

# Dottrina e attualità giuridiche

## La giustizia predittiva

a cura di Enrico Gabrielli e Marco Dell'Utri\*

Introduzione, Marco Dell'Utri

La giustizia predittiva: una visione realistica, Amedeo Santosuosso e Giovanni Sartor

### Introduzione

Marco Dell'Utri

A distanza di circa tre anni, *Giurisprudenza Italiana* torna a riflettere sui temi dell'Intelligenza Artificiale e dei suoi incontri con il diritto.

La raccolta di saggi che, nell'estate del 2019, si propose alla meditazione del lettore<sup>1</sup> aveva cercato di ordinare, nelle forme di una possibile rassegna, una sorta di generale ricognizione di questioni e di ambiti di incidenza dell'impiego dell'Intelligenza Artificiale nei settori tradizionali del diritto privato.

Dall'esame del destino dei diritti della persona nella prospettiva dello *human enhancement* e del transumanesimo<sup>2</sup>, all'incidenza dell'uso delle macchine nel trattamento dei dati personali<sup>3</sup>; dalla descrizione dell'attitudine delle macchine intelligenti allo svolgimento dell'attività contrattuale<sup>4</sup>, ai riscontri dell'azione dell'Intelligenza Artificiale sul terreno del danno e dei diversi settori della responsabilità civile<sup>5</sup>; dallo studio degli ambiti in cui si pone il problema della protezione e della sicurezza della persona<sup>6</sup>, a quello sulle ricadute dell'elaborazione dell'Intelligenza Artificiale in relazione ai temi della creatività e del diritto d'autore<sup>7</sup>; dalla tutela della concorrenza<sup>8</sup>, all'atteggiamento da assumere, sul piano tributario<sup>9</sup>, su quello amministrativo<sup>10</sup> e su quello giudiziario<sup>11</sup>, nei confronti dell'azione delle macchine.

Il testo che qui si introduce, dietro le forme di un esame analitico dei momenti e dei diversi passaggi che descrivono i modi della predizione affidata alle macchine, invita, in termini generali, a sollecitare la consapevolezza e la responsabilità umana di fronte alle prospettive dell'Intelligenza

Artificiale, con particolare riguardo alle sue possibili applicazioni sul terreno della giustizia predittiva.

Si tratta di un richiamo a rendersi coscienti dell'urgenza di assolvere in modo puntuale ai doveri che alla persona competono (individualmente e socialmente) con riguardo all'uso della ragione; di un monito a riconsiderare, in chiave contemporanea, e proprio nei confronti dell'Intelligenza Artificiale, il valore di quel *sapere aude* che la modernità illuministica aveva assunto a simbolo della lotta contro ogni forma di pigrizia intellettuale, e del rifiuto di qualunque trascurata concessione culturale all'idiosincrasia o all'oscurantismo.

Il testo di Giovanni Sartor e di Amedeo Santosuosso segnala, in definitiva, al di là della semplicità e dell'accessibilità del linguaggio, per la preziosa esortazione alla sobrietà dell'approccio, lontano, tanto dall'irragionevole ossessione di una certa lettura 'apocalittica' della contemporaneità tecnologica, quanto dal vagheggiamento di irrealistiche prospettive di onnipotenza, naturalmente inesigibili dal lavoro di qualunque macchina.

Seguendo la strada di questo richiamo alla semplicità e alla realtà delle cose, all'Intelligenza Artificiale (applicata alla giustizia predittiva) è, dunque, possibile riconoscere il potere formidabile di assicurare, attraverso gli apporti di una sempre più sofisticata capacità di elaborazione dell'esperienza, un contributo di elevatissimo significato e valore per l'attività umana: quello per cui è possibile tornare a scoprirsi capaci di emanciparsi dalla passività della ripetizione e di riattivare, dietro quella dura 'scorza', il '*pathos* incontenibile' di un'insostituibile creatività.

\* I contributi sono stati sottoposti, in forma anonima, alla valutazione di un referee.

<sup>1</sup> E. Gabrielli e U. Ruffolo (a cura di), *Intelligenza Artificiale e diritto*, in *Giur. It.*, 2019, fasc. 7 (luglio 2019), 1657 e segg.

<sup>2</sup> U. Ruffolo e A. Amidei, *Intelligenza Artificiale e diritti della persona: le frontiere del "transumanesimo"*, *loc. cit.*, 1658 e segg.

<sup>3</sup> G. Finocchiaro, *Intelligenza Artificiale e protezione dei dati personali*, *loc. cit.*, 1670 e segg.

<sup>4</sup> F. Di Giovanni, *Attività contrattuale e intelligenza artificiale*, *loc. cit.*, 1677 e segg.

<sup>5</sup> M. Costanza, *L'intelligenza artificiale e gli stilemi della responsabilità civile*, *loc. cit.*, 1686 e segg.; U. Ruffolo, *Intelligenza Artificiale, machine learning e responsabilità da algoritmo*, *loc. cit.*, 1689 e segg.; U. Ruffolo ed E. Al Mureden, *Autonomous vehicles e respon-*

*sabilità nel nostro sistema ed in quello statunitense*, *loc. cit.*, 1704 e segg.; A. Amidei, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, *loc. cit.*, 1715 e segg.

<sup>6</sup> M. Gambini, *Algoritmi e sicurezza*, *loc. cit.*, 1726 e segg.

<sup>7</sup> M. Capparelli, *Intelligenza Artificiale e nuove sfide del diritto d'autore*, *loc. cit.*, 1740 e segg.

<sup>8</sup> A.M. Gambino e M.C. Manzi, *Intelligenza Artificiale e tutela della concorrenza*, *loc. cit.*, 1744 e segg.

<sup>9</sup> A. Uricchio, *Robot tax: modelli di prelievo e sportive di riforma*, *loc. cit.*, 1749 e segg.

<sup>10</sup> E. Picozza, *Politica, diritto amministrativo and Artificial Intelligence*, *loc. cit.*, 1761 e segg.

<sup>11</sup> R. Bichi, *Intelligenza Artificiale tra "calcolabilità" del diritto e tutela dei diritti*, *loc. cit.*, 1772 e segg.

## La giustizia predittiva: una visione realistica

Amedeo Santosuosso e Giovanni Sartor

Una valutazione realistica delle prospettive dell'impiego di strumenti di intelligenza in ambito giuridico e giudiziario richiede un'informazione corretta, una presa di coscienza da parte degli stessi operatori del processo, giudici e avvocati, senza eccessi tanto nella speranza quanto nel timore. Gli autori, in primo luogo, offrono un quadro ampio e tecnicamente accurato delle tecnologie e degli usi della predizione automatica in ambito giuridico, dai sistemi esperti fino alle reti neurali: una mappa non facile da reperire altrimenti. Sulla base di queste conoscenze esaminano, poi, i timori maggiormente diffusi dividendoli in tre gruppi: il primo, riguarda le critiche che si basano su assunti irrealistici o paradossali; il secondo, riguarda alcune questioni reali, che, però, potrebbero essere risolte con accorgimenti pratici o tecnici; il terzo, è dedicato ad alcune questioni obiettivamente importanti e che richiedono ulteriore ricerca e riflessione.

### Introduzione: il punto di partenza

Sono questi tempi nei quali sembra prevalere nei media, nella popolazione generale e anche nelle professioni una visione molto incerta rispetto al futuro, così come dominata dalla tecnologia e, in particolare, dall'Intelligenza artificiale (IA) di cui vengono annunciati progressi straordinari<sup>1</sup>. Come di fronte ad altri sviluppi tecnologici, si confrontano spesso posizioni estreme. Per usare la terminologia introdotta anni fa da Umberto Eco, anche rispetto alle più recenti tecnologie si confrontano apocalittici e integrati, i primi a sottolineare i rischi che esse comportano, i secondi a enfatizzare i benefici attesi. Nei tempi più recenti, sono forse i timori ad attirare maggiormente l'attenzione del pubblico, rispondendo, sembrerebbe, a un aspetto profondo della psicologia umana (quanto più numerosi sono gli esempi distopici nella letteratura o nella cinematografia fantascientifica rispetto a quelli in cui si prospettano utopiche società del futuro!).

La popolare rivista Wired, solo un esempio tra i tanti, in un articolo del 27 dicembre 2021, mette in guardia circa i pericoli che corre la libertà dell'essere umano, in un mondo nel quale le "profezie algoritmiche" sono onnipresenti e arrivano a formulare predizioni non solo sugli eventi ma anche sulle persone, sulla possibilità che esse ottengano un prestito, un lavoro, un appartamento, un contratto di assicurazione, la libertà vigilata, e altro ancora<sup>2</sup>. Se all'individuo fossero precluse, in modo pervasivo, attività future sulla base dell'aspettativa probabilistica che egli ne farebbe cattivo uso, fondata sul comportamento passato suo e di altri a lui simili, egli verrebbe privato della possibilità, e quindi della libertà, di smentire quelle aspettative, grazie alle proprie scelte e al proprio impegno. Ciò non significa che si debba rinunciare a ogni valutazione di affidabilità degli individui prima di esporli (o esporre i terzi) ai rischi che l'interazione sociale comporta (si pensi alle difficili valutazioni del giudice di sorveglianza circa la concessione di permessi e libertà anticipata a chi abbia commesso gravi reati), ma richiede di ripensare il modo in cui tali valutazioni – molto più frequenti, economiche, e precise – possono essere utilizzate.

I temi giuridici e giudiziari, che occupano un posto no-

tevole nell'immaginario collettivo, non sono esenti da queste suggestioni, alimentate talora da notizie imprecise, come quella che annunciava che *In Cina un magistrato-software decide condanne e pene*<sup>3</sup>. In realtà si tratta soltanto di un sistema sperimentato dagli uffici dell'accusa per selezionare i casi da portare prioritariamente in giudizio. Nessuno nega che la selezione dei casi da sottoporre a giudizio possa essere un problema delicato, ma sicuramente lo è in misura differente rispetto all'automazione della stessa *decisione*. Il problema della determinazione delle priorità potrebbe essere piuttosto quello dei rapporti tra magistratura inquirente e giudicante in quel paese e dei rapporti di forza tra le due, che, però, è questione assai diversa. Comunque, resta il segnale del clima culturale intorno ai rapporti tra giustizia e tecnologia, specie su IA e predizione.

A nostro parere, un'informazione corretta, una presa di coscienza da parte degli stessi operatori del processo, e in primo luogo dei giudici, può consentire una valutazione accurata delle prospettive dell'impiego di strumenti di intelligenza, senza eccessi tanto nella speranza quanto nel timore (*nec spe nec metu*, come si suol dire). Nelle pagine seguenti ci proponiamo di offrire un quadro sufficientemente ampio delle tecnologie e poi degli usi della predizione automatica in ambito giuridico, che consenta di valutare le possibilità oggi disponibili, e di prospettare quelle del prossimo futuro. Infine, cercheremo di dare risposta ad alcuni dei timori maggiormente diffusi.

### Il problema della predizione nel diritto

In questa sezione introduciamo la tematica della predizione in ambito giuridico. Dopo aver introdotto il dibattito odierno sull'uso di tecnologie predittive, esaminiamo le ragioni che inducono da sempre il giurista a cercare di prevedere l'esito delle controversie.

*Predizione e decisioni giuridiche: illusioni, speranze e timori.*

Il 3-4 dicembre 2018, a Strasburgo, la Commissione europea per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) approva il documento *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and*

<sup>1</sup> Saltlux, *The Third Wave of Artificial Intelligence: Neuro Symbolic AI*, <https://medium.com/@saltluxus/the-third-wave-of-artificial-intelligence-neuro-symbolic-ai-666ccec8c85c>.

<sup>2</sup> *If AI Is Predicting Your Future, Are You Still Free?*, Wired, <https://www.wired.com/story/algorithmic-prophecies-undermine-free-will/> Tra i lavori che più ampiamente illustrano i rischi della sorveglianza e la manipolazione mediante tecniche di intelligenza artificiale, si veda Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Hachette.

<sup>3</sup> Corriere della Sera, pagina online "Il mondo visto dagli altri", 29

dicembre 2021, articolo di Alessandro Trocino, che riprende un articolo apparso sul sito Formiche con il titolo più corretto "Il pubblico ministero? È l'algoritmo. L'esperimento cinese che fa paura", Gabriele Carrer, 27 dicembre 2021, <https://formiche.net/2021/12/intelligenza-artificiale-cina/>, che a sua volta riprende un articolo del South China Morning Post, dal titolo ancora più preciso "Chinese scientists develop AI 'prosecutor' that can press its own charges" <https://www.scmp.com/news/china/science/article/3160997/chinese-scientists-develop-ai-prosecutor-can-press-its-own>. La catena informativa è interessante per la progressiva perdita di accuratezza della notizia.

*their Environment*. È la prima volta, ci pare, che in un testo ufficiale europeo, che riguarda i 47 Paesi del Consiglio d'Europa, le parole 'giudice' e 'intelligenza artificiale' vengono accostate.

È da ricordare che il tema dell'intelligenza artificiale in ambito giudico è oggetto di studi da ormai più di cinque decenni. Risalgono agli anni '70 del secolo scorso i primi contributi che hanno applicato tecniche di intelligenza artificiale (allora si trattava dei cosiddetti sistemi esperti: si veda avanti) al diritto, in particolare, al fine di ottenere indicazioni sulla soluzione giuridica di casi concreti, mentre fin dal 1983 si susseguono, ogni due anni, le edizioni della *International Conference on Artificial Intelligence and Law*<sup>4</sup>. Anche nel nostro paese non sono mancate le elaborazioni teoriche, le analisi di scenari futuri, e le sperimentazioni in tema di intelligenza artificiale applicata al diritto<sup>5</sup>, ma fino ai nostri giorni sono state assai limitate le concrete implicazioni per la giustizia. Ricordiamo gli scritti e l'azione di un personaggio straordinario, come Renato Borruso, che hanno segnato lo sviluppo dell'informatica giuridica, specialmente in Italia. Nell'opera di Borruso si possono trovare alcune intuizioni non solo circa l'utilizzo avanzato di algoritmi di ricerca documentale, ma anche l'idea che sia possibile ricorrere all'automazione nell'applicazione della legge, pur riservando all'essere umano, quale legislatore o giudice, le scelte di valore. Tuttavia, pur con tutti i pregi del Centro Elettronico della Corte di Cassazione (CED) e del sistema di ricerca Italgire-Find per la ricerca automatica, sarebbe una forzatura dire che, al momento della scomparsa di Borruso, nel 2014, l'IA applicata alla giustizia fosse una realtà consolidata, o anche solo avviata, e che tra professionisti come giudici e avvocati vi fosse una consapevolezza circa le tecnologie di IA e le prospettive della loro applicazione.

Qualcosa di nuovo nei rapporti tra intelligenza artificiale e giustizia comincia ad accadere solo nella seconda metà della scorsa decade quando alcuni bandi europei finanziano progetti dedicati a tematiche di intelligenza artificiale e diritto, non solo per quanto riguarda il profilo della regolazione dell'intelligenza artificiale, ma anche per quello dello sviluppo di sistemi di intelligenza artificiale destinati alla giustizia<sup>6</sup>.

Invece, il documento CEPEJ accosta, come se fosse una realtà scontata, IA e sistemi giudiziari e parla di uso della prima nei secondi. Ora, vero o non vero che fosse vero questo assunto (e non lo era), sicuramente da quando quel documento ha infranto il tabù dell'accostamento tra IA e giurisdizione è stato tutto un susseguirsi di riflessioni e spesso di allarmi per le conseguenze di quell'uso.

In questo contesto è degno di nota il focus che, in modo obiettivamente tempestivo, la rivista online *Questione Giustizia* (n. 4 del 2018) dedica al tema *Una giustizia (im)prevedibile?*<sup>7</sup> Gli articoli hanno titoli significativi, come *Giustizia predittiva. La qualità della giustizia in due tempi*, op-

pure *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, oppure *Ti faresti giudicare da un algoritmo?* E altri ancora. Una sintesi dei timori e delle cautele è fornita dall'Introduzione al focus a firma di Andrea Natale<sup>8</sup>, dalla quale traiamo una sintesi:

- il risultato fornito dagli algoritmi predittivi è necessariamente influenzato dalla qualità dei dati che vengono posti come input; bisogna quindi curarne la qualità, l'indipendenza della fonte, e l'accessibilità;

- l'algoritmo può avere un esito discriminatorio fondato su dati personali sensibili, tra cui la razza e l'estrazione sociale;

- la struttura dell'algoritmo non è neutra e deve essere verificabile; cosa difficile se l'algoritmo è protetto da diritti di proprietà intellettuale; vi possono essere anche errori di progettazione;

- l'algoritmo, ove usato come mero supporto alla decisione del giudice, richiede formazione del personale giudiziario;

- l'algoritmo predittivo – muovendo da una elaborazione della giurisprudenza e dei casi precedenti – può indicare non "il risultato" esatto di una certa controversia, ma il suo possibile esito, con vari rischi: l'algoritmo non è in grado di "riconoscere" che quello a lui sottoposto non è un caso simile; vi sono delle singolarità che un decisore umano forse rilevarebbe e che lo porterebbero a operare una distinzione (*distinguishing*);

- l'algoritmo può favorire quello che Garapon chiama effetto gregge (*moutonnier*)<sup>9</sup>: vi sarebbe il rischio di indurre il giudice pigro ad adagiarsi sulla proposta dell'algoritmo senza assumere su di sé l'autentica responsabilità del giudizio che egli emette, con effetto di cristallizzazione della giurisprudenza, che risulterebbe meno sensibile ai cambiamenti sociali.

Va ricordato, inoltre, un lungo capitolo, apparso nel volume su *Intelligenza artificiale e diritto* curato da Ugo Ruffolo, nel quale Roberto Bichi esprime tutte le sue perplessità circa l'utilizzo delle tecnologie in generale e in specie di quelle di IA nel campo della giustizia. Il focus critico è esattamente sulla giustizia predittiva, e sugli algoritmi decisori, e riguarda in particolare il rischio che le macchine si sostituiscano agli umani, che la decisione sia delegata alla macchina e che un "inevitabile" ammodernamento tecnologico "emargini l'uomo-giurista", che invece deve conservare il ruolo di soggetto, e, alla fine, la preoccupazione di tipo culturale di "trovarsi rinchiusi in una nuova *normatività informatica*"<sup>10</sup>.

Preoccupazioni di questo genere sono condivise e amplificate da alcuni esponenti dell'avvocatura e dell'accademia. Gustavo Ghidini e Daniele Manca così si esprimono:

"Proviamo a pensare le insane applicazioni di 'giustizia predittiva'. Sono quelle che anziché limitarsi ad offrire un ampio quadro informativo di precedenti e opinioni, si spingano a 'predire/suggerire' la soluzione giurisdizionale: l'at-

<sup>4</sup> Per un quadro aggiornato sugli sviluppi più recenti, si veda Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics*. Cambridge University Press.

<sup>5</sup> Il rapporto tra diritto e cibernetica (oggi diremo intelligenza artificiale) è già discusso da Wiener, N. (1950). *The human use of human beings: cybernetics and society*, le cui considerazioni saranno riprese e sviluppate da Frosini, V. (1968). *Cibernetica, diritto, e società*. Comunità. La prima monografia dedicata all'intelligenza artificiale nel nostro paese è quella di Sartor, G. (1995). *Intelligenza artificiale e diritto. Un'introduzione*, Milano.

<sup>6</sup> Tra gli esempi più recenti, ricordiamo i progetti Interlex (in tema di diritto internazionale privato), Crossjustice (in tema di cooperazione giudiziaria nel penale), Adele (in tema di analisi automatica delle

sentenze).

<sup>7</sup> È il n. 4 del 2018. L'editoriale di Renato Rordorf è accessibile presso [https://www.questionegiustizia.it/rivista/articolo/editoriale\\_571.php](https://www.questionegiustizia.it/rivista/articolo/editoriale_571.php).

<sup>8</sup> Andrea Natale, Introduzione. Una giustizia (im)prevedibile?, <https://www.questionegiustizia.it/rivista/articolo/introduzione-una-giustizia-imprevedibile-572.php>.

<sup>9</sup> A. Garapon e J. Lasségue, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, 2018, 279 e segg.

<sup>10</sup> Roberto Bichi, *Intelligenza digitale, giurimetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione* in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020, 423-447.

tribuzione di ragioni e torti. Tollerandole, si rafforzerebbe un commercio di ‘pacchetti’ algoritmici per incoraggiare, deresponsabilizzando i giudici, atteggiamenti decisorii ‘passivi’ (e pigri), ancorati al passato: ‘dato’ è participio passato. E così dunque, pure, disincentivando interpretazioni evolutive.”<sup>11</sup>

Vanno ricordati, per finire (ma non certo per importanza e qualità degli interventi), gli incontri svoltisi presso l’Accademia dei Lincei nel 2018 sul tema della calcolabilità del diritto. Tra questi ve ne è uno di particolare interesse, che prende poi la forma di un ampio articolo di Massimo Luciani su *La decisione giudiziaria robotica*<sup>12</sup>. L’illustre autore, pur partendo dall’ammissione di non avere “le conoscenze scientifiche ‘dure’ necessarie per dare risposte dotate di un minimo di attendibilità”, ritiene comunque di poter enunciare i dubbi “di un giurista (che trascendono largamente il profilo tecnico-scientifico)”. Essi riguardano i criteri di selezione dei fatti rilevanti sui quali si basa la decisione robotica, il ruolo e la conoscibilità dell’errore robotico, cioè della macchina, il tipo di dottrina giuridica (e quindi di metodo) che la macchina ha incorporata, il rapporto con i precedenti (come selezionati e organizzati), il modo in cui il robot organizza e risolve il raffronto o conflitto tra norme, la sorte che in una decisione robotica hanno la difesa tecnica, l’appello, il ricorso per cassazione e la stessa motivazione, se e quale discrezionalità il giudice-robot eserciti, il possibile passaggio dal giudice-robot al robot-legislatore e mediatore.

A queste critiche di giuristi italiani, relativamente estranei a quelle che Luciani chiama *conoscenze scientifiche ‘dure’*, si aggiungono i dubbi che sull’uso della *predizione* giuridica sono avanzati da indiscussi esperti in materia. Pensiamo a Floris Bex e Henry Prakken, dell’Università di Utrecht, entrambi docenti presso facoltà di informatica olandesi. Quegli studiosi rivolgono due principali critiche ai sistemi predittivi, e in particolare ai sistemi predittivi opachi, che formulano una predizione ma non ne forniscono una spiegazione comprensibile e significativa per l’uomo (il concetto di sistema opaco sarà illustrato nel seguito). Innanzitutto, le indicazioni proposte dal sistema (per esempio, la conclusione che l’attore probabilmente vincerà la causa) non darebbero un’indicazione attendibile circa la futura decisione del giudice. La previsione che il giudice deciderà una specifica causa in un certo modo consisterebbe solo in una generalizzazione statistica, che si fonda sull’insieme limitato delle caratteristiche trattate dal sistema, valutate in relazione a casi simili, e non terrebbe conto delle maggiori conoscenze accessibili al giudice. Inoltre, la mera previsione di una decisione, senza fornire altresì un’adeguata motivazione della stessa, mediante ragioni giuridicamente significative, non fornirebbe indicazioni utili ai giudici e potrebbe essere anzi fuorviante e generare pigrizia intellettuale”<sup>13</sup>.

A questo punto, anche solo sulla base di questo veloce e certo incompleto excursus, è evidente che l’idea di giustizia predittiva solleva numerose perplessità, che meritano attenta considerazione. Esse sono di diverso genere e, pertanto,

potranno avere risposte diverse a seconda della loro natura. Alcune sono legate a questioni meramente fattuali, come, per esempio, quale sia il reale livello di IA applicato presso i nostri (e altri) uffici giudiziari o praticato dagli studi professionali oppure cosa sia possibile immaginare con la tecnologia disponibile oggi o in un futuro prevedibile. Altre perplessità riguardano la convenienza o meno di spendere tempo e risorse in un’automazione che non porta alcun beneficio apprezzabile. Vi è, poi, un ulteriore livello, dove si collocano quelle critiche che si giustificano solo in ragione di un punto di vista teorico particolare, come per esempio una visione più aperta al ruolo dell’interpretazione o meno, oppure che riguardano alcuni aspetti specificamente teorici legati ai sistemi tecnici usati.

Discuteremo e proveremo a organizzare questi temi in modo ordinato nei paragrafi finali. Qui ci limitiamo a sgombrare il campo da una questione fattuale, che è di più facile soluzione, evitando così di portarla avanti come inutile fardello. La questione può essere così posta: i giudici utilizzano tecniche di IA? Se sì, quali? E i giudici italiani? Tutto ciò incide sulle loro decisioni o sulla motivazione dei loro atti? Come interagiscono gli usi da parte dell’avvocatura con quelli dei giudici?

Sul punto la stessa CEPEJ dichiara apertamente che, nel 2018, l’uso di algoritmi di intelligenza artificiale per previsioni “giuridiche” rimane principalmente un’iniziativa commerciale del settore privato rivolta a compagnie assicurative, dipartimenti legali, avvocati e singoli individui. Per il momento i giudici negli Stati membri del Consiglio d’Europa non sembrano fare alcun uso pratico e quotidiano di software predittivi. Test locali e lavori accademici sono stati condotti per esplorare il potenziale di queste applicazioni, ma non sono ancora state usate su vasta scala<sup>14</sup>. Se questo è il punto di partenza realistico (e non vi è da dubitarne), sarebbe forse stato meglio che la Commissione avesse dato un titolo meno altisonante, come, per esempio, “Carta etica sugli usi, *possibili o futuri*, dell’intelligenza artificiale”.

La situazione in Italia è ancora più o meno a questo punto. Significativamente la Relazione sullo stato della Giustizia telematica – anno 2021, rilasciata il 15 dicembre 2021 dalla Settima commissione del CSM, può riferire, nel paragrafo dedicato a *Intelligenza artificiale e giustizia predittiva*, che “sono in corso, negli uffici di alcune Corti d’Appello (tra le quali Venezia, Brescia e Genova) progetti finalizzati a sperimentare l’utilizzo di sistemi di intelligenza artificiale nell’attività giudiziaria con particolare riferimento alla sistematizzazione del patrimonio giurisprudenziale”<sup>15</sup>. Ma, purtroppo, non è esattamente vero neanche questo, perché se si va a verificare sui siti ufficiali delle Corti d’appello di Venezia e di Brescia, si scopre che si tratta di raccolte di piccoli gruppi di sentenze massimate e organizzate per materia e per data, con un’aggiunta dei tempi di definizione per materia secondo le statistiche ufficiali<sup>16</sup>. Senza voler negare che possa trattarsi di iniziative utili e meritorie, va detto con chiarezza, però, che è del tutto fuorviante che esse siano presentate come utilizzi o esperimenti di utilizzo di tecniche di IA alla giustizia, non solo per la distanza che

<sup>11</sup> Gustavo Ghidini e Daniele Manca, *Intelligenza artificiale: dubbi (e rischi) in Europa*, Corriere della sera, 13 dicembre 2021.

<sup>12</sup> Massimo Luciani, *La decisione giudiziaria robotica*, in *Nuovo Diritto Civile*, III, 2018, 1.

<sup>13</sup> Floris Bex-Henry Prakken, *On the relevance of algorithmic decision predictors for judicial decision making*, ICAIL’21, June 21-25, 2021, São Paulo, Brazil, <https://doi.org/10.1145/3462757.3466069>.

<sup>14</sup> Il documento CEPEJ riporta con un certo scetticismo un esperimento condotto per tre mesi in due Corti d’appello francesi: “Douai

and Rennes Courts of Appeal in France conducted a three-month trial in 2017 with a software programme labelled ‘predictive’ by a panel of judges”.

<sup>15</sup> Consultabile presso <https://www.csm.it/web/csm-internet/-/relazione-sullo-stato-della-giustizia-telematica-anno-2021>.

<sup>16</sup> Per la Corte d’appello di Venezia si consulti il sito ufficiale [https://www.corteappello.venezia.it/giurisprudenza-predittiva-per\\_198.html](https://www.corteappello.venezia.it/giurisprudenza-predittiva-per_198.html). Per la Corte d’appello di Brescia si consulti [http://www.giustizabrescia.it/giustizia\\_predittiva.aspx](http://www.giustizabrescia.it/giustizia_predittiva.aspx).

separa le raccolte in questione da qualsiasi, sia pure embrionale, utilizzo di quelle tecniche, ma soprattutto perché questo tipo di informazione, in quanto non rispondente alla realtà, genera paure e resistenze irrealistiche e ingiustificate.

Tornando al documento CEPEJ, va notato che esso è basato su dati dichiaratamente scarsi e su una letteratura internazionale proveniente prevalentemente dagli USA. Rimane, alla fine, l'importanza dei principi etici enunciati, sia pure in via ipotetica, ma ci sembra che quel documento, proprio in virtù dell'autorevolezza della fonte, possa contribuire a un'attitudine allarmistica sui rapporti tra IA e diritto nelle corti. A nostro parere il tema richiede una discussione pacata e realistica, nella quale si distinguano chiaramente le tecnologie, le pratiche e i problemi di oggi rispetto a quanto possiamo prospettare in scenari futuri.

### *Predizione, previsione e certezza del diritto prima delle applicazioni di IA*

L'antico, e a tratti ancora vivo, dibattito sulla certezza del diritto è in buona parte innervato proprio dalla prevedibilità delle decisioni sulle controversie che sorgono tra cittadini. In questo senso è qualcosa che va ben oltre il, pur ragionevole, interesse dell'avvocato a dare risposte al cliente sulle probabilità di successo di un'azione giudiziaria, in quanto riguarda il funzionamento complessivo del sistema giuridico, di cui l'avvocato è parte.

Come è stato notato, il diritto è di vitale importanza per la società, promuovendo la giustizia e la stabilità e avendo effetti su molte persone in aspetti importanti della loro vita privata e pubblica. Poiché il diritto ha obiettivi sociali ed effetti sociali, deve essere compreso da coloro che ne sono interessati, e la sua applicazione deve essere spiegata e giustificata. Da qui l'importanza della chiarezza di significato e della solidità del ragionamento, e quindi l'importanza della logica per il diritto e per le applicazioni giuridiche dell'IA<sup>17</sup>.

La certezza del diritto, tema evidentemente connesso con la prevedibilità delle decisioni, ha molte facce: è un principio e un valore, in collegamento con quello di uguaglianza; vive nell'interpretazione che la dottrina e i pratici ne danno; cambia contenuto secondo il contesto e l'idea di diritto in esso prevalente, tanto che in un sistema ispirato a un'ideologia legicentrica (alla francese, per intenderci) diventa certezza della legge e, in un sistema dove maggior rilievo (istituzionale o di fatto) ha la formazione giurisprudenziale, diventa certezza dell'interpretazione/creazione giurisprudenziale del diritto.

Nel common law, conoscere il diritto significa innanzitutto essere capaci di prevedere le decisioni dei giudici (e la loro successiva attuazione), cioè di formulare, come osservava Oliver Wendell Holmes "the prediction of the incidence of the public force through the instrumentality of the

courts."<sup>18</sup> Si tratta di una visione del diritto che sembra fare propria la prospettiva di una professione, quella dell'avvocato, che viene pagato per consigliare se valga la pena di andare o resistere in giudizio o stare alla larga dai giudici, in società nelle quali il comando della forza pubblica è conferito in certi casi ai giudici e alle loro pronunce. In quest'ottica "positivistico-realista", che riconduce il diritto a predizioni fattuali (previsioni concernenti il futuro comportamento dei giudici), la prevedibilità/predizione della decisione è non tanto un valore aggiunto al diritto, ma il suo aspetto centrale: "a body of dogma or systematized prediction which we call the law"<sup>19</sup>. Quindi, la predicibilità è coesistente al diritto e un diritto che non sia predicibile nemmeno in termini probabilistici non può esistere. Semplicemente non è diritto.

Alla luce di questi precedenti storici centrati sulla predizione come aspetto essenziale della professione dell'avvocato, appare un po' singolare che in Italia la riflessione sulla predizione veda una partecipazione importante e un ruolo da protagonista da parte della magistratura, più che da parte degli avvocati<sup>20</sup>.

La tematica del rapporto tra conoscenza del diritto, predizione del diritto, e attività del giudice, ci conduce a tematiche di teoria generale che non possiamo affrontare in questa sede<sup>21</sup>. Possiamo tuttavia osservare che l'attività del giudice, se certamente non può consistere nella predizione del proprio comportamento, non può neppure fondarsi sulla predizione dei comportamenti dei colleghi. Tale attività include la determinazione dei fatti del caso, l'interpretazione delle norme e dei precedenti e la costruzione di un quadro normativo che consenta di giustificare la decisione del caso, il che può richiedere scelte talvolta ispirate a giudizi di valore. Tuttavia, anche nella prospettiva del giudice, la possibilità di anticipare l'esito delle controversie ha un certo rilievo.

La predizione dell'esito delle controversie è infatti importante perché le norme giuridiche vigenti, quale che ne sia l'origine, sono soggette ad applicazione autoritativa/coercitiva da parte degli organi a ciò deputati, cioè gli organi giurisdizionali e in qualche misura gli organi amministrativi. I cittadini, nel valutare la propria e l'altrui azione secondo il diritto, tengono in considerazione il fatto che possibili controversie daranno determinazioni coercitive da parte delle autorità competenti. La predizione che una possibile controversia su un certo tema sarebbe decisa dall'autorità competente in un certo modo (per esempio, sanzionando un comportamento come illecito), induce le possibili parti ad assumere che il diritto corrisponda a tale predizione (per esempio, a considerare il comportamento illecito e, quindi, ad astenersene).

Le autorità stesse, nella decisione autoritativa/coercitiva di casi concreti, possono considerare come quegli stessi casi (o casi simili) avrebbero potuto essere decisi da altre auto-

<sup>17</sup> Prakken, H.; Sartor, G. Law and logic: A review from an argumentation perspective. *Artif. Intell.* 2015, 227, 214-245 (214 in particolare): [testo originale] "Law is of vital importance to society, promoting justice and stability and affecting many people in important aspects of their private and public life. [...] Since law has social objectives and social effects, it must be understood by those affected by it, and its application must be explained and justified. Hence the importance of clarity of meaning and soundness of reasoning, and hence the importance of logic for the law and for legal applications of AI".

<sup>18</sup> Oliver W. Holmes, *The Path of Law*, in *Harvard Law Review*, Mar. 25, 1897, Vol. 10, No. 8 (Mar. 25, 1897), 457-478 (in particolare 457) "la predizione dell'incidenza della forza pubblica attraverso lo strumento delle corti".

<sup>19</sup> Holmes, 458.

<sup>20</sup> La rivista "Questione Giustizia" dell'associazione di magistrati "Magistratura Democratica", dedica ampia parte del Fascicolo 4/2018 al tema "Una giustizia (im)prevedibile?", pubblicando ben diciotto contributi di notevole interesse. La rivista è liberamente accessibile online.

<sup>21</sup> Secondo lo stesso Holmes, l'attività del giudice comprende un aspetto politico-teleologico, quello del perseguimento di obiettivi e valori sociali, integrando o anche superando la giurisprudenza del passato. Come è noto, il rapporto tra dimensione fattuale e normativa nel diritto e nella conoscenza del diritto è il tema fondamentale nella teoria del diritto, cui diversi studiosi hanno dato diverse risposte, nel confronto tra indirizzi realistici e normativistici, così come nelle diverse posizioni sul rapporto tra diritto e morale.

rità giuridiche dello stesso livello, o di livello superiore. La predizione di decisioni dello stesso livello può avere un rilievo epistemico: se altri decisori avrebbero adottato una certa decisione, ed essi adottano di regola decisioni corrette, allora presumibilmente la decisione prevista è anche quella giusta. Se io, persona cui è affidata la decisione del nuovo caso, tendo ad adottare una decisione diversa, è probabile che mi sbagli, e quindi, la differenza tra la mia opinione e quella prevista giustifica, quanto meno, un'indagine ulteriore e una motivazione più accurata. La predizione può, inoltre, avere un rilievo "convenzionale" per il decisore, ispirata al valore della certezza del diritto: nella misura in cui la predizione del decisore coincide con la previsione dei cittadini, una decisione conforme alla previsione corrisponderebbe alle aspettative dei cittadini stessi. Infine, la predizione della decisione di livello superiore, ovviamente, indica al decisore quale soluzione giuridica sarebbe presumibilmente confermata, anziché riformata in caso di impugnazione.

Tuttavia, la prevedibilità della decisione giuridica non sempre e non principalmente si basa sul fatto che i giudici si impegnino a prevedere come decideranno i colleghi. Essa può invece basarsi sulla vicinanza teorica e ideale tra i giudici, cioè al fatto che essi facciano riferimento alle stesse fonti del diritto, adottino gli stessi metodi nell'interpretazione e la ricostruzione dei fatti, e si ispirino agli stessi obiettivi e valori fondamentali. Inoltre, la prevedibilità si fonda sul fatto, egualmente importante, che le fonti del diritto siano formulate con chiarezza, in modo da ridurre conflitti, vaghezze, imprecisioni e inutili complessità. Infine, alla certezza contribuisce la tendenza dei giudici a conformare le proprie decisioni ai precedenti, anche quando non vi sia un dovere giuridico di farlo.

L'importanza dei precedenti trova conferma nel fatto che, nel tempo, la distanza tra civil law e common law si è arricchita di sfumature che la rendono meno netta rispetto al passato, così come la distanza e la distinzione tra "precedente" e "massima" delle sentenze della Corte di cassazione. Come nota Enzo Vincenti, attuale Direttore del CED, non può più essere sostenuta l'idea che la "massima" del nostro diritto sia nettamente contrapposta al "precedente" e che il principio di diritto intercettato dalla massima sia estraneo alla logica del precedente, anzi "si può affermare che la massimazione, anche nel nostro ordinamento, crea il 'precedente'. [...] In questa direzione si sono mosse le riforme processuali succedutesi dal 2006 al 2016<sup>22</sup>, valorizzando eminentemente la funzione nomofilattica della Cassazione e, in tal modo, accreditando il precedente di legittimità come regola 'forte' di decisione per controversie giudiziarie future e, dunque, di quella capacità a farsi proiezione tendenziale degli esiti del contenzioso in cui trova applicazione. [...] il legislatore ha inteso implementare lo spazio vitale della funzione nomofilattica sem-

pre più nell'ottica valoriale della certezza giuridica, costituendo il precedente come criterio e misura della prevedibilità e calcolabilità [enfasi aggiunte] riguardo alla decisione dei casi futuri"<sup>23</sup>.

Bisogna, però, distinguere il rispetto del precedente dalla previsione delle decisioni future. Per rispettare il precedente il giudice deve guardare al passato, piuttosto che al futuro: si deve chiedere che cosa abbia deciso una corte nel passato, sulla base di quali ragioni, piuttosto che chiedersi che cosa deciderà quella stessa corte in futuro. La prevedibilità è un effetto indiretto del rispetto del precedente – nella misura in cui i giudici rispettino gli stessi precedenti, le loro decisioni saranno maggiormente prevedibili – piuttosto che il risultato di un'attività previsionale.

È peraltro da osservare che l'esigenza che il giudice consideri i precedenti – e se ne discosti solo in presenza di adeguate ragioni – non è collegata solo all'esigenza della prevedibilità; risponde anche all'esigenza che il giudicante si confronti con le soluzioni adottate nel passato da altri giudicanti, ne abbia conoscenza e spieghi le ragioni per cui eventualmente intenda discostarsene, così da partecipare a un dialogo a distanza dal quale possano emergere le soluzioni migliori.

### *Prevedibilità e memoria*

La prevedibilità delle decisioni come base di un trattamento uniforme dei cittadini dinanzi al giudice è un bene e, al tempo stesso, un problema, perché se è vero che in termini generali, e indipendentemente dall'organizzazione giuridica e dal retroterra culturale dei diversi ordinamenti, rimane centrale la necessità per un cittadino di fare affidamento sul sistema e sulle sue istituzioni, è anche vero che essa è diventata problematica negli ultimi decenni, a causa della complessità dei processi di produzione del diritto, della globalizzazione giuridica e della transnazionalità del diritto<sup>24</sup>. Si tratta di un punto molto delicato sul quale torneremo nei paragrafi finali.

Questo quadro circa la prevedibilità e la predizione giuridica è preesistente all'emergere della cosiddetta *new prediction*, che è al centro della ricerca sin dal 1974 (sui casi di controversie fiscali), che ha la pretesa di essere quantitativa e di essere il risultato dell'applicazione di tecniche di IA<sup>25</sup>. Di questo, e dei profili tecnici attuali, ci occuperemo nel prossimo paragrafo. Prima può essere utile una breve incursione in un campo, quello delle neuroscienze, per cogliere un aspetto interessante della predizione e dei suoi rapporti con la memoria.

In un volume pubblicato presso MIT Press, due studiosi italiani della memoria, Tommaso Vecchi e Daniele Gatti, sostengono che la capacità di conservare informazioni (quella che comunemente è intesa come memoria del passato) è una proprietà collegata al futuro, tanto da potersi

<sup>22</sup> In particolare: art. 360-bis, n. 1, c.p.c., sull'inammissibilità del ricorso per cassazione allorché questo non offra elementi per ripensare l'orientamento consolidato della giurisprudenza di legittimità al quale il provvedimento impugnato si conforma; art. 363 c.p.c., sull'enunciazione del principio di diritto nell'interesse della legge, allorché il ricorso sia inammissibile o il provvedimento non ricorribile; art. 374, 3° comma, c.p.c., sull'obbligo, per le sezioni semplici, di rimettere il ricorso alle sezioni unite quando ritengano di non condividere un principio di diritto da queste ultime già enunciato; art. 384, 1° comma, c.p.c., sull'enunciazione del principio di diritto anche nei casi in cui non si provveda alla cassazione con rinvio.

<sup>23</sup> Enzo Vincenti, *Massimazione e conoscenza della giurisprudenza nell'era digitale* [https://www.questionegiustizia.it/rivista/articolo/massimazione-e-conoscenza-della-giurisprudenza-nell-era-digitale\\_588.php](https://www.questionegiustizia.it/rivista/articolo/massimazione-e-conoscenza-della-giurisprudenza-nell-era-digitale_588.php),

in *Questione Giustizia*, 2018/4.

<sup>24</sup> Si vedano, tra i molti, Mark Fenwick – Wulf Kaal – Erik P.M. Vermeulen, *Legal education in a digital age: why coding for lawyers matters*, in *Legal Tech, Smart Contracts and Blockchain (Perspectives in Law, Business and Innovation)*, Springer Singapore, 2018; Siems, M. Wrbka, S. (Eds.) *The Shifting Meaning of Legal Certainty in Comparative and Transnational Law*; Hart Publishing: Oxford, UK, 2017. In Italia si veda Paolo Grossi, *Sulla odierna 'incertezza' del diritto*, in *Giustizia Civile: rivista giuridica trimestrale*, n. 4, 2014, 921-955. Sia consentito rinviare anche a A. Santostuo, *Intelligenza artificiale e diritto*, Milano, 2020, specie il capitolo finale.

<sup>25</sup> Ashley, K. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*; Cambridge University Press: Cambridge, UK, 2017, 107, par. 4.1.

dire che “il vero scopo della memoria è fare predizioni”. Secondo autorevoli studi in materia “il sistema cognitivo anticipa continuamente e attivamente ciò che sta per accadere sulla base della memoria”. Insomma, l’informazione sul passato è intrecciata con i processi predittivi, siano essi umani o strettamente biologici o ambientali. Il sistema cognitivo mette in pratica i risultati dell’elaborazione delle relazioni causali, e dei rapporti tra eventi, basati sull’apprendimento precedente, cioè sulla memoria. Tale apprendimento può avvenire estraendo regolarità statistiche (costruendo sequenze o modelli di eventi) o facendo un uso più o meno esplicito di inferenze e analogie tra gli eventi; questi modelli consentono, quindi, di fare delle previsioni<sup>26</sup>.

Se questo è quello che accade nella nostra attività mentale in generale, qualcosa del tutto simile accade nella mente del giurista dell’accademia o del pratico del diritto, giudice o avvocato. Lo studio dei testi di dottrina o delle raccolte cartacee di giurisprudenza è il modo in cui sono state fatte le predizioni giuridiche dopo l’invenzione della stampa a caratteri mobili, usando carta, penna, annotazioni personali, memoria di studio e di esperienze pratiche. Oggi, quella stessa attività può essere svolta con l’aiuto di strumenti di Legal Analytics (LA), di cui parliamo nel prossimo paragrafo, cogliendo aspetti comuni e differenze.

Tornando al campo del diritto e alle prime applicazioni di tipo informatico si può ricordare che l’idea che uno scopo fondamentale dell’applicazione dell’informatica al diritto consista nelle predizioni si può ritrovare fin dalle origini dell’informatica giuridica<sup>27</sup>, ancor prima che si realizzassero le prime vere applicazioni predittive (il primo sistema per la ricerca documentale risale agli anni ‘50). Solo negli anni più recenti l’idea di usare sistemi informatici per prevedere l’esito di controversie è divenuta reale. Nelle pagine seguenti si illustrerà dapprima il modo in cui le predizioni possono avvenire usando tecnologie informatiche, partendo dal modello dei sistemi basati sulla conoscenza (sistemi esperti), per poi passare a considerare la predizione nei sistemi per l’apprendimento automatico.

## Le tecnologie della predizione

Nella presente sezione si illustreranno le tecnologie che possono essere utilizzate nella predizione in ambito giuridico. Si partirà dai sistemi esperti, il paradigma prevalente fino agli anni ‘90, per poi passare all’apprendimento automatico.

### *La predizione attraverso sistemi basati sulla conoscenza (sistemi esperti)*

Fino a pochi decenni fa, si assumeva che al fine di realizzare un sistema intelligente, gli umani dovessero fornire tutta la conoscenza necessaria all’operatività del sistema, espressa in un linguaggio costruito con precisione, cioè in un linguaggio formale. Non esistono a tutt’oggi sistemi informatici in grado di comprendere il significato del linguaggio umano, di avere accesso, in senso pieno, alla “semantica”. Pertanto, le operazioni che un sistema può fare con delle parole dipendono solo dalle strutture sintattiche impiegate per collegare quelle parole, strutture sintattiche che costituiscono appunto la forma delle espressioni del linguaggio.

Il sistema ragiona solo sulla base delle forme (non dei contenuti), e in questo senso il suo ragionamento è formale, o se si preferisce, è un “ragionamento cieco”, come diceva uno dei massimi filosofi (e giuristi) del passato, G.W. Leibniz, nel quale vengono manipolati dei simboli senza avere consapevolezza del loro significato. Per chiarire il concetto di sistema formale consideriamo la classica inferenza sillogistica che dalle premesse (1) “Ogni uomo è mortale”, (2) “Socrate è un uomo”, deriva la conclusione (3) Socrate è mortale. Per effettuare questa inferenza non abbiamo bisogno di sapere che cosa significa uomo, che cosa significa mortale, che cosa significa Socrate. Ci basta riconoscere la struttura linguistica (la forma) delle premesse, e applicare una regola di inferenza agli elementi racchiusi in quella struttura. Anche un sistema automatico, una volta riconosciuta la forma delle premesse (ogni X è Y, z è un X, dove X e Y esprimono concetti generali, e z è il nome di un individuo), sarà in grado di trarre la conseguenza (z è Y): a tal fine non occorre capire i significati delle parole, basta unire il terzo termine nelle premesse (z) al secondo attraverso la copula “è”. Allo stesso modo consideriamo due semplici regole in materia di cittadinanza, che varrebbero se il cosiddetto *ius soli* fosse introdotto nel nostro paese, e due asserti di fatto relativi a due persone, Antonio e Mara.

SE Y è nato in Italia ALLORA Y è italiano

SE Y è genitore di X E Y è italiano ALLORA X è italiano.

Aggiungiamo due fatti specifici

Antonio è nato in Italia

Antonio è genitore di Mara.

Anche in questo caso per trarre la conclusione che Mara è italiana, non ci serve conoscere il significato delle parole “genitore”, “cittadino”, “italiano”. Basta sostituire il nome Antonio al posto della Y nella prima regola così da ottenere la regola specifica “SE Antonio è nato in Italia, ALLORA Antonio è italiano”, e usare quest’ultima regola in combinazione con il primo fatto (Antonio è nato in Italia), per trarre la conseguenza “Antonio è italiano”. A questo punto possiamo sostituire alla Y e alla X nella seconda regola rispettivamente Antonio e Mara, e dato che Antonio è genitore di Mara (il secondo fatto) e che Antonio è italiano (dal ragionamento precedente), concludere che Mara è italiana (o meglio, italiana). Anche in questo caso, il ragionamento è puramente formale e quindi automatizzabile: si basa sulle strutture sintattiche e prescinde dal significato delle parole (tranne quelle usate, secondo regole precise e univoche, per esprimere strutture sintattiche, come SE ... ALLORA ed E, nel nostro esempio, che possono essere infatti sostituite dai simboli usuali della logica matematica.

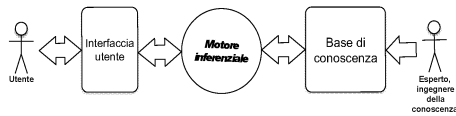
Numerosi formalismi sono stati sviluppati per la rappresentazione della conoscenza (regole, concetti, schemi, logiche classiche, modali, descrittive, *defeasible*, schemi – *frame*-, ontologie, ecc.) combinati con algoritmi per compiere diversi tipi di inferenze sulla base di tali rappresentazioni (inferenze deduttive, presuntive, induttive, probabilistiche, argomentative, basate sui casi, ecc.).

La struttura di un sistema esperto è rappresentata nella Figura 1. Si noti che gli umani compaiono sia come utenti del sistema, sia come creatori della base di conoscenza del sistema stesso. La creazione della conoscenza è affidata a

<sup>26</sup> Tomaso Vecchi-Daniele Gatti, *Memory as Prediction*. From Looking Back to Looking Forward, MIT Press, 2020, 45 e passim.

<sup>27</sup> Lee Loewinger. *Jurimetrics: The next step forward*. Minnesota Law Review, 33:455-493, 1949.

esperti nella materia su cui verte l'uso del sistema, possibilmente aiutati da esperti nella rappresentazione formale della conoscenza (i cosiddetti ingegneri della conoscenza).  
Figura 1. Sistemi basati sulla conoscenza



I sistemi esperti in ambito giuridico hanno condotto a risultati applicativi significativi, ma non così dirimpenti come ci si aspettava, comunque non tali da cambiare le regole del gioco. Possono fornire risposte adeguate solo nei casi previsti da chi ne ha realizzato la base di conoscenza. Infine, tali sistemi richiedono interventi frequenti per aggiornare la loro base di conoscenza, seguendo lo sviluppo del diritto (nuove norme, abrogazioni esplicite o implicite, ecc.).

Si possono distinguere due tipi di sistemi basati sulla conoscenza utilizzati in ambito giuridico. Quelli che consistono di regole e concetti, e quelli basati sui casi.

I sistemi basati sulla conoscenza che *applicano regole e concetti* sono stati impiegati soprattutto nell'applicazione amministrativa di regole nel campo della sicurezza e previdenza sociale e della tassazione. Questi contesti sono governati da vasti insiemi di regole connesse, ciascuna delle quali ha un significato preciso e disciplina circostanze usualmente ben specificate, non controverse. Si pensi, ad esempio, alle regole che stabiliscono che, per avere diritto a certe prestazioni previdenziali, sono richieste certe condizioni personali (età, stato di famiglia, ecc.) e certe condizioni patrimoniali (reddito, proprietà, ecc.), ulteriormente specificate in altre norme. Sistemi basati sulla conoscenza possono affrontare in modo soddisfacente il compito di applicare diligentemente le regole rilevanti per il caso in esame, una volta che il caso stesso sia stato formulato usando il linguaggio impiegato nelle regole.

Sistemi di questo tipo (che applicano regole e concetti) hanno invece avuto meno successo nel supportare l'analisi giuridica negli ambiti in cui i fatti, i concetti, o le regole siano controversi, come nel caso della decisione giudiziale. Anche in ambiti amministrativi, tali sistemi hanno un uso limitato qualora non sia comunque possibile codificare in anticipo, nelle regole del sistema (strutture del tipo SE ... ALLORA, come quelle usate nell'esempio precedente), informazioni sufficienti per sussumere casi specifici, o quando comunque si richieda di tener conto, con discrezionalità, delle diverse situazioni di fatto alla luce dei valori giuridici rilevanti e degli obiettivi politici perseguiti. Possiamo forse dire che i limiti di questi sistemi sono quelli propri della cosiddetta "giurisprudenza meccanica", cioè dell'idea secondo cui ogni caso possa essere risolto mediante la precisa applicazione di regole precostituite in anticipo<sup>28</sup>. Ciò non esclude che in molti ambiti regole di questo tipo possano condurre a soluzioni appropriate nella grande mag-

gioranza dei casi, e che quindi possa essere vantaggioso ricorrere a sistemi informatici per la loro applicazione, salva la possibilità di ricorrere all'intervento umano per valutare se adattamenti o eccezioni siano giustificate in casi particolari, secondo principi giuridici o regole prevalenti. Quale esempio di un sistema che ha avuto un certo successo possiamo ricordare il *Oracle Policy Automation*, mediante il quale sono stati realizzati sistemi basati sulla conoscenza in tema di immigrazione, prestazioni sociali e imposte (in Australia, nel Regno Unito, negli Stati Uniti e in Francia). Nel nostro paese possiamo ricordare il sistema ReMida, sviluppato da Gianfranco D'Aiuti.

Anche un sistema basato su regole e concetti può essere usato per fare predizioni. Se al sistema si forniscono tutte le caratteristiche di un nuovo caso, espresse nel linguaggio usato per specificare le regole del sistema, il sistema stesso sarà in grado di trarre la conclusione che risulta dall'applicazione delle regole a quel caso (come nell'esempio precedente, da regole e fatti si è potuto concludere che Mara è cittadina italiana). Tale conclusione potrà essere intesa come una predizione di una decisione futura da parte del decisore competente. Per esempio, un sistema basato su regole potrà consentire al cittadino di determinare se la richiesta di una prestazione previdenziale potrà aver successo. La predizione sarà affidabile nella misura in cui chi ha inserito regole e concetti nel sistema e chi ha qualificato i fatti corrispondenti nei termini di quelle regole sia riuscito ad anticipare, nel formulare le regole inserite nel sistema, il modo nel quale i decisori competenti avrebbero interpretato il diritto e qualificato i fatti.

È il caso di precisare che i sistemi informatici che operano inferenze sulla base di una rappresentazione logica della conoscenza vanno al di là dell'effettuazione di inferenze deduttive. Esistono sistemi in grado di trarre conclusioni presuntive<sup>29</sup>, sviluppare argomentazioni, proporre interpretazioni e analogie, di rappresentare l'evoluzione di normative nel tempo, di ragionare sulla base di casi e così via. Si tratta, tuttavia, di applicazioni ancora prototipali, di cui non sono al momento disponibili applicazioni significative per la pratica del diritto<sup>30</sup>.

Quale esempio di sistema per il ragionamento che opera sulla base di *rappresentazioni di casi*, anziché di regole, possiamo ricordare il sistema HYPO, applicato in particolare nell'ambito della violazione dei segreti industriali<sup>31</sup>. La base di conoscenza del sistema è costituita da un insieme di precedenti. Ciascun precedente è descritto (annotato) con il suo esito e mediante un insieme di *fattori*, cioè di circostanze a favore o contro quell'esito. Per esempio, nei casi in cui il convenuto sia venuto a conoscenza di informazioni sulle attività dell'attore, alcuni fattori, come i seguenti portano alla conclusione che vi sia stata una violazione del segreto industriale: l'attore aveva adottato specifiche iniziative intese a mantenere il segreto, il prodotto cui afferiva il segreto era unico, il convenuto si era impegnato a rispettare il segreto. Invece altri fattori, come i seguenti, favoriscono la conclusione che non vi sia stata violazione: l'attore aveva

<sup>28</sup> Sul rapporto tra *rules* e *legal system* (inteso come 'insieme di norme') è interessante la visione di Frederick Schauer, secondo il quale la qualità di *rule* va riservata solo alla norma di origine legislativa o (ma a determinate condizioni, 175-7) anche giurisprudenziale, ben scritta e a contenuto specifico, generale e aperto, ragion per cui non sono *rules*, criteri decisionali come *best interest* e *equity*: Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*, Clarendon Press, Oxford, 1991 10-11.

<sup>29</sup> Suscettibili di essere revocate sulla base di informazioni ulteriori

(per esempio, la conclusione che l'imputato è punibile, avendo commesso un fatto costituente reato, potrà essere revocata quando risulti che egli ha agito in presenza di una causa di giustificazione. Per indicare questo tipo di conclusioni si usa la parola inglese "defeasible", che letteralmente significa "sconfiggibile").

<sup>30</sup> Henry Prakken and Giovanni Sartor. *Law and logic: A review from an argumentation perspective*. *Artificial Intelligence*, 227:214-45, 2015.

<sup>31</sup> Kevin D. Ashley. *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. MIT, 1990.



comunicato il segreto nel corso di trattative con il convenuto, l'informazione era ottenibile mediante *reverse-engineering*, etc. Una volta che il nuovo caso sia stato rappresentato mediante un insieme di fattori, è possibile ottenere indicazioni sulla sua possibile decisione ricorrendo al ragionamento analogico. Per esempio, i precedenti che siano stati decisi a favore dell'attore, e che condividano con il nuovo caso dei fattori a favore della violazione del segreto, possono essere prodotti a sostegno della stessa conclusione, mentre il convenuto può distinguere il nuovo caso dal precedente, adducendo fattori contro la violazione (assenti dal precedente) o la mancanza di fattori a favore della stessa (compresi nel precedente).

Anche il ragionamento per fattori può essere utilizzato per fare previsioni, nella misura in cui la descrizione dei precedenti mediante fattori sia riuscita a cogliere gli aspetti che hanno determinato la decisione di quei precedenti e la loro relativa importanza. Lo sviluppo di sistemi che adottano questo indirizzo è ostacolato dalla difficoltà di associare, in modo coerente, tutti i fattori rilevanti a un ampio insieme di casi. Si tratta, infatti, di un compito che richiede molto tempo, rischia di riflettere le idiosincrasie degli esperti incaricati di questo compito e comunque di non cogliere tutti gli aspetti del caso che hanno influito sulla sua decisione. Per superare questi limiti si sta sperimentando l'uso di tecnologie per l'elaborazione del linguaggio naturale al fine di assegnare fattori ai casi<sup>32</sup>. Altre ricerche combinano il ragionamento per fattori, e inferenze concettuali e argomentative<sup>33</sup>.

Per concludere l'analisi dei sistemi esperti, possiamo dire che i sistemi basati sulla rappresentazione umana della conoscenza, sia essa costituita da regole e concetti o da fattori associati ai casi, sinora non hanno condotto a risultati significativi nel campo della predizione delle decisioni dei giudici. Ciò si può spiegare con la difficoltà di tradurre la complessità delle problematiche giuridiche e le sfumature della conoscenza giuridica in una rappresentazione formalizzata di regole e concetti. Creare una rappresentazione formalizzata della conoscenza giuridica è un compito difficile e faticoso, i cui risultati sono parziali e riflettono le selezioni e le scelte interpretative di chi opera la formalizzazione, nonché il tempo in cui essa è avvenuta. Inoltre, il ragionamento giuridico procede secondo l'applicazione di regole e concetti, ma richiede anche la capacità di adottare soluzioni adeguate a realizzare valori e interessi, di cogliere analogie tra precedenti e nuovi casi, richiede la consapevolezza dei problemi sociali e umani implicati nell'applicazione del diritto. Più che strumento per la previsione dei giudici, i sistemi basati su regole possono essere visti come mezzi per conoscere il diritto e applicarlo ai casi concreti (tenendo in considerazione che il diritto è visto attraverso la prospettiva di chi ha creato la base di conoscenza del sistema).

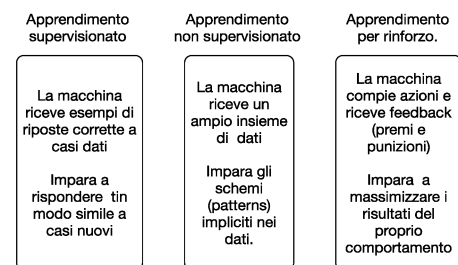
### L'apprendimento automatico

L'intelligenza artificiale ha fatto enormi passi avanti da quando si è concentrata sull'applicazione dell'apprendimento automatico a grandi masse di dati. Nei sistemi basati sull'apprendimento automatico la conoscenza relativa al-

l'ambito di applicazione del sistema non è più fornita dall'uomo, ma piuttosto è costruita dalla macchina, sulla base dei dati cui essa ha accesso. Questo indirizzo ha condotto a un grande numero di applicazioni di successo in molti settori, dalla traduzione automatica, all'ottimizzazione industriale, al marketing, alla visione robotica, al controllo dei movimenti, e altro ancora. Grazie ai metodi di apprendimento automatico, le macchine possono estrarre le informazioni rilevanti dai dati di input e così imparare a svolgere in modo adeguato le funzioni a esse affidate. In questo modo, come già osservava Alan Turing negli anni '50<sup>34</sup>, una macchina capace di apprendere potrà operare in modi che non sono stati anticipati dai suoi creatori e addestratori, e anche senza che questi conoscano i dettagli del funzionamento interno della macchina.

Si distinguono tre principali indirizzi nell'apprendimento automatico: l'apprendimento supervisionato, l'apprendimento per rinforzo e l'apprendimento non-supervisionato (vedi Figura 2).

Figura 2. Tipi di apprendimento automatico



*L'apprendimento supervisionato.* Nell'apprendimento supervisionato la macchina apprende mediante supervisione, cioè attraverso una fase di istruzione o addestramento in cui le viene fornito un ampio insieme di esempi – soluzioni corrette a casi particolari – sulla base dei quali essa costruisce un modello generale, applicabile anche a casi nuovi, parzialmente diversi da quelli nell'insieme di addestramento.

Più esattamente, il sistema riceve un insieme di coppie, ciascuna della quale combina la descrizione di un problema alla risposta corretta allo stesso. Ecco alcuni esempi: in un sistema destinato a riconoscere oggetti (per esempio, a classificare animali) che compaiono in immagini online, ogni immagine nell'insieme di addestramento è etichettata con il tipo di oggetto che contiene (e.g., cane, gatto, coniglio, ecc.); in sistemi per la traduzione di testi, ogni (frammento) di documento nel linguaggio sorgente è appaiato alla sua traduzione; nei sistemi per la selezione del personale, la descrizione di candidati passati (età, esperienza, studi, ecc.) è collegata all'esito della selezione (o alla valutazione della prestazione lavorativa del candidato) nel caso in cui questi sia stato assunto; nei sistemi che propongono suggerimenti (*recommender system*) sugli acquisti, le caratteristiche dei consumatori passati vengono associate agli oggetti acquistati dagli stessi; nei sistemi che valutano richieste di mutuo, i dati raccolti sulle richieste già valutate (in particolare le informazioni sul richiedente) sono associati alla decisione adottata (o anche, nel caso di richieste accettate

<sup>32</sup> Kevin D. Ashley and Stephanie Brueninghaus. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17:125-165, 2009; L.K. Branting, C. Pfeifer, B. Brown, L. Ferro, J. Aberdeen, B. Weiss, M. Pfaff, e B. Liao, Scalable and explainable legal prediction. *Artificial Intelligence and Law*, 29:213-38, 2021.

<sup>33</sup> Henry Prakken and Giovanni Sartor. Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214-45, 2015.

<sup>34</sup> Alan M. Turing. Computer machinery and intelligence. *Mind*, 59:433-60, 1950.

con informazioni sul successivo adempimento o inadempimento del richiedente).

Come illustrano questi esempi, l'addestramento di un sistema non sempre richiede un istruttore umano con il compito di fornire esempi di risposte corrette al sistema. In molti casi, l'insieme di addestramento può essere raccolto "nella natura" (*in the wild*), cioè dal web aperto. Per esempio, immagini o facce disponibili sulle reti sociali (nelle pagine aperte al pubblico) possono essere prelevate ("raschiate" o *scraped*, in inglese) e usate per l'addestramento di classificatori automatici. Nel caso prototipico di giustizia predittiva gli esempi consisteranno in casi passati, e ciascun esempio assocerà la descrizione di un caso alla risposta allo stesso, costituita dalla decisione del giudice in quel caso. Lo scopo di sistemi intesi a "predire" la decisione dei giudici non è, quindi, quella di fornire sempre la decisione corretta, ma piuttosto quella di anticipare le decisioni che di fatto i giudici potrebbero adottare.

**L'apprendimento per rinforzo.** L'apprendimento per rinforzo è simile all'apprendimento supervisionato, nel senso che entrambi comportano l'addestramento mediante esempi. Tuttavia, nel caso di addestramento per rinforzo, il sistema apprende dai risultati delle proprie azioni, cioè dalle ricompense o penalità (punti guadagnati o persi) che sono collegati ai risultati a quelle azioni. Per esempio, nel caso di un sistema destinato a partecipare a un gioco (per esempio, gli scacchi), le ricompense possono essere collegate alle vittorie e le penalità alle sconfitte; in un sistema che apprende a giocare in borsa, le ricompense possono essere collegate ai guadagni ottenuti e le penalità alle perdite; in un sistema che impara a inviare messaggi pubblicitari mirati, le ricompense possono essere collegate ai click degli utenti. In ogni caso, il sistema basato sul rinforzo osserverà i risultati delle proprie azioni e si autosamminerà le ricompense e penalità appropriate. Essendo diretto a massimizzare il proprio punteggio complessivo, il sistema apprenderà a compiere le azioni che più probabilmente conducono a risultati collegati a ricompense (vittorie, guadagni, click) e a evitare le azioni che più probabilmente conducono a risultati collegati a penalità (sconfitte, perdite, nessun click).

Nel campo della giustizia predittiva non sembra operino già ora sistemi basati sull'idea dell'apprendimento per rinforzo. A tale fine bisognerebbe che fosse possibile determinare le conseguenze delle decisioni del sistema e valutarle come positive e negative. Per esempio, il rinforzo potrebbe essere positivo (ricompensa) o negativo (punizione) a seconda che il giudice umano accolga il suggerimento, o lo respinga, oppure a seconda che la proposta del sistema venga confermata o rigettata in appello.

**L'apprendimento non supervisionato.** Nell'apprendimento non-supervisionato, infine, il sistema impara senza ricevere istruzioni, né da fonti esterne (apprendimento supervisionato), né dai risultati delle proprie attività (apprendimento per rinforzo). Le tecniche dell'apprendimento non-supervisionato sono usate in particolare per il raggruppamento (cosiddetta clusterizzazione - *clustering*), cioè per riunire insieme di oggetti che presentano somiglianze o connessioni rilevanti (i documenti che riguardano gli stessi oggetti o problemi, le persone con simili caratteristiche, le parole che hanno simili significati o funzioni). Per esempio, nell'ambito di un'indagine, può essere utile individuare i raggruppamenti nei documenti elettronici disponibili, per individuare quelli che attengono al caso in esame. Oppure in un'indagine su precedenti giudiziari finalizzata all'adozione di nuove politiche criminali, si potrebbe chiedere al sistema di raccogliere i casi simili, e quindi esaminare, per

esempio, le connessioni tra certi reati e l'uso di droghe o di armi. Quindi l'apprendimento non-supervisionato può essere utilizzato, sulla base delle tecniche oggi disponibili solo per compiti ausiliari rispetto all'obiettivo di prevedere le decisioni dei giudici.

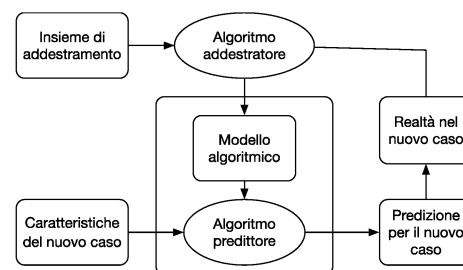
#### *Addestramento e costruzione di un modello nell'apprendimento supervisionato*

Mettiamo ora a fuoco alcuni aspetti fondamentali dei sistemi di apprendimento automatico. In tutti questi sistemi si possono distinguere due componenti fondamentali: l'algoritmo che apprende (*learning*, detto anche algoritmo addestratore, *training*) usando gli esempi forniti al sistema, e l'algoritmo o modello appreso (il risultato dell'apprendimento).

La Figura 3 raffigura il processo dell'apprendimento supervisionato. L'algoritmo che apprende usa l'insieme di addestramento per costruire un modello del compito che il sistema dovrà svolgere. Tale modello può essere visto, in generale, come una funzione matematica, cioè come un meccanismo che collega possibili input a output corrispondenti. Per esempio, il modello potrebbe collegare possibili immagini di animali alle parole che denotano la specie animale corrispondente (immagini di gatto alla parola "gatto", di cane alla parola "cane", ecc.) o potrebbe collegare descrizioni dei "fatti" (naturali o giuridici) presenti in possibili casi giudiziari all'indicazione di decisioni corrispondenti.

Ogni modello non riproduce gli esempi presenti nell'insieme di addestramento, ma effettua una generalizzazione: esso può essere applicato anche a nuovi casi, che differiscono in qualche aspetto da ciascuno degli esempi su cui si è basato l'addestramento.

Figura 3. Apprendimento supervisionato



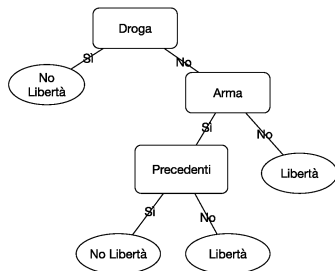
Il modello è poi usato da un diverso algoritmo, l'*algoritmo predittore*, per fornire risposte sperabilmente corrette a casi nuovi. Se gli esempi più simili al caso nuovo (rispetto alle caratteristiche suscettibili di influenzare l'esito) hanno avuto una certa risposta, l'algoritmo predittore potrà proporre la stessa risposta per il caso nuovo. Per esempio, se nell'insieme di addestramento le immagini di animali più simili (negli aspetti correlati alla classificazione) alla nuova immagine proposta al sistema sono etichettate come immagini di gatto, anche la nuova immagine sarà classificata nello stesso modo; se nell'insieme di addestramento i richiedenti di prestiti le cui caratteristiche si avvicinano a quelle del nuovo candidato sono classificati come inaffidabili, il sistema classificherà nello stesso modo il nuovo candidato; se in passato i lavoratori aventi caratteristiche più simili al nuovo candidato sono stati assunti, il sistema predirà l'assunzione del nuovo candidato.

Infine, nell'ambito per noi di interesse, se i casi maggiormente simili al nuovo caso, rispetto agli aspetti suscettibili di influenzare la decisione nell'una o nell'altra direzione

(secondo il modello costruito dal sistema), hanno condotto a una certa decisione, la stessa sarà proposta anche nel nuovo caso.

Figura 4. Apprendimento supervisionato: libertà vigilata

Insieme di addestramento					
Caso	Predittori				Esito
	Lesione	Droga	Arma	Precedenti	Decisione
1	Nessuna	No	No	SI	SI
2	Grave	SI	No	Gravi	No
3	Nessuna	No	SI	No	SI
4	Grave	SI	No	SI	No
5	Leggera	SI	SI	SI	No
6	Nessuna	SI	SI	Gravi	No
7	Nessuna	No	SI	SI	No



Quale semplice esempio di apprendimento supervisionato, la Figura 4 mostra un piccolo insieme di addestramento (alcune decisioni in tema di libertà vigilata) e di sotto l'albero di decisione che può essere appreso da quell'insieme di addestramento (si tratta di un albero a orientamento invertito, nel quale dalla radice in alto si dipartono i rami verso il basso). L'albero di decisione riprende e generalizza l'informazione implicita nell'insieme di addestramento nella forma di una combinazione di test, da effettuarsi sequenzialmente, seguendo uno dei rami dell'albero, dall'alto verso il basso. Il primo test chiede se l'imputato sia stato implicato in un reato correlato alla droga. Se la risposta è positiva abbiamo raggiunto il fondo dell'albero (una sua foglia) con la conclusione che la libertà vigilata è negata. Se la risposta è negativa (non si tratta di droga), si passa al secondo test, che chiede se l'imputato abbia usato un'arma, e così via.

Si osservi che l'albero di decisione non include informazioni riguardanti il tipo di lesione causata dal reato. Ciò è dovuto al fatto che tutte le decisioni contenute nell'insieme di addestramento possono essere spiegate senza far riferimento a questa informazione: applicando l'albero di decisione otteniamo, per ciascun caso nell'esempio, la decisione adottata dai giudici in quel caso. L'algoritmo per la costruzione dell'albero di decisione ha costruito il modello esplicativo più semplice tra quelli che consentono di spiegare tutti i casi, omettendo le informazioni inutili a tal fine.

In questo esempio possiamo chiaramente distinguere gli elementi raffigurati nella Figura 4. La tabella è l'*insieme di addestramento*, e l'albero di decisione è il *modello*. Il software che costruisce l'albero di decisione a partire dalla tabella è l'*algoritmo addestratore*. In questo caso, l'albero di decisione codifica la logica (così come ricostruita dal software addestratore) delle decisioni umane rappresentate nell'insieme di addestramento. Il software che predice la decisione nuovi casi, usando l'albero di decisione, e compie predizioni basate sulla caratteristica di tali casi è l'*algoritmo predittore*.

L'albero di decisione riflette le attitudini dei decisori umani le cui decisioni sono comprese nell'insieme di addestramento, riproducendo le virtù e i vizi di quelle decisioni. Per esempio, nell'albero di decisione, il fatto che l'imputazione concerna un reato collegato agli stupefacenti è sufficiente perché sia negata la libertà vigilata. Ci potremmo chiedere se questo criterio di decisione sia strettamente legale e, se no, se esso sia equo o rifletta dei pregiudizi dei decisori rispetto a chiunque sia coinvolto nel traffico illegale di stupefacenti. Inoltre, si osservi che l'albero di decisione fornisce risposte anche rispetto ai casi che non coincidono esattamente con alcun caso compreso nell'insieme di addestramento. Per esempio, non c'è alcun caso, nell'insieme di addestramento, che riguardi un reato connesso all'uso di droga nel quale non siano state usate armi e non ci siano precedenti penali. Tuttavia, il sistema fornisce una risposta anche in questo caso: la libertà vigilata dovrebbe essere negata, poiché ciò accade in tutti i casi compresi nell'insieme di addestramento in cui ci sia uso di droga. Ci si può inoltre interrogare se questa generalizzazione sia corretta. In un caso in cui ci sia uso di droghe ma non vi siano armi né lesioni, potrebbe invece essere preferibile concedere la libertà vigilata. Inoltre, si consideri come la predizione del sistema si basi solo sui fattori da esso considerati, e non consideri altri aspetti possibilmente rilevanti per determinare l'opportunità di concedere la libertà vigilata, come il carattere dell'imputato, le sue connessioni con ambienti criminali, ecc.

L'esempio illustra come i suggerimenti forniti da un sistema automatico si basino sempre su generalizzazioni estratte da casi simili. I critici delle decisioni automatizzate<sup>35</sup> affermano infatti che tali decisioni non sarebbero mai veramente individualizzate: anche i sistemi più avanzati si limiterebbero a mettere nello stesso secchio i casi simili (*bucketing*) senza considerare sul serio le circostanze di ogni caso: la decisione su una particolare persona si baserebbe solo sull'appartenenza di quella persona a una certa classe. Tale classe ricomprende infatti le persone che, nell'insieme di addestramento sono assimilate all'individuo in questione, condividendo con lo stesso caratteristiche emerse come significative nei precedenti già considerati (per esempio, per essere stati coinvolti in reati attinenti al possesso o traffico di stupefacenti). All'individuo coinvolto nel nuovo caso viene associata una predizione (il diniego della libertà vigilata) sulla base del fatto che in passato a persone simili a lui (anch'esse coinvolte nel traffico di stupefacenti) tale libertà è stata negata.

A questa critica si può rispondere osservando che le proposte del sistema diventano più individualizzate quanti più sono i predittori che esso prende in considerazione e quanto più ampio il suo insieme di addestramento. Inoltre, anche i giudizi umani si basano su generalizzazioni basate su esempi passati, in cui il giudicante è stato coinvolto o di cui ha avuto notizia, ragion per cui l'assoluta individualizzazione sfuggirebbe anche al decisore umano<sup>36</sup>.

Un'altra critica può far riferimento alla possibilità che i dati stessi siano incoerenti, non in quanto errati, ma perché riflettono un contrasto reale: come potrebbe un modello che fornisce un'unica soluzione a ciascun caso, dare un'immagine adeguata di una realtà caratterizzata da conflitti di giudizi, come tipicamente accade in ambito giuridico. Immaginiamo, per esempio, che nell'insieme di addestramen-

<sup>35</sup> Cathy O'Neil. Weapons of math destruction: how big data increases inequality and threatens democracy. Crown Business, 2016.

<sup>36</sup> Frederick Schauer. Profiles, Probabilities and Stereotypes. Belknap, 2003.

Dottrina e attualità giuridiche | **Giustizia predittiva**

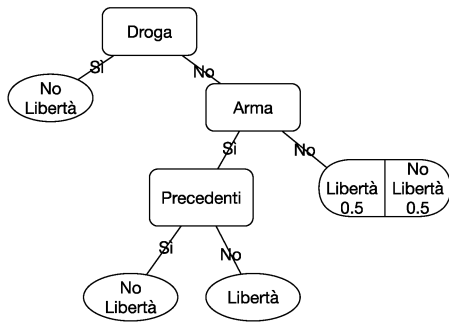
to sia inserito un nuovo caso, con valori identici al primo caso della lista, ma con decisione opposta (Figura 5).

Figura 5. Un nuovo caso

Caso	Lesione	Droga	Arma	Precedenti	Decisione
8	Nessuna	No	No	Si	No

Per catturare questo conflitto nell'albero di decisione, potremmo specificare che il ramo "Droga=no, Arma =no" da luogo a decisioni alternative, ciascuna associata a una probabilità che corrisponde alla proporzione delle decisioni corrispondenti.

Figura 6. Albero di decisione con incertezza



Usando questo albero di decisione, il sistema potrà dare, a chi chiede una predizione per il caso in cui una persona abbia commesso un reato in cui non vi siano stupefacenti né armi, la previsione che vi è sono 50% probabilità per il rilascio e 50% per la detenzione.

È possibile altresì estrarre regole alternative da un insieme di addestramento valutandone l'affidabilità. Nel nostro caso un sistema per l'analisi di un dataset mediante l'estrazione di regole di associazione, applicato all'insieme di addestramento esteso con il nuovo caso) ci direbbe che tanto la regola "SE Droga=no, Armi=no ALLORA Libertà=sì", quanto la regola "SE Droga=no, Armi=no ALLORA Libertà=no" hanno un supporto (rapporto tra il totale dei casi e quelli in cui sono soddisfatti sia l'antecedente sia il conseguente della regola) pari a 1/8=0.15, e una confidenza (rapporto tra i casi in cui è soddisfatto l'antecedente e quelli cui oltre all'antecedente è soddisfatto anche il conseguenti) è 0.5. Infatti, solo un caso su 8 ha i valori della prima regola (Droga=no, Armi=no, Libertà=sì), mentre tra i casi che soddisfanno l'antecedente di quella regola (Droga=no, Armi=no), una metà (uno su due) ha il valore che soddisfa il conseguente (Libertà=sì). Il supporto ci dice quindi quale proporzione di casi esemplifica l'applicazione della regola, mentre la confidenza ci dice in quale proporzione di casi, tra quelli in cui la regola risulta applicabile, dà luogo alla conseguenza stabilita dalla regola stessa.

Questo semplice esempio conferma come l'uso di tecniche di analisi automatica (data analytics) non necessariamente spinge al paventato effetto gregge (vedi avanti par. 6.2). Esso può invece condurre l'operatore consapevole ad approfondire l'esame delle alternative che emergono dalle esperienze passate.

*Le tecnologie per l'apprendimento automatico: sistemi trasparenti e opachi*

L'apprendimento automatico usa diversi metodi: gli alberi di decisione, la regressione statistica, le macchine a vettori di supporto (support vector machine), gli algoritmi evolutivi, le reti neurali, ecc. Tali metodi differiscono non solo nelle prestazioni predittive ma anche nella capacità di fornire spiegazioni, e spesso vi è una tensione tra i due

obiettivi: i sistemi che forniscono le prestazioni più accurate sono più opachi, cioè meno capaci di giustificare le proprie decisioni. Per esempio, il risultato di un albero di decisione può essere spiegato attraverso la sequenza di test che conducono a quel risultato. Con riferimento all'esempio della Figura 7 (il punto di domanda nell'elemento decisione indica che si tratta di un caso nuovo, non contenuto nell'insieme di addestramento, cui il sistema deve dare una risposta):

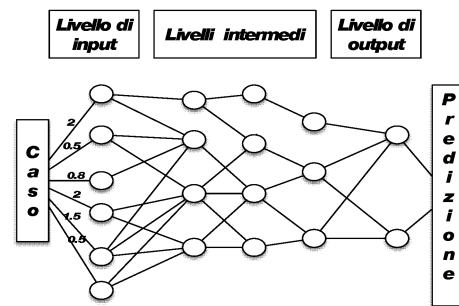
Figura 7. Valutazione di un nuovo caso

Caso	Lesione	Droga	Arma	Precedenti	Decisione
9	Leggera	Sì	no	no	?

La predizione del sistema in questo caso sarebbe negativa (no, niente libertà vigilata). Una spiegazione può essere fornita ripercorrendo il cammino sull'albero che ha condotto a tale risposta: "La libertà vigilata è stata negata nel caso in esame perché il caso era connesso al traffico di stupefacenti (nella spiegazione non compare il predittore "Lesione", poiché come abbiamo visto, esso è stato considerato irrilevante dal sistema, essendo possibile spiegare tutti i casi nell'insieme di addestramento prescindendo da esso).

Alcuni dei modelli appresi in modo automatico sono caratterizzati dal fatto che il loro funzionamento non si basa su una sequenza di passi ciascuno dei quali collega premesse dotate di significato a conclusioni sostenute da tali premesse (come nei sistemi basati su regole o gli alberi di decisione). Invece, l'attivazione di quei modelli comporta calcoli complessi intesi a riprodurre le correlazioni statistiche tra caratteristiche di input e risultati da predire. Oggi il modello più influente è rappresentato probabilmente dalle reti neurali, sistemi informatici che consistono di nodi (i cosiddetti neuroni) collegati da link ai quali sono assegnati pesi numerici (nella Figura 8, per semplicità si è omessa l'indicazione dei pesi, tranne che nel primo livello, per non appesantire l'immagine).

Figura 8. Rete Neurale



Quando il caso da valutare è sottoposto alla rete, le caratteristiche del caso sono tradotte in valori numerici. Per esempio, ogni frase di una sentenza può essere trasformata in un vettore di numeri binari (una sequenza di cifre uno e zero) che indicano quali parole siano presenti o assenti nella frase. Allo stesso modo possono essere codificate informazioni sulla presenza o assenza delle caratteristiche rilevanti di un caso (per esempio, se vi siano droga, armi, ecc.). Questi valori numerici vengono acquisiti dai neuroni di input e trasmessi attraverso i link ai neuroni connessi (dopo essere stati aumentati o diminuiti a seconda dei pesi delle connessioni), i quali, su questa base potranno inviare ulteriori valori numerici ad altri neuroni. Ogni neurone effettua calcoli numerici che determinano se e quali valori saranno trasmessi ai neuroni connessi. Qui non possiamo approfondire l'esame delle tecnologie delle reti neurali. Ba-

sti osservare che le reti neurali, originariamente concepite secondo il modello del cervello, effettuano un'elaborazione "subsimbolica". Una rete non consiste di simboli – quali entità linguistiche che esprimono concetti e fanno riferimento a certi tipi di oggetti, come le parole del linguaggio umano – ma piuttosto di vettori di numeri. Le elaborazioni di quei numeri determinano tanto l'apprendimento della rete, quanto la sua risposta a nuovi casi.

Per creare una rete neurale per predizioni giuridiche, bisogna innanzitutto identificare le caratteristiche dei casi in esame che possono essere rilevanti per l'esito di quei casi e rappresentare numericamente quelle caratteristiche. Bisogna quindi costruire una rete intesa a collegare (attraverso più livelli di nodi intermedi) quelle caratteristiche, rappresentate nei nodi di input agli esiti possibili, rappresentati dai nodi di output. La rete è addestrata su casi reali e ipotetici (l'insieme di addestramento) finché non fornisce risultati (normalmente) corretti, cioè quando all'input di un caso reale, la rete risponde con l'esito verificatosi in quello stesso caso.

A differenza di un albero di decisione, una rete neurale non si presta a fornire direttamente spiegazioni comprensibili. È possibile determinare in quale modo il sistema abbia raggiunto una certa decisione, esaminando come l'input fornito abbia determinato l'attivazione di certi neuroni, e come l'output di questi abbia determinato l'attivazione di altri neuroni (in base ai pesi assegnati alle connessioni neurali in seguito all'addestramento della rete). Tuttavia, questa informazione non rende esplicite le ragioni per le quali è stata data una certa risposta, non è una giustificazione comprensibile dalla mente umana. La difficoltà di comprendere il funzionamento di una rete cresce all'aumento del numero dei nodi, dei livelli in cui essi sono organizzati e della complessità delle loro connessioni. Si possono così realizzare reti per il cosiddetto apprendimento profondo (*deep learning*), in grado di apprendere anche da dati non strutturati. Le reti neurali, come altri modelli del cui funzionamento non si riesca a dare spiegazioni, sono dette "opachr" o anche "scatole nere" (*black box*).

Numerose ricerche si propongono di realizzare tecnologie che consentano spiegazioni comprensibili all'uomo anche per l'attività di tali sistemi, ma i risultati di tali ricerche sono ancora molto limitati. Di qui la necessità di bilanciare efficienza (accuratezza nelle predizioni) e spiegabilità, nella scelta del sistema predittivo da utilizzare in un determinato ambito. In molti casi (per esempio, nei sistemi robotici o nella diagnosi medica o manutentiva) è tendenzialmente l'efficienza a prevalere: preferiamo un sistema che commetta meno errori rispetto a un sistema più fallibile, ma capace di fornire delle ragioni per le proprie scelte. L'esigenza di fornire spiegazioni è, invece, fondamentale quando il sistema sia usato per l'esercizio di funzioni di rilevanza pubblica, specialmente quando siano in gioco interessi in contrasto e le esigenze del controllo siano preminenti, come accade rispetto all'azione amministrativa e ancor più nell'ambito della giurisdizione.

### Fenomenologia delle predizioni giuridiche

In questo paragrafo chiariamo come tutti i sistemi basati sul machine learning in ambito giuridico possono considerarsi macchine predittive, ma solo a patto che il termine

predizione sia inteso in un senso assai ampio, così da applicarsi a qualsiasi inferenza intesa a espandere le informazioni disponibili su un certo problema, inferenze che possono riguardare non solo il futuro, ma anche il passato e il futuro.

Anche le predizioni giuridiche sul futuro (le previsioni) debbono peraltro essere distinte dalle decisioni giuridiche, cioè dalle determinazioni autoritative finalizzate alla decisione di una controversia: una previsione diverrebbe una decisione automatica solo quando, sulla base della previsione il sistema adottasse autonomamente una decisione corrispondente (sulla base della previsione che l'imputato probabilmente sarà recidivo, il sistema gli nega la libertà vigilata; sulla base della previsione che un giudice "medio" deciderebbe in un certo modo una controversia, il sistema emana una statuizione corrispondente, dotata di effetto giuridico). Poiché un assetto di questo tipo non è adottato nel nostro sistema giuridico (né in altri, per quanto sappiamo), né dovrebbe a nostro parere essere adottato, è, quindi, corretto parlare dell'uso di macchine predittive nel diritto, ma non di "decisioni" automatiche (questione che riprenderemo più avanti).

### Le predizioni giuridiche

Alcune applicazioni prototipiche delle reti neurali e di altre tecnologie per il machine learning al diritto sono state realizzate fin dagli anni '80, per trattare concetti giuridici indeterminati o per suggerire la divisione del patrimonio in casi di divorzio. Negli anni più recenti vi è stata una vera esplosione di queste applicazioni, con l'obiettivo di replicare nel diritto il successo straordinario che l'apprendimento automatico aveva ottenuto in altri settori<sup>37</sup>. Tale obiettivo è facilitato dalla digitalizzazione delle fonti del diritto (legislazione, giurisprudenza, atti amministrativi, contratti) e di altri documenti giuridicamente rilevanti, che possono essere usati per creare insiemi di addestramento giuridico. L'applicazione delle tecnologie di machine learning al diritto ha già condotto a risultati importanti dal punto di vista sperimentale, ma per ora non si può parlare di una rivoluzione nella pratica giuridica, e in particolare dell'effettiva possibilità di prevedere con grande accuratezza gli esiti dei nuovi casi, o addirittura di sostituire queste previsioni alle decisioni di giudici umani.

Per valutare la situazione attuale è necessaria innanzitutto un'attenta descrizione dei sistemi oggi esistenti, per chiederci se le loro operazioni consentano realmente, e in quale misura, di prevedere decisioni giudiziarie. Infatti, come si è detto in apertura, tutti i sistemi basati sul machine learning in ambito giuridico possono considerarsi macchine predittive, ma solo se il termine predizione è inteso in un senso assai ampio. Si possono distinguere diversi tipi di predizioni che sono rilevanti in ambito giuridico/giudiziario:

1. La predizione può riguardare (i) il risultato finale di un caso giudiziario (la sentenza o altro atto emesso dal giudice) o (ii) altro evento o circostanza che possa contribuire a determinare quel risultato (per es., la predizione può riguardare il comportamento futuro di una parte, quale la recidiva del condannato, la fuga dell'accusato, l'inadempimento del debitore, etc.).

2. La predizione può riguardare (i) un evento passato

<sup>37</sup> Per un'introduzione e riferimenti bibliografici, si veda Kevin D. Ashley. *Artificial Intelligence and Legal Analytics*. Cambridge University Press, 2017.

(per es., il risultato di un caso già deciso) o (ii) un evento futuro (per es., la decisione di un caso nuovo).

3. La predizione può essere usata (i) per anticipare il comportamento altrui (per es., le parti sono interessate a prevedere il comportamento del giudice nel caso che le riguarda) o a (ii) offrire indicazioni di comportamento al decisore stesso (il giudice di un caso può usare la predizione del sistema quale guida alla decisione del caso).

4. La predizione può essere fondata su (i) elementi normativamente rilevanti del caso in esame, che il giudice dovrebbe considerare per adottare o giustificare la propria decisione o (ii) elementi normativamente irrilevanti, che il decisore non dovrebbe prendere in considerazione (identità delle parti, loro condizione economico-sociale, identità dei loro avvocati, del giudice, ecc.).

5. La predizione può essere fondata (i) sul contenuto testuale della sentenza o di altra decisione del giudice (quindi la predizione si fonderà sulle tecniche per l'elaborazione del linguaggio naturale, il cosiddetto Natural Language Processing-NLP), oppure (ii) su dati non testuali (e.g., parole chiave o fattori associati al testo, informazioni sui fatti del caso e le caratteristiche delle parti).

6. Il sistema può (i) essere capace di fornire una spiegazione delle sue decisioni, adducendo ragioni comprensibili agli umani, o (ii) può rimanere opaco rispetto alla richiesta di spiegazioni.

Il termine “giustizia predittiva” è stato applicato a sistemi basati sull'apprendimento automatico che presentano diverse combinazioni delle funzionalità appena elencate. Nella terminologia usata per il machine learning si tratta sempre di “predizioni”, ma questo termine può essere fuorviante, perché suggerisce che il sistema in questione preveda l'esito di casi futuri, così da anticipare, guidare o sostituire il decisore umano. Ciò, va detto chiaramente, non si verifica nella maggior parte delle applicazioni finora sviluppate.

Taluni sistemi basati sull'apprendimento automatico prevedono eventi futuri, ma tali eventi non consistono in decisioni di giudici (ipotesi 1.ii). È il caso dei sistemi, già in uso presso i tribunali statunitensi, che quantificano il rischio di recidiva o di violazioni della libertà vigilata. Si tratta di elementi che influiscono sull'attività di *sentencing*, cioè l'attività del giudice quando quantifica la pena da irrogare o da inasprire per il rischio di recidiva in concreto, basandosi anche sull'indicazione fornita dal sistema di assistenza. Tali sistemi, quali ausili informativi del giudice, hanno sollevato importanti questioni etiche o giuridiche, con riferimento all'accuratezza delle loro previsioni, alla loro imparzialità, all'assenza di discriminazioni, al rispetto del principio del giudizio individuale<sup>38</sup>. Va detto però che si tratta di problemi che, pur di grande importanza, sono diversi rispetto a quelli sollevati dalla possibilità di prevedere la decisione del giudice stesso.

Inoltre, la “predizione” che il giudice deciderà in un certo modo una controversia potrà essere considerata (A) dall'avvocato come la previsione dell'esito più probabile (ipotesi 3.i), mentre (B) dal giudice come un mero suggerimento per la decisione che deve prendere, sempre che non voglia discostarsi dall'orientamento prevalente (3.ii).

Taluni sistemi “predicono” il risultato di una decisione

già adottata, sulla base di parti della stessa decisione (ipotesi 2.ii). Ad esempio, un sistema potrebbe indicare il dispositivo di una sentenza sulla base della motivazione della stessa, o potrebbe indicare l'esito di una sentenza sulla base della sola porzione della sentenza che descrive i fatti della causa. Questa funzione è stata realizzata nello studio sulle decisioni della Corte europea dei diritti dell'uomo: sulla base di porzioni della sentenza il sistema classifica la sentenza, indicando se si tratta di una sentenza che riconosce o esclude la violazione di uno dei diritti riconosciuti dalla Convenzione<sup>39</sup>. In tutti questi casi non si tratta della previsione di un fatto futuro, ma piuttosto dell'inferenza di un'informazione (la decisione della controversia) su un evento passato (la controversia già decisa), sulla base di informazioni sullo stesso evento che non includono il dato inferito (per es., la parte fattuale della decisione). I risultati ottenuti con tali sistemi possono soprattutto essere utili per futuri sviluppi nella direzione della vera previsione, che si avrebbe solo quando certi dati su un caso (per esempio, gli atti di parte, o la descrizione dei fatti del caso, come fornita dalle parti, o anche l'intera decisione di primo grado rispetto all'esito dell'appello), disponibili prima della decisione, siano utilizzati per prevedere la decisione futura. Tuttavia, il passaggio da sistemi usati per estrarre informazioni da casi passati (studi retrospettivi a scopo sperimentale) a sistemi che prevedono decisioni di casi futuri comporta incertezze aggiuntive, correlate alla mancanza di informazioni rilevanti, al fatto che il soggetto che decide sia diverso da quello che ha predisposto i dati per effettuare la previsione, ecc.

Vi sono alcuni sistemi basati sull'apprendimento automatico che possono fornire valutazioni giuridiche su problemi controversi che non si sono ancora presentati in sede contenziosa, cosicché le indicazioni possano essere usate dalle parti per prevedere i risultati di controversie future. Per esempio, il sistema CLAUDETTE “predice” quali clausole contenute nei contratti al consumo possano ritenersi illegali o inique sulla base delle norme a tutela dei consumatori. Il sistema opera questa predizione sulla base di un insieme di addestramento che comprende circa 150 contratti nei quali le clausole inique sono state identificate da esperti giuristi, tenendo conto di precedenti decisioni dei giudici<sup>40</sup>. Esso opera sulla base della somiglianza formale (vocabolario, strutture sintattiche) tra le frasi etichettate come inique nell'insieme di addestramento e quelle rinvenute nei nuovi contratti sottoposti al sistema. Ciò che il sistema prevede direttamente è in questo caso la valutazione dell'analista giuridico (che funge da proxy rispetto alla decisione del giudice). Ovviamente un sistema di questo tipo potrà fornire indicazioni sufficientemente affidabili sugli esiti di possibili controversie solo nella misura in cui le informazioni nell'insieme di addestramento (la classificazione delle clausole come abusive) si basino effettivamente su precedenti decisioni dei giudici in materia.

Alcuni sistemi basati sull'apprendimento automatico usati a fini predittivi non fondano le proprie decisioni su aspetti giuridicamente significativi del contenuto testuale delle decisioni giudiziali, ma si basano su aspetti che, pur strettamente correlati all'esito delle controversie, non posso-

<sup>38</sup> Per una discussione del caso Loomis, e riferimenti bibliografici, si veda Francesca Lagioia, Riccardo Rovatti, and Giovanni Sartor. Fairness through group parities? the case of COMPAS-SAPMOC. AI and Society, 2022.

<sup>39</sup> N. Aletras, D. Tsarapatsanis, D. Preo, tiuc-Pietro, and Lamos. Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ Computer Science, 2: e 93, 2016.

2: e 93, 2016.

<sup>40</sup> Marco Lippi, Giuseppe Contissa, Agnieszka Jablonowska, Francesca Lagioia, Hans-Wolfgang Micklitz, Przemyslaw Palka, Giovanni Sartor, and PaoloTorrioni, Claudette: an automated detector of potentially unfair clauses in online terms of service, Artificial Intelligence and Law 27:117–139, 2019.

no fornire una giustificazione giuridicamente rilevante (ipotesi sub 4.i). Per esempio, sono state effettuate delle previsioni sulle decisioni dei giudici della Corte Suprema statunitense sulla base di informazioni extratestuali, come l'oggetto della causa, la procedura che ha condotto alla decisione, l'orientamento politico e le esperienze professionali dei giudici, ad esclusione di ogni informazione sulle ragioni di merito addotte dalle parti<sup>41</sup>. Allo stesso modo, un sistema commerciale prevede l'esito di casi in tema di brevetti sulla base delle caratteristiche delle parti, degli avvocati e dei giudici<sup>42</sup>. Sistemi di questo tipo possono essere utili alle parti interessate ad anticipare l'esito di una controversia, ma ovviamente lo sono molto meno per il giudice che deve valutare il merito della causa che è chiamato a decidere.

Un criterio misto, che combina quello testuale e quello extratestuale, è stato adottato nello studio di Nay sui disegni di legge presentati negli USA e riguardante la previsione della loro trasformazione in leggi approvate dal potere legislativo<sup>43</sup>. Ma in questo caso si tratta di una previsione che, se può essere utile per imprese e cittadini, è del tutto estranea al contesto di una decisione contenziosa.

#### *Contenuti giuridici e calcoli numerici: una precisazione importante*

Quando la previsione dell'esito di una controversia si fonda su contenuti testuali concernenti la controversia, essa si fonda, in linea di principio, su contenuti giuridicamente rilevanti (vedi sopra, ipotesi 4.i), ma dobbiamo ricordare che, nell'apprendimento, i contenuti linguistici vengono tradotti in vettori numerici, ed elaborati con calcoli numerici. Il sistema non possiede una rappresentazione computabile dei significati giuridici che sia comprensibile all'uomo. Per il sistema stesso la descrizione dei fatti di una controversia diventa un insieme di vettori numerici, che debbono essere collegati al possibile esito della stessa, secondo le correlazioni statistiche elaborate dall'algoritmo di apprendimento automatico sulla base delle combinazioni tra descrizioni di fatti ed esiti presenti nell'insieme di addestramento. Di solito questi sistemi sono opachi rispetto all'utilizzatore: essi forniscono una risposta, ma non forniscono spiegazioni<sup>44</sup>.

È da osservare che la previsione sulla base di dati testuali può configurarsi in modi diversi.

Come nell'esempio riportato nella Figura 4 i testi nell'insieme di addestramento possono essere corredati da una caratterizzazione (nella forma di keywords, fattori, citazioni, ecc.) che ne descrivono, in modo uniforme (usando la stessa terminologia), i contenuti giuridicamente rilevanti, grazie al lavoro di analisti umani, possibilmente con il supporto di strumenti automatici. Per esempio, nell'esempio appena citato, hanno questa natura le specificazioni delle caratteristiche dei casi (se il reato abbia comportato lesioni, sia stato commesso con l'uso di armi, ecc.). Sulla base di queste informazioni, il sistema costruisce poi il proprio modello che correla le caratteristiche dei casi agli esiti possibili. In questa ipotesi, quindi, è necessaria l'opera dell'uomo per identificare le caratteristiche rilevanti e associarle ai

casi nell'insieme di addestramento e ai nuovi casi da valutare (seppure anche queste attività possano in linea di principio, avvalersi di strumenti automatici).

Alternativamente, l'insieme di addestramento potrà consistere di documenti testuali (le descrizioni dei fatti del caso contenuta nella motivazione delle sentenze, o negli atti di parte), correlati al risultato corrispondente. In questo caso il sistema dovrà correlare direttamente le strutture testuali (il vocabolario, la sintassi, ecc.) di quei documenti (le descrizioni dei fatti delle cause o gli atti delle parti, ecc.) alla statuizione del giudice, e effettuerà la propria previsione in un nuovo caso sulla base di un testo dello stesso tipo. In questa ipotesi l'intervento umano sarà molto più limitato, limitandosi alla selezione dei testi delle sentenze e all'indicazione del loro esito.

I sistemi che operano sulla base di descrittori comprensibili all'uomo si prestano maggiormente a fornire spiegazioni, indicando le caratteristiche giuridicamente rilevanti (così come rappresentate dai descrittori), che hanno determinato la previsione proposta per il caso. Per esempio, nel caso sopra indicato il sistema potrà indicare quali caratteristiche sono state determinanti, nel senso che se i valori per quelle caratteristiche fossero stati diversi (per esempio, il reato non avesse riguardato gli stupefacenti), il risultato sarebbe stato diverso. L'inserimento dei descrittori, tuttavia, richiede un lavoro molto maggiore da parte dell'uomo (per l'identificazione dei descrittori, e poi la loro applicazione ai documenti). È possibile, peraltro utilizzare tecniche per l'apprendimento automatico per individuare i descrittori e associarli ai testi, alleggerendo il lavoro umano.

#### *Predizioni e proxy*

Al fine di cogliere il senso delle predizioni giuridiche è importante ricordare, che un sistema automatico può solo imparare la correlazione tra diversi dati concernenti gli esempi compresi nell'insieme di addestramento (vedi sopra par.3.3). Rispetto ai nuovi casi alcuni di quei dati fungono da elementi di input nella previsione (*predittori*), altri costituiscono l'output della previsione (*obiettivi*). Nel caso paradigmatico della previsione giurisprudenziale, ciascun esempio nell'insieme di addestramento comprende sia la descrizione di aspetti del caso (i fatti del caso stesso, gli atti delle parti) sia la decisione del caso. Il sistema, data la descrizione del caso, prevede quale potrebbe esserne la decisione.

In questo, come in altri casi, è utile distinguere l'obiettivo diretto della previsione automatica (quale decisione futura è maggiormente probabile?), e un possibile obiettivo finale per il quale il sistema viene utilizzato, che può essere diverso dall'obiettivo diretto (quale sarà/sarebbe una decisione "corretta"). Anche quando nella maggioranza dei casi l'obiettivo diretto tenda ad allinearsi all'obiettivo finale dell'utilizzo del sistema l'allineamento può essere imperfetto. Per esempio, se un sistema per la diagnosi medica viene addestrato sulla base delle diagnosi effettuate da medici umani (senza correggere le diagnosi che poi si rivelino errate), il suo obiettivo diretto sarà fornire indicazioni corri-

<sup>41</sup> Daniel Martin Katz, Michael J. Bommarito, and Josh Blackman. A general approach for predicting the behavior of the Supreme Court of the United States. *PLoS ONE*, 12, 2017.

<sup>42</sup> Surdeanu, R. Nallapati, G. Gregory, J. Walker, and C. Manning. Risk analysis for intellectual property litigation. In *ICAIL-2011*, pages 116-20. ACM, 2011.

<sup>43</sup> J.J. Nay, Predicting and understanding law-making with word vectors and an ensemble model, in "PLoS ONE" 12(5): e0176999.

<https://doi.org/10.1371/journal.pone.0176999>. A. Santosuosso-G. Pionotti, Bottleneck or crossroad? Problems of legal sources annotation and some theoretical thoughts, in *STATS* (ISSN 2571-905X), 376-395; <https://doi.org/10.3390/stats3030024> – 09 Sep 2020.

<sup>44</sup> Riccardo Guidotti, Anna Monreale, Franco Turini, Dino Pedreschi, and Fosca Giannotti. A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(Article 93):1-42, 2018.

spondenti a quelle che avrebbe adottato il medico medio, obiettivo che si avvicina (almeno così sperano i pazienti), ma non si identifica con quello di effettuare diagnosi corrette. Similmente, se un sistema automatico per valutare l'affidabilità creditizia viene addestrato sulla base delle decisioni passate, il sistema avrà l'obiettivo diretto di prevedere le valutazioni del funzionario medio, che si avvicinano (in misura maggiore o minore), ma non si identificano con l'obiettivo finale di determinare la probabilità con cui il richiedente restituirà il debito. Un sistema destinato a prevedere l'esito delle sentenze può aver l'obiettivo diretto di prevedere la sentenza del giudice medio, obiettivo che si avvicina a quello dell'utilizzatore avvocato o cittadino, solitamente interessati a prevedere il comportamento del giudice della propria causa, così da impostare di conseguenza il proprio comportamento processuale; quell'obiettivo ha una importanza più limitata per il giudice, il cui compito è quello di adottare una decisione giuridicamente corretta. Come abbiamo osservato sopra, per il giudice può essere utile aver un sistema che non si limiti a fornire una sola decisione possibile, ma indichi piuttosto, nel caso di divergenze nella giurisprudenza, accanto all'ipotesi maggioritaria, anche quelle che trovino meno seguito.

Nelle applicazioni di giustizia predittiva, va individuato con precisione quale sia l'obiettivo diretto delle predizioni del sistema, per valutarne la rilevanza. Per esempio, possiamo distinguere due tipi di sistemi impiegati al fine di determinare se concedere o meno la libertà vigilata a un imputato. Il primo sistema potrebbe essere addestrato sulla base di un insieme di addestramento che associa richieste di libertà vigilata alle corrispondenti decisioni dei giudici. La predizione diretta del sistema riguarderebbe in questo caso la decisione del giudice medio (o anche le diverse soluzioni risultanti da diversi indirizzi presenti nella giurisprudenza). Il secondo sistema potrebbe invece effettuare le proprie predizioni sulla base di esempi che riportino il comportamento successivo di un insieme di imputati che abbiano goduto della libertà vigilata. In questo caso il sistema predirebbe la probabilità che l'imputato reiteri il reato o si sottragga alla pena, fornendo un'indicazione più "oggettiva" di cui il giudice potrebbe tener conto.

La possibilità di una predizione "oggettiva" non esiste quando, come nel caso della predizione dell'esito di controversie, l'obiettivo da prevedere sia proprio la decisione umana e non vi sia un riscontro obiettivo della correttezza della decisione. In questo caso, il sistema si limiterà a "predire" quale nuova decisione potrà allinearsi alle decisioni passate, riproducendo le virtù (accuratezza, imparzialità) così come i vizi (imprecisioni, iniquità) di quelle decisioni.

È importante chiarire il rapporto tra decisione *media* e decisione *corretta*. Assumiamo che il sistema sia in grado di prevedere la decisione media dei giudici. Per esempio, se il 50% dei giudici decida cause di un certo tipo accordando un risarcimento di circa 1000 euro, mentre il 25% dei giudici accorda un risarcimento di 500 euro e altri di 1500, il sistema prevederà un risarcimento atteso di 1000 euro. Qualora, una volta introdotto il sistema, tutti i giudici ne seguissero le indicazioni, accordando sempre un risarcimento di 1000 euro a tutti i nuovi casi, si sarebbe raggiunto un risultato importante. Si sarebbe ridotto il "rumore" o la dispersione nei risultati del sistema: il sistema non tratterebbe in modo diverso casi eguali<sup>45</sup>. Ciò faciliterebbe la previsione da parte di cittadini, che saprebbero quale ri-

sposta attendersi dal sistema giuridico. Ma si potrebbe dire che in tal modo sia aumentata la "giustizia" del sistema? Ciò dipende da dove stia la decisione più giusta o giuridicamente corretta. Se essa è compresa tra 500 e 1500, consistendo nel valore di 800, sembra che la risposta debba essere positiva: convergendo su tale cifra si limita l'ingiustizia media, cioè la divergenza media tra il risarcimento attribuito e quello ideale (essa ammonterà a 200 per ogni persona, piuttosto che  $200 \cdot 0.5 + 300 \cdot 0.25 + 700 \cdot 0.25 = 100 + 75 + 175 = 350$ ). Tuttavia, è da osservare che il diritto possiede altri metodi per rimediare alle divergenze (come il meccanismo del precedente), e che il disaccordo può essere utile ad avviare un dibattito che conduca a convergere verso le soluzioni migliori.

A questo riguardo, bisogna peraltro osservare che, come già ricordato, non necessariamente un sistema dovrà essere destinato a fornire solo la soluzione "media". Questo orientamento sembra ragionevole quando dal sistema ci si attenda un'unica soluzione, possibilmente per dare esecuzione ad essa senza un ulteriore intervento umano. Quando invece il sistema sia destinato a interagire con l'uomo, fornendogli elementi di conoscenza – a questo riguardo si parla talvolta di computazione cognitiva, *cognitive computing* – può essere utile che il sistema fornisca indicazioni alternative, alla luce degli esempi divergenti contenuti nel suo insieme di addestramento (o delle regole divergenti nella sua base di conoscenza), come si è evidenziato nei paragrafi precedenti.

### **I dati nella predizione: algoritmi e dataset**

Si è visto come nei sistemi basati sull'apprendimento automatico la conoscenza relativa all'ambito di applicazione del sistema non sia più fornita dall'uomo, ma piuttosto sia costruita dalla macchina, sulla base dei dati cui essa ha accesso. Di qui l'importanza dei dati nella predizione. Per il giurista l'obiettivo realizzabile mediante la predizione automatica, sia pure per vie diverse, corrisponde a una fondamentale esigenza conoscitiva presente anche nel passato: rivolgersi alla memoria pregressa per formulare ipotesi attendibili sul futuro. Così il giudice, che si chiede qual è il rischio che un imputato commetta nuovi reati (rischio recidiva), in futuro potrà chiedere indicazioni a un sistema predittivo; l'avvocato potrà interrogare un sistema su quale decisione potrà essere probabilmente adottata dal giudice sul caso del suo cliente; il giudice di primo grado sulla probabilità che la propria sentenza sia confermata in appello, e altro.

Certo, non tutto è uguale a prima. Innanzitutto, l'uso di tecnologie predittive consente di ridurre il costo delle previsioni e di aumentare la loro precisione, il che potrebbe condurre a un uso pervasivo delle predizioni attraverso macchine. Di qui l'esigenza di interrogarsi sui contesti nei quali e le finalità per le quali sia opportuno ricorrere a predizioni automatiche su persone e gruppi, sulle modalità per il loro uso proficuo, e sulle precauzioni da adottare per prevenire i rischi di un cattivo uso delle stesse, tale da condurre a scelte inaccurate, discriminatorie o inique.

Vi è poi una differenza importante tra le predizioni automatiche e quelle basate su teorie costruite dall'uomo, che vale sia per le teorie scientifiche sia per le dottrine giuridiche (nella misura in cui si faccia riferimento ad esse per anticipare comportamenti futuri degli operatori del diritto,

<sup>45</sup> Sul tema si veda Kahneman, D., Sibony, O., and Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgement*. Collins.



piuttosto che per avere indicazioni normative sulle scelte preferibili alla luce di certe interpretazioni norme e valori). Mentre nell'impostazione usata tradizionalmente in approcci sperimentali o quasi-sperimentali si andava dalla costruzione di un modello verso (*forward*) i dati osservabili, dove si cercavano conferme o smentite, nell'apprendimento automatico si usano, invece, i dati osservabili per costruire il modello (*inverse*)<sup>46</sup>. In altri termini, il rapporto ipotesi-test, alla base del metodo scientifico moderno, è (almeno inizialmente) invertito ed è l'analisi con tecnologie *machine learning* di un appropriato insieme di dati a far emergere un possibile modello (tratto dall'insieme di addestramento) da poi testare, applicandolo a nuovi dati.

In ogni caso, dato che le predizioni di un sistema di AI si basano sui dati, quanto maggiori saranno la quantità e la qualità dei dati e quanto migliore il livello di analisi dei dati stessi, tanto maggiore sarà l'accuratezza della predizione possibile.

In Italia, limitandosi alle sole realtà in cui si fa uso di tecniche di *machine learning*, è in corso presso la Scuola Superiore Sant'Anna di Pisa, in collaborazione con i Tribunali di Genova e di Pisa, uno studio avente come oggetto la giurisprudenza in materia di alimenti per la separazione e per il divorzio, e sono stati avviati due studi condotti dall'Università di Bologna e dall'Università di Pavia all'interno del progetto PRIN "LAILA" e del progetto europeo ADELE, aventi a oggetto l'analisi (e relativa predizione) delle decisioni in materia di proprietà intellettuale e diritto tributario. È possibile che altro sia in movimento e ci scusiamo con chi non si vedesse citato.

Per procedere su questa strada è, dunque, necessaria la disponibilità ampia di dati. In realtà, nel dibattito sull'applicazione di tecniche di IA al settore giuridico molta attenzione viene dedicata al ruolo degli algoritmi, che sono oggetto di un'attenzione che ne fa entità di cui talora sembra perdersi la reale consistenza, quale sequenza logico-matematica che consente di estrarre conoscenza da dati. Molti hanno subito il fascino del neologismo *algocrazia* sull'onda del saggio di John Danaher sulla minaccia dell'algocrazia, appunto<sup>47</sup>. Persino la Pontificia Accademia per la Vita si è fatta promotrice, in un incontro a Roma (28 febbraio 2020<sup>48</sup>) insieme a Microsoft, IBM, FAO e Governo italiano, di un documento teso a promuovere una *algor-etica* (algo-ethics). Ora, poco importa che nel documento si precisi, correttamente, che lo scopo del testo romano è quello di promuovere lo sviluppo e l'utilizzo dell'Intelligenza Artificiale nel suo insieme, secondo principi fondamentali di una buona innovazione e si parli anche di dati. Il problema è che, in questi e altri casi, gli algoritmi vengono eretti a emblema dei pericoli delle tecnologie di IA, come se, da soli, avessero la possibilità di fare tutto il male (o il bene) che viene loro attribuito.

Ma un algoritmo è, come si diceva prima, solo una sequenza logico-matematica: "qualunque schema o procedimento matematico di calcolo; più precisamente, un procedimento di calcolo esplicito e descrivibile con un numero finito di regole che conduce al risultato dopo un numero finito di operazioni, cioè di applicazioni delle regole. [...] In informatica, insieme di istruzioni che de essere applicato

per eseguire un'elaborazione o risolvere un problema"<sup>49</sup>. È certamente possibile che un algoritmo sia iniquo o persino illegale, e a tal fine non è necessario ricorrere a tecnologie di intelligenza artificiale. Si pensi, a un algoritmo che, nel valutare le domande di assunzioni, respinga tutte quelle che provengano da donne con figli (SE Sesso=femminile E NumeroFigli > 0 ALLORA Decisione=rigetto), o per un esempio più realistico, si consideri la piattaforma per consegne a domicilio, che applicava penalità agli addetti che rifiutavano di eseguire richieste (SE Risposta=no ALLORA Punti=Punti-Penalità), senza distinguere i casi in cui il rifiuto fosse motivato dalla partecipazione ad azioni sindacali (tema affrontato in una recente decisione del Tribunale di Bologna).

Nelle applicazioni di intelligenza artificiale vengono di regola utilizzati algoritmi suscettibili di impiego generale, per realizzare modelli che consentono di effettuare con accuratezza predizioni in conformità ai dati contenuti nell'insieme di addestramento. Nell'addestramento supervisionato, il modello viene infatti costruito sulla base di un'ampia porzione dell'insieme di addestramento, e le sue predizioni sono poi verificate sulla porzione rimanente (per esempio, se l'insieme di addestramento consta di 1000 esempi, ciascuno costituito dalla descrizione dei fatti di un caso e dalla decisione dello stesso, si potranno usare 900 esempi per la costruzione automatica del modello, e 100 per misurarne le prestazioni). Tanto il modello usato per le previsioni quanto la sua valutazione si basano quindi sui dati forniti al sistema. Se quei dati non sono adeguati (assumiamo per esempio, che l'insieme di addestramento contenga solo i casi che vertono su una certa materia, o che riflettano un orientamento minoritario), il sistema riprodurrà quella inadeguatezza. Come dicevamo sopra, la risposta del sistema (o quella cui il sistema stesso assegna un punteggio più elevato, qualora il sistema fornisca decisioni alternative), consisterà nella previsione della decisione del giudice medio, ma del giudice medio tra quelle le cui sentenze compaiono nell'insieme di addestramento, che può non corrispondere al giudice medio in relazione alla totalità dei giudici.

Dobbiamo, quindi, non perdere di vista che l'algoritmo ha un partner necessario e indispensabile: il dataset. Il punto critico è l'applicazione dello schema matematico a un insieme di dati. Tanto che, a fronte di un risultato pericoloso o socialmente inaccettabile che sia prodotto da un sistema di IA, non sempre è facile stabilire se ciò sia dovuto all'algoritmo o al dataset o a entrambi. È possibile, infatti, che un algoritmo in sé corretto produca errori sistematici (bias) per averli acquisiti dal dataset sul quale è stato addestrato (tipico il caso di algoritmi allenati su archivi di polizia che riflettono i pregiudizi degli operatori, che, per esempio, hanno sottoposto a controlli più pervasivi certe porzioni della popolazione). E si può anche ipotizzare che non sia colpa di nessuno dei due, preso isolatamente, mentre la tendenza all'errore è frutto di un accostamento non appropriato.

L'MIT EmTechDigital ha organizzato il 29-30 marzo 2022 un evento, esattamente su questo tema, tanto che l'intervento di apertura è così presentato:

<sup>46</sup> A. Santosuosso, *Intelligenza artificiale e diritto*, Milano, 2020, 99.

<sup>47</sup> John Danaher, 'The Threat of Algocracy: Reality, Resistance and Accommodation' (2016) 29 *Philosophy & Technology* 245.

<sup>48</sup> <https://www.academyforlife.va/content/pav/it/events/workshop-intelligenza-artificiale.html>. Tra i tanti esempi si può citare una recente prestigiosa raccolta di ottimi saggi presso Cambridge Univer-

sity Press, che è dedicata esattamente alla società algoritmica come sfida costituzionale <https://www.cambridge.org/core/books/constitutional-challenges-in-the-algorithmic-society/constitutional-law-in-the-algorithmic-society/969E0889109C8092AD0AB57019E507E3>.

<sup>49</sup> <https://www.treccani.it/vocabolario/algoritmo/>.

“Data is the most under-valued and de-glamorized aspect of AI. Learn why shifting the focus from model/algorithm development to quality of the data is the next and most efficient, way to improve the decision-making abilities of AI”<sup>50</sup>.

E, nello stesso evento, vi è chi propone un approccio *Data-First AI*, che può trasformare l’inferenza dell’IA, traducendosi in migliori modelli che producono migliori decisioni.

Si aprono a questo punto due tipi di problema, strettamente connessi, di cui il primo è tecnico/teorico e riguarda il modo in cui sono impostate le attività di ricerca, gli assunti teorici sui quali sono basate e l’influenza che essi hanno sulla prestazione (performance) in relazione al dataset utilizzato, mentre il secondo è empirico/giuridico e riguarda la provenienza, natura e qualità del materiale sul quale è possibile operare, tenuto conto dei limiti normativi.

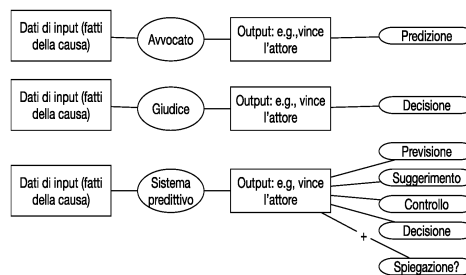
### *Il problema della predizione*

Possiamo a questo punto riprendere alcune considerazioni sviluppate nei paragrafi precedenti per considerare la predizione sotto un profilo teorico-giuridico.

Innanzitutto, come esposto nel par. 4, nell’intelligenza artificiale il termine predizione è usato ogni qualvolta si traggano nuovi dati dai dati già disponibili. Quindi la predizione può coprire attività molto diverse da una prospettiva teorico-giuridica: la ricerca documentale (predire quali dati siano più rilevanti per chi formula un quesito, e quindi fornirli in risposta allo stesso), la creazione automatica di abstract/massime (predire, e quindi formulare, un estratto che corrisponda al testo), l’associazione di descrittori al testo (predire, e quindi aggiungere, parole chiave o classificazioni documenti), l’anticipazione di condotte future di imputati (predire se l’imputato commetterà nuovi reati, se turberà le prove o fuggirà in seguito alla libertà vigilata) o parti (predire quale accordo le parti stipulerebbero consensualmente, se dovessero accordarsi per le implicazioni finanziarie di una separazione o un divorzio), il supporto alla creazione di documenti (predire quali frammenti di testi passati possono essere utili per la redazione di un nuovo documento, per esempio, una sentenza), il supporto allo studio empirico del diritto (predire correlazioni, tendenze, attitudini), e altro ancora.

Ciascuna di tali predizioni solleva diverse questioni teorico-giuridiche ed etiche, che non possiamo considerare in questa sede. Ci concentriamo pertanto sull’ipotesi paradigmatica di previsione giuridica, alla quale si fa di solito riferimento con il tema “giustizia predittiva”, la previsione dell’esito di controversie presenti o future, benché anche le altre predizioni appena elencate possano essere utilizzate nell’ambito degli strumenti informatici che supportano la giurisdizione, e anzi forse in esse, più che nella predizione degli esiti delle controversie.

Figura 9. Tipi di previsioni giuridiche



La figura 8 indica come tanto l’avvocato, quanto il giudice, quanto un sistema predittivo possono formulare delle predizioni (in senso ampio) circa l’esito della causa. È, però, essenziale notare che giudizi aventi il medesimo contenuto (per esempio, “vince la causa l’attore al quale è riconosciuto un risarcimento pari all’intero ammontare del danno”) possono avere un significato pragmatico assai diverso (dipendente dal contesto in cui il giudizio viene formulato, e dalle attitudini di chi lo formula e riceve).

Il prototipico giudizio dell’avvocato o della parte sull’esito di una controversia consiste di una previsione di fatto, che sarà confermata o smentita dal successivo comportamento del giudice. La previsione di come potrebbe essere decisa una futura controversia determina le aspettative e, quindi, il comportamento delle parti, anche prima dell’insorgere della controversia, mentre poi determina il loro comportamento processuale (in particolare la scelta se addivenire a una transazione).

Il prototipico giudizio del giudicante ha un effetto istituzionale. Si tratta di una statuizione che determina le posizioni giuridiche oggetto della controversia tra le parti, diventa incontestabile con il passaggio in giudicato e dà alla parte vittoriosa il potere di ottenere l’esecuzione.

Avendo presente che il giudizio di un sistema predittivo ha un significato diverso a seconda del contesto in cui tale giudizio è formulato, si può notare che se il sistema presentasse le ragioni per le quali suggerisce una decisione, cosicché il giudice potesse considerare se aderire o meno a quelle ragioni, vi sarebbe un reale progresso rispetto alla mera presentazione di un’ipotesi decisoria, che non consentirebbe al giudice un’adeguata valutazione giuridica. Inoltre, ricordiamo come un sistema predittivo non necessariamente deve limitarsi a proporre un’unica soluzione (la soluzione corrispondente alla decisione del giudice medio). Esso potrebbe presentare altresì decisioni alternative, cui sia associato un rinvio alle ragioni dei diversi orientamenti.

### *Un problema empirico: il caso Italia tra Cina e USA*

Il problema dei dataset, della loro appropriatezza, per quantità e qualità, ha carattere universale.

La Cina, tanto per fare un esempio, rende milioni di dati liberamente accessibili via web al sito, che, affermano con vanto, “è diventato il più grande database di sentenze del mondo”, dal quale si può arrivare ad accedere a *China Judgments Online*, dove “chiunque può leggere quasi tutti i documenti delle sentenze pronunciate dai tribunali cinesi dal 2013”<sup>51</sup>.

Si potrà dire che non è il caso di competere con questo

<sup>50</sup> “I dati sono l’aspetto più sottovalutato e *de-glamorized* dell’IA. Scopri perché spostare l’attenzione dallo sviluppo di modelli/algoritmi alla qualità dei dati è il prossimo e più efficiente modo per migliorare le capacità decisionali dell’IA” [traduzione nostra] ([https://event.technologyreview.com/emtech-digital-2022/agenda?utm\\_source=event\\_email&utm\\_medium=email&utm\\_campaign=emtech\\_digital\\_2022\\_unpaid\\_acquisition&discount=EMAIL011850&mc\\_cid=bb255161cb&mc\\_eid=24673f4347](https://event.technologyreview.com/emtech-digital-2022/agenda?utm_source=event_email&utm_medium=email&utm_campaign=emtech_digital_2022_unpaid_acquisition&discount=EMAIL011850&mc_cid=bb255161cb&mc_eid=24673f4347)).

<sup>51</sup> <https://www.chinajusticeobserver.com/a/you-can-view-almost-all-the-chinese-court-judgments-online-for-free> e <http://wenshu.court.gov.cn/>.

grande paese asiatico per quantità di dati e per qualità democratica. Ma questa obiezione non dovrebbe valere per il *Caselaw Access project* della *Harvard Law School*, che è consultabile, completamente aperto e scaricabile, e non anonimizzato. Sul portale del progetto si apprende che esso garantisce un accesso pubblico e gratuito a oltre 6,5 milioni di decisioni pubblicate dalle corti statali e federali nel corso della storia degli Stati Uniti: “l’accesso al nostro common law – le decisioni scritte emesse dai nostri tribunali statali e federali – sostiene l’uguaglianza e permette l’innovazione nei servizi legali”. Questo è il risultato di una collaborazione sviluppata a partire dal 2013 con la startup giuridica Ravel Law e che ha portato alla digitalizzazione di oltre 40 milioni di pagine di decisioni dei tribunali statunitensi trasformandole in un set di dati di grandi dimensioni. Un aspetto interessante è che il sistema mette a disposizione un servizio bulk che consente di porre “questo importante set di dati alla portata dei ricercatori, dei membri della comunità legale e del pubblico in generale”<sup>52</sup>.

In Italia vi è certamente una gran quantità di dati potenzialmente disponibili. Per avere idea della dimensione basti pensare alla storia dell’informatica giuridica in Italia e all’accumulo quotidiano delle decisioni e degli atti di parte che il processo civile telematico (PCT) produce. Secondo i dati disponibili presso il sito del Ministero della Giustizia<sup>53</sup>, da luglio 2014 a luglio 2020 il sistema ha ricevuto un totale di 50.140.847 di atti, di cui 2.340.552 ricorsi per decreto ingiuntivo, 8.897.156 atti introduttivi e di costituzione, 37.678.637 atti endo-procedimentali, con un andamento crescente delle medie mensili. Questo riguarda anche le decisioni dei giudici (sentenze, decreti e ordinanze, decreti ingiuntivi) che nell’arco di tempo 2014-2020 raggiungono un totale di 30.178.442 di atti.

A dispetto di questa realtà, che potrebbe consentire uno sviluppo importante delle attività di ricerca con tecnologie moderne (e pur al netto dei problemi tecnici legati alla qualità di questi dati), in Italia non vi è niente di paragonabile al libero accesso che il governo cinese o il progetto statunitense della *Harvard Law School* consentono. Solo ora, la spinta data dal PNRR sembra dover portare alla creazione di un dataset liberamente accessibile delle sentenze civili entro la fine del 2023, compatibilmente con le norme sulla privacy.

Non è questa la sede per affrontare il delicato tema dell’anonimizzazione, ma ci sembra che le norme in materia di tutela della riservatezza debbano essere correttamente collocate nel quadro giuridico internazionale, europeo e italiano, dove si trova più di un riferimento giuridico alla fondamentale libertà di accesso alle decisioni. Ci limitiamo a ricordare come l’art. 6 CEDU, quando stabilisce che “la sentenza deve essere resa pubblicamente”, pone una regola che è rivolta sia alla tutela delle parti di quel processo, sia a garantire l’interesse generale alla correttezza della giurisdizione e alla conoscibilità dell’ordinamento: una sentenza resa pubblicamente è una decisione presa in modo legale in nome dell’intera comunità<sup>54</sup>. Non va infatti dimenticato che la giurisdizione è attività pubblica e deve essere sotto il controllo dell’opinione pubblica.

Inoltre, passando a una questione di dettaglio, ma assai

significativa, la necessità di avere una banca dati delle decisioni, che sia completa e liberamente accessibile, è diventata una stretta necessità da quando i giudici possono motivare le sentenze “anche con riferimento a precedenti conformi” (art. 118 disp. att. c.p.c.). Ora, o il precedente è conoscibile dalle parti in giudizio, e allora potrà dirsi garantita la condizione di parità tra le parti (art. 111, 2° comma, Cost.), oppure la non piena e paritaria accessibilità ai precedenti può portare a decisioni potenzialmente prese in violazione di quella garanzia costituzionale di parità e, alla fine, a una violazione del diritto di difesa (art. 24 Cost.), specie se il giudice fa riferimento a un precedente non versato in giudizio dalle parti.

Senza scendere in ulteriori dettagli del problema, si può ricordare che il vecchio *codice privacy* (l’art. 52, D.Lgs. n. 196/2003, non modificato dal successivo regolamento) dava facoltà alla parte che ritenga di avere “motivi legittimi” di chiedere tempestivamente (e cioè prima del deposito della decisione che definisce il grado di giudizio) che sia apposta “sull’originale della sentenza o del provvedimento, un’annotazione volta a precludere, in caso di riproduzione della sentenza o provvedimento in qualsiasi forma, per finalità di informazione giuridica su riviste giuridiche, supporti elettronici o mediante reti di comunicazione elettronica, l’indicazione delle generalità e di altri dati identificativi del medesimo interessato riportati sulla sentenza o provvedimento”. Leggendo la norma in controluce (e in linea con il comma 7, secondo il quale “fuori dei casi indicati nel presente articolo è ammessa la diffusione in ogni forma del contenuto anche integrale di sentenze e di altri provvedimenti giurisdizionali”) si può affermare che, quando il diretto interessato non abbia esercitato quella facoltà, si espanda l’interesse generale di rilievo costituzionale alla piena conoscibilità delle decisioni giudiziarie.

Inoltre, la Corte di cassazione ha affermato (anche di recente: Cass. civ., Sez. V, ord., 10 agosto 2021, n. 22561) che è necessario un “bilanciamento tra le esigenze di riservatezza del singolo e il principio della generale conoscibilità dei provvedimenti giurisdizionali e del contenuto integrale delle sentenze, quale strumento di democrazia e di informazione giuridica” (Cassazione civile sez. trib., 07/08/2020, n. 16807/2020). Non può, inoltre, essere trascurato che l’attività di ricerca in materia giuridica rientra pienamente sotto l’ombrello della Costituzione, laddove protegge e promuove la libera ricerca scientifica (artt. 9 e 33 Cost.).

Dunque, ferma restando la facoltà prevista dal nostro ordinamento che una parte chieda tempestivamente l’anonimizzazione (che potrebbe risolversi in una pseudonimizzazione, coerente con le finalità del trattamento), la regola generale dovrebbe essere quella della piena conoscibilità delle decisioni giudiziarie. Oggi solo la pubblicazione e messa a disposizione dell’intero patrimonio delle decisioni adottate dai giudici italiani può rendere possibile che gli studi italiani sulle applicazioni di tecniche di *machine learning* possano evolvere dallo stadio puramente sperimentale attuale e diventare reale e affidabile strumento a disposizione dei giudici, degli avvocati e dei ricercatori italiani.

<sup>52</sup> <https://il.law.harvard.edu/projects/caselaw-access-project/>.

<sup>53</sup> [https://pst.giustizia.it/PST/resources/cms/documents/PCT\\_Dati\\_lug2014\\_lug2020\\_v1.pdf](https://pst.giustizia.it/PST/resources/cms/documents/PCT_Dati_lug2014_lug2020_v1.pdf).

<sup>54</sup> Enrico Consolandi, *Conoscenza informatica della giurisprudenza e privacy*, in *Il libro dell’anno del diritto 2019* (2019), disponibile presso [https://www.treccani.it/enciclopedia/conoscenza-informatica-della-giurisprudenza-e-privacy\\_%28altro%29/](https://www.treccani.it/enciclopedia/conoscenza-informatica-della-giurisprudenza-e-privacy_%28altro%29/) Si veda anche *Access*

to Court Decisions, A legal Analysis of relevant international and national provisions, OSCE 2008, 3. A. Santosuosso, *Giustizia predittiva*, ecco i tre pilastri per capire gli impatti della tecnologia sul diritto, In *Agenda Digitale*, 1° marzo 2021, <https://www.agendadigitale.eu/documenti/giustizia-digitale/giustizia-predittiva-ecco-i-tre-pilastri-per-capire-gli-impatti-della-tecnologia-sul-diritto/>.

*Un'esperienza recente (e in corso)*

Un'occasione in tal senso è offerta da alcuni dei progetti presentati da consorzi di università italiane nel bando pubblicato nell'agosto 2021 dal Ministero della giustizia<sup>55</sup>. Il bando indica già nel titolo l'obiettivo dell'innovazione tecnologica e fissa, tra i risultati da raggiungere, "l'aumento dell'utilizzo delle tecnologie da parte di tutti gli stakeholder e in particolare dei software utilizzati dai giudici, dalle cancellerie e dagli 'assistenti dei giudici' ossia dai componenti dell'Ufficio per il Processo".

Nella Macroarea 1 (Distretti di Milano, Brescia, Torino, Genova) si è costituito un consorzio unico che ricomprende tutte le università pubbliche di quel territorio. La componente che si occupa dell'informatizzazione e telematizzazione degli uffici giudiziari<sup>56</sup> si è posta il problema di quali possano essere i modi più efficaci perché le infrastrutture digitali facilitino lo sviluppo quantitativo e qualitativo della produzione degli Uffici per il processo e ha individuato i seguenti passaggi: l'Ufficio del processo ha bisogno di avere accesso al patrimonio giurisprudenziale e legislativo che ha già una forma digitale; a tal fine è necessario *valorizzare l'esistente digitale* costituito da milioni di atti digitali che si trovano presso i server ministeriali; è necessario che questa grande quantità di documenti assuma la forma di insieme di dati (*dataset*) efficacemente esplorabile con tecniche per l'estrazione di informazione (*information extraction*), la creazione di massime e abstract (*automated summarization*), la costruzione di documenti (*document builder*), la rappresentazione grafica (*visual graphics*) e, in generale, con tecniche di analitica giuridica (*legal analytics*). Un dataset ben organizzato e alimentato con dati di buona qualità può essere interrogato con algoritmi che possano aiutare il giudice in molteplici modi.

In sintesi, il problema che si è posto (e che è comune a tutto il territorio nazionale) è se un gruppo di università possa svolgere un'attività di ricerca applicata su materiale giurisprudenziale esistente, con la finalità di rispondere alle esigenze dell'amministrazione della giustizia e secondo l'indicazione contenuta nell'Avviso di valorizzare le risorse digitali.

Ad oggi (data della consegna del presente manoscritto) sono in corso contatti con il Ministero della giustizia e con la DGSIA per poter avere accesso a tale patrimonio di dati. Certo è che, anche sotto questo profilo, la questione dei dataset è centrale per lo sviluppo delle tecniche di IA in ambito giuridico, predizione e *legal analytics* in generale.

**Alcune risposte ai timori più diffusi**

In questo paragrafo proviamo a rispondere alle principali critiche e ai dubbi avanzati all'uso di tecniche predittive in ambito giuridico e, soprattutto, giudiziario. Difficile essere esaurienti, rispondiamo sulla base delle nostre esperienze e delle considerazioni svolte nei paragrafi precedenti. Per comodità raggruppiamo i temi in tre gruppi: il primo, riguarda le critiche che si basano su assunti che ci appaiono irrealistici o paradossali; il secondo, riguarda alcune questioni reali, che, però, potrebbero essere risolte con accorgimenti pratici o tecnici; il terzo, è dedicato ad alcune questioni obiettivamente importanti e che richiedono ulteriore ricerca e riflessione.

*Assunti paradossali o irrealistici: le macchine soppiantano gli umani*

Alcuni dei timori che abbiamo richiamato nel par. 2 riguardano il rischio che le macchine si sostituiscano agli umani, o che la decisione sia delegata alla *macchina*, o che ci si ritrovi rinchiusi in una nuova *normatività informatica*. Ora, con tutta la considerazione verso gli autori che hanno formulato quei timori, ci pare che le ipotesi da essi paventate si basino su assunti irrealistici, se non, in qualche caso, persino paradossali, se visti alla luce dell'ampia descrizione che abbiamo dato nei paragrafi precedenti su quello che effettivamente è stato fatto e su quello che è all'ordine del giorno.

*Macchine che sostituiscono gli umani.* Circa la sostituzione degli umani con macchine decidenti, si è visto in apertura del paragrafo 4, come non sia corretto parlare di "decisioni" automatiche, perché i sistemi basati sul machine learning in ambito giuridico possono considerarsi *macchine predittive* solo a patto che il termine predizione sia inteso in un senso assai ampio, tanto da poter essere applicato a qualsiasi inferenza intesa a espandere le informazioni disponibili su un certo problema.

A questa spiegazione tecnica, si può ora aggiungere un motivo in più, di tipo sociale e istituzionale. Come la teoria degli atti linguistici (*speech acts*) si è incaricata di chiarire ormai da tempo<sup>57</sup>, una stessa frase o enunciato (*utterance*), come "Antonio ha ragione", ha valore e significato diverso a seconda che, facendo quattro ipotesi semplificate, sia detta i) dagli avventori di un bar davanti a un aperitivo, ii) da uno studente o da un professore in ambito accademico, iii) da un avvocato a colloquio con un suo cliente o iv) da un giudice nell'esercizio delle sue funzioni. Solo nell'ultimo caso quella frase sarà una decisione di una controversia, perché sarà pronunciata da chi, secondo le regole sociali e giuridiche stabilite, è nella posizione di (ha titolo per) "dire il diritto"

<sup>55</sup> Avviso "Finanziamento di interventi a regia nell'ambito dell'Asse I – Obiettivo Specifico 1.4 – Azione 1.4.1 "Azioni di miglioramento dell'efficienza e delle prestazioni degli Uffici Giudiziari attraverso l'innovazione tecnologica, il supporto organizzativo alla informatizzazione e telematizzazione degli Uffici Giudiziari, disseminazione di specifiche innovazioni e supporto all'attivazione di interventi di change management". (accessibile presso [https://www.giustizia.it/giustizia/it/mg\\_1\\_4\\_1.page?contentId=SBG343205&previousPage=mg\\_1\\_4](https://www.giustizia.it/giustizia/it/mg_1_4_1.page?contentId=SBG343205&previousPage=mg_1_4)).

<sup>56</sup> Prof. Guido Boella (Direttore del Dipartimento di informatica dell'Università di Torino, coordinatore), la Scuola Universitaria Superiore IUSS Pavia, oltre i dipartimenti di informatica di tutte le altre università. Vi è anche un precedente. Tra il dicembre 2015 e il dicembre 2017 il Ministero della giustizia, Direzione generale per i servizi informativi e per l'automazione (DGSIA, Direttore P. Liccardo) ha promosso e finanziato un progetto pilota di collaborazione tra uffici giudiziari milanesi e università dell'area milanese per la promozione dell'informatizzazione nel processo penale (Amedeo Santosuosso è

stato il Coordinatore del progetto su delega del Presidente della Corte d'appello di Milano, Dott. G. Canzio). La centratura era sul processo e sulla funzionalizzazione ad esso di tutti i collegamenti con i diversi uffici, dal Portale notizie di reato fino alle impugnazioni della procura generale. Una linea di ricerca approfondì la struttura delle sentenze, sia civili sia penali, giungendo alla conclusione che la struttura logica e giuridica d'entrambe le tipologie fosse essenzialmente la medesima e che questo costituiva una grande possibilità per l'elaborazione di modelli in grado di aprire all'applicazione di tecniche di IA.

<sup>57</sup> La letteratura sul punto è sconfinata. Ci limitiamo a richiamare i testi alla base della teoria degli atti linguistici a partire dal lavoro di J. L. Austin, *How to do things with words*, Oxford University Press, Oxford 1962, che raccoglie le lezioni tenute a Harvard nel 1955, e John R. Searle, *Speech acts*, Cambridge University Press, Cambridge 1969. Le considerazioni di questo paragrafo vanno viste in collegamento con quelle di cui sopra, al paragrafo *Fenomenologia delle predizioni giuridiche*.

del caso in cui è coinvolto Antonio. E ha questo titolo perché, a seguito di una selezione pubblica, ha acquisito la posizione di giudice e, secondo specifiche norme organizzative e procedurali, è chiamato a decidere quella specifica causa e può, quindi, pronunciare la frase “Antonio ha ragione” con la forza di una decisione autoritativa e vincolante per le parti. Tutti gli altri enunciati (*utterances*) saranno, la prima, un’opinione da bar, la seconda, un’esercitazione didattica su un caso ipotetico o, la terza, una mera predizione di un difensore nell’interesse di una parte/cliente.

Se proviamo a collocare in questo piccolo schema esplicativo ciò che viene “detto” o “predetto” da un sistema predittivo (ciò che potremmo chiamare un “enunciato della macchina”, *machine utterance*) dobbiamo preliminarmente chiarire da quale posizione, in quale contesto e “a che titolo” il sistema dica “Antonio ha ragione”, e quindi *chi* lo abbia interrogato e in *quale* posizione fosse l’interrogante e *quale auctoritas* l’interrogante riconosca al sistema predittivo, e altro ancora.

In conclusione, le macchine e i sistemi di ML non decidono alcunché, perché per decidere bisogna essere titolati a farlo e, a quel che ci risulta, nessuno ha intitolato questi sistemi a decidere. Quindi, le macchine non prendono il posto degli umani, a meno che non siano proprio gli umani a sottrarsi ai propri compiti, conferendo una malaugurata *auctoritas* alle macchine, di diritto (cioè prevedendo l’efficacia giuridica della predizione automatica) o anche solo di fatto (adeguandosi passivamente a tale predizione)<sup>58</sup>.

Se a ciò si aggiunge quanto abbiamo notato sopra circa il fatto che la possibilità di una predizione “oggettiva” non esiste, quando l’obiettivo da prevedere sia proprio la decisione umana in una controversia e non vi sia un riscontro obiettivo della correttezza della decisione, il cerchio si chiude sul timore che le macchine rimpiazzino gli umani. Come visto sopra, vi è anche la possibilità di sviluppare sistemi che non si limitino a prevedere una singola soluzione, ma possano formulare diverse ipotesi alternative, con l’effetto di potenziare le capacità di scelta del giudice e accelerare l’evoluzione del diritto (di queste tecniche esistono al momento solo realizzazioni prototipali).

Vi è una sola eccezione, nella quale una parziale delega alle macchine può essere persino saggia. Nel caso delle *online dispute resolution*, avere un sistema automatico che si limiti a “dire” quale sia la soluzione della lite che sia maggiormente allineata alle decisioni passate in casi simili, può essere un’utile guida per le parti, che possono decidere di aderirvi o meno. Diciamo che può essere una soluzione saggia, in primo luogo, perché il sistema non decide nulla, limitandosi solo a indicare lo stato dell’arte al momento, e, poi, perché la decisione di aderire o meno al “detto” della macchina (*machine utterance*) ha carattere negoziale tra le parti, senza alcuna valenza autoritativa (se non quella implicita e conseguente alla stipulazione di un contratto).

*La decisione robotica travolge le istituzioni giuridiche.* Secondo altra preoccupazione, una volta che la decisione è robotica perderebbero ogni valore la difesa tecnica, l’appello, il ricorso per cassazione e la stessa motivazione, e si potrebbe passare dal giudice-robot al robot-legislatore. Provando a esplicitare l’assunto alla base di questa idea, si scoprirebbero due cose: una visione mitica, e del tutto irrealistica, della potenza delle macchine e una visione del

tutto meccanicistica del diritto e del suo funzionamento. Sulla prima visione mitica, veramente ha senso esercitarsi sull’ipotesi che un sistema artificiale possa riprodurre con esatta precisione ogni elemento dell’esperienza giuridica in un modo che sia unitario, unico e indiscutibile, tanto che da qualunque parte la si guardi dia sempre la stessa risposta allo stesso problema? Può bastare, per fugare il profilo di un tale *monstrum*, il richiamo alla molteplicità di approcci tecnici e teorici che abbiamo descritto sopra<sup>59</sup>? Oppure ricordare che, anche quando l’insieme di dati su cui si basa una predizione è condiviso, l’utilizzo di diversi modelli o tecniche di interrogazione di quella stessa base dati fornisce per definizione risposte diverse?

E sul piano giuridico (prima ancora di quello tecnico), veramente si pensa che l’intera esperienza giuridica possa coerentemente rappresentarsi con precisione matematica e “meccanica”? E che questo sia possibile a livello soprannazionale? O anche solo nazionale, ma tenendo conto delle universalmente riconosciute e variegate interferenze transnazionali e sovranazionali?

E allora, se tutto ciò è vero, e cioè che la rappresentazione unitaria e totalitaria del diritto è una visione da notte insonne, allora bisogna riconoscere che la (presunta) meccanica del diritto si scioglierà nella ricchezza delle dinamiche processuali, del contraddittorio, dell’*audiatur et altera pars*, tutti sistemi e modi che garantiscono la pluralità dei punti di vista (anche se ciascuno assistito tecnologicamente) e l’originalità della decisione finale (anch’essa eventualmente assistita tecnologicamente).

#### *Questioni risolvibili con accorgimenti pratici o tecnici*

In questo breve paragrafo parliamo dell’effetto gregge e della composizione della base dati per la predizione.

*Effetto gregge.* Alcuni autori paventano che lo sviluppo di sistemi predittivi porti all’impigrirsi dei giudici, che troverebbero più comodo adeguarsi a quello che suggerisce un sistema automatico rispetto all’elaborazione di nuove soluzioni. La questione non è nuova. Qualsiasi raccolta di precedenti, anche meramente cartacea, può provocare il temuto effetto gregge.

Allora, delle due una, o si abolisce ogni supporto documentale o persino la memoria professionale, perché condizionanti, tanto da dover coerentemente tessere l’elogio dell’ignoranza, perché ogni conoscenza è fonte di precomprensione e, quindi, di possibile pregiudizio (quanto più profonda la non conoscenza – ignoranza-, tanto maggiore la libertà del giudicante?), oppure non vi è altra risposta diversa dal dialogo e dalla formazione.

Il problema è quello di acquisire le conoscenze sui sistemi predittivi, sulle loro caratteristiche e sul loro uso appropriato nella pratica professionale (formazione). Come osservavamo sopra, il modello prodotto dal sistema può dare risultati accurati pur senza essere comprensibile all’uomo. Ciò può generare una dipendenza rispetto alla macchina, di cui si accettano le conclusioni, pur senza comprenderne le ragioni. Nell’ambito del diritto, l’idea del controllo del modello sulla base di dati osservabili non sarebbe peraltro realizzabile quando le previsioni del sistema fossero esperite dal giudice come indicazioni sulla sentenza da emettere: conformandosi al sistema il giudice ne confermerebbe le

<sup>58</sup> Si veda anche A. Carcaterra, *Machinae autonome e decisione robotica*, in A. Carleo (a cura di), *Decisione robotica*, Il Mulino, Bologna, 2019, 51 in particolare.

<sup>59</sup> Si pensi ai limiti storicamente verificati dei “sistemi esperti” (si veda sopra sub Par. 3.1).

predizioni. Infatti, se le indicazioni del modello dovessero essere considerate prescrizioni da seguire (anziché predizioni fattuali tutte da verificare!) si verificherebbe un corto circuito tra predizione e fatti mediante i quali testarla: l'indicazione del sistema ("puoi decidere secondo questa previsione") genera, una volta che il decisore vi si adegui, una realtà che conferma la predizione. Un sistema che genera predizioni automatiche sul comportamento futuro dei giudici potrà essere soggetto a verifiche empiriche solo se i giudici stessi sceglieranno di decidere senza adeguarsi passivamente alle previsioni del sistema. Che è esattamente un problema di formazione.

Inoltre, un importante rimedio all'effetto gregge può venire dalla predisposizione di sistemi che non si limitino a fornire solo un'indicazione che rifletta il risultato più probabile (la decisione del giudice medio), ma piuttosto riflettano l'esistenza di diversi indirizzi all'interno dei precedenti usati nell'insieme di addestramento, associando a ciascuno di essi un richiamo alle ragioni a sostegno dei diversi orientamenti. Spetterà poi al giudice, avvertito dell'esistenza del conflitto, decidere se allinearsi alla tendenza prevalente o invece a un orientamento minoritario.

La Scuola superiore della magistratura ha fatto e sta facendo un importante lavoro, che è l'unico antidoto ai rischi paventati. D'altra parte, così come non si può guidare un'automobile senza avere acquisito le conoscenze e la destrezza che le scuole guida danno, altrettanto vale per i nuovi sistemi. I giudici devono conoscere cosa sono effettivamente i sistemi di IA e cosa possono offrire realmente e devono maturare una visione realistica della motivazione, improntata a una razionalità che Patrizia Borsellino dice di "tipo logico o 'quasi-logico'" e che può manifestarsi come un aggregato temporaneo di dati<sup>60</sup>. In definitiva, devono imparare a coltivare la parte nobile del loro lavoro, come quella di cui parliamo più avanti a proposito delle questioni di costituzionalità.

*Quali dati per la predizione.* Altro problema è quello dei dati su cui si basa la predizione. Secondo alcuni, un sistema di predizioni che sia basato solo su precedenti giurisprudenziali sarebbe inconsistente con il nostro sistema giuridico, che non è di common law ma di civil law, nel quale le decisioni devono essere basate sulla legge e non sui precedenti. La critica ci pare che colga solo parzialmente nel segno, in primo luogo perché certe nette separazioni tra sistemi non possono più essere richiamate in modo meccanico, visto che ormai (e per effetto di modifiche sia culturali, sia legislative) il precedente (specie di legittimità) tende a porsi come regola 'forte' di decisione per controversie giudiziarie future e, dunque, dotato di quella capacità a farsi proiezione tendenziale degli esiti del contenzioso in cui trova applicazione<sup>61</sup>. E, d'altra parte, se la predizione è predizione della decisione è giocoforza che alle decisioni si faccia riferimento.

Vi è però un aspetto che merita attenzione. Gli avvocati lamentano che, nelle sentenze, i giudici riportino talora le loro posizioni in modo non esauriente (se non "infedele"), con la conseguenza che vanno perse visioni e interpretazioni nuove, che vengono proposte in giudizio e che meriterebbero di entrare nel circuito della conoscenza giuridica. Non ci nascondiamo che ciò possa accadere, quale che ne sia il motivo, e che possa avere effetti anche sulla qualità del

dataset che è alla base della predizione. La soluzione, però, non ci sembra che possa essere la creazione di un dataset unitario che contenga tutte le decisioni dei giudici e tutti gli atti degli avvocati dei rispettivi processi. A parte l'uniformità del dataset, che potrebbe essere pregiudicata, vi sono dei limiti posti dalle norme sulle privacy, circa la disponibilità degli atti dei difensori.

Una soluzione pratica, ma con effetti positivi proprio sulla qualità dei dati, è quella proposta dal progetto della Macroarea 1 di cui abbiamo parlato sopra al par. 5.3. In quel progetto, riprendendo importanti esperienze precedenti tra le quali gli accordi tra Consiglio Nazionale Forense e CSM, nonché un orientamento che si è andato affermando anche nel CSM, si propone di creare modelli di provvedimenti decisori, che riflettano la struttura logico-giuridica essenziale, che si rapportino a un modello dell'atto dell'avvocato che sia ispirato alla medesima visione del processo e che differisca non tanto per i dati normativi e fattuali, quanto per la necessaria specificità del ruolo difensivo dell'avvocato. I "modelli di nuova generazione" saranno file nativi digitali, capaci di raccogliere dati strutturati e capaci di acquisire automaticamente le parti delle difese degli avvocati che gli avvocati stessi avranno indicato come essenziali (sorta di abstract). In tal modo, le nuove raccolte di giurisprudenza saranno costituite sì da atti decisori, che però conterranno anche le parti essenziali e potenzialmente pregiate degli atti di parte, così come selezionate (è bene ripeterlo) dagli stessi avvocati.

#### *Questioni obiettivamente importanti*

In questo paragrafo ci occupiamo di alcune questioni obiettivamente importanti, come quella della trasparenza e della spiegabilità (*black box*), quella dei diritti di privacy industriale che coprono alcuni sistemi in uso e quello della dimensione temporale nel machine learning.

*Black boxes, tecniche e giuridiche.* Nella Sezione 3.4 si è visto come una rete neurale non si presta a fornire direttamente spiegazioni comprensibili circa il modo in cui il sistema abbia raggiunto una certa decisione e a rendere esplicite le ragioni per le quali è stata data una certa risposta. Si parla a tale proposito di sistemi "opachi" o di "scatole nere" (*black box*). Di qui la necessità di bilanciare efficienza (accuratezza nelle predizioni) e spiegabilità, nella scelta del sistema predittivo da utilizzare in un determinato ambito. Quando sono in gioco interessi in contrasto e le esigenze del controllo siano preminenti, come nel caso dell'azione amministrativa e ancor più nell'ambito della giurisdizione, l'esigenza di fornire spiegazioni è fondamentale.

Tutto ciò è sicuramente vero e va tenuto nella massima considerazione. Tuttavia, la questione può essere circoscritta tenendo conto di alcuni aspetti. In primo luogo, è bene ricordare quanto detto sopra sul fatto che i sistemi non decidono alcunché: allora, si potrà dire che la *machine utterance* (l'enunciato della macchina), non spiegabile, è solo uno tra i plurimi stimoli che il giudicante riceve nel corso del processo decisionale, dagli atti di parte, a proprie letture, ad apporti di assistenti e altro, e che poi richiedono una sintesi dell'umano, spiegata dall'umano. È la differenza tra il *contesto della scoperta* e quello della sua *giustificazione/spiegazione*<sup>62</sup>. In secondo luogo, se si esclude che la

<sup>60</sup> Sul punto di cosa sia la motivazione e dei profili attuali che essa assume, sia consentito il rinvio a A. Santosuosso, *Intelligenza artificiale e diritto*, Milano, 2020, 101-120 (dove è spiegata la visione della motivazione come aggregato temporaneo di dati).

<sup>61</sup> Vincenti, *op. cit.*, *loc. cit.* citato al par. 2.2.

<sup>62</sup> A. Santosuosso, *Intelligenza artificiale e diritto*, cit., 107 e seguenti.

motivazione sia il rapporto sui percorsi psicologici del giudicante e si accetta che essa è un atto di responsabilità sociale che rende la decisione criticabile per via di impugnazione, si potrà allora contestualizzare il problema della spiegabilità notando che vi sono certo scatole nere tecniche (quali sono quelle dell'apprendimento profondo o *deep learning*), ma che noi viviamo immersi in e conviviamo anche con scatole nere giuridiche, come per esempio il divieto di opinione dissenziente nel nostro ordinamento, che non consente di esplicitare i diversi percorsi argomentativi che hanno portato a una certa decisione, e varie altre. Tra le black box giuridiche vi sono anche quelle indotte dai sistemi di privativa industriale, di cui parliamo ora.

*Proprietà intellettuale e conoscibilità delle ragioni delle decisioni.* Si è parlato sopra (alla Sezione 4.1) del notissimo caso del signor Loomis, che si è visto infliggere un gravoso aumento di pena per il rischio di recidiva, senza poter conoscere le ragioni che avevano portato a quella misura. La particolarità di questo caso va messa a fuoco correttamente. Se il problema sono i possibili bias del sistema (dataset e/o algoritmi applicati), essi potrebbero essere oggetto di discussione preventiva e di continuo monitoraggio applicativo, rimanendo fermo che il giudice può giustificare i fattori aggiuntivi e particolari che portano a una determinazione diversa (come ha ricordato anche la Corte suprema del Wisconsin nel caso Loomis). Se invece il problema è quello di affidarsi a un aiuto tecnologico nella quantificazione della pena, ci si può chiedere: dovremmo fare più affidamento su un ragionamento giuridico esclusivamente umano, nel quale possono annidarsi bias cognitivi del giudicante (coperti talora da motivazioni apparenti), piuttosto che su dati quantitativi, che possono essere ulteriormente discussi prima della sentenza finale<sup>63</sup>?

Gli aspetti da tenere in conto sono il diritto dell'imputato al giusto processo e all'esercizio del diritto di difesa contro la segretezza dei dataset e dei della loro elaborazione, coperti dal segreto industriale. Ma se il problema è la proprietà privata del sistema COMPAS (o simili), con i conseguenti diritti di privativa che impediscono di conoscere nel caso specifico la logica e il modo di funzionamento del sistema, ci si può chiedere quanto segue. Se (in via meramente teorica) il Ministero della giustizia italiana fornisse anche un sistema del genere COMPAS, sistema preventivamente conosciuto nella sua logica e architettura, discusso e approvato da ordini degli avvocati e dai giudici, verificato e corretto alla luce delle applicazioni, sarebbe esso accettabile? Se il sistema fosse in mano pubblica e la sua architettura fosse conosciuta e condivisa dalla magistratura e dall'avvocatura, sarebbe comunque da rifiutare?

Uno sviluppo di tal genere, che era stato ipotizzato qualche anno fa<sup>64</sup>, vi è stato proprio negli Stati Uniti con l'introduzione del sistema pubblico PATTERN (che sta per *Prisoner Assessment Tool Targeting Estimated Risk and Needs*), che ha segnato un importante cambiamento di rotta verso il rispetto della trasparenza, non solo perché il nuovo sistema algoritmico non è di proprietà di società private, e quindi è svincolato dalle regole sul segreto indu-

striale, ma anche per le revisioni previste dal *First Step Act*, effettuate da organismi indipendenti nominati dal *Department of Justice* (DOJ) e i cui primi risultati sono stati divulgati al pubblico, ma anche per il coinvolgimento, nella fase "genetica" del nuovo strumento, di enti e associazioni rappresentative degli interessi in gioco, che inaugura un nuovo *modus operandi*<sup>65</sup>. Naturalmente il dibattito è ancora aperto, ma lo sviluppo pubblico indica una strada interessante e dimostra come i sistemi non vadano visti come entità non suscettibili di cambiamenti sulla base dell'esperienza.

*Dimensione temporale e machine learning.* Come si osservava nelle pagine precedenti, nell'ipotesi di apprendimento supervisionato, il sistema apprende da un insieme di addestramento, costituito da decisioni passate, e la soluzione proposta dal sistema, dovrebbe essere, di regola, quella che riflette la maggioranza dei giudizi passati. Se i giudici, seguendo le indicazioni del sistema si allineassero a quell'indirizzo, lo sviluppo del diritto risulterebbe bloccato. Come affermava Roscoe Pound, contemporaneo al sopra citato Holmes, nella sua critica alla "giurisprudenza meccanica", esso sarebbe "pietrificato", incapace di rispondere alle sfide del presente, rispondendo "ai bisogni della vista sociale odierna"<sup>66</sup>. Inoltre, una prospettiva collegata esclusivamente alle decisioni passate disconoscerebbe le innovazioni legislative, e potrebbe quindi fornire informazioni fuorvianti al giudice di casi futuri (e incompatibili con il principio del vincolo del giudice alla legge).

Non mancano le risposte tecnologiche a tale problema. Si possono sviluppare sistemi nei quali un maggior peso sia attribuito, all'interno dell'insieme di addestramento, agli esempi più vicini a noi rispetto a quelli passati. Si possono altresì collegare i casi passati alle norme legislative che essi applicano, così da tener conto delle ipotesi in cui le norme applicate siano state derogate o abrogate da norme successive. Tutte tali soluzioni, tuttavia sono fallibili. Ad esse deve affiancarsi l'unico rimedio generale, che consiste nella consapevolezza e competenza dell'utilizzatore del sistema. Nella prospettiva dell'informatica cognitiva (*cognitive computing*), l'utilizzatore deve considerare il sistema come un utile strumento informativo, piuttosto che come un sostituto al quale delegare il proprio compito.

*La comparazione tra norme.* Alcuni autori (Massimo Lucciani con particolare chiarezza) hanno sollevato il problema del tipo di dottrina giuridica (e, quindi, di metodo) che la macchina ha incorporata, di come essa organizzi il rapporto con i precedenti e di come risolva il raffronto o conflitto tra norme. Su questo punto, non abbiamo alcuna difficoltà a dire, non solo che non ci risulta che, a oggi, vi sia un sistema capace di far ciò, ma anche che si tratta di un'attività talmente piena di sfumature complesse che sia bene che rimanga nel dominio umano (senza spreco di risorse nel tentativo di riprodurla in un sistema di IA).

Si pensi al caso di un giudice che può rimettere alla Corte costituzionale una norma per sospetto contrasto con una norma costituzionale. Capita che la norma sospettata sia

<sup>63</sup> Sui bias dei giudicanti è fondamentale D. Kahneman, O. R. Sibony, C. R. Sunstein, *Noise, A flaw in Human Judgment*, Little, Brown and Company 2021 (ora anche in traduzione italiana come D. Kahneman, O. R. Sibony, C. R. Sunstein, *Rumore. Un difetto del ragionamento umano*, Torino, 2021).

<sup>64</sup> Si veda A. Santosuoso, *Intelligenza artificiale e diritto*, cit., 91-93.

<sup>65</sup> Si veda DOJ, Office of the Attorney General, *The First Step Act*

of 2018: Risk and Needs Assessment System – Update, gennaio 2020, 1. Per un'ampia illustrazione e discussione su questi temi, molto interessante il lavoro D. Zingales, *Risk assessment: una nuova sfida per la giustizia penale?* in *Diritto Penale e Uomo*, 9 dicembre 2021, [https://dirittopenaleuomo.org/contributi\\_dpu/risk-assessment-una-nuova-sfida-per-la-giustizia-penale/](https://dirittopenaleuomo.org/contributi_dpu/risk-assessment-una-nuova-sfida-per-la-giustizia-penale/).

<sup>66</sup> Roscoe Pound, *Mechanical Jurisprudence*, *Columbia Law Review*, 8 605-623, 1908.

stata applicata mille volte, anche dalla Cassazione, in un certo modo. Ma al giudice, che dovrebbe applicarla la 1001° volta, all'improvviso (di sua iniziativa o sollecitato da un difensore) quella norma appaia in contrasto con la Costituzione per alcuni buoni motivi, che esistevano anche prima, ma che nessuno aveva notato. Ecco, crediamo che questo sia un esempio di qualcosa, una facoltà intellettuale umana che un sistema di IA difficilmente potrà sostituire. Questo è qualcosa che il giudice deve imparare a coltivare.

Salvatore Settis, in un bellissimo libro dedicato al rapporto tra innovazione e tradizione nell'arte, a un certo punto nota: "vi sono *Pathosformeln* [espressione coniata da Warburg] inventate dagli Antichi, tramandate poi per secoli in modo via via più stanco, fino a disseccarsi e irrigidirsi in passiva ripetizione. Eppure, in quella scorza cova ancora il fuoco di un *pathos* incontenibile: dopo secoli di formularità senz'anima, talora un artista può riconoscere d'un colpo il potenziale patetico di una svigorita formula, farla propria e riattivarla rilanciandola entro una propria creazione artistica"<sup>67</sup>.

Ora, non vi è bisogno di arrivare a paragonare il giudice all'artista per apprezzare l'idea dell'apparizione improvvisa di qualcosa di vitale in un contesto senz'anima. È qualcosa che somiglia a quello che accade per i sospetti di incostituzionalità, che a volte portano il giudicante umano a dire:

*come non averci pensato finora!* Noi pensiamo che questo vada preservato.

### **Una previsione per la predizione? *Trial and error* per giuristi**

Il prossimo periodo potrebbe essere segnato dal passaggio da sistemi predittivi sperimentali (al livello di prototipi) alla creazione di sistemi che possano essere sperimentati all'interno della pratica giudiziaria. Crediamo che questo passaggio richieda una consapevolezza e un'attitudine nuova.

La consapevolezza è che l'applicazione dell'IA in ambito giudiziario è un'attività ad alto rischio (*high risk*, secondo il gergo dello *EU Artificial Intelligence Act*<sup>68</sup>), che richiede la massima cautela e condivisione tra le professioni coinvolte.

L'attitudine riguarda lo sviluppo di una familiarità con il metodo sperimentale che l'uso di tali sistemi richiede. Non vi è alcun sistema, per accurato che sia nella progettazione, che non richieda di essere testato, verificato in pratica, sottoposto a riesame critico, dal punto di vista sia tecnico, sia giuridico. *Trial and error* anche per i giudici? Diremmo di sì, partendo dalla consapevolezza che anche *non tentare un'innovazione* può essere un errore e provocare danni.

<sup>67</sup> Salvatore Settis, *Incursioni. Arte contemporanea e tradizione*, Milano, 2020, 19.

<sup>68</sup> Interessante il lavoro L. Floridi et al, *capAI. A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act*, disponibile presso <https://ssrn.com/abstract=4064091>.

A proposito dell'affidabilità e qualità dei dataset interessante anche la proposta di Timnit Gebru di istituire *Datasheets for Datasets* <https://cacm.acm.org/magazines/2021/12/256932-datasheets-for-datasets/fulltext#>.