contratti – ha consentito di applicare tecnologie per l'analisi testuale e l'apprendimento automatico in numerosi progetti, alcuni dei quali hanno raggiunto risultati operativi (per esempio, Lippi et al. 2019; sulla "analitica giuridica", si veda Ashley 2017). Tra gli obiettivi la previsione delle decisioni giudiziarie e amministrative sulla base di decisioni passate, il raggruppamento (*clustering*) di testi che affrontano le medesime tematiche, la previsione dei comportamenti delle parti e degli imputati (Larson et al. 2016).

Sotto il secondo profilo, la ricerca si è concentrata su come regolamentare lo sviluppo e l'uso di applicazioni dell'intelligenza artificiale, sul modo in cui esse influiscono sui valori e i principi giuridici, nel tentativo di adattare norme e istituzioni giuridiche ai cambiamenti sociali derivanti dall'intelligenza artificiale.

I primi contributi dottrinali furono realizzati negli anni Settanta e Ottanta, e furono dedicati soprattutto alla protezione della proprietà intellettuale sulle basi di dati e alla responsabilità derivante da errori nei sistemi esperti. L'interesse per tali questioni, e conseguentemente l'ampiezza del dibattito dottrinale e il suo impatto sulla pratica giuridica furono circoscritti, riflettendo l'utilizzo ancora assai limitato dell'intelligenza artificiale. Una vera cesura si è verificata a partire dagli anni Duemila, quando l'intelligenza artificiale, grazie in particolare alle tecnologie di *machine learning*, ha dato luogo ad applicazioni pervasive, di grande impatto sociale. È emersa la consapevolezza che per rispondere ai problemi generati dall'intelligenza artificiale non è sufficiente fare riferimento a specifiche norme in vigore, che potrebbero non fornire soluzione adeguate; ma bisogna considerare anche i principi etici e giuridici più generali, che includono diritti fondamentali e valori sociali (Hildebrandt 2015).

I problemi giuridici legati all'intelligenza artificiale diventano più pressanti man mano che i sistemi intelligenti abbandonano gli ambienti controllati e limitati dei laboratori e delle fabbriche, e condividono con gli esseri umani spazi fisici e virtuali come le strade, i cieli, e i mercati. L'urgenza della regolazione dell'intelligenza artificiale è confermata da documenti di policy su specifici domini applicativi, come gli aerei senza pilota, i veicoli autonomi e i robot di servizio (*service robot*). La creazione di un'adeguata cornice normativa è certamente un elemento essenziale per indirizzare l'intelligenza artificiale verso il bene degli individui e della società. È necessario garantire protezione adeguata ai molteplici interessi messi a rischio, salvaguardando al tempo stesso le molte applicazioni utili dell'intelligenza artificiale. Le tecnologie dell'intelligenza artificiale e dei big data operano in settori già caratterizzati da grandi squilibri di potere, che esse contribuiscono ad accentuare. Tali tecnologie infatti creano nuove conoscenze (capacità di analisi e previsione) e poteri (capacità di controllo e direzione) e le mettono a disposizione di chi governa le stesse tecnologie, vale a dire delle grandi imprese e dei poteri pubblici che cooperano con esse. Non solo l'individuo isolato, ma anche le sue organizzazioni, prive di analoghe risorse, si trovano in una crescente posizione di svantaggio, nella quale diventa difficile avvalersi delle tutele giuridiche astrattamente disponibili.

3. Big data, conoscibilità degli algoritmi e predizioni giuridiche

Nella sezione precedente si è illustrato come le ricerche di intelligenza artificiale e diritto si siano focalizzate su due obiettivi complementari: da un lato, l'elaborazione di modelli computabili del diritto e dall'altro, le questioni etico-giuridiche e sociali legate all'intelligenza artificiale.

In questo focus si affronterà questo secondo aspetto, concentrandosi su una delle tematiche al centro dell'odierno dibattito sui profili etici e giuridici dell'intelligenza artificiale, vale a dire il tema del trattamento di dati personali e quello connesso delle decisioni automatiche.

Nei tre contributi che compongono il focus vengono presentati tre aspetti interconnessi delle decisioni e delle predizioni automatiche: il ruolo dei big data, la conoscibilità e spiegabilità dei processi di inferenza, e l'uso di predizioni e decisioni automatiche in ambito giudiziario.

Nel primo contributo Monica Palmirani introduce il concetto di big data e illustra il ruolo dei dati nella decisione automatica, distinguendo dati personali, non personali, misti, statistici e sintetici. In particolare, l'autrice osserva che solo la conoscibilità di tutti i dati, di ogni tipo, su cui è basata una decisione ne garantisce la spiegabilità. Su questa base identifica i requisiti da soddisfare per garantire un'adeguata spiegazione *ex ante* ed *ex post*. La spie-

gazione *ex ante* dovrebbe comprendere informazioni sui modelli matematici e informatici adottati e sull'infrastruttura prescelta, i dati usati nell'addestramento, lo sviluppo e la verifica del sistema, i dati di contesto, ontologie e possibili risposte. La spiegazione *ex post* dovrebbe comprendere una giustificazione della decisione finale accompagnata da indicazioni sul flusso logico, i metadati, la sorgente e la natura dei dati impiegati.

Nel secondo contributo, Ugo Pagallo esamina i problemi della conoscibilità e della spiegabilità degli algoritmi sulla base del Regolamento sulla Protezione dei Dati (GDPR) e alla luce di principi etici. A questo riguardo, l'autore osserva che il principio di spiegabilità riveste un ruolo fondamentale, quale necessaria integrazione dei principi etici fondamentali, già elaborati in ambito bioetico: beneficenza, non-maleficenza, autonomia e giustizia. Infatti, la spiegabilità rappresenta una condizione necessaria affinché quei principi siano realizzati. Una spiegazione del funzionamento di un sistema è necessaria per capire se il suo uso sia stato o possa essere benefico o invece malefico, se il sistema operi rispettando l'autonomia degli individui e se il suo comportamento si conformi a criteri di correttezza. Egli afferma che il titolare del trattamento ha l'obbligo di assicurarne l'intelligibilità e comprensibilità, il che comporta, in particolare, l'obbligo di documentare tutte le operazioni di elaborazione di dati personali. La spiegabilità delle decisioni algoritmiche consente di affrontare in modo più equo e rigoroso non solo aspetti di protezione dei dati ma anche altre questioni giuridiche, come quelle attinenti alla responsabilità civile e penale.

Infine, Francesco Romeo affronta la trasformazione che le scienze cognitive comportano per le scienze sociali e in particolare per le discipline giuridiche. Egli osserva che il tentativo di creare sistemi intelligenti richiede modelli logico-matematici, ma quei modelli danno luogo in molti casi a risultati probabilistici e suscettibili di errore. Romeo si sofferma sulle reti neurali artificiali, sistemi capaci di apprendere da esempi automodificando la propria struttura (le connessioni tra neuroni che le compongono). In particolare, reti neurali addestrate su precedenti casi giuridici possono acquisire competenze giuridiche e applicarle a nuovi casi, meglio dei modelli puramente logici. Le reti neurali, avendo appreso a riprodurre il passato, non possono però cogliere l'orientamento verso il futuro, la ricerca dell'innovazione, che contribuisce a caratterizzare il pensiero umano. A questo riguardo, Romeo osserva come gli algoritmi possano guardare anche al futuro: è il caso degli algoritmi equitativi, che propongono alle parti i risultati che meglio soddisfano gli interessi di cui le stesse sono portatrici, secondo un equo bilanciamento.

I tre saggi inclusi nel focus colgono quindi diversi aspetti dell'uso di decisioni giuridiche automatiche: l'importanza dei dati, la necessità di spiegazioni/comprensibilità, e le prospettive della decisione automatica in ambito

giuridico. Nel loro insieme danno un'ampia visione degli sviluppi dell'intelligenza artificiale e dei nuovi problemi e opportunità che tali sviluppi aprono al diritto.

Giovanni Sartor
European University Institute
Law Department
Via dei Roccettini 9
I-50014 San Domenico di Fiesole (FI)
giovanni.sartor@EUI.eu

Università degli Studi di Bologna
CIRSFID
Via Galliera 3
40121 Bologna
giovanni.sartor@unibo.it

Riferimenti bibliografici

- Ashley, Kevin D. 1990. *Modeling Legal Argument. Reasoning with Cases and Hypotheticals*. Cambridge (MA): MIT Press.
2017. *Artificial Intelligence and Legal Analytics*. Cambridge: Cambridge University Press.
- Dayal, Surendra, e Peter Johnson. 2000. "A Web-Based Revolution in Australian Public Administration." *Journal of Information, Law and Technology* 1. <http://elj.warwick.ac.uk/jilt/00-1/dayal.html>>. New citation as at 1/1/04: <http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2000_1/austlii/dayal/
- Frosini, Vittorio. 1968. *Cibernetica, diritto, e società*. Roma: Comunità.
- Gordon, Thomas F. 1995. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Alphen aan den Rijn: Kluwer.
- Governatori, Guido, e Antonino Rotolo. 2010. "Changing Legal Systems: Legal Abrogations and Annulments in Defeasible Logic." *Logic Journal of IGPL* 18: 157-194.
- Hage, Jaap C. 1997. *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic*. Alphen aan den Rijn: Kluwer.
- Hildebrandt, Mireille. 2015. *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Cheltenham and Camberley: Elgar.
- Horty, John. 2019. "Reasoning with Dimensions and Magnitudes." *Artificial Intelligence and Law* 27: 309-345.
- Larson, Jeff, et al. 2016. "How We Analyzed the COMPAS Recidivism Algorithm." <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>>.
- Lippi, Marco, et al. 2019. "CLAUDETTE: An Automated Detector of Potentially Unfair Clauses in Online Terms of Service." *Artificial Intelligence and Law* 2: 117-139.

- Losano, Mario. 1969. *Giuscibernetica: Macchine e modelli cibernetici nel diritto*. Torino: Einaudi.
- Marcus, Gary, e Davis, Ernest. 2019. *Rebooting AI. Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.
- McCarty, L. Thorne. 1982. "A Computational Theory of Eisner vs Macomber." In *Artificial Intelligence and Legal Information Systems*, a cura di Costantino Ciampi, 329-355. Amsterdam: North Holland.
- 2002. "Ownership. A Case Study in the Representation of Legal Concepts." *Artificial Intelligence and Law* 10: 135-161.
- Nilsson, Nils J. 2010. *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.
- Prakken, Henry. 1993. *Logical Tools for Modelling Legal Argument*. Amsterdam: Free University.
- Prakken, Henry, e Giovanni Sartor. 2015. "Law and Logic: A Review from an Argumentation Perspective." *Artificial Intelligence* 227: 214-245.
- Roth, Bram, et al. 2007. "Strategic Argumentation: A Game Theoretical Investigation." In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, 81-90. Doi: 10.1145/1276318.1276333.
- Sartor, Giovanni. 2006. "Fundamental Legal Concepts: A Formal and Teleological Characterisation." *Artificial Intelligence and Law* 21: 101-142.
- 2013. "The Logic of Proportionality. Reasoning with Non-Numerical Magnitudes." *German Law Journal* 14: 1419-1457.
- Turing, Alan M. (1951) 1996. "Intelligent Machinery, a Heretical Theory." *Philosophia Mathematica* 4: 256-260.
- Van Harmelen, Frank, Vladimir Lifschitz, e Bruce Porter (a cura di). 2008. *Handbook of Knowledge Representation*. Amsterdam: Elsevier.
- Waterman, Donald, e Mark Peterson. 1980. "Rule-Based Models of Legal Expertise." In *Proceeding National Conference on Artificial Intelligence*, 272-275. <https://dl.acm.org/doi/abs/10.5555/2876590.2876667>.