

RECENSIONE A GIOVANNI SARTOR, *L'INTELLIGENZA ARTIFICIALE E IL DIRITTO*, GIAPPICHELLI, TORINO, 2022 *

STEFANO PIETROPAOLI**

* Contributo non sottoposto a referaggio in conformità al Regolamento della Rivista.

** Professore associato di Filosofia del diritto nell'Università degli Studi di Firenze.
Contatto: stefano.pietropaoli@unifi.it

«Siamo entrati nell'era dell'intelligenza artificiale». Questa affermazione riecheggia ormai costantemente, ovunque, a ogni livello culturale e sociale, suscitando un misto di inquietudine, speranze, aspettative e paure. Queste reazioni, tanto dissonanti tra loro, sono generate dalla confusione che segna il dibattito sull'intelligenza artificiale e dalla diffusa incomprensione del fenomeno. In altre parole, dire che siamo entrati nell'era dell'intelligenza artificiale, può significare tutto e il contrario di tutto (ma, soprattutto, può non significare nulla).

Se è vero che siamo nell'era dell'intelligenza artificiale, dobbiamo prima di tutto chiarire che questa nuova epoca non è iniziata oggi e neppure negli ultimi anni, ma intorno alla metà del Novecento, quando quel (fortunatissimo) termine è stato coniato e si sono elaborati i concetti fondamentali di un nuovo campo di ricerca, sulla scia dei successi dell'informatica di Turing e della cibernetica di Wiener. E se oggi stiamo vivendo una "primavera" dell'intelligenza artificiale, non dimentichiamo che nella sua breve storia essa ha già vissuto non pochi "inverni": fasi di stallo determinate dal mancato raggiungimento degli obiettivi promessi e segante dal conseguente definanziamento da parte di Stati e aziende.

Tutto ciò premesso, rimane la questione: che cosa significa essere entrati nell'era dell'intelligenza artificiale, e in particolare che cosa può voler dire per un giurista e, in generale, per chi studia, usa, applica il diritto?

Per rispondere a questo interrogativo in una fase, come quella attuale, segnata dal pullulare di sedicenti esperti di nuove tecnologie, l'unica via sicura è quella di affidarsi a chi ha iniziato ad occuparsi di queste tematiche in tempi in cui l'intelligenza artificiale non era ancora la giostra dalle mille luci abbaglianti che conosciamo oggi. In altre parole, credo (e il vecchio buon senso vorrebbe) che per comprendere le reali implicazioni di qualcosa che non capiamo, la strada migliore sia quella di lasciarci guidare da chi è in grado di guardare a quel fenomeno da una posizione più elevata rispetto a quella del non esperto: posizione, si badi bene, guadagnata con la lunga e tenace passione per la ricerca, il rigore degli studi, la felice intuizione di campi di indagine destinati a diventare centrali.

In una simile prospettiva, il nome di Giovanni Sartor è sin troppo facile da individuare come quello dell'autore di riferimento sul rapporto tra intelligenza artificiale e diritto: tematica che egli ha iniziato ad affrontare in un'epoca che potremmo definire pionieristica (basti pensare che *Intelligenza artificiale e diritto: un'introduzione* è il titolo di una sua monografia del 1993), e che ha continuato incessantemente a indagare fino ad oggi. Di questo lungo percorso di studi il volume qui recensito è testimonianza: tentativo – e di tentativo pienamente riuscito si tratta – di trovare un difficile equilibrio, evitando sia una monografia iperspecialistica poco gestibile da parte di chi si accosti per la prima volta a queste problematiche, sia un'epitome meramente divulgativa.

Il testo si articola in quattro capitoli. Il primo di questi svolge quella funzione definitoria la cui assenza, come si è già ricordato, produce confusione e genera incomprensioni (e alimenta paure). L’A. mostra, prima di tutto, la complessità del concetto di IA, mettendo in evidenza come tale espressione non indichi tanto una disciplina monolitica, quanto piuttosto un’ampia gamma di metodi e tecniche applicate a una straordinaria varietà di obiettivi scientifici e tecnologici.

Dopo aver esaminato le diverse concezioni che hanno segnato la storia dell’IA (IA debole e IA forte *in primis*), Sartor affronta alcuni snodi teorici fondamentali (come il senso stesso dell’intelligenza e la questione della comprensione dei significati da parte di enti dotati di IA), non sottraendosi all’onere di chiarire termini e concetti troppo spesso impiegati in maniera disordinata o errata: algoritmi, big data, robotica. Ma il passaggio forse più significativo del capitolo riguarda un possibile “concetto giuridico” di IA.

Il giurista, ricorda Sartor, ha bisogno di strumenti concettuali di grande precisione, che consentano distinguere i fenomeni cui si applicano determinate disposizioni, da quelli in cui esse quelle stesse non si applicano. Se non siamo in grado di stabilire in modo sufficientemente chiaro che cosa sia intelligente e che cosa non lo sia, non possiamo neppure distinguere tra sistemi intelligenti e sistemi privi di intelligenza al fine di applicare agli uni e agli altri le rispettive norme. In questa prospettiva, l’A. propone una concezione giuridica dell’IA declinata in senso sostanziale e teleologico: ciò che conta non è la qualificazione formale di un sistema come intelligente, ma è il fatto che esso presenti i rischi e le opportunità delle applicazioni comunemente considerate intelligenti.

Nel secondo capitolo vengono passate in rassegna le principali tecnologie dell’IA, con particolare riferimento alla rappresentazione della conoscenza (e ai limiti del modello logico applicato al diritto) e all’apprendimento automatico. Nella prima fase di sviluppo dell’IA, infatti, l’assunto di partenza era che per realizzare un sistema intelligente fosse indispensabile fornire al sistema stesso tutta la conoscenza necessaria per affrontare i compiti affidatigli, e che tale conoscenza dovesse essere rappresentata mediante linguaggi formali ed elaborata mediante ragionamenti automatici. Ma negli ultimi due decenni si è imposta prepotentemente una diversa linea di sviluppo, basata sulle tecniche per l’apprendimento automatico.

L’A. distingue così con estrema chiarezza tra apprendimento supervisionato, apprendimento non supervisionato e apprendimento per rinforzo, ricorrendo a una serie di esempi che aiutano il lettore non soltanto a comprendere i meccanismi che animano l’IA (come gli alberi di decisione, la regressione statistica, gli algoritmi evolutivi e le reti neurali), ma, soprattutto, chiarisce un punto essenziale soprattutto per il diritto: i sistemi di apprendimento automatico che

forniscono le prestazioni più puntuali sono anche i più “opachi” (*black box*), cioè sono i meno capaci di giustificare le proprie decisioni.

Quest’ultimo punto solleva la questione della possibile discriminatorietà dei sistemi di IA e introduce il tema delle opportunità e dei rischi dell’IA, affrontato nel terzo capitolo.

Il potenziale impatto – etico, ancor prima che giuridico – espresso dai sistemi di IA è straordinario. Basti pensare alle capacità predittive che alcuni sistemi hanno già mostrato di avere, e al loro possibile impiego in settori delicatissimi, dall’accesso al lavoro all’amministrazione della giustizia.

L’impressionante quantità di dati che questi sistemi sono in grado di processare crea inedite opportunità di influenzare e manipolare la cittadinanza, grazie a elaborate strategie di profilazione connotate da una pervasività finora impensabile. E a chi volesse etichettare questi rischi come remote eventualità di un futuro distopico, l’Autore ricorda la nuova dimensione della sorveglianza che si è manifestamente concretizzata nel sistema di credito sociale cinese.

Di fronte a queste nuove sfide, è urgente sviluppare strumenti etici e giuridici capaci di evitare derive profondamente lesive dei diritti fondamentali. Come ricorda l’A., l’IA costituisce forse la principale sfida alla quale l’umanità dovrà far fronte nei prossimi decenni.

Grandi opportunità di progresso individuale e sociale sono affiancate da gravi rischi, anche nel campo della formazione dell’opinione pubblica e del funzionamento dei processi democratici. La circolazione della conoscenza promossa dall’avvento del web non si è dimostrata garanzia di una migliore formazione politica e culturale, e neppure di una maggiore consapevolezza: profilazione, decisioni automatiche, polarizzazione delle opinioni, disinformazione, manipolazione e controllo di massa sono problemi già attuali.

Il realismo con cui passa in rassegna i pericoli presenti e futuri non conduce certo l’A. a un rifiuto delle tecnologie dell’IA. L’A., al contrario, sottolinea con forza come tali tecnologie possano portare enormi benefici all’umanità, a patto che il loro uso sia indirizzato in maniera eticamente, giuridicamente, politicamente ed economicamente sostenibile.

L’attuale quadro etico-giuridico, rispetto ai nuovi problemi posti dall’IA, si mostra spesso inadeguato e comunque incapace di dare indicazioni puntuali ai cittadini e alle istituzioni pubbliche e private. Promuovere pratiche eticamente positive nell’uso dell’IA significa, secondo l’A., assicurare che lo sviluppo e l’impiego dell’IA avvenga in un contesto nel quale interessi individuali e valori sociali siano rispettati e promossi. In questa direzione, tre sono gli obiettivi minimi di un’etica dell’IA: tutelare l’autorealizzazione umana, preservando le abilità tipiche degli esseri umani; continuare a dare un senso alla capacità di azione umana, prevedendo nuove forme di responsabilità; promuovere la

coesione sociale, custodendo allo stesso tempo l'autodeterminazione degli esseri umani.

Anche sulla scorta di documenti elaborati da esperti a livello internazionale (come l'High-Level Expert Group on Artificial Intelligence), l'A. sottolinea la necessità di sviluppare una IA “degnata di fiducia” (*trustworthy AI*), all'insegna di alcuni principi fondamentali.

In primo luogo, l'IA deve essere progettata per aumentare, integrare o rafforzare le abilità di tipo cognitivo, sociale o culturale degli esseri umani, assicurando il pieno rispetto per l'autonomia umana ed evitare qualsiasi forma di ingiustificata manipolazione o condizionamento degli esseri umani.

In secondo luogo, i sistemi di IA e gli ambienti nei quali essi operano devono essere “sicuri”, nel senso che non debbono causare o esacerbare danni o avere altri impatti avversi sugli esseri umani. L'IA deve essere sviluppata con la consapevolezza dei rischi che può comportare, in maniera da prevenire esiti lesivi dell'integrità fisica e mentale e della dignità stessa degli esseri umani.

In terzo luogo, l'IA deve evolversi all'insegna della *fairness*: non soltanto essa deve assicurare una distribuzione eguale e giusta sia dei benefici sia dei costi sociali che può comportare, ma deve contrastare la riproduzione di pregiudizi, di atteggiamenti discriminatori e di condotte stigmatizzanti. Inoltre, occorre garantire il diritto di contestare le decisioni adottate da sistemi di IA e prevedere rimedi effettivi contro tali decisioni.

Infine, i sistemi di IA devono tendere alla “trasparenza”, nel senso che le loro funzionalità e gli obiettivi che consentono di conseguire devono essere dichiarati apertamente e in maniera comprensibile per tutti i potenziali interessati. Le decisioni prese da questi sistemi devono essere, per quanto possibile, “spiegabili”.

Nel quarto e ultimo capitolo, l'A. passa in rassegna le principali applicazioni giuridiche dell'IA: i sistemi basati su regole (composti da una base di conoscenza costituita da regole e da un motore inferenziale che applica tali regole ai casi concreti), il ragionamento *defeasible* (dove un'inferenza si impone solo in modo provvisorio, cioè solo a condizione che non emergano eccezioni, contro-esempi, argomenti contrari di importanza preminente), il ragionamento basato sui casi (in cui la base di conoscenza del sistema è costituita da un insieme di precedenti, come il sistema HYPO comunemente applicato nell'ambito della violazione dei segreti industriali), l'apprendimento automatico (come il sistema Claudette, dedicato all'individuazione di clausole contrattuali abusive e illegali), la giustizia predittiva.

Quest'ultimo tema è oggetto di particolare attenzione. Gli ultimi paragrafi sono infatti dedicati alla previsione (di fattori e basata su fattori) e alle predizioni (e ai loro obiettivi e proxy). L'A. enumera i vantaggi che un sistema “predittivo” potrebbe offrire alla pratica del diritto, se sviluppato in modo da

fornire indicazioni accurate e spiegazioni adeguate, e purché applicato con piena consapevolezza dei suoi limiti. Per giudici, avvocati e privati potrebbe trattarsi – così sostiene l’A. – di uno strumento capace di offrire una più profonda conoscenza del “diritto in azione”, ovvero di un ordinamento colto nella sua vitalità.

Il testo di Giovanni Sartor aiuta il lettore ad ammirare con maggiore consapevolezza lo spettacolare scenario creato dall’IA, smorzandone tanto gli eccessivi entusiasmi quanto le paure infondate. Al giurista in particolare il volume offre la possibilità di comprendere in quale misura le nuove tecnologie possano rendere più efficace ed efficiente il diritto, contribuendo a realizzare valori di razionalità e giustizia, allontanando allo stesso tempo lo spettro della sostituzione dell’attività umana con la decisione automatica e l’asservimento del giurista umano alla macchina.

Il volume, rendendo comprensibili termini e concetti troppo spesso avvolti da un alone quasi mistico, anche attraverso il ricorso a numerosi esempi e casi pratici, offre una nuova “attitudine di speranza” nei confronti dell’IA.

Se moltissimi sono i rischi che le tecnologie intelligenti sembrano dischiudere, e se è impossibile prevedere quali saranno i loro sviluppi nel prossimo futuro, una comprensione sufficientemente precisa delle tecnologie dell’IA e dei problemi che esse sollevano rappresenta, anche per il giurista, non soltanto la via di salvezza da esiti potenzialmente catastrofici, ma anche la concreta possibilità di realizzare un futuro migliore. Solo attraverso questo sforzo conoscitivo si potrà contribuire a costruire una “buona” regolazione dell’IA e, più in generale, a “salvare” un sapere – la scienza giuridica – che trova la ragione della propria esistenza nella centralità dell’essere umano.