

## 1. Questioni definitorie

### Amedeo Santosuosso

Docente di Diritto, Scienza, Nuove tecnologie, Università di Pavia e Scuola Universitaria Superiore di Pavia. E-mail: [a.santosuosso@unipv.it](mailto:a.santosuosso@unipv.it)

Vi sono più differenze che cose in comune nel modo in cui si parla di IA nel mondo dell'educazione, della ricerca, dell'industria e dei media: il campo dell'Intelligenza Artificiale (IA) manca di un significato che sia universalmente accettato.

Tuttavia, vi sono alcune questioni, che si potrebbero dire di confine, sulle quali vi è un ampio consenso.

Un limite, che a oggi appare invalicabile, è la comprensione di quello che esse stesse, le macchine, riescono a fare. Roger Penrose, matematico illustre, emerito dell'Università di Oxford e vincitore del Premio Wolf assieme al suo amico e collaboratore Stephen Hawking, è molto netto. In un'intervista rilasciata in occasione dell'evento "AI for Good" (2018) tesse le lodi delle incredibili capacità di calcolo dei sistemi di IA, della loro utilità ai fini della progettazione, ma è molto chiaro nel porre la distinzione tra «agire sulla base dell'esperienza», che è quella che può portare una macchina a vincere persino nel Go contro l'umano più esperto (come già accaduto per gli scacchi con Kasparov), e la «comprensione di quello che sta facendo»<sup>1</sup>. Così le macchine sviluppano abilità senza avere la conoscenza teorica di quale sia il fondamento di quelle azioni.

<sup>1</sup> Le riflessioni contenute in questo scritto sono tratte dall'Introduzione e dal cap. 1 del mio A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Mondadori Università, 2020, al quale mi permetto di rinviare per uno sviluppo più ampio del mio pensiero e per i riferimenti bibliografici.

Ci si può allora chiedere, come mai in articoli, libri, convegni, siti web, social, aperitivi, discussioni sobrie o da bar, viene usata pervasivamente la parola "intelligenza", con ad essa associato l'attributo artificiale? E se tutti lo fanno con toni ora allarmati, ora banalizzanti, ora stupiti, ora pensosi (ma non sempre con pensiero), allora non sembrano che darsi due possibilità, o sono tutti fuori di senno (cosa che di solito è bene escludere) oppure qualcosa che effettivamente sia intelligente, oltre che artificiale, deve esistere per davvero.

Ma, quello che viene chiamato IA non è altro che lo sviluppo ampio, potente, talora sofisticato di capacità computazionali settoriali. Niente che abbia a che fare con le entità alle quali normalmente connettiamo l'attributo di intelligenza nella nostra vita di tutti i giorni, sin da prima dell'IA. A coloro che hanno esperienza di insegnamento sarà capitato di dire che uno studente sia molto intelligente. E, a veder bene, questo ha riguardato quegli studenti che mostrano la capacità di porre a frutto le nozioni acquisite e di fare collegamenti, spaziando da un campo all'altro, talora in modo inatteso. Ricordo una studentessa di fisica che allo IUSS di Pavia, classi interdisciplinari, rispondeva molto meglio degli studenti di giurisprudenza alle questioni di diritto che ponevo loro. Oppure quella studentessa che, dopo una brillante laurea in medicina, ha conseguito il diploma IUSS con una tesi di diritto sulla possibilità della coscienza in entità artificiali (poi reclutata per il dottorato presso un'università londinese molto prestigiosa). E potrei continuare. Bene, cosa avevano in comune Pia e Laura, e molti altri come loro? Una straordinaria capacità generale, quella per cui, a un certo punto di intelligenza umana, arrivi a dire che quella persona è intelligente tanto da poter fare molto bene qualsiasi cosa in qualsiasi campo. Questa è una qualità tipicamente

umana, che gli esperti di IA chiamano *general purpose intelligence*, e che, a oggi, non esiste in nessuna macchina in nessun angolo del mondo, benché fosse già nelle intenzioni di chi mosse i primi passi nell'intelligenza artificiale negli anni Cinquanta del secolo scorso. Menti alacri vi stanno lavorando e certamente porteranno ad avanzamenti importanti. Ma, a voler essere semanticamente precisi, solo dal momento in cui vi saranno quegli avanzamenti saremo autorizzati a parlare effettivamente di IA, mentre fino a quel momento l'uso della parola "intelligenza" riferita ad artefatti tecnologici sarà un (piccolo?) abuso, un gioco di etichette. Questo potrebbe essere la fine del discorso, perché se la cosiddetta IA non esiste, allora non abbiamo neanche da definirla: non possiamo definire il nulla.

Tuttavia, stiamo al gioco e vediamo qualche altro aspetto. Le macchine oggi, pur meravigliose, fanno cose diverse da quelle che fa un umano con la sua intelligenza, ammesso che le macchine debbano (e perché mai? Se non per un'irrefrenabile tendenza all'antropomorfizzazione) fare esattamente quello che facciamo noi umani. Perché non prendere atto che la nostra intelligenza, per straordinaria che sia (e credo che lo sia) non è l'unico tipo di intelligenza possibile, così come la nostra coscienza non è l'unico tipo di coscienza che in natura si dà? Il nostro amico 'polpo' ci può dire molte cose circa la distribuzione e il modo di funzionamento dell'intelligenza in un sistema non accentrato come quello umano!

Sembra di essere ancora alla convocazione del seminario del 1956, quando John McCarthy in collaborazione con Marvin Minsky, Claude Shannon e Nathaniel Rochester, coniano il termine IA e formulano, in modo che appare oggi un po' naif, la seguente proposta:

*«Proponiamo di svolgere uno studio sull'intelligenza artificiale per due mesi, con dieci persone, durante l'estate del 1956 al Dartmouth College di Hanover, nel New Hampshire. Lo studio procederà sulla base della congettura che ogni aspetto dell'apprendimento, o qualsiasi altra caratteristica dell'intelligenza, possa in linea di principio essere descritto con precisione tale che sia possibile costruire una macchina per simularlo. Si tenterà di scoprire come costruire macchine in grado di utilizzare il linguaggio, formare astrazioni e concetti, risolvere tipi di problemi che oggi sono di esclusiva competenza degli uomini, migliorare sé stesse. Riteniamo che sia possibile ottenere un significativo progresso in uno o più di questi problemi dedicando un'intera estate al lavoro collettivo di un gruppo di scienziati selezionati».*

Vi è un limite conoscitivo che porta gli umani a riprodurre quella che considerano la migliore intelligenza disponibile, cioè quella umana. Questo limite porta a usare un linguaggio antropomorfo

e, quindi, ad antropomorfizzare le macchine che produciamo. Visto in questa ottica, anche lo «Human Brain Project», il colossale piano di ricerca finanziato dall'Unione europea, che ha lo scopo di riprodurre *in silico* l'attività cerebrale, non appare altro che uno sviluppo di qualcosa intravisto già molti decenni prima e che è diventato possibile (anche se pur sempre controverso) proprio per lo sviluppo delle capacità computazionali delle macchine. Ma l'antropomorfismo è capace di vendicarsi svelando i limiti del concetto di intelligenza umana che si assume e che si proietta sulle macchine. Per esempio, è interessante che Russell e Norvig, nell'espone l'approccio razionalista in IA e nel distinguere tra «comportamento umano» e «comportamento razionale», debbano precisa-

re, in modo che pare un po' imbarazzato, che «non stiamo implicando che gli esseri umani siano “irrazionali nel senso di emozionalmente instabili” o siano pazzi». Ne emerge una visione riduttiva dell'intelligenza e della stessa razionalità umana, che sembra ignorare gli sviluppi delle ultime decadi della ricerca neuroscientifica a proposito del ruolo delle emozioni nei processi cognitivi.

Ecco, quindi, un'oscillazione tra IA antropomorfa e modelli distribuiti di IA, che necessariamente non pongono come centrale la visione soggettiva centralizzata tipica dell'identità, prima ancora dell'intelligenza, umana. Oscillazione che pone non pochi problemi alle ambizioni definitorie.

Vi è una certa circolarità delle questioni, che lascia irrisolta la domanda centrale: si può definire cosa sia IA? Domanda resa ancora più acuta considerando l'incremento importante del numero di pubblicazioni in materia negli ultimi anni. Da 10.000 a 60.000 dal 1998 al 2017, secondo un rapporto Elsevier. Lo stesso rapporto indica che, a livello globale, la ricerca sull'intelligenza artificiale ha accelerato, crescendo di oltre il 12% ogni anno negli ultimi cinque anni (2013-2017), rispetto a meno del 5% nei cinque anni precedenti (2008-2012). La produzione complessiva della ricerca, a livello globale in tutte le aree tematiche connesse a IA, è cresciuta dello 0,8% ogni anno negli ultimi cinque anni (2013-2017).

Tuttavia, se volessimo dire esattamente quale sia l'oggetto di tutti questi investimenti di risorse e di intelligenze, saremmo in seria difficoltà. Lo è stata anche Elsevier, tanto che per redigere il suo *Rapporto sull'IA*, ha scelto di far ricorso a tecniche di IA per far emergere dal campo dell'IA stessa un pattern, un'idea, che attraversasse tutti gli studi: usando IA per definire IA.

Se dai temi generali si passa a questioni specifiche, ma non meno rilevanti, emerge un'ulteriore ragione della difficoltà definitoria: nel campo dell'IA viene utilizzato un complesso di sistemi e tecniche in continua evoluzione e adattamento. Per esempio, è comunemente notato che l'avvento di internet abbia consentito all'IA di uscire da un periodo di stagnazione, così come ora è la prospettiva della connettività ultraveloce (5G), che in sé nulla avrebbe a che fare con l'IA, che promette uno sviluppo ampio. Così come la tecnologia blockchain, che poggia concettualmente su Internet, si intreccia con l'IA perché trasforma ogni nodo in un luogo non di mera circolazione di dati, ma di elaborazione degli stessi.

Nell'ambito più specifico della ricerca sull'IA sono presenti tre principali orientamenti: un'impostazione logicista, una non logicista con un approccio simbolico e una non logicista che fa uso di approcci connessionisti e neurocomputazionali. L'impostazione logicistica dell'IA, nel contesto della creazione di un agente intelligente, è stata così sintetizzata: «un agente intelligente riceve percezioni dal mondo esterno sotto forma di formule in alcuni sistemi logici e deduce, sulla base di queste percezioni e della sua base di conoscenze, quali azioni dovrebbero essere eseguite per garantire gli obiettivi dell'agente». Ma, se si prova ad applicare il modello di tipo logicistico non solo a set limitati di dati e decisioni, ma ad ampi settori di attività cognitiva, la complessità di calcolo diventa tale da rendere difficile da capire «come potrebbe un tale modo di ragionare applicarsi a compiti come quelli che un falco affronta quando piomba verso il basso per catturare una preda in corsa (*scurrying prey*)».

All'interno dell'approccio non logicista neurocomputazionale si collocano i ricercatori che utilizzano i network neurali artificiali (*artificial neu-*

ral networks). L'aspetto interessante è che questi diversi orientamenti, per quanto presenti sin dalle origini della moderna IA e pur partendo da presupposti e impostazioni diversi l'uno dall'altro, sono sempre più spesso usati congiuntamente, in modo da essere interoperabili. È acquisito il concetto che, se si vogliono costruire dispositivi funzionanti in modo efficiente e conveniente, bisogna usare tutti gli approcci e le tecniche disponibili. Per esempio, è questo il tipo di impostazione seguito con l'architettura di DeepQA di Watson (IBM) e da Google DeepMind AlphaGo.

Quello che mi pare sia ancora più interessante è che questo approccio con più sistemi (concettualmente eterogenei) è alla base della recente ripresa della ricerca sull'IA di livello umano (quella *generale* per intenderci), cui si è fatto cenno sopra. A questo punto sorge una domanda per noi giuristi: una combinazione dei tre approcci disomogenei può essere usata anche nel diritto? Si può adempiere all'obbligo di motivazione se si usano i *neural networks*? L'IA può organizzare anche attività che non appaiano immediatamente spiegabili in termini razionali? L'approccio logicista tende a escluderlo, perché non è spiegabile, così come sembra escluderlo tutto quel movimento che attualmente punta sulla spiegabilità delle scelte fatte o proposte da sistemi di IA. Cosa può dire, allora, a noi giuristi il modo di lavorare dei *computer scientist* sull'IA?

A me pare che questo modo di essere dell'intelligenza umana, aperto, plurale, composito, sempre in cambiamento sia una sfida non tanto alla definizione dell'IA, quanto ai tradizionali settori disciplinari che non mostrano una simile plasticità di strutture e apertura mentale. Per il diritto, vuol dire aprire il cantiere di una concezione molecolare, che sia relativista, plu-

ralista e policentrica, sempre mutevole dei fenomeni giuridici?

## 2. L'impatto dell'AI sulla società

### Roberto Battiston

Professore ordinario di Fisica sperimentale, Università degli studi di Trento. Mail: [roberto.battiston@unitn.it](mailto:roberto.battiston@unitn.it)

Nel 1942 un professore di chimica americano, figlio di una famiglia di mugnai ebrei immigrati dalla Russia, scrisse in un racconto di fantascienza dal titolo *Runaround*, che potremmo tradurre con "girare a vuoto" o "circolo vizioso" in cui introduceva per la prima volta le Tre Leggi della Robotica:

**Prima Legge:** Un robot non può ferire un essere umano o, per inazione, permettere ad un essere umano di mettersi in pericolo.

**Seconda Legge:** Un robot deve obbedire agli ordini dati da esseri umani tranne quando questi ordini confliggano con la Prima Legge

**Terza Legge:** Un robot deve proteggere la propria esistenza nella misura in cui questa protezione non confligga con la Prima o con la Seconda legge.

Alla fine del racconto l'autore, un certo Isaac Asimov poi diventato leggendario come scrittore di fantascienza e pensatore contemporaneo, aggiunse un'altra legge ancora più elementare:

**Legge 0:** Un robot non può danneggiare l'umanità o, per inazione, permettere che l'umanità si metta in pericolo.

Queste leggi hanno un particolare fascino ed efficacia fondativa, legata alla loro semplicità.

Esse affermano con chiarezza quale sia il posto dei robot (che qui assumo essere la logica estensione del concetto di AI) nel mondo degli umani: si tratta di un ruolo subordinato, caratterizzato da un'etica al servizio dell'uomo, di tutti gli uomini, dell'umanità. Un'altra caratteristica molto interessante è il grado di ricorsività delle leggi: sono le leggi primarie, la Legge 0 e la

Prima Legge, a determinare gli effetti delle altre due leggi.

A quel tempo la proposta di Asimov, raccontata con una tecnica di scrittura che ricorda il *Dialogo sopra i Due Massimi Sistemi del Mondo* di Galileo, stimolò la riflessione alcuni scienziati, tra cui Marvin Minsky, John McCarthy, Herbert Simon, Allen Newell, che avrebbe portato negli anni '50 alla nascita dell'Intelligenza Artificiale (AI).

L'entusiasmo per i risultati iniziali, ottenuti con programmi che vincevano contro l'uomo al gioco della dama o si esprimevano ragionevolmente in inglese, scemò e si riprese almeno due volte nei decenni successivi: alterne vicende prodottesi a causa di un misto di ragioni tecnologiche (potenza e costo dei sistemi di calcolo) e teoriche (sviluppo di algoritmi in grado di autoapprendere).

Sta di fatto che poco più di mezzo secolo dopo la nascita dell'AI, assistiamo oggi ad una crescita sbalorditiva del potenziale di queste tecniche. I traguardi dei sistemi artificiali "intelligenti" consistono, comprensibilmente, nel raggiungere e battere le migliori capacità umane in un dato settore. Simbolica è la crescita di potenza nei settori dei giochi di strategia ad informazione completa come la dama (1959), gli scacchi (1996), il go (2017), fino ai recenti risultati in giochi ad informazione incompleta come il poker (2017). Si tratta dei risultati maggiormente visibili a fronte della diffusione sempre più capillare delle tecniche AI nei settori più disparati. A fronte di questa crescita esponenziale dell'AI e della rivoluzione tecnologica verso cui siamo diretti, non c'è stata un'analogha intensità del dibattito sullo sviluppo del pensiero fondativo relativo agli aspetti etici e giuridici di questo tipo di tecnologie, alle loro conseguenze sociali ed individuali.



Ci troviamo oggi in una condizione di anarchia, un po' come accade con i fenomeni devianti presenti sul web e sui social: fake news, haters, trolls, influenza mirata, diffamazione etc. Nel caso del web e dei social, però, la controparte consiste in uomini e donne aventi interessi diversi, ma che in linea di principio devono sottostare alle stesse leggi ed essere giudicati sulla base di principi morali più o meno condivisi. Interventi specifici, come ad esempio l'obbligo di registrazione delle proprie generalità prima di potere partecipare ai dibattiti in rete, possono riportare in un alveo di normalità comportamenti intollerabili che sfruttano l'anonimato a fini diffamatori (ben comprendendo l'importanza dell'anonimato per evitare che regimi autoritari e totalitari possano esercitare ulteriori controlli sulla libertà delle persone).

Nel caso dell'AI la controparte è invece "qualcosa" di artificiale che però punta a comportarsi "come", e sempre più spesso "meglio", dell'uomo. Da un punto di vista etico e giuridico si tratta di una vera novità, un terreno in cui nuovi principi e fondamenti devono essere identificati e condivisi, come potrebbe accadere nel caso della scoperta di altre forme di vita "intelligente" nell'universo.

La velocità con cui l'AI si diffonde e si potenzia, spinta da interessi commerciali e militari, rende questa riflessione assolutamente urgente. Nel 2011, due comitati inglesi, il Consiglio della Ricerca in Ingegneria e Scienze Fisiche (EPSRC) ed il Consiglio delle Arti e delle Scienze Umane (AHRC), introdussero cinque nuovi principi etico-giuridici relativi alla robotica. Li elenchiamo, non tanto per confrontarli alle leggi di Asimov, bensì per sostenere l'urgenza di una definizione che permetta di individuare limiti e potenzialità relativamente alle capacità della robotica intelligente, in un contesto normativo che supera i

confini fisici-territoriali e pone la necessità di una legislazione globale:

- 1) I robot non devono essere progettati solamente o primariamente per uccidere o danneggiare esseri umani.
- 2) Gli uomini, non i robot, sono agenti responsabili. I robot sono strumenti disegnati per raggiungere obiettivi umani.
- 3) I robot devono essere disegnati per assicurare la loro protezione e sicurezza.
- 4) I robot sono dei manufatti: non devono essere progettati per sfruttare le debolezze degli utilizzatori evocando risposte emotive o creando dipendenza. Deve sempre essere possibile distinguere un robot da un essere umano.
- 5) Deve sempre essere possibile identificare chi è giuridicamente responsabile per un robot.

Se pensiamo alla situazione attuale (2020) vengono in mente molti esempi in cui questi principi sono palesemente violati, sottolineando ancora una volta l'estensione del problema e l'urgenza di una riflessione. Droni intelligenti progettati per uccidere gli avversari, impiegati con regolarità nei teatri bellici, violano come minimo il primo ed il terzo principio. Assistenti robotici online e nei centralini telefonici vengono migliorati senza sosta per imitare i comportamenti umani, violando il quarto principio. L'identificazione dei volti applicata in modo massiccio e connessa con la rete viola il quinto principio, e così via.

Ogni meta raggiunta dall'AI rappresenta un miglioramento rispetto alle corrispondenti capacità umane. Considerato quanto sia forte la spinta competitiva verso il potenziamento di queste tecniche nei settori della ricerca, della finanza, del mercato nonché in quello militare, assistiamo quindi ad un attacco concentrico dell'umanità verso quello che caratterizza mag-

giormente l'essere umano: la sua intelligenza, originalità, creatività, curiosità.

È difficile dire quale sarà l'esito di questo attacco, ma alcuni dei protagonisti non nascondono la propria determinazione a puntare al superamento della specie di cui fanno parte, la nostra specie. È come se osservassimo un abile boscaiolo seduto su un ramo e forsennatamente intento a tagliare il ramo dalla parte sbagliata.

È questo il motivo principale, secondo me, per cui una profonda riflessione sui principi etici e giuridici dell'AI è necessaria ed urgente: la prateria delle innumerevoli possibilità che si aprono di fronte a noi contiene alcune direzioni che, in prospettiva, vedono l'uomo al centro e altre che lo vedono invece in periferia, se non addirittura scomparso dalla scena.

È necessario fare delle scelte fondanti e drastiche, consci della posta in gioco: rimanere gli attori principali della nostra storia o passare ad "altri" questo ruolo, preparandoci ad una uscita di scena. Non sappiamo fra quanto tempo questo possa accadere, ma dovremmo essere allertati dalla velocità esponenziale con cui si sviluppano le tecnologie, in particolare quelle informatiche. Consci del fatto che la maggior parte degli addetti ai lavori non riesce ad analizzare il problema etico-giuridico nella sua complessità e ha bisogno di linee guida.

Temi particolarmente importanti di questa riflessione potrebbero riguardare, ad esempio:

- le applicazioni dell'AI che potenziano le attività umane rispetto a quelle che tendono a sostituire l'uomo;
- l'identificazione e la trattazione di diverse tipologie di AI capaci di:
- classificare e trattare dati o documenti (repertori, dizionari etc.);
- comunicare e diffondere informazioni (reti, social, transazioni online, media online etc.);

- collegare aspetti cognitivi con aspetti operativi su precise indicazioni umane;
- sostituire l'attività umana secondo una precisa programmazione (algoritmi);
- autoapprendere con intervento sugli aspetti interpretativi fino a raggiungere ed includere anche la definizione dei fini.

Mai come oggi abbiamo bisogno di un nuovo umanesimo in grado di dialogare autorevolmente dei risultati e dei prodotti del nostro ingegno e della nostra curiosità. A fronte di prospettive straordinarie abbiamo a che fare con altrettanto straordinarie sfide che richiedono una nuova definizione delle basi della nostra convivenza, non solo con gli altri esseri umani ma anche con le nuove macchine intelligenti della nostra civiltà prossima futura.

**Malik Ghallab**

*Research Director at LAAS-CNRS, University di Toulouse. Mail: [malik.ghallab@laas.fr](mailto:malik.ghallab@laas.fr)*

Artificial Intelligence and computational technologies are significantly affecting humanity. They correspond to a revolution as disruptive as the deployment of writing and printing, which have significantly extended human cognition and interaction capabilities. Writing and printing have deeply changed laws, cities, economy, and science; they have affected human values, beliefs, and religions. We are witnessing a similar but much faster revolution, in which every person today is a contributing actor. There are different levels of responsibility, but everyone needs to be a well informed, responsible actor.

We already observe the positive effects of AI in almost every field, from agriculture, industry, and services, to social interaction, knowledge dissemination, health, and sciences, as well as in the arts and culture. We foresee AI potential to help addressing our sustainable development goals and the urgent challenges for the preservation of the environment.

However, it is important to stress that no technology is without risks, and that its social acceptability is not equivalent to its market acceptance. The social acceptability of a technology is much more demanding and requires, among other things, (i) to take into account the long term, including possible impacts on future generations; (ii) to worry about social cohesion, in terms of employment, resource sharing, inclusion and social recognition; (iii) to integrate the imperatives of human rights, as well as the historical, social, cultural and ethical values of a community; and (iv) to consider global constraints affecting the environment or international relations.

The AI risks related to the safety, security, confidentiality, and fairness of AI systems are frequently exposed. The threats of manipulative systems to human free will are raising legitimate concerns. The impacts of AI on the economy, employment, human rights, equality, diversity, inclusion, and social cohesion need to be better addressed.

Decision support tools can be biased. Some systems are intentionally designed as unbalanced, e.g., for a recommender system integrating commercial or propaganda. Users should be explicitly warned about the underlying objectives of systems that may distort their recommendations. The hidden and non-intentional biases of systems required to be neutral and fair are more problematic. Numerous cases of gender, ethnical or seniority biases have been reported in decision support systems for health, banking, insurance, recruitment, career assessment, or even in public services such as legal assessment and city surveillance applications. This is generally the case because these systems lack transparency, intelligibility and rely on training data which is biased in hidden ways difficult to uncover and mitigate. There is a need for further research in techniques for auditing the fairness of a system, and in regulations requiring their use for certification mechanisms.

It has been known for ages that individuals can be manipulated. AI technologies augment their vulnerability, in particular with the large deployment of playful devices that implement powerful communication, sensing, processing and decision-making functions. Manipulation capabilities are illustrated by the increasingly more effective techniques for social monitoring, debate steering, behavior modeling and shaping, and market driving. The incentives for using available techniques toward profitable purposes are high. Dubious practices with social, political

and economic risks will remain in use as long as they are unregulated. To support regulations, further research in AI may contribute to methods for detecting manipulation attempts.

The political risks, illustrated by the Cambridge Analytica scandal, are analyzed by several authors as a serious threat to democracy. Studies show that AI presents opportunities as well as risks on the full range of human rights, with already observed impacts.

Economic risks correspond to several AI uses, e.g., High Frequency Trading (HFT) or algorithmic pricing. The possible destabilization effects of HFTs are far from being well understood. Algorithmic pricing uses learning, profit optimization and indirect interactions between computational agents; it can lead, even without explicit agreement, to artificially higher prices, as with the illegal price cartel mechanisms. The real time observation, learning, modeling and feedback control capabilities permitted with AI tools are in clear contradiction with the assumption of a supposedly neutral free market. Regulations to mitigate the corresponding risks are urgently needed.

AI contributes to the increasing automation of services, industry and agriculture, which brings progress, as well as important risks for employment. Available studies, which remain insufficient, converge toward a substantial reduction of jobs in the short to medium term. According to an OECD study for its 21 countries, 9% of jobs have a high risk of automation; a higher percentage of 20 to 25% of jobs have a medium risk (other studies conclude to more alarming risk levels). Furthermore, AI developments are strongly suspected to contribute to the observed increase in social inequalities, which reduce social involvement.

AI in weapons and military systems correspond to another area of worry, which raises ethical

concerns, as well as risks of international instability and conflicts. AI technologies enhance the military capabilities of perception, surveillance, intelligence, fighting, and intervention. AI is naturally a dual-use technology, easily transposed from the commercial to the military domain. This makes impracticable control procedures such as those used for nuclear weapons containment. This also makes weapons and devices with integrated AI relatively more “affordable” than other heavy military technologies. These weapons may be more easily accessible to rogue groups. In addition, international arm trade agreements, including the recent Arms Trade Treaty, do not cover digital weapons, such as drones, robots, ROV and AUV. The widely supported Open Letter for a ban on autonomous weapons is an excellent initiative which needs to be pursued into studies and regulations.

With respect to the above risks, some consider implementing moral appraisal capabilities in AI. We certainly need machines that are provably safe, secure, intelligible, unbiased, respectful of privacy, and meeting the constraints and rules demanded by society. These properties can be reasonably well understood, formalized, and machine implementable. However, it is unclear what might be the specification of an automated weapon, or an automated trader, capable of resolving ethical choices on the basis of moral principles. Approaches to the notion of an “artificial moral agent” are criticized as being philosophically illegitimate and technically misleading. Instead, developers should clarify and disseminate widely the knowledge about the capabilities and limitations of their tools. They should integrate the social involvement and assessment of their potential uses as an essential component in the research and design methodology.

The needs for responsible AI developments with respect to the social risks correspond in particular to political and legal measures and to international agreements.

However, the required measures are part of the regulatory mechanisms of society. These mechanisms have a quite long response time: decades are needed to better understand, educate, spread the awareness and build up the social forces required to impose regulations. But the momentum of technology is much faster. The discrepancy between the two dynamics demands for proactive approaches. However, no predictive model of the possible effects of AI readily available. A proactive approach must rely on social experiments, and on integrative research about social risks and mitigation measures. A change of paradigm is required to fund and develop joint investigations between AI and social scientists, to give a better understanding of AI to the former, and of social and economic mechanisms to the latter. It should be noted that simulation, based on elementary models, is emerging in a few areas of social sciences. AI can actively contribute to its development and effectiveness. Finally, let us remark that social experiments prior to some deployments reduce the discrepancy between the technology momentum and the social dynamics.

**Paolo Traverso**

Direttore del Centro per le Tecnologie dell'Informazione e della Comunicazione (ICT), Fondazione Bruno Kessler (FBK). Mail: [traverso@fbk.eu](mailto:traverso@fbk.eu)

Negli anni 50, quando è nata, al seminario presso il Dartmouth College, i padri dell'Intelligenza Artificiale (AI), guidati da John McCarthy, si diedero come obiettivo ... un sogno: creare una macchina in grado di simulare il ragionamento, l'apprendimento, il pensare proprio della mente umana e l'agire proprio delle persone. Si trattava definitivamente di una "AI forte", di una "Strong AI".

Ma il grande "boom dell'AI" è ben più recente, dovuto principalmente alle moderne tecniche di apprendimento automatico, quali il cosiddetto "Apprendimento Profondo", l'ormai famoso "Deep Learning" - in realtà una rete neurale dotata di parecchi "strati". Queste tecniche, utilizzando l'accresciuto potere computazionale e la disponibilità di enormi quantità di dati, opportunamente addestrate con tantissimi esempi (il famoso "training"), sono riuscite a fare cose che un tempo erano appannaggio solo dell'essere umano, come capire le immagini e le parole, creare autonomamente nuove immagini e testi, generare film e scrivere libri, comporre musica e dipingere quadri, costruire robot in grado di interagire con le persone e a volte comportarsi come le persone. E in certi compiti specifici, hanno dimostrato di replicare e addirittura superare le capacità umane, anche in azioni importanti e delicate. Persino i medici riconoscono ormai che gli algoritmi di AI sono migliori di loro nell'analizzare le radiografie, nel capire da una immagine di una retina se si tratta o meno di una retinopatia. Ma sono comunque compiti specifici, e non è affatto chiaro se, anche quando replicano o superano le capacità umane, questa AI si avvicini al sogno dei suoi padri: pensare come la mente umana e agire come le per-

sone. Si tratta ancora, nonostante i risultati stupefacenti, di una AI debole, una "Weak AI".

E tutto questo sembrerebbe rispondere alla domanda che ci viene posta con un secco "no", non arriveremo ad una "Strong AI", al sogno dei padri. Ma io non credo che sia così!

Non è così perché ... "non è finita". La maggior parte dei grandi player industriali (e sapete a chi mi riferisco) stanno investendo sempre di più le loro enormi risorse spingendo verso il limite massimo le tecnologie che al momento sono di grande successo nello svolgimento di compiti specifici (la cosiddetta "Weak AI"), ovvero tecnologie basate, come dicevo, sull'addestramento di algoritmi di apprendimento automatico con enormi quantità di dati disponibili. Ma nei centri di ricerca più prestigiosi (inclusi alcuni in cui si è inventato il "Deep Learning") si "guarda oltre". In questi centri si studiano, ad esempio, tecniche di "zero shot learning", ovvero di apprendimento a partire da pochi dati, di "imitation learning", ovvero algoritmi che imparano imitando semplici istruzioni ed esempi forniti dalle persone e cercano poi di generalizzare i loro comportamenti ad altri casi. Si studiano tecniche di "behavioural learning", cioè tecniche che imparano dal comportamento umano. Anche l'ormai vecchia idea del "deep learning" sta lasciando spazio in certi casi al "deep symbolic reinforcement learning", ovvero sistemi che imparano agendo nel mondo e imparando dai propri errori. Si studiano "modelli di apprendimento generativi e non discriminativi", in cui si possono anche generare esempi invece che solamente imparare da essi. Si studiano tecniche che integrano "planning and learning", in cui tecniche di apprendimento automatico vengono unite a tecniche di pianificazione (di azioni) che possono analizzare innumerevoli possibili scenari futuri.

Ma soprattutto, si parla di *"Integrative AI"*, ovvero di "mettere nel computer" diverse tecniche di "rappresentazione" e "ragionamento" per far sì che il computer sia in grado di svolgere un insieme di diversi "task cognitivi" (!). Ad esempio, trovare il modo di combinare rappresentazioni *"simboliche"* con rappresentazioni *"sub simboliche"*. Le prime sono ad esempio le ontologie, le quali rappresentano esplicitamente concetti e conoscenza sul mondo, o la logica che cerca di formalizzare tale conoscenza. Le seconde sono ad esempio le informazioni numeriche, quali quelle che arrivano dai sensori di un robot, o le rappresentazioni probabilistiche o quelle che sono frutto della statistica. Si studiano tecniche di ragionamento che uniscano l'apprendimento dai dati, dagli esempi e dall'esperienza con tecniche di ragionamento deduttivo, che a partire da fatti inferiscano e deducano conclusioni.

Se questi studi avranno successo, si potrà risolvere ad esempio l'ormai famoso problema delle attuali tecniche di *"deep learning"* (e delle reti neurali in generale), che non sanno distinguere fra relazione e causalità, uno dei principali fattori alla base della *"non spiegabilità"* dei risultati delle reti neurali, che ancora appaiono come scatole nere.

Si tende quindi (giustamente, perché la ricerca deve farlo) a spingersi veramente vicino al sogno dei padri dell'AI, ovvero di poter costruire macchine in grado di simulare il ragionamento, l'apprendimento, il pensare proprio della mente umana e l'agire proprio delle persone, ovvero a spingerci verso una AI forte (*"strong AI"*), non solo in grado di risolvere compiti specifici a partire da un addestramento con enormi quantità di dati, addestramento che spesso è innaturale, faticoso, a volte impossibile e in un certo senso ... *"inumano"*.

Questa lunga premessa per rispondere alla domanda "ritenete possibile lo sviluppo di una strong AI, capace di replicare, e addirittura superare, le abilità umane in ogni settore del pensare e dell'agire?". Rispondere in modo ideologico a questa domanda dicendo che le macchine "non saranno mai capaci di fare certe cose", credo possa essere pericoloso: qualche anno fa si diceva che le macchine non sarebbero mai state capaci di fare cose che adesso fanno e in questo hanno superato le abilità umane. Qualche anno fa nessuno l'avrebbe detto. Io credo che l'AI del futuro ci sorprenderà ancora, se il mondo avrà ancora il coraggio di finanziare la ricerca in AI dopo che ne avrà capito i limiti attuali.

La domanda che viene dopo è però ancor più interessante: "E tale sviluppo sarebbe opportuno?". Io credo che l'AI, come ogni tecnologia potente, vada regolamentata perché sia una AI per le persone (una "human centric AI"). E questo vuol dire anche essere coscienti e evitare che si avverino i rischi che certe scoperte scientifiche portano, e in particolare i rischi di possibili utilizzi dell'AI. Rischi di pregiudizi nelle decisioni e nei giudizi (i famosi "bias"), rischi di affidarsi a macchine che fanno errori, rischi addirittura di andare contro diritti vitali per l'umanità, quali i diritti sanciti dalla Costituzione. Rischi per la democrazia e la sovranità popolare derivati dalle tecniche (quali quelle che si usano con le reti sociali ma non solo) che alcuni big player industriali usano per avere i nostri dati, per profilarci, per capire i nostri gusti, influenzarci (addirittura nel voto come Cambridge Analytica insegna); i rischi di un uso di sistemi che sostituiscano del tutto le persone nell'interazione con le persone, togliendo il diritto ad un persona di avere una spiegazione da un proprio simile e non da una macchina (come nei casi delicatissimi della medicina e della giustizia); il rischio

che venga violato il diritto delle persone a conoscere la natura, umana o artificiale, del proprio interlocutore (il successo pieno della macchina sul vecchio Test di Turing); o il pericolo che venga meno il diritto delle persone alla riservatezza, alla privacy, il rischio per antonomasia in questo settore.

Credo che combattere e evitare che questi rischi diventino realtà dipenda fortemente da noi tutti, ricercatori, imprenditori, giuristi, da chi lavora nelle istituzioni, dai policy makers etc. Dobbiamo tutti lavorare assieme per impedire che questo accada. Ma dobbiamo evitare anche il rischio opposto. Fra tutti gli interventi alla conferenza sullo Statuto Etico e Giuridico dell'Intelligenza Artificiale, organizzato dalla Fondazione Leonardo "La civiltà delle Macchine", mi ha colpito particolarmente un intervento di una giovane ricercatrice di Oxford. Il suo intervento, in sintesi e come lo ricordo diceva: *"la privacy è chiaramente un diritto delle persone e va tutelata. Ma cosa diremo ai nostri figli? Non abbiamo guarito il diabete perché altrimenti violavamo la privacy dei cittadini?"* E l'esempio della giovane ricercatrice si può facilmente estendere un po' a tutti i rischi e ai diversi aspetti della vita sociale dove l'AI sta avendo un enorme impatto e, io credo, lo avrà ancor di più nel futuro. Il confine fra "regolare" e "ingessare" è sottile, e trovare il giusto equilibrio è probabilmente il nostro compito più difficile.

Mi piace finire con una considerazione che ho imparato da una osservazione che ha fatto un mio caro amico e stimato professionista ad una mia presentazione divulgativa sull'AI e l'impatto che può avere sulla società. La mia presentazione si concludeva con una frase del "quasi premio Nobel" Stephen Hawking: *"L'Intelligenza Artificiale sarà la cosa migliore o peggiore mai successa all'umanità"*. Il mio amico mi ha fatto

notare che probabilmente avrei dovuto sostituire la congiunzione "o" con la congiunzione "e". Sicuramente ci sarà il bene e il male provocato dall'AI, ci saranno entrambi. Allora dico: *Facciamo di tutto per far pendere la bilancia verso il bene, per quanto difficile sia.*

## 2.1 Essere umano e macchina

### Maria Chiara Carrozza

*Istituto di Biorobotica, Scuola Superiore Sant'Anna di Pisa, Dipartimento di Eccellenza in Robotica e Intelligenza Artificiale, Scuola Superiore Sant'Anna di Pisa e Fondazione Don Carlo Gnocchi Onlus. Mail: [m.c.carrozza@santannapisa.it](mailto:m.c.carrozza@santannapisa.it)*

**2.1.a** L'intelligenza artificiale è una delle tecnologie abilitanti che caratterizzano la quarta rivoluzione industriale, ma la sua influenza andrà ben oltre il mondo della produzione di beni e della offerta di servizi, essa ha dimostrato di possedere una ricaduta sociale e culturale dirompente generata principalmente dal carattere di pervasività con cui è entrata nelle vite di tutti noi ponendo la natura umana e le abilità ad essa connesse al centro. A differenza delle precedenti rivoluzioni tecnologiche, non riguarda solo la forza, il lavoro ma la stessa cognizione: la capacità tipicamente ed esclusivamente umana di pensare e di compiere delle scelte. Per cui sì, certamente l'AI sarà e in parte è in grado di replicare alcuni comportamenti umani, tuttavia non posso immaginare che questa tecnologia, per quanto estremamente potente, possa sostituire l'uomo in ogni settore del pensare e dell'agire. È vero, oggi i sistemi di AI sono capaci di adattarsi e adeguarsi alle mutevoli condizioni in cui operano, simulando ciò che farebbe una persona. Si parla, a questo proposito di AI "bioispirata" poiché in essa vengono riportati schemi tipici del ragionamento umano ma l'AI, per poter funzionare, ha comunque bisogno di dati, che alimenteranno degli algoritmi, frutto di una pre-impostazione/imposizione di un essere umano.

**2.1.b** Stiamo assistendo ad un mutamento significativo caratterizzato dalla simbiosi uomo macchina, un processo di forte integrazione nel quale emergono sinergie e differenze: ciò che carat-

terizza l'uomo e lo differenzia dalla macchina è la capacità di compiere scelte che siano eticamente orientate e quindi il saper discernere il bene dal male. Quello che manca alla macchina per poter completamente assumere una decisione al posto dell'essere umano è la componente etica che, per poter essere implementata in un sistema di AI, dovrebbe essere tradotta in algoritmo (algor-etica). Questo pone una serie di interrogativi, oggetto oggi di grandi dibattiti nel mondo della scienza, della filosofia, della teologia e del diritto. Finché dunque non saremo in grado di tradurre l'etica in un valore trasferibile dall'uomo alla macchina, la funzione dell'AI non potrà essere sostitutiva di quella umana, ma potrà però certamente semplificare e supportare il comportamento umano, agevolare alcune azioni, soprattutto attraverso la capacità predittiva tipica dell'AI.

## Federico Faggin

*Federico & Elvia Faggin Foundation*

**2.1.a** The recent successes of AI have prompted a renewed enthusiasm for Strong AI, the idea that a properly organized “robot” may develop a higher intelligence than a human being.

The most advanced framework within which this feat is considered possible is to give a computer a “body” endowed with human-like sensory-motor subsystems, and let it learn like a child in interaction with the animate and inanimate world. The expectation is that such a robot may then develop a language and a common sense like a human.

I think this approach, enticing as it may sound, cannot succeed because it disregards entirely the crucial functions that consciousness, emotional intelligence, and intuition accomplish in the developmental process of a child. It also disregards the essential and complementary *emotional* mentoring that the mother and father perform at different stages of the child’s development.

Furthermore, the basic idea that human intelligence is reducible to a finite set of “mathematical functions” that can be simulated by a computer, is also fundamentally flawed. Intelligence is not separable from the emotions and the higher mental functions of a human being, and these require consciousness. Only the mechanical-logical portion of human intelligence may be imitated by a computer simulation, and this is only a small fraction of the real human intelligence. Understanding the nuances of language, emotions, and human situations requires *conscious comprehension*, something digital computers do not and cannot have.

The difference between the two intelligences is the same one that exists between a quantum and a classical system – at the very least – when

we further assume that consciousness is the capacity to “experience” the quantum state of the physical body. However, this is not an assumption currently made by science.

I think that the crucial feature of possessing an inner semantic conscious state and an outer symbolic public state can only exist for a system that is partly quantum and partly classical, just like all living organisms are.

The materialistic and reductionistic stance that informs the idea of strong AI is that living organisms are classical systems like computers. This is not true, however, even though cells and animals have been almost exclusively studied as classical systems (often after killing them). Only recently a small group of scientists have given birth to *quantum biology*, a new discipline that recognizes the crucial role played by quantum physics in the ontology of living organisms. I believe that the mystery of life goes far beyond what we currently know.

By equating intelligence to a purely mechanical function that can be imitated by a sufficiently powerful classical computer, AI has *ipso facto* taken off the table as irrelevant to human intelligence, the consciousness we all have and the “lived experience” that only consciousness can provide. This *reductive* definition of intelligence has also eliminated a priori and by *fiat* the crucial aspects of the emotions and the higher intelligence with which human beings are endowed: intuition, imagination, invention, insights, and comprehension.

Human beings are holistic rather than reductionistic systems, making it impossible to separate the physical, emotional, mental, and spiritual aspects that interplay in the human behavior and inner experience. Furthermore, comparing a man with a computer over the vast range of human abilities is impossible because we couldn’t even make a sensible list of those



abilities, never mind properly defining them – a prerequisite to be able to measure them and say that “robot” is greater than “human.” Robotic intelligence and human intelligence are sort of complementary properties, like the particle and the wave properties of quantum mechanical systems.

The crucial aspects that computers will never possess is the free will and the qualia perception and comprehension that are inherent in conscious beings. Machines find correlations between data sets; humans find ever-deeper comprehensions, i.e. new connections among previous comprehensions. This feat requires a more abstract type of information than Shannon’s information. I claim this information exists only in the quantum domain as a result of the superposition and entanglement properties of quantum systems, properties that are totally absent in classical systems.

Let me give you a simple example of this human capacity: When a person has learned through meditation to quiet his mind and feels a thought enter the “theater” of his mind, he may also recognize the urge to automatically translating that pure “multidimensional image” into mental words. When the person learns to “stay” with the pure thought as an image, “contemplating” without reifying it, he may find that some thoughts simultaneously contain many possibilities in superposition. This feature cannot exist when the pure thought has been translated into mental words because this symbolic reification would work like the “collapse of the wavefunction” of a quantum system; the process that brings into physical reality only one of the many states that exist in superposition in the quantum domain.

The bottom line is that strong AI is impossible, even in principle, with classical computers, no matter how powerful and complex a computer

may be. Whether strong AI may be possible with *quantum* computers as currently conceived, is an open question. I think that living organisms are using powerful “quantum tricks” we have yet to discover. And since living organisms are conscious, it should be in principle possible to create “strong AI” via biology. However, if the ethical issues raised by AI are thorny, the manipulation of living organisms raises the ethical bar one hundred-fold. This subject is outside the scope of this article, nonetheless, I’d like to say that this is a far more disquieting possibility to me than strong AI based on classical computers.

I don’t think that my discussion will convince any believer in strong AI not to “grow” a child-robot in a laboratory. The narrow scientific and business interests of our time will demand the investigation of this new path of inquiry. Nonetheless, I predict that this approach will eventually (30-50 years from now, perhaps more) show the impossibility of developing strong AI, vindicating the common sense of simple men in touch with their intuitive mind *and heart*, who know that there must be more to reality than mechanisms.

**2.1.b** This question is more urgent than the previous one, and it must be answered very well and very soon because AI has reached a level of performance that requires strong ethical and pragmatic criteria to be properly regulated.

Let’s take an example with self-driving cars: A child is suddenly crossing the street in front of an autonomous vehicle (AV) and braking at full power cannot avoid a collision. Inside the AV is a pregnant woman and the situation calls for an immediate decision.

Within a few milliseconds, the AV’s algorithms have properly evaluated the only two possible options: (1) fully braking; in which case the AV will hit the child with 80% probability of killing



him and will cause some bruises to its passenger; or (2) swerving to the left and braking, in which case the AV will avoid the child but hit the approaching SUV on the other side of the street, with 60% probability of killing its own passenger and wounding the SUV driver. Asking for human assistance is out of the question because there is no time to do so: the decision must be algorithmic.

In fact, if a human driver found himself in a similar situation, he too would have to *react* the best way he can, hopefully without panicking. Notice that I said “react” rather than “decide” because a rational decision would require the ability to identify and evaluate the possible alternatives and then choose the best course of action. But in this case, there is no time for a typical human to rationally assess the situation. The brain is much too slow to make a conscious decision and the unconscious reaction is unpredictable. Here the machine has a definite speed advantage over the human.

What can be done? If we want autonomous driving – and there are definitely major benefits to it – a government appointed body with representatives of all stakeholders (insurance companies, legal experts, AV manufacturers, consumers, etc.) need to evaluate this class of situations and provide *deterministic* rules that will be implemented by the algorithms of the AV. If this task is properly done, given the AV’s ability to rapidly evaluate the possible options and its short reaction time, the AV’s decisions may likely surpass, on average, the human reactions.

Another important aspect of autonomous driving is human access to the full record of the accident, including the AV’s “reasoning” that went into the decision. These data can help to more fairly settle the legal responsibilities, as well as improving the algorithms for future use. I think that on average, stakeholders-regulated algo-

rithmic decisions may fare better than human unpredictable reactions in these kinds of situations.

In many other cases, especially in medical or military decisions, the temptation to abdicate decisions to AI systems for the same reasons discussed earlier, i.e. the lack of time for responsible human intervention, is very problematic to me. There is also a dark cloud here, which is the opportunity to dodge responsibility by claiming that “the machine did it,” when in fact the humans made the machine do it by the way the machine was designed and controlled.

I think that there should always be a person, or a human organization, totally responsible for the actions of the machine. No machine should be held responsible for its own actions because AI has no consciousness, no comprehension, and no ethics. It simply does what it is “coerced” to do. A dog is certainly more responsible (and thus accountable) than a robot.

The bottom line is that, as in most complex situations, there isn’t a magic black or white answer. Before allowing a robot to autonomously make decisions in situations that involve the possible loss of human life or great damage to property or to the ecosystem, the decisional tree should be evaluated and *transparently* vetted by a panel of experts and stakeholders, like in the example of a self-driving car.

A possible criterion to determine if a robot should be allowed to autonomously decide is to prove that the robot will *always* cause less harm than a decision (or reaction) made by a competent person. Decisions that involve subtle ethical standards that haven’t yet been settled must always be supervised by qualified human organizations or persons who are strictly held responsible.

**Oliviero Stock**

*Istituto per la Ricerca Scientifica e Tecnologica (IRST),  
Fondazione Bruno Kessler (FBK). Mail: [stock@fbk.eu](mailto:stock@fbk.eu)*

**2.1.a** Si parla molto di Intelligenza Artificiale (AI), spesso evocando scenari non realistici, alternando aspettative mirabolanti, o di pericolo esistenziale per l'uomo. Molto frequentemente viene poi assimilata l'AI a un approccio particolare, in questi tempi a tecniche statistiche basate su apprendimento, o addirittura più specificatamente a una tecnica di buon successo chiamata apprendimento profondo. Non aiuta in tutto questo la stessa sensibilità dei ricercatori alle mode, e in questa fase, per molti, magari nuovi nel campo, anche il voler trarre vantaggio immediato dalla popolarità del termine AI. Prima di tutto non è univoca la definizione del campo Intelligenza Artificiale: spesso dico che se chiediamo a cento ricercatori di definirlo probabilmente otterremo almeno cento definizioni diverse. La mia favorita ad esempio – la dico subito – è “far fare alle macchine lo sforzo di capire noi, anziché viceversa”. Nei fatti poi vi è stata tradizionalmente una segmentazione in settori di ricerca sostanzialmente indipendenti, come oggetto della ricerca e – per molti casi – come metodologie adottate. È del tutto marginale nella ricerca in AI una visione complessiva o integrata dell'intelligenza. Quest'ultima però è quella che spesso viene evocata nel mondo esterno, soprattutto nei film di fantascienza o quando si tratta di enfatizzare i pericoli che l'AI potrebbe comportare. Bisogna dire tuttavia che a questo hanno anche contribuito la stessa ambizione dell'AI di avere un test di raggiungimento di intelligenza non distinguibile da quella umana (quello più conosciuto anche se a sua volta controverso e per molti aspetti inadeguato, è il test di Turing) e, se vogliamo, una spinta recente di ricerca tecnologica orientata a robot

antropomorfi, che necessariamente coinvolge più aspetti dell'intelligenza, puntando a integrarne alcuni in qualche modo.

La dicitura “strong AI” poi, almeno per molto tempo, richiamava l'obiettivo non solo di realizzare compiti che se svolti dall'uomo richiedono capacità che comunemente vengono intese denotare intelligenza, ma di farlo con modalità simili a quelle umane. Quindi non solo il risultato, ma anche il come, se non altro per garantire flessibilità e adattabilità del sistema. Lavorando a realizzare macchine che tendono a essere ispirate dal funzionamento della mente umana, si ottengono due risultati: si costruiscono appunto sistemi flessibili e che possono integrare vari aspetti dell'intelligenza – come vediamo realizzato nella mente – e si entra in una dinamica fruttuosa con le scienze cognitive, contribuendo anche agli studi sulla comprensione del funzionamento della mente.

Ora la prospettiva metodologica prevalente si è molto allontanata da questa ambizione.

Certamente vi sono stati sviluppi importanti nell'ultimo paio di decenni. Sono successe alcune cose determinanti.

Innanzitutto, per alcuni settori di ricerca si sono rese disponibili grandissime quantità di dati, grazie anche alle opportunità offerte dal web. Ad esempio, per quanto riguarda il linguaggio umano, quantitativamente quasi tutto quello che è stato scritto dall'uomo è disponibile in forma digitale.

In secondo luogo, il continuo avanzamento della tecnologia degli elaboratori ha consentito di rendere pratici metodi di calcolo basati su grandi quantità di dati. A questo si aggiunge un affinamento di algoritmi per l'elaborazione di grandi quantità di dati, già studiati da tempo, senza che esso dovesse richiedere un grande sviluppo matematico.



A mio modo di vedere, l'approccio che si sta seguendo in questi anni consegue risultati notevolissimi nel breve periodo, ma per molti aspetti ci allontana dall'ambizione di realizzare una "strong AI". Intendo proprio che il risultato, nell'ottica di questa grande ambizione, non è nullo, ma è sostanzialmente negativo. Le ragioni sono proprio la mancanza dei due elementi che citavo prima. Quindi ritengo non vi possa essere una "strong AI" per molto tempo e per il progresso sarà necessaria almeno un'integrazione tra approcci basati su conoscenza e approcci basati esclusivamente su dati.

Un aspetto strettamente legato a questa perplessità riguarda in particolare la capacità di spiegazione del proprio comportamento da parte del sistema. Restiamo pure nell'ambito di compiti molto circoscritti, quindi lontani dalla riproduzione di capacità intelligenti complessive e integrate in un unico sistema. Se ci interessiamo a compiti che normalmente richiedono qualche forma cognitiva superiore - andando oltre d esempio a quelli legati alla semplice percezione e al riconoscimento - risulta difficile accettare sistemi che non spieghino le linee di ragionamento in termini affini a quelle che seguiremmo noi umani. Risulta piuttosto forzato e quasi velleitario l'impegno attualmente diffuso tra i ricercatori nel voler trovare una maniera di spiegare a posteriori un comportamento della macchina che arriva a conclusioni accettabili sì ma per vie non comprensibili.

**2.1.b** Per me, paradossalmente, la prospettiva è opposta a quella spesso evocata. I sistemi artificiali già ora o in tempi vicini sono in grado di prendere decisioni autonome - in certi ambiti circoscritti. È molto più ambizioso e difficile l'obiettivo di realizzare sistemi che siano davvero in grado di prendere decisioni in modo collaborativo con l'uomo; quindi sistemi che operino non in modo totalmente autonomo, ma in gra-

do di mantenere la delega fino ad un certo punto, di spiegare le proprie linee di ragionamento, criticare la decisione umana e accettare critiche, in una discussione dialettica. Questo è quello che noi umani abbiamo imparato a fare quando si tratta di prendere decisioni, a meno che si tratti di ambiti la cui natura e il tipo di decisione richiesta siano estremamente semplici.

Tra i tanti temi necessari per lo sviluppo, in particolare ritengo di grande importanza che l'AI cominci seriamente a occuparsi di etica. Non intendo solamente che si affrontino i problemi etici del campo, similmente ai problemi che i ricercatori di altre discipline, ad esempio biologia, o anche fisica, devono porsi in relazione ai risultati delle loro ricerche o alla sperimentazione adottata. Intendo proprio il dotare i sistemi della capacità di prendere decisioni su basi etiche, a fronte di situazioni nuove, un tema che è solo concepibile per l'intelligenza artificiale. Il tema naturalmente si aggancia a una lunga tradizione di studi filosofici, basta pensare a Kant, ma in AI si tende ad affrontarlo anche in collegamento a studi cognitivi, di solito a carattere empirico. Anche in questo caso ritengo ci possano essere ambiti circoscritti e semplici in cui decisioni autonome potranno essere prese dalle macchine, ma gli ambiti più complessi dovranno coinvolgere la capacità di collaborare, anche dialetticamente, alla presa di decisione.

Quanto detto precedentemente ha ovvie implicazioni per l'ambito legale e giuridico. Ritengo che solo per casi molto semplici si possa automatizzare il giudizio, e quasi esclusivamente nell'ambito amministrativo. Sarà invece di grandissimo impatto qualitativo oltre che quantitativo, nella sfera civilistica e penale, la prospettiva di sistemi di supporto e critica per il giudizio umano.

Prima di concludere, mi permetto di ampliare l'ambito di questi brevi note riguardo il tema



della prospettiva dell'AI per il Diritto. A mio modo di vedere, un fondamentale apporto dell'AI potrà avvenire a monte, e consistere nel rendere le leggi comprensibili ed adattate al livello culturale del lettore. Sostanzialmente un tema educativo, che potrebbe contribuire a superare la difficoltà insita nel Diritto costituzionale, dove si afferma che l'ignoranza della legge non è una scusante. La riformulazione dell'art. 5 del Codice Penale attesta che "L'ignoranza della legge penale non scusa tranne che si tratti d'ignoranza inevitabile." E la Corte costituzionale ha chiarito che questa ignoranza può essere considerata scusabile solo se "riconducibile ad una oggettiva oscurità (attestata, eventualmente, da persistenti contrasti interpretativi) della norma violata." La giurisprudenza sembra non considerare "inevitabile" l'ignoranza di proibizioni stabilite da lungo tempo, come succede per la maggior parte dei delitti trattati nel codice penale. È differente invece il caso della giustizia civile e amministrativa, dove il diritto spesso risulta meno connesso al senso comune e consolidato storicamente, e le leggi sono più tecniche e tendono a essere variate con maggiore frequenza. Ritengo che il ruolo dell'AI nell'adattare la formulazione e la spiegazione delle leggi a diversi gruppi e livelli culturali possa aiutare a rendere la nostra società migliore e realizzare quanto inteso dalla Costituzione. Anche in questo caso la complessità dei meccanismi necessari – qui di tipo linguistico, ad esempio l'integrazione tra semantica e pragmatica - vanno ben oltre quanto realizzabile con gli attuali paradigmi, ma sono concepibili in un nuovo ciclo di AI più "profonda".

### 3. L'impatto della AI sul diritto e sui diritti.

#### Massimo Luciani

*Professore ordinario di Istituzioni di diritto pubblico, Università degli Studi di Roma – La Sapienza. Mail: [massimo.luciani@uniroma1.it](mailto:massimo.luciani@uniroma1.it)*

**3.a** Il quesito si articola, in realtà, in due interrogativi diversi, che meritano altrettanto diverse risposte.

Quanto all'esigenza di nuove regole, sarebbe addirittura assurdo negarla. Non ci troviamo di fronte, infatti, a innovazioni tecnologiche marginali, ma a vere e proprie ridefinizioni di stili di vita, di posizionamenti sociali, di letture del mondo. Si tratta di fenomeni ancor più radicali dell'avvento - poniamo - dell'elettricità, che (anche a causa della minore rapidità delle trasformazioni tecnologiche) poté essere affrontato mantenendo largamente in vita le regole all'epoca vigenti. Conseguentemente, i principi generali (o le stesse norme di dettaglio, eventualmente da applicare in via estensiva) possono servire solo in una fase transitoria, occorrendo a regime regole nuove.

Discorso diverso quanto alle categorie giuridiche. È, questo, un equivoco diffuso e sempre rinnovato: al mutare dei rapporti sociali dovrebbero mutare - si dice - le categorie di analisi (anche quelle dell'analisi giuridica). Non è così. Le categorie di cui disponiamo, elaborate nel corso di un processo più che bimillenario, sono perfettamente in grado di inquadrare e comprendere i nuovi fenomeni. E non si devono confondere gli adattamenti (sovente infelici: si pensi a "beni comuni") delle categorie tradizionali con la creazione di categorie davvero nuove.

**3.b** Alcuni principi generali appaiono indispensabili (penso a quelli in materia di responsabilità), ma l'estrema varietà del fenomeno suggerisce l'adozione di discipline specifiche per i vari

campi di applicazione dell'AI. Ovviamente, la presenza di principi generali faciliterà la regolazione delle future novità, probabilmente profonde e in rapida successione, che una regolazione analitica sarebbe costretta (sempre in ritardo) a inseguire.

**3.c** Sono contrario alla moltiplicazione inutile dei diritti. Non comprendo perché il diritto all'educazione all'AI dovrebbe essere un diritto specifico e non una semplice porzione del generale diritto all'istruzione e all'educazione, che - per stare al caso italiano - la Costituzione già garantisce. Temo che la moltiplicazione dei diritti serva sovente, più che a rafforzare le tutele, a moltiplicare le monografie utili a ottenere l'abilitazione scientifica nazionale.

**3.d** Ne elenco tre, che a me paiono le principali:

- a) la penuria di giuristi con adeguata formazione scientifica, specie in questo dominio (il che può lasciare "in balia" dei tecnici);
- b) la relativa lentezza del diritto a fronte dell'estrema velocità di trasformazione della tecnica (il che costringe all'"inseguimento" cui ho accennato);
- c) la natura globale dei problemi e il perdurante radicamento statale del diritto "di tutti i giorni" (il che rende la regolazione su scala statale poco efficace e quella su scala globale difficile da negoziare).



**Alessandro Pajno**

Presidente Emerito del Consiglio di Stato e docente di diritto processuale amministrativo, Luiss Guido Carli.  
Mail: [a.pajno@giustizia-amministrativa.it](mailto:a.pajno@giustizia-amministrativa.it)

Il diritto non anticipa, ma segue le grandi trasformazioni della società. L'AI non fa eccezione a questa regola. Il progresso tecnologico offre inedite opportunità, ma può dare origine a rischi per le libertà fondamentali, ove non adeguatamente governato. Spetta, di conseguenza, ai pubblici poteri delineare modalità di intervento che minimizzino le esternalità negative garantendo, nel contempo, il nucleo duro dei diritti individuali.

Negli ultimi anni sono stati elaborati strumenti e carte a livello sovranazionale allo scopo di fornire alcune linee generali a proposito dell'AI e si è registrata una tendenza verso la codificazione dei diritti digitali (si pensi all'art.22 del GDPR e al d. lgs. n. 80/2005). Le iniziative susseguitesi documentano una frammentarietà nella declinazione di linee guida tendenzialmente settoriali e la carenza di interventi di carattere organico. Ricorrere ad una regolamentazione orizzontale in "autonomia" (*self regulation*) non sembra costituire una situazione idonea allo scopo, capace di costruire una disciplina unificante. Sembra, pertanto, opportuno ricorrere a forme di *hard regulation*, preferibilmente a livello sovranazionale, anche al fine di predisporre per i diritti digitali un sistema affidabile di *enforcement*, in assenza del quale vi è il rischio di trovarsi dinanzi a indicazioni e prescrizioni senza potere di vincolatività.

Le osservazioni sopra esposte rendono palese l'esigenza di nuove discipline volte a regolare, a livello appropriato, le AI. Le discipline primarie esistenti (soprattutto in tema di protezione dati) costituiscono un punto di partenza ma non esauriscono l'esigenza di regolamentazione posta dai nuovi fenomeni.

Quanto alla questione dei principi generali, occorre ricordare che la normativa costituzionale e quella della Carta dei diritti forniscono un quadro sufficientemente definito. Potrebbe tuttavia, essere utile declinare ulteriormente questo quadro in una serie di indicazioni volte a specificare i diritti della persona proprio con riferimento al mondo dell'AI e dei settori in cui esso si articola.

Quanto alla eventuale creazione di nuovi istituti e all'elaborazione di nuove categorie giuridiche, il punto di partenza non può che essere costituito da singoli istituti e dalle categorie già esistenti, di cui va attentamente sperimentata la capacità di ordinare e spiegare i nuovi fenomeni. Solo dopo una sperimentazione del genere, può essere presa in considerazione l'elaborazione di nuovi strumenti. Allo stato sembra che le categorie tradizionali conservino una capacità ordinatrice della nuova realtà.

Condivisibile appare la configurazione di un diritto/dovere di educazione all'AI. Misure di questo genere sono decisive non solo per un uso consapevole e responsabile, ma per la qualità stessa della democrazia.

### Oreste Pollicino

Professore ordinario di diritto costituzionale, Università Bocconi. Mail: [oreste.pollicino@unibocconi.it](mailto:oreste.pollicino@unibocconi.it)

Il dilemma legato all'introduzione di nuove regole per fronteggiare le nuove sfide tecnologiche o, al contrario, all'applicazione in via analogica delle regole esistenti non è nuovo, tutt'altro. A metà anni 90 il Giudice Easterbrook si chiedeva, provocatoriamente, nel suo saggio "Cyberspace and the law of the horse"<sup>1</sup>, se fosse sufficiente applicare i principi giuridici che classicamente si applicano al mondo degli atomi, e quindi rifarsi alla categoria unitaria del "diritto dei animali" o se, invece, fosse indispensabile un "diritto del cavallo", un *law of the horse* che rompesse con il passato dei principi generali del diritto, così come internet aveva di fatto stravolto lo status quo della tecnologia esistente. «Now you can see the meaning of my title "cyberspace and the Law of the Horse" When asked to talk about "Property in Cyberspace," my reaction was, "Isn't this just the law of the horse?" This leads directly to my principal conclusion: Develop around law of intellectual property, then apply it to internet». La sua risposta alla domanda, come si evince da questa citazione, è univoca: non servono regole nuove ma è sufficiente applicare analogicamente quelle esistenti.

Una risposta del genere, data oggi, sia con riguardo al cyberspace in generale che a quel particolare ecosistema digitale rappresentato dall'intelligenza artificiale non può che essere assai semplicistica. Il sistema è molto più complesso rispetto alla logica binaria nuove regole vs applicazione analogica di quelle esistenti. La

<sup>1</sup> F. H. EASTERBROOK, *Cyberspace and the Law of the Horse*, in *University of Chicago Legal Forum*, 1996, 207.

domanda che sembra doversi porre a monte è: quando ed in che modo alcune forme di iniziativa economica, specialmente in ambito digitale, si trasformano tecnicamente in potere e, conseguentemente, quale strumentario offre oggi il diritto, e in particolare il diritto costituzionale per reagire a tale trasformazione?

Impossibile dare una risposta esaustiva, ma tre indicazioni prospettiche possono forse azzardarsi.

La prima. Mai come in questo scenario un orientamento che guardi in modo autoreferenziale ai vari regimi giuridici rilevanti, risulta essere, per definizione, perdente. Per esempio, la leva del diritto antitrust, nel passaggio, che si è cercato di descrivere, da esercizio di attività economica a concretizzazione di veri e propri poteri privati in concorrenza con quelli pubblici, non è più sufficiente. Da una parte la leva antitrust deve, a sua volta, poggiarsi sulla forza d'urto del GDPR, dall'altra, una seconda leva va azionata: quella di un consapevole utilizzo dello strumentario di diritto costituzionale a disposizione, a cominciare dalla possibilità di poter applicare orizzontalmente, nei confronti dei nuovi poteri privati, il catalogo dei diritti fondamentali previsti dalla Carte europee.

Qui si passa alla seconda considerazione che si vuol fare.

Si potrebbe sostenere, come fa Robert Alexy<sup>2</sup> che la questione relativa agli effetti orizzontali dei diritti fondamentali previsti dalle Carte costituzionali o dai *Bills of rights* non possa essere concettualmente scissa dal problema più generale del riconoscimento di un effetto diretto agli stessi diritti. In altre parole, se è riconosciuto ad un diritto fondamentale effetto diretto, tale riconoscimento dovrebbe essere caratterizzato da una doppia dimensione: quella verticale (au-

<sup>2</sup> R. ALEXY, *Teoria dei diritti fondamentali*, Bologna, 2012, pp. 570-571.

torità vs. libertà) e quella orizzontale (nei rapporti tra privati).

Nella pratica, come noto, la situazione è assai diversa. Con riferimento allo scenario statunitense, per quanto riguarda l'applicazione dei diritti previsti dagli Emendamenti della Costituzione USA introdotti a partire dal 1791, e con particolare riferimento al I Emendamento che riconosce la libertà di espressione (*right to free speech*), l'efficacia orizzontale (*inter privatos*) dello stesso è negata dall'applicazione della c.d. *state action doctrine*, in forza della quale le garanzie previste dai diritti sanciti dal *Bill of Rights* federale possono essere fatte valere soltanto nei confronti dei poteri pubblici e non dei privati. La ragione alla base della resistenza all'accettazione di una generale efficacia orizzontale dei diritti previsti dalla Costituzione federale si spiega perché l'architettura e, ancor prima, il terreno culturale che fa da *humus* al costituzionalismo statunitense è basato sui valori di *liberty* e *individual freedom* che costituiscono il fondamento dell'autonomia privata.<sup>3</sup>

Un tale *humus* costituzionale rende di fatto impossibile l'attecchimento in terra statunitense della teoria di matrice tedesca dell'effetto orizzontale (*Drittwirkung*) che, dopo aver preso il via dalla giurisprudenza coraggiosa, per l'appunto, delle corti tedesche a partire dagli anni '50, è diventata familiare per altre corti costituzionali europee, ma anche per la Corte europea dei diritti dell'uomo, nonché per la Corte di giustizia dell'Unione europea che ha utilizzato, di fatto gli art. 7 ed 8 della Carta dei diritti fondamentali, nella sua giurisprudenza, quale

<sup>3</sup> G. ROMEO, *Federalising and constitutionalising a legal order: the case for (still) entrusting the bill of rights with the task of full integration*, in *Georgia Journal of International and Comparative Law*, forthcoming.

arma contundente nei confronti, proprio, dei nuovi giganti del web.

Quindi in Europa abbiamo un motivo in più, nel momento in cui si ha a che fare con una nuova sfida tecnologica come quella rappresentata dall'intelligenza artificiale, per utilizzare le categorie e gli strumenti classici di tutela dei diritti fondamentali che possono essere applicati in una dimensione orizzontale, per contrastare i nuovi poteri privati.

Terza riflessione. Una volta (ri)scoperte l'attualità e l'efficacia dell'arsenale europeo in materia di diritti fondamentali nell'epoca dell'intelligenza artificiale, è necessario ed opportuno pensare alla enucleazione, in ogni caso, di nuovi diritti che rafforzino la posizione (al momento assai debole) degli individui-utenti rispetto ai nuovi poteri (privati) nella società algoritmica?

Sarei piuttosto prudente sul punto. La previsione di nuovi diritti non sempre si traduce in una protezione più adeguata dei diritti in gioco, al contrario. L'inflazione di diritti può comportare anche una maggiore conflittualità e, alla resa dei conti, un abbassamento della tutela complessiva che viene riconosciuta alla posizione (già debole, come si diceva) degli individui.

A mio avviso, ciò che serve rafforzare è la dimensione non tanto sostanziale (quindi nuovi diritti che si aggiungono a quelli esistenti), ma procedurale della protezione da riconoscere a detti individui. E dimensione procedurale vuol dire affrontare la questione del *data due process* ai tempi dell'intelligenza artificiale, in modo che i diritti al contraddittorio, ad essere ascoltato, alla parità di armi da un punto di vista, per l'appunto, procedurale, non vengano meno quando la posizione degli individui vada tutelata non nei confronti dei poteri pubblici, ma, dei nuovi, non meno intrusivi, poteri privati.

**Stefano Quintarelli**

Associazione Copernicani. Mail: [stefano@quintarelli.it](mailto:stefano@quintarelli.it)

Anche il fenomeno giuridico nel suo complesso è fortemente coinvolto nell'applicazione della AI che sta trasformando profondamente, fra l'altro, la distribuzione dei poteri pubblici e privati, la tenuta dei diritti e delle libertà fondamentali e molti principi giuridici anche di livello costituzionale.

È possibile applicare le discipline esistenti in via analogica solo fino ad un certo punto.

Bisogna premettere che sotto l'etichetta "AI" vengono inclusi un ampio spettro di metodi tecnici, alcuni dei quali esibiscono comportamenti deterministici, altri probabilistici, alcuni spiegabili, altri parzialmente spiegabili, altri non spiegabili (o comunque la cui spiegabilità avrebbe un costo finanziario tale da impedire l'adozione della tecnica stessa, privando la collettività di effetti eventualmente benefici – certamente da misurare).

Grazie all'aumento della capacità di calcolo ed all'accumulo di dati tramite la rete, la metodica di maggiore visibilità oggi è il *machine learning* che consiste – in estrema sintesi – nella distillazione di modelli statistici a partire da correlazioni riscontrate nei dati esaminati, e l'utilizzo di tali modelli per effettuare predizioni. Tali predizioni, derivando da modelli statistici, sono caratterizzate da un certo livello di precisione (nel caso della valutazione della pericolosità sociale, ad esempio, la percentuale di "veri" criminali) e di recupero (la percentuale di criminali riconosciuti).

È generalmente una coperta corta: ad un guadagno di recupero corrisponde una riduzione della precisione. Si pongono pertanto delle questioni etiche circa la funzione obiettivo che si desidera massimizzare e la sua compatibilità

con i valori che ispirano la nostra civiltà giuridica.

Questa natura di sistema probabilistico introduce una nuova categoria concettuale: quella della macchina non difettosa che commette predizioni errate.

Si pensi all'esempio di un esoscheletro deambulatore che getti a terra un anziano rompendogli il femore: tale comportamento non è conseguenza di un difetto o anomalia; la macchina è perfettamente funzionante e si sa che, statisticamente, qualche volta ciò accadrà. Come comportarsi? Privare gli anziani della possibilità di muoversi autonomamente o accettare un determinato livello di rischio? Quali responsabilità civili e penali prevedere, e in capo a chi? Quali soluzioni normative elaborare per non privare la società di un'utile innovazione tecnologica?

Per certi versi si tratta di una situazione analoga alla farmacologia, dove sappiamo che determinate sostanze sono generalmente utili ma possono determinare, statisticamente, effetti collaterali anche gravi. In questo settore osserviamo un continuum regolatorio che va dal cibo, agli integratori alimentari, ai farmaci, le cui modalità di introduzione in commercio e di attribuzione di responsabilità avvengono con metodi diversi, di gravosità crescente in funzione dell'intensità del rischio di effetti indesiderati.

Tornando all'esempio dell'esoscheletro deambulatore, se da un punto di vista civile la questione può essere affrontata in via assicurativa, dal punto di vista penale è necessaria qualche infrastruttura ulteriore.

Essendo un valore statistico, il "determinato livello" di rischio va riscontrato con la frequenza di incidenti accaduti su tutto l'insieme di deambulatori prodotti, prima di poter affermare una negligenza da parte del produttore o una difettosità del prodotto, con le inerenti (eventuali) conseguenze penali per i responsabili aziendali.



D'altro canto, non si può accettare l'introduzione nel mercato di deambulatori che regolarmente provochino le cadute dei propri utenti. Ne consegue la opportunità di creare meccanismi, graduati in funzione degli effetti possibili, che possano prevedere autodichiarazioni ed ispezioni, certificazioni, autorizzazioni preventive, eccetera, analogamente a quanto avviene nelle industrie alimentari, nella nutra-ceutica e nella farmaceutica. Va infine notato che, in senso più specifico, le applicazioni della AI possono essere varie e non compatibili con questi meccanismi. Talune applicazioni, come ad esempio i casi di gestione del personale, pur se tecnicamente immaginabili, potrebbero violare una molteplicità di norme. Inoltre, il GDPR consente l'opt-out da decisioni automatiche che violano i diritti fondamentali.

In conclusione, attualmente l'ordinamento complessivo pare risultare poco attrezzato e per attrezzarlo occorre anche definire un perimetro di legittimità dell'uso e, soprattutto, i rimedi azionabili in caso di abuso o errato uso (e cosa ciò significhi).

Certamente sì. Anche nell'ambito di un più ampio diritto/dovere all'educazione al pensiero computazionale. L'Intelligenza Artificiale ricomprende sia algoritmi deterministici che sistemi probabilistici. In entrambi i casi stiamo parlando di modalità di scrittura del software.

Acquisirne i principi base consente alle persone – in ogni ambito, dallo studio al lavoro, alla vita di tutti i giorni – di affrontare questi sistemi con consapevolezza, sfruttandoli al meglio, senza eccedere in ansiogene visioni irrealistiche.

Formare gli individui consente di affrontare in modo consapevole la raccolta ed il trattamento dei dati che influenzano le modalità di training e funzionamento del software.

Inoltre, educare all'AI (a qualsiasi età) significa anche trasmettere una comprensione delle pos-

sibilità, dei rischi e delle opportunità che la tecnologia offre nei vari scenari e modalità d'impiego.

È necessaria questa educazione all'AI non solo per far capire come funziona il software, ma anche quali sono le implicazioni che il software avrà nel contesto nel quale verrà utilizzato. Per esempio: è sensato usare un software per decidere il tasso di recidiva? Come devono essere gestite le responsabilità per azioni o decisioni prese da un individuo aiutato da AI?

Una pubblicità televisiva chiedeva "ti piace vincere facile?" e mostrava una partita di calcio con in campo una squadra, composta da undici giocatori come di prammatica, che veniva affrontata da una squadra composta da centinaia di giocatori che invadevano la metà campo avversaria come uno stormo di cavallette. In una simile condizione, poco rileva il confronto sulla tecnica individuale, le strategie di gioco, eccetera. Ogni schema salta perché la quantità muta i termini della questione in modo radicale.

Questo è un primo punto fermo. La digitalizzazione di un'attività ne determina un incremento di scala estremamente significativo, tale da sconvolgere il significato originario. Una volta che una attività è dematerializzata e trattata immaterialmente con strumenti digitali, può beneficiare di diverse proprietà di base: riprodurre un bene materiale ha un costo variabile non trascurabile, così come immagazzinarlo e manipolarlo; inoltre, trasferirlo di luogo presenta costi elevati e richiede tempi non brevi.

Di contro, un "oggetto" immateriale può essere trasferito istantaneamente da un capo all'altro del globo con un costo marginale nullo, può essere "manipolato" da un computer, può essere copiato ed immagazzinato una infinità di volte con un costo marginale ancora nullo. Stiamo entrando (o siamo già immersi) in un contesto in cui le applicazioni informatiche sono istantanee,

mentre non lo è l'eventuale gestione giuridica: la dimensione immateriale contrae la dimensione temporale.

A ben vedere, quando la possibilità di diffusione delle informazioni era limitata dalla materialità dei supporti, solo pochi soggetti potevano permettersi di creare e gestire dei canali di comunicazione (editori) per i quali era prevista una speciale responsabilità che prevedeva (e prevede), tra l'altro, misure legali specifiche, come l'obbligo di rettifica e sanzioni rafforzate in caso di abuso dello strumento.

Di contro, la libertà di espressione dell'individuo, per sua natura confinata ad una platea ristretta, era massima. Non poteva estendersi molto al di là delle esternazioni al bar. Con l'avvento della digitalizzazione, forte delle caratteristiche ricordate sopra, si annulla questo divario tra la comunicazione individuale/di gruppo e la comunicazione di massa; in questo modo, qualunque individuo può raggiungere ad un costo marginale nullo ed in tempo reale platee persino più ampie di quelle raggiungibili da un editore ed il sistema regolamentare, che era disegnato per un mondo materiale, si trova messo alle strette per effetto del cambiamento di scala della velocità e della quantità. E così discorsi d'odio e bufale prosperano, con un significativo impatto sulla società e sulla politica.

L'Intelligenza Artificiale gioca un ruolo in questo scenario perché, grazie a sistemi di classificazione usati per fare raccomandazioni, determina la distribuzione e visibilità dei contenuti e delle interazioni degli utenti con questi contenuti.

#### **Rosso di sera, bel tempo si spera**

Questo proverbio condensa il principio di funzionamento del machine learning che alimenta le principali applicazioni di intelligenza artificiale: si osserva il colore del cielo in un particolare momento del giorno, per predire che tempo fa-

rà il giorno dopo. Il tutto però è racchiuso da un senso di precarietà e speranza, ad evidenziare che nulla è certo al 100% ma che anche le più semplici predizioni sono legate ad una risposta che deve racchiudere una percentuale di errore. "Rosso di sera, bel tempo si spera" è un modello statistico generato dall'osservazione di dati dalla realtà, usato per fare predizioni.

Se anziché il colore del cielo considerassimo le caratteristiche di un appartamento (mq., piano, classe energetica, quartiere) e come al variare di queste, una casa cambi il suo prezzo, possiamo immaginare un modello statistico che parta dall'osservazione di dati della realtà per essere usato per fare predizioni e quindi rispondere alla domanda "un appartamento con queste caratteristiche, quanto costa?". O ancora, se anziché il colore del cielo o il valore della superficie, partissimo da caratteristiche di una cellula (densità, forma della membrana cellulare, forma e densità degli organelli in essa contenuti) e trovassimo le correlazioni con la probabilità che essa sia malata, siamo in grado di usare questo modello per predire i tumori.

È bene evidenziare un punto: sebbene si dica spesso che Internet sia uno strumento di disintermediazione, non è così. Internet è uno strumento di reintermediazione con caratteristiche diverse (quelle ricordate sopra). L'intermediario è il social network con il suo sistema AI, invisibile ma sempre presente, che gestisce la diffusione dei contenuti e le interazioni delle persone. I nuovi intermediari hanno un obiettivo – la massimizzazione dei profitti – strettamente legato da una AI all'aumento di polarizzazione e conflittualità, contrariamente all'obiettivo di ogni comunità sociale che è tendere alla coesione ed alla tranquillità. Per questo, presto o tardi, la politica finirà per regolamentare questi intermediari e l'uso che essi fanno dell'IA.

Un altro esempio preclaro di come un cambiamento di scala cambi la natura dell'attività si ha con i cartellini dei fotosegnalati. Le forze dell'ordine conservavano in modo materiale circa 50mila cartellini di persone che per qualche ragione erano reputate meritorie di particolare attenzione. Le fotografie fisiche erano raccolte in un archivio con dei cartelli recanti i dati caratteristici. In caso di necessità, tali dati venivano confrontati con quelli di un sospetto per procedere al reperimento della foto. Questa attività è stata naturalmente digitalizzata e la gara d'appalto per il sistema prevedeva di poter gestire 100mila cartellini. Una dimensione congrua con il fatto che nelle carceri italiane vi sono circa 60mila detenuti.

Dato che il costo marginale di elaborazione è nullo, così come il costo di archiviazione, le ultime notizie ottenute con una richiesta di accesso civico (il cd. "FOIA", introdotto in Italia nel 2013 sul modello del *Freedom Of Information Act* statunitense) dicono che, dopo varie iterazioni, il sistema è stato oggetto di un semplice "upgrade" hardware (è immaginabile un costo assai contenuto) e che il numero di soggetti presenti nell'archivio è pari a 9 milioni<sup>1</sup>: un numero pari a circa un sesto della popolazione italiana adulta. Gestire a mano 9 milioni di schede è praticamente impossibile. Farlo con un computer che grazie all'AI può aumentare le nostre funzioni (di sorveglianza, in questo caso) è decisamente cosa fattibile e poco costosa.

I britannici hanno dei furgoni posti nelle vicinanze di luoghi sensibili che riprendono i passanti a loro insaputa con delle telecamere nascoste. Le immagini acquisite vengono confrontate in tempo reale con quelle presenti in un database di criminali fotosegnalati e le persone così individuate vengono fermate.

<sup>1</sup> Risposta dell'On. Sibillia all'interrogazione 4/04528, presentata dall'On. Sensi il 21/1/2020

Ma questi sistemi funzionano? Sì e no. Dipende dai punti di vista.

Come è già stato spiegato, i sistemi che si basano su tecniche statistiche vanno valutati da due punti di vista chiamati recupero e precisione. Chiaramente, se fossimo in un regime che non rispettasse i diritti civili, si potrebbe tarare il sistema per avere un alto recupero, ovvero per identificare un gran numero di criminali. Anche se prenderebbe nella rete tanti innocenti.

Viceversa, se si configurasse il sistema per minimizzare o annullare la percentuale di innocenti erroneamente ritenuti criminali, più in linea con i valori europei, molti criminali non verrebbero riconosciuti.

I risultati di una richiesta FOIA di accesso civico ai dati di performance del sistema britannico parlano chiaro: il 96% delle persone fermate erano innocenti<sup>2</sup>. Un ragazzo che si copriva la faccia per evitare di farsi controllare è stato fermato per ostacolo alle forze dell'ordine e multato.

I politici spesso hanno aspettative taumaturgiche nei confronti della tecnologia e, installando sistemi di riconoscimento facciale, possono dire agli elettori di avere fatto qualcosa per risolvere un problema di sicurezza (magari inesistente, la cui percezione hanno contribuito a creare, peraltro assumendo come implicita l'equiparazione tra sorveglianza e sicurezza). In alcuni luoghi, politici più competenti e lungimiranti hanno vietato l'utilizzo di queste tecnologie. Ad esempio, a San Francisco, nel cuore della Silicon Valley che vive di tecnologia.

<sup>2</sup> TIM CUSHING, *New Data On London Metro Police Facial Recognition Tech Shows It's Still Wrong 96 Percent Of The Time*, 9 maggio 2019, in <https://www.techdirt.com/articles/20190506/13533942150/new-data-london-metro-police-facial-recognition-tech-shows-still-wrong-96-percent-time.shtml>.

In ragione di velocità e scala diversa, come è cambiata la natura della comunicazione da individuale a personalizzata di massa, così è cambiata la natura del controllo delle persone, da individuale a massiva, senza una riflessione politica sulle conseguenze del passo che si stava compiendo e senza un adeguamento regolamentare, quanto mai opportuno secondo il gruppo di esperti ad alto livello sull'Intelligenza Artificiale istituito dalla Commissione Europea, che scrive:

«L'IA permette a organismi pubblici e privati di identificare in modo sempre più efficiente le singole persone.

Esempi degni di nota di una tecnologia scalabile di identificazione tramite IA sono il riconoscimento facciale e altri metodi di identificazione involontaria che utilizzano dati biometrici (ad esempio il rilevamento della menzogna, la valutazione della personalità in base a micro-espressioni e il rilevamento automatico della voce).

Identificare una persona talvolta è un atto auspicabile e rispettoso dei principi etici (ad esempio nei casi di frode, riciclaggio di denaro o finanziamento del terrorismo). L'identificazione automatica tuttavia desta enormi preoccupazioni di natura sia giuridica che etica, in quanto può avere effetti non previsti sotto molti aspetti a livello psicologico e socioculturale. Per salvaguardare l'autonomia dei cittadini europei è necessario ricorrere alle tecniche di controllo tramite l'IA in modo proporzionato. Definire chiaramente se, quando e come l'IA può essere utilizzata per l'identificazione automatica degli individui e differenziare tra l'identificazione di un individuo e la sua tracciatura e localizzazione, e tra sorveglianza mirata e sorveglianza di massa,

sarà fondamentale per ottenere un'IA degna di fiducia<sup>3</sup>».

Di contro, la Cina viene spesso indicata come il luogo dove la AI viene sfruttata per il controllo di massa, andando ad attribuire punteggi o schedare persone che compiono atti illeciti o degni di biasimo.

In conclusione, le nuove possibilità del software di fare – grazie alla AI – attività di percezione e classificazione spostano il perimetro di ciò che prima era realizzabile dagli umani. Una volta portate tali attività nel dominio delle macchine, grazie all'aumento della velocità di elaborazione e grazie ai costi marginali nulli di archiviazione ed elaborazione, si determina un cambiamento di scala che non è più confrontabile con quanto avveniva in precedenza e che conseguentemente mette alle corde il quadro regolamentare pensato per un regime diverso, più lento e limitato.

Se i principi fondamentali rimangono attuali, devono essere riscritte le norme che li attuano. Ad esempio, non è in discussione il principio di "libertà di manifestazione del pensiero", ma mancano le norme che lo attualizzano al quadro tecnologico. Se ciò non accade, l'adattamento sarà molto lento e avverrà per opera della giurisprudenza, come già è avvenuto per responsabilità su Internet, copyright, ecc., però con costi economici e sociali rilevanti dato il prevedibile moltiplicarsi delle questioni da dirimere e la relativa incertezza.

<sup>3</sup> GRUPPO DI ESPERTI AD ALTO LIVELLO SULL'IA, *Orientamenti Etici per un'IA Affidabile*, Bruxelles, 2019, p. 41.

**Andrea Simoncini**

Professore ordinario di Diritto costituzionale, Dipartimento di Scienze Giuridiche (DSG) dell'Università degli Studi di Firenze. Mail: [andrea.simoncini@unifi.it](mailto:andrea.simoncini@unifi.it).

Il rischio più grave dinanzi alla fase di trasformazione tecnologica che ci troviamo a vivere è quella di cedere alla ipersemplicizzazione manichea che oggi divide il dibattito sull'AI tra *tecnocentusiasti* e *tecnocatastrofisti* (o se vogliamo, tra "apocalittici e integrati" per echeggiare il famoso saggio di Umberto Eco del 1964). I primi, certi della proporzionalità diretta tra progresso tecnico e sviluppo della libertà, i secondi, invece, altrettanto certi della relazione inversa. Non si può negare, d'altronde, che la "narrazione collettiva", come si usa dire, abbia seguito queste due polarità. Siamo passati da una enfasi quasi "messianica" sulla capacità della tecnologia digitale di potenziare la libertà e, conseguentemente, le basi stesse della democrazia, ad una fase – quella attuale – in cui prevale una percezione distopica delle nuove tecnologie, sempre più avvertite come forze senza controllo e responsabilità.

Per avviare una riflessione seria e razionale su questi temi, dunque, occorre, innanzitutto, rifiutare questa dicotomia come assoluta e inevitabile. La modernità – per riprendere un importante saggio di Bauman del 1993 – è intrinsecamente *ambivalente*. Non possiamo, infatti, disconoscere l'utilità rivoluzionaria della trasformazione digitale avviata da Internet, dai *big data*, dalle tecniche di *machine learning*, con la conseguente "rinascita" dell'AI. D'altra parte, come è ben noto alla antropologia (si pensi ai lavori di Leroi-Gourhan), ogni nuovo sviluppo tecnico lascia una impronta culturale indelebile e, dunque, giuridica. Ed è qui che si deve aprire la nostra riflessione. Difatti, alcuni di questi sviluppi stanno producendo effetti negativi pro-

prio sul piano di quelle libertà che intendono potenziare e, questione ancor più delicata, altri effetti potenzialmente negativi oggi non sono neppure chiaramente percepiti ovvero percepibili in tutta la loro portata. La rapidità della evoluzione tecnica è tale da eccedere – in alcuni settori – la stessa capacità umana di spiegazione ovvero di controllo. Per riprendere un principio giuridico nato – non a caso – nel diritto dell'ambiente, ci troviamo in una situazione di incertezza, di deficit conoscitivo, dinanzi a possibili conseguenze indesiderate derivanti dallo sviluppo tecnologico. Non è quindi possibile adottare un approccio dicotomico (libertà assoluta o divieto assoluto), ma occorre affidarsi ad un principio *precauzionale*. Ritengo, per venire alla nostra questione, che l'impatto della AI sul diritto e sui diritti sia ambivalente. In parte, essa può potenziare le facoltà individuali e collettive e, dunque, porsi come un amplificatore della libertà umana, ma allo stesso tempo essa può recare danni anche molto rilevanti. Orbene, laddove questi danni siano chiari ed evidenti, è altrettanto chiaro quale debba essere la reazione dell'ordinamento: il divieto. Il caso più diffuso e problematico da esaminare è quando non abbiamo informazioni sufficienti per determinare se tali effetti negativi si produrranno ovvero in che misura, ovvero ancora quando questi effetti, negativi a livello individuale, comportino delle utilità collettive su larga scala.

È qui, a mio avviso, che si pone la questione fondamentale per il diritto costituzionale contemporaneo; questione che deve essere affrontata affidandosi ad un approccio *precauzionale*, intendendo con questo termine quel principio per cui l'incertezza sui possibili effetti negativi di una determinata tecnologia, ovvero l'incertezza sulla sua possibile utilità, non possono essere argomenti sufficienti a giustificare l'assenza di regolazione.



Detto altrimenti, occorre sempre prevedere una forma di regolazione *giuridica* per quelle tecnologie delle quali ancora non è chiaro ovvero quantificabile il potenziale offensivo.

Due precisazioni si impongono. La prima, è che tale principio, avendo come obiettivo la protezione delle libertà fondamentali, si pone ad un livello intrinsecamente *costituzionale*; nel senso che è destinato a precedere e guidare la legislazione ovvero le altre forme di regolazione.

La seconda è che con il termine “forma di regolazione” non si intende necessariamente la “legislazione nazionale” ovvero il “diritto europeo” ovvero altra forma di normazione pubblica; ben si può utilizzare tutto lo spettro delle diverse modalità attraverso le quali l’interesse “politico” superindividuale può ordinare i comportamenti individuali (dal cosiddetto *hard-law* alla *soft-law*).

In ogni caso, però, il principio che chiamerei di “precauzione costituzionale”, impone che non si possa lasciare alla spontanea “mano invisibile” del mercato ovvero alla “sola” auto-regolazione, la definizione delle responsabilità o dei limiti nel mondo della tecnica.

In questo senso, la grande “esplosione” – anche numerica – di “documenti”, “carte”, “linee guida” ovvero di altri strumenti riguardanti l’AI, accomunati dalla qualificazione “etica” (si veda l’utile *Sourcebook* recentemente predisposto da Marc Rotemberg per conto di EPIC.ORG) è un fenomeno che deve far molto riflettere. Esso esprime, infatti, quella ambivalenza di cui parlavamo. Da un lato, dimostra che oggi sono chiaramente percepiti i rischi connessi allo sviluppo di certe tecnologie, richiedendo l’adozione di *standards* uniformi di comportamento rispettosi dei valori fondamentali; dall’altro, si ha timore – o difficoltà – ad immaginare risposte di natura dichiaratamente *giuridica*, dunque, anche, *giustiziabile*.

Chiarita questa premessa penso di poter più rapidamente rispondere agli altri temi sollecitati.

A mio avviso la regolazione dell’AI deve avvalersi di un mix di strumenti.

Da un lato, è certamente possibile estendere interpretativamente fattispecie esistenti alle nuove possibilità offerte dalle tecnologie, dall’altro occorre “inventare” nuove strategie di regolazione.

Un esempio, può aiutarci a comprendere; pensiamo alla vicenda della disinformazione *online* (nota come *fake news*); da un lato, dinanzi ad un soggetto che “dolosamente” renda disponibili in rete dei contenuti falsi, ovvero ne falsifichi il contesto, per ottenere un qualsiasi tipo di vantaggio politico, economico, o anche solo ideologico, mi pare si possano estendere senza problemi fattispecie criminose esistenti (si pensi alla diffamazione, alla truffa, al procurato allarme, etc.). Il problema pratico è che spesso non è semplice trovare chi ha consapevolmente caricato la notizia o l’informazione sapendo che è falsa, pur di ottenere tali scopi; ma laddove si riuscisse a rintracciarlo, non necessitiamo di ulteriori fattispecie penali. Il problema del tutto nuovo nasce dinanzi alla cosiddetta “viralità”, cioè al fenomeno per cui migliaia di altre persone, che ben possono essere ignare della falsità originaria, replicano la comunicazione semplicemente perché questa rafforza convinzioni (o pregiudizi) che la persona ha già a prescindere dalla informazione, creando un effetto “valanga” che nei motori di ricerca automatizzati genera l’ulteriore effetto *loop* di rendere più visibile la stessa notizia, amplificando così il suo effetto “virale”.

Questo fenomeno, estremamente pericoloso, profondamente impattante sulla corretta formazione dell’opinione pubblica, è reso possibile dalle nuove tecnologie, ma eccede le possibilità



di una estensione interpretativa; richiede, dunque, una normazione specifica.

Quale normazione?

Questo è il problema più serio per i giuristi.

Molto spesso per poter regolare questi ambiti occorre una conoscenza tecnico-scientifica che coloro che sono incaricati di scrivere le regole non hanno.

Questo ha fatto sì che in molti casi si sia chiesto ai tecnici stessi (ingegneri, fisici, informatici, *computer scientists*) di scrivere le proprie regole; analogamente, alcuni grandi soggetti privati che detengono la maggior parte dei dati nel mondo necessari per poter elaborare gli algoritmi di intelligenza artificiale (le cosiddette Big Tech), hanno deciso di adottare proprie strategie di regolazione (si pensi allo Ethical Committee di Google – fallito appena nato – o all’Oversight Board di Facebook o analoghe strutture).

È evidente il limite intrinseco di queste forme di *self-regulation*: la *cattura* della regolazione da parte del regolato. Anche volendo concedere la massima buona fede a questi tentativi, essi mancano di un requisito fondamentale: l’indipendenza e la terzietà di chi pone le norme di garanzia e di chi le fa rispettare. Penso, quindi, che la prospettiva più costruttiva sia quella di una eteronormazione di matrice politica (europea-nazionale-locale) che ponga la disciplina *di principio* e lasci l’attuazione a forme di autoregolamentazione assistite, però, da Autorità di garanzia indipendenti.

In buona sostanza, un meccanismo analogo a quello realizzato dalla *prima* legislazione sulla privacy (oggi il GDPR, avendo voluto sostituire la legislazione di principio con norme auto-applicative ha finito per creare molti più varchi): uno schema normativo in cui, a norme giuridiche di principio poste da autorità politiche pubbliche, segua una fase di attuazione normativa in cui vengano coinvolti gli stessi destinatari del-

le norme a fianco di autorità indipendenti tecnicamente esperte dotate di poteri di controllo e co-regolazione.

In questo senso, quindi, penso che se i principi debbano essere comuni le declinazioni non possano che essere differenziate per settore: l’impiego della AI nei sistemi sanitari non può avere le stesse regolazioni che nel campo dei trasporti, della prevenzione del crimine o delle tecniche in agricoltura.

In conclusione, tutte queste proposte regolative hanno in comune un problema di natura eminentemente educativa. Il problema nasce dalla separazione dei saperi che ormai ha destrutturato i sistemi educativi e soprattutto quelli superiori.

Molto spesso i valori e i principi che il diritto cerca di difendere – pensiamo all’area delle libertà fondamentali – sono del tutto ignorati nella formazione di chi poi professionalmente si troverà a progettare e sviluppare quegli stessi strumenti tecnologici di cui stiamo parlando.

Oggi, la maggior parte dei problemi di “compatibilità” della tecnica con il diritto si pongono *dopo* che la tecnica stessa è stata creata (e spesso commercializzata). A questo punto è troppo tardi.

L’unico modo realmente effettivo per impedire la violazione dei diritti fondamentali è intervenire in via preventiva, alla fonte, cioè nel momento stesso in cui vengono pensate, strutturate, elaborate quelle macchine, quei programmi o quei codici che chiamiamo AI.

È quello che nel settore della protezione dei dati personali è stata chiamata “privacy by design”, ma che, a mio avviso, come principio va generalizzato ed anticipato come una più ampia tutela “by education”. I valori connessi alle libertà fondamentali debbono essere patrimonio comune non solo di chi professionalmente lavora

nei settori giuridici o giudiziari, ma, almeno per i principi basilari, anche in chi si occupa di tecnologia. È solo questa forma di educazione che può, da un lato, costituire un argine effettivo alle “involuzioni” e, dall’altro, potenziare la ricerca di quelle applicazioni davvero utili al genere umano.

*W. S. - Forum*



## 4. Medicina e giustizia: ambiti paradigmatici

### 4.1 La medicina

#### Maria Chiara Carrozza

*Istituto di Biorobotica, Scuola Superiore Sant'Anna di Pisa, Dipartimento di Eccellenza in Robotica e Intelligenza Artificiale, Scuola Superiore Sant'Anna di Pisa e Fondazione Don Carlo Gnocchi Onlus. Mail: [m.c.carrozza@santannapisa.it](mailto:m.c.carrozza@santannapisa.it)*

**4.1.a** Uno degli ambiti di riferimento a maggior impatto per lo sviluppo dell'IA è senza dubbio la medicina a partire dalla cosiddetta 'trasformazione digitale' incentrata sull'utilizzo dei dati disponibili attraverso opportune infrastrutture.

Certamente tale trasformazione inciderà sulla figura del medico e del ricercatore che opera in ambito clinico: in primis sul profilo formativo, le figure che opereranno nella sanità del futuro dovranno essere in grado di utilizzare la tecnologia, selezionare il dato, conoscerne le potenzialità.

La raccolta di "buoni dati" è alla base del funzionamento dei sistemi di AI in medicina: più un data base è ben strutturato e contiene un gran numero di informazioni, più è possibile cercare relazioni tra i vari set di dati per trovare correlazioni. Questo tipo di analisi ha un importante valore predittivo, il suo utilizzo in medicina può permettere di prevedere condizioni, situazioni e avvenimenti del futuro, con previsioni statistiche a livello di popolazione o anche in modo puntuale a livello di singolo individuo (la così detta 'medicina personalizzata').

Vi sono difatti numerosissime possibilità di impiegare algoritmi di IA nella medicina: dalla sperimentazione e ricerca traslazionale, alla medicina personalizzata, dalla diagnostica, dalla

tele-assistenza e tele-riabilitazione alla chirurgia robotica, dal supporto fino al potenziamento funzionale del paziente mediante robotica e sensori indossabili o impiantabili nel corpo umano. L'applicazione della IA in ambito medico ha avuto delle ricadute anche in termini di scelte di mercato promuovendo diverse realtà di imprese che hanno portato dell'innovazione in campo healthcare. Per esempio, nel contesto europeo stanno emergendo le cosiddette 'terapie digitali' che dimostrano come sia possibile un buon utilizzo dell'IA somministrando terapie integrate che possono cambiare la vita e migliorare le condizioni anche in presenza di gravi malattie croniche, guidando i pazienti verso comportamenti ottimizzati a seconda delle loro condizioni fisiologiche e cognitive. L'accesso alle terapie potrà avvenire attraverso opportune piattaforme che si presenteranno come applicazioni sugli smartphone.

Anche in medicina, forse soprattutto in medicina, è indispensabile che l'utilizzo di sistemi di AI venga validato dal contesto normativo, tutelando pienamente i diritti del paziente e che avvenga in senso eticamente orientato. La medicina è infatti, a mio avviso, uno dei settori in cui maggiormente emerge la necessità di dotare l'AI di una sovrastruttura valoriale, la cui bussola dovrebbe essere rappresentata dai principi di equità e uguaglianza, che si traducono in uguale accesso alle risorse, parità di trattamento.

**4.1.b** Nonostante la diffusione ampia e, per molti aspetti, positiva dell'utilizzo di sistemi di Intelligenza Artificiale in campo healthcare è importante precisare che la medicina resta comunque una prerogativa umana.

Tra medico e tecnologia deve piuttosto instaurarsi un rapporto di collaborazione orientato all'ottimizzazione del trattamento del

paziente in senso globale: dall'accuratezza e tempestività della diagnosi (data la capacità predittiva che caratterizza l'AI), alla somministrazione di una terapia sino alla riabilitazione post evento traumatico. Il rapporto medico – paziente, prima diretto, oggi è in un certo senso mediato dalla tecnologia: il robot utilizzato dal terapeuta durante una seduta di riabilitazione, l'app con cui il medico monitora lo stato di salute di un paziente, sono alcuni esempi di questo genere di mediazione. Entrando in modo così diretto in una dinamica relazionale tipicamente umana, è ancora più forte la necessità di costruire un perimetro normativo ed etico che tenga conto della centralità del paziente nel percorso di cura.

L'applicazione dell'IA a dati sensibili come sono certamente quelli connessi allo stato di salute di un paziente, ci impone una riflessione profonda sul senso del rispetto della privacy in connessione con lo sviluppo di sistemi che, come abbiamo ripetuto in diverse occasioni, si nutrono di dati. Bisogna essere in grado di ponderare gli interessi, tutelare i diritti senza limitare lo sviluppo tecnologico: tale assunto presuppone un bilanciamento delicatissimo. Sul tema sono intervenute le Istituzioni e lo stesso regolamento generale UE sulla protezione dei dati prevede un livello massiccio di tutela della riservatezza delle informazioni personali fornite in formato digitale. Alla luce degli sviluppi sul fronte dell'applicazione dell'Intelligenza Artificiale, è fondamentale monitorare con attenzione che nella pratica clinica, nei protocolli di ricerca, si mostri particolare attenzione alla garanzia non soltanto del diritto al consenso informato e alla libertà di scelta nella trasmissione di dati, ma anche quello di accedere ai dati trasmessi e di poterli modificare e verificare, conoscere in maniera completa, chiara ed esaustiva la finalità di

utilizzo del dato, la destinazione, se questi verranno inviati a terzi e soprattutto garantire il reale ed effettivo diritto del cittadino/paziente di revocare il proprio consenso.

**Silvio Garattini**

Presidente dell'istituto di ricerche farmacologiche  
 "Mario Negri" IRCCS. Mail:  
[silvio.garattini@marionegri.it](mailto:silvio.garattini@marionegri.it)

L'intelligenza artificiale (IA) è ormai una realtà con cui dovremo fare i conti molto presto anche in campo biomedico. Attraverso l'IA sarà possibile, ad esempio, interpretare immagini con maggior probabilità di precisione rispetto a quella dei migliori esperti. Oltre a facilitare e migliorare la fase diagnostica, l'IA permetterà di adattare alla persona gli eventuali interventi terapeutici o riabilitativi, lasciando al medico relativamente poco spazio d'azione, se non quello di trasferire le informazioni al paziente con pochi margini di autonomia.

Si profila quindi un grande cambiamento che richiede urgenti interventi di varia natura: educativi, normativi e politici. Qui si discute degli aspetti su cui può intervenire il diritto per evitare che la giurisprudenza arrivi in ritardo e spesso nell'impossibilità di mettere "paletti" a situazioni ormai ampiamente acquisite.

Ecco alcune riflessioni sul tema:

Per quanto riguarda la medicina, l'IA fa parte oggi di attività di ricerca di tipo essenzialmente privato, mentre in futuro agirà non solo nella medicina privata, ma essenzialmente in quella pubblica, rappresentata dal Servizio Sanitario Nazionale (SSN). È quindi importante esercitare uno stretto controllo pubblico sull'IA, tenendo conto che molti algoritmi su cui si basa l'IA sono brevettati e segreti. Bisogna perciò essere in grado di ottenere la massima trasparenza anche sulle modificazioni che un determinato "oggetto" dell'IA subisce nel corso del tempo.

Poiché i vantaggi dell'IA in medicina sono evidenti, sarà bene concentrare l'attenzione sui pericoli, verso i quali è necessario attuare forme di prevenzione. Un primo problema è rappresentato dal "fai da te" che è già esploso con

l'avvento di alcuni interventi. La grande disponibilità di informazioni accessibili a tutti grazie a internet ha indotto molti ammalati - o presunti tali - a dettare al medico le prescrizioni di cui avevano bisogno, qualche volta magari senza sbagliare troppo. Di fronte a questa situazione, è facile da parte del medico ribadire che su internet si trova di tutto e quindi spetta al medico, che ha una specifica formazione, decidere che cosa sia giusto fare. Con l'IA il rapporto cambia, perché la selezione fra dati giusti e sbagliati è già stata fatta ed è difficile dire che i ragionamenti sui dati non siano stati fatti con sufficiente "intelligenza"!

Un secondo problema riguarda gli interessi ideologici non disgiunti da quelli economici che possono inserirsi nell'IA. Ad esempio, nell'ambito delle prescrizioni si possono privilegiare interventi o prodotti privi di evidenze scientifiche, come quelli promossi dalla cosiddetta medicina "alternativa" o integrativa.

Ancor più importante nell'ambito della medicina "ufficiale" è la possibilità di favorire interventi diagnostici non strettamente necessari, suggerire trattamenti farmacologici funzionali a specifici interessi industriali, prescrivere interventi chirurgici in eccesso per sostenere attività private o gli interessi di compagnie assicurative. L'avvento dell'IA sarà molto più veloce della capacità del medico di adattarsi alla nuova condizione della sua professione. È urgente perciò che i problemi posti dall'introduzione della IA vengano presentati e discussi fin dal primo anno della scuola di medicina, avendo cura di farne apprezzare i lati positivi – ad esempio quando si tratta dell'interpretazione di immagini complesse – ma al tempo stesso rendere edotti i futuri medici degli aspetti negativi. Non solo, è urgente anche un'adeguata formazione di chi già esercita la professione. Occorre sviluppare nel mondo medico un atteggiamento positivo per



ricercare nuovi compiti e ruoli che oggi non trovano spazio nella professione, come ad esempio una maggior attenzione allo studio e alla formazione permanente o ai problemi della prevenzione che non sono delegabili all'IA perché richiedono una motivazione e una volontà che solo il medico può indurre o agevolare nei suoi assistiti. Non si deve dimenticare che la prevenzione, attraverso la proposizione dei buoni stili di vita, è uno degli interventi essenziali per evitare malattie e quindi contribuire alla sostenibilità del SSN.

Un ultimo aspetto è che non si deve dare per acquisito il fatto che l'IA migliori necessariamente la medicina e i suoi risultati. È perciò importante mettere in atto al più presto adeguate sperimentazioni per stabilire il livello di salute in coorti di popolazione, in cui si segue la IA rispetto a coorti che non la utilizzano. Saranno necessari studi epidemiologici a vasto raggio per valutare l'andamento delle varie patologie, della durata di vita, della qualità di vita, nonché l'andamento dei consumi diagnostici, terapeutici e riabilitativi. La numerosità delle visite dei medici, del numero e della durata delle ospedalizzazioni saranno altri aspetti di studi che dovrebbero essere obbligatori. Tale controllo deve essere effettuato nel tempo, anche nel caso in cui l'impiego dell'IA sia generalizzato.

I problemi sono molti e non facili da risolvere, ma sono urgenti e pertanto è benvenuta la possibilità di discuterli.

### **Giovanni Muto**

*Dipartimento di Scienze Biomediche, Humanitas University.*

### **Lorenzo Tosco**

*Dipartimento di Urologia, Ospedale Humanitas Gradenigo.*

L'intelligenza artificiale (AI) è definita come una scienza ed un insieme di tecniche informatiche ispirate alla maniera in cui gli esseri umani utilizzano il loro sistema nervoso ed il corpo per percepire, imparare, ragionare e portare avanti azioni<sup>1</sup>. I campi di applicazione di tali tecnologie sono molteplici come nel mondo della salute. L'AI è un tema sempre più diffusamente discusso nel mondo contemporaneo e non si tratta di un concetto puramente teorico o filosofico, bensì di una realtà in cui si sta investendo attivamente. L'interesse è talmente elevato che si è pianificata una strategia europea di sviluppo coordinato nel campo della AI. Alcuni stati membri dell'Unione Europea hanno già implementato delle strategie e stanziato dei fondi dedicati alla AI<sup>2</sup>. Nel campo della medicina le applicazioni della AI possono essere molte tra cui la valutazione del rischio e della sua gestione, i sistemi diagnostici automatizzati, la personalizzazione dei piani di gestione clinica, la valutazione dell'efficacia terapeutica, l'assistenza automatizzata del malato, l'analisi epidemiologica per la valutazione del rischio della comunità, i sistemi di traduzione linguistica immediata per interagire con pazienti multilingue e non ultima l'automazione chirurgica. Si pensa che la AI potrebbe sostituire gradualmente l'essere umano in molte attività pratiche; nel campo della medicina si è previsto che intorno al 2050 potrebbero essere già disponibili tecnologie in grado di effettuare interventi chirurgici in autonomia<sup>3</sup>, ponendo a breve termine un traguardo che fino a non molto tempo fa veniva considerato fantascienza.

La professione medica ha come scopo per sua natura la risoluzione dei problemi di salute di uno o più soggetti. Essa si estrinseca attraverso dei passaggi obbligati. La patologia, infatti, deve essere primariamente definita attraverso l'interpretazione semeiologica (studio dei segni e dei sintomi), laboratoristica e/o strumentale tale da produrre una ipotesi diagnostica. Dalla diagnosi si passa alla valutazione delle possibili terapie (mediche e/o chirurgiche) allo scopo di ottenere un beneficio per il soggetto o la comunità affetta da tale patologia. La decisione terapeutica è basata sull'applicazione dei trattamenti con miglior efficacia basandosi sui dati di letteratura disponibili (evidence-based medicine, EBM). Tale iter rappresenta il cuore del processo decisionale ed operativo del medico-chirurgo che in molte delle sue componenti potrebbe essere sostituito dalla AI. Ci sono progetti di AI che hanno previsto l'analisi di enormi volumi di dati scientifici e la loro integrazione per ottenere una sintesi organizzata dei risultati al fine di avere una decisione basata sull'evidenza o perfino generare delle ipotesi nel campo clinico e della ricerca<sup>4</sup>. La professione medica dovrà necessariamente modularsi nel tempo, modificando il ruolo del professionista da attore principale a supervisore e coordinatore della AI nella gestione del paziente anche sotto l'aspetto chirurgico. Una figura nuova nel panorama medico risulterà il *trainer*, ossia un soggetto medico che "formi", immettendo dei dati basilari, il sistema di AI. La formazione accademica del medico potrà essere influenzata dalla AI, adattandosi ad essa sin dai primi anni di studio, rivolgendo l'attenzione più alle tecniche di controllo che all'esecuzione diretta della pratica clinica e tutto ciò potrebbe portare ad una involuzione di quello che oggi viene chiamato "intuito clinico". Non ultimo è il ruolo affettivo ed umano che il medico ha nei confronti del paziente.

L'avvento dei robot potrebbe compensare anche questa componente del rapporto medico-paziente, per mezzo dell'assistenza alla persona, basandosi sull'implementazione di robot anche antropomorfi. L'autonomia del medico è un punto essenziale nella gestione dei pazienti e la AI potrebbe ipoteticamente alterarne gli equilibri. L'autonomia del medico potrebbe variare entro un intervallo che va dalla totale autonomia umana, passando per un sistema che utilizzi la AI a supporto del medico, al medico che supporti la AI fino ad un grado di indipendenza tale da delegare completamente un'attività medica o chirurgica specifica alla AI. A questo climax ascendente di autonomia della AI probabilmente corrisponde la tendenza verso cui il progresso potrà orientarsi nel corso dei prossimi decenni. Tuttavia, bisognerà dimostrare la sicurezza della AI, il vantaggio in termini di costo-efficacia nonché la diffusibilità e la sostenibilità di tali tecnologie sul territorio e nel tempo.

Nel campo della ricerca l'AI renderà possibile la gestione di enormi flussi di dati (big data) altrimenti non approcciabili dalla mente umana o da sistemi informatici più semplici. L'aspetto interessante potrebbe essere l'integrazione di dati, anche per loro natura differente (biologia molecolare, dati ambientali, biometrici e clinici), per cercare dei parametri singoli piuttosto che dei raggruppamenti (cluster) di biomarker che risultino clinicamente significativi ai fini diagnostici e terapeutici. A tale proposito l'AI permetterà di studiare nuove molecole che possano diventare candidabili come terapie, nonché personalizzare il trattamento valutando il profilo dei pazienti secondo variabili multiple anche sotto l'aspetto molecolare. Nel campo della ricerca clinica, la possibilità di ottenere dati biometrici, preferenzialmente longitudinali, permetterebbe di studiare condizioni fisiologiche o patologiche nonché l'effetto delle terapie non

più da popolazioni selezionatissime di soggetti in condizioni notevolmente standardizzate, bensì dalla popolazione reale con la possibilità di controllare i numerosi effetti confondenti con nuove metodiche non necessariamente statistiche.

Anche il rapporto medico-paziente dovrà necessariamente evolvere con l'implementazione della AI. La crescente autonomia dei sistemi di AI sposterà l'equilibrio fiduciario del paziente dal medico al sistema medico-AI ed infine probabilmente all'AI. Il ruolo della comunità scientifica sarà quello di dimostrare l'attendibilità delle decisioni e delle procedure eseguite dall'AI, nonché l'affidabilità di tali sistemi trovando delle metodologie di certificazione che ne garantiscano l'utilizzabilità nella pratica clinica. Solo in questo modo si potranno delegare mansioni anche complesse in maniera da supportare il rapporto fiduciario tra paziente ed AI. Il paziente si troverebbe a formalizzare un consenso informato che preveda la presa di coscienza dei rischi nell'affidarsi all'AI, i quali potrebbero anche essere in parte imprevedibili. Tale imprevedibilità risulterebbe tanto maggiore quanto più la AI risultasse autonoma e dotata di auto-evolutività, diminuendo così l'intelligibilità degli algoritmi (black-box effect)<sup>5</sup>. A tale proposito, gli sviluppatori che si occuperanno della creazione e dell'implementazione di tali tecnologie dovranno prevedere anche un'attenta valutazione del rischio clinico. Tuttavia, quando pensiamo alla AI con un grado elevato di autonomia, capace per esempio di apprendere ed evolvere autonomamente (Machine Learning), ci troviamo davanti ad un problema di responsabilità. Il danno derivato dall'applicazione della AI potrebbe non essere più imputabile allo sviluppatore, perché il risultato della AI non deriverebbe più da una programmazione umana ma da una evoluzione indipendente della tecnologia. A tale

proposito, è difficile conciliare attualmente il concetto di consenso informato con l'AI e quindi sarà necessario che l'assetto legale evolva insieme all'AI, colmando le lacune legislative attualmente esistenti<sup>6</sup>. Nel percorso di maturazione dell'autonomia di tali tecnologie, i processi decisionali del medico e dell'AI resterebbero regolamentati, in maniera analoga a quanto avviene oggi, dalle linee guida condivise, con le conseguenti ricadute di carattere medico legale. Come già avviene per lo sviluppo e l'implementazione dei farmaci e delle nuove tecnologie/devices, le istituzioni nazionali e sovranazionali avranno un ruolo chiave di selezione e controllo a garanzia della popolazione. L'avvento dell'AI mette in questione il concetto dell'autodeterminazione della persona, intesa come indipendenza ed autonomia nel prendere una decisione. A tale proposito bisogna considerare che l'autodeterminazione viene garantita primariamente da una informazione completa e trasparente erogata al paziente che, nei limiti della propria capacità intellettuale e critica, potrà fare una scelta personale e libera. Sarebbe importante, altresì, garantire al paziente delle alternative non legate necessariamente a tale tecnologia, evitando un atteggiamento monopolistico della AI sulla società. In questa maniera si manterrebbe il ruolo del medico al centro della scelta, tutelando la libertà individuale. La AI cambierà radicalmente la medicina creando un nuovo modo di intenderla e viverla. Sarà un processo progressivo con molti rischi ma che potrebbe portare immensi benefici qualora il cambiamento avvenga nel rispetto dell'etica e della centralità dell'essere umano. Per tale motivo le comunità medica, giuridica e civile dovranno evolvere parallelamente all'AI per garantirne una costruttiva integrazione nella società.

**Elettra Stradella**

Professoressa associata di Diritto pubblico comparato, Università degli Studi di Pisa. Mail: [elettra.stradella@unipi.it](mailto:elettra.stradella@unipi.it)

Crede che per rispondere a questi quesiti occorra partire in primo luogo dallo stato dell'arte tecnologico, e questo mostra il crescente e soddisfacente impiego dell'AI in ambito biomedico: dalle svariate applicazioni in ambito diagnostico (interpretazione dei tracciati elettrocardiografici, previsione di eventi correlati a patologie quali il diabete, o malattie autoimmuni), fino all'elaborazione di paradigmi Brain-Computer Interface (BCI) per l'interpretazione dell'attività cerebrale, settore ancora prevalentemente appannaggio della ricerca più che dell'applicazione medica.

Le condizioni di invecchiamento generale della popolazione, da cui deriva la necessità di confrontarsi con un ampio numero di persone anziane e non-autosufficienti, in molti casi affette da patologie croniche, rendono particolarmente utile il potenziamento di modelli diagnostici e di sistemi di monitoraggio, che in parte già beneficiano, e di più beneficerebbero, degli enormi vantaggi computazionali dell'AI, incidendo (anche) sulla sostenibilità dei sistemi sanitari.

D'altra parte, credo che sia essenziale, prima di concentrarmi sulla questione del rapporto medico-paziente, dell'autonomia della professione medica, e sulle trasformazioni del principio del consenso informato, sottolineare una prima criticità che l'utilizzo dell'AI in ambito sanitario pone, e che è più in generale rilevabile nei sistemi di intelligenza artificiale. In particolare, la presenza di bias derivanti dai dati messi a disposizione dei sistemi, e dunque i fenomeni di amplificazione delle esistenti discriminazioni, rischiano di produrre in ambito sanitario effetti di penalizzazione dei soggetti in relazione a caratteristiche personali ad alto grado di vulnerabili-

tà e stereotipizzazione, quali l'identità sessuale o di genere, l'origine etnica, l'appartenenza religiosa.

Ci sono esempi concreti di questo fenomeno: basti pensare agli algoritmi di previsione utilizzati negli Stati Uniti per l'individuazione di pazienti con esigenze sanitarie complesse (Obermeyer, 2019). In questo caso è stato recentemente mostrato come a un determinato livello di rischio i pazienti neri stiano peggio, siano più malati dei pazienti bianchi, e risolvere questa disuguaglianza sociale e assistenziale comporterebbe un aumento dei pazienti neri che ricevono ulteriori cure fino a quasi il 50%. Ma perché l'algoritmo utilizzato per scegliere verso quali pazienti destinare la spesa sanitaria, in questo caso, produce in maniera così evidente una disparità? Perché automaticamente prevede, e dunque richiede, minori costi per l'assistenza sanitaria per i pazienti neri rispetto ai pazienti bianchi. È probabile che questo derivi, ancora una volta, da una storia di disparità e discriminazioni: la tendenziale minore attenzione alla salute dei pazienti neri, sottoposti ad un'intensità di cura inferiore a fronte di sintomatologie o patologie analoghe o corrispondenti a quelle dei pazienti bianchi, si traduce in un algoritmo che paradossalmente, considerandoli meno bisognosi di cure, li condanna a peggiori condizioni di salute.

D'altra parte, le questioni relative alla qualità dei dati, e alla presenza di bias, sembrano assumere sempre maggiore rilevanza, anche nella riflessione dei tecnologi, ragione per cui non sembra irrealistico ritenere che nei prossimi anni sarà possibile progettare sistemi antidiscriminatori *by default* e *by design*, in grado di superare le criticità delineate.

Una volta limitate o risolte queste criticità, non resteranno dubbi sull'importante funzione predittiva dell'AI in ambito sanitario. Da questo

punto di vista, la capacità predittiva dei sistemi di AI sarà sempre più essenziale, a mio avviso, quando sarà utilizzata per una funzione che definirei: *ingegneristica, sistemica, epidemiologica*.

Basti pensare alla vicenda su cui si concentra in questi giorni tutta la nostra attenzione, il dilagare dell'epidemia di Covid-19, e all'impatto che su fenomeni del genere potrebbe avere l'utilizzo dell'Intelligenza Artificiale.

Come è stato sottolineato da tecnologi e operatori dell'innovazione (Lambrecht, *Global Health and Life Sciences Practice – SAS*) l'AI può agire su diversi versanti: dalla rilevazione dei primi sintomi indicativi di una nuova epidemia prima che questa sia conclamata, con gli ovvi effetti di limitazione della diffusione del virus responsabile (sintomi che non sono necessariamente clinici), al completamento per i risultati clinici, le specifiche complicanze e le caratteristiche di una nuova epidemia, essenziale per consentire al sistema sanitario di rispondere con efficacia, fino all'utilizzo in attività di automazione rivolte a medici e cittadinanza (ad esempio *chatbot* per la segnalazione di sintomi, con capacità di raccolta di informazioni immensamente superiori a quelle che possono mettere in campo gli strumenti tradizionali quali “numeri verdi” o simili). Un simile utilizzo dell'AI in sanità, ed anche, in senso lato, nella ridefinizione del rapporto medico-paziente può condurre a risultati virtuosi quando inserito all'interno di un generale sistema di e-health.

Il discorso cambia quando da un utilizzo “ingegneristico”, “sistemico” e “epidemiologico” si passa alla configurazione dell'AI come vera e propria “assistente cognitiva”, che affianca il medico nelle attività diagnostiche e di cura.

Esempi già ne esistono: basti pensare a Watson, di IBM, programma di intelligenza artificiale rivolto agli oncologi per la definizione di cure (o-

sannato nel 2017 per la sua capacità di diagnosticare un tumore in dieci minuti, e vituperato un anno dopo in un'inchiesta della rivista *Stat* che ne metteva in luce la notevole fallibilità derivante, probabilmente, ancora una volta dai dati, perché le sue decisioni sarebbero state assunte sulla base di dati riguardanti un numero limitato di medici e pochi ospedali), o all' algoritmo “Medical Sieve”: un sistema di “assistenza cognitiva automatizzata” dedicata a radiologi e cardiologi, disegnato per aiutarli proprio nella loro attività diagnostica e decisionale. Il sistema, raccogliendo dati clinici, documentali e derivanti da tecniche di *imaging* dei pazienti, li analizza rintracciando, se ve ne sono, anomalie. I risultati della sua ricerca alimentano un algoritmo che utilizza elementi provenienti sia da risultanze cliniche indipendenti del singolo paziente, sia da statistiche relative a pazienti simili, ovviamente nella misura resa possibile dalla potenza computazionale del sistema, per giungere ad una potenziale diagnosi differenziale che orienti il medico nella sua decisione.

Ma un recente stato dell'arte pubblicato su *Nature* mostra la molteplicità delle possibili applicazioni assistenziali e diagnostiche, tra cui il progetto di diagnosi umana *Human Dx*, strumento decisionale basato sull'AI che si avvale dei Big Data, accumulabili attraverso il *crowdsourcing* che coinvolge a livello globale medici e ricercatori. Il progetto ha ricevuto centinaia di migliaia di contributi da decine di migliaia di professionisti distribuiti in oltre 80 Paesi: l'idea è che l'intelligenza collettiva, potenziata dalla capacità computazionale che solo l'AI può mettere in campo, possa condurre ad un successo diagnostico incomparabile rispetto a quello conseguibile dai singoli medici posti di fronte al caso. (McCall, 2020).

È di queste ore la notizia di algoritmi (Alibaba) sviluppati proprio per l'individuazione



dell'infezione da Covid-19 in grado di diagnosticare in 20 secondi la malattia, grazie a sistemi di AI alimentati dai dati campione di più di 5.000 casi confermati di infezione: l'algoritmo batterebbe sul tempo dunque il medico, più lento nell'interpretare un'indagine tomografica e comunque bisognoso di numerose immagini.

Si potrebbero fare molti altri esempi, ma bastano questi, nello spazio a disposizione, per comprendere che l'interazione tra medico e AI pone una serie di problemi che riguardano in particolare la responsabilità del primo, da diversi punti di vista: quello privatistico della responsabilità derivante da danni provocati nei confronti di terzi; quello dell'*accountability* del medico, a cui si collega il rischio della deresponsabilizzazione che, in questo ambito, potrebbe tradursi in un appiattimento decisionale e diagnostico; e infine quello della responsabilità nella relazione di cura, da cui deriva il valore attribuito all'autodeterminazione del paziente.

Il primo punto di vista è quello più "tradizionale", sempre toccato quando ci si interroga sull'utilizzo della robotica e dell'AI, nei vari settori. Questo non significa che sia di immediata soluzione. Com'è noto la dottrina, con particolare riferimento all'ambito medico, utilizza l'art. 2050 del codice civile, cioè la fattispecie della responsabilità per l'esercizio di attività pericolose (Topi, 2016; Parziale, 2019), che pure potrebbe non ritenersi del tutto adeguata a descrivere le peculiarità dell'AI, la cui "pericolosità" non deriva certamente dall'oggetto dell'attività che, in questo caso, altro non sarebbe che la stessa attività diagnostica a cui è chiamato il medico, ma semmai dalle modalità con cui quell'attività è realizzata (statistiche, probabilistiche, matematiche, basate su dati le cui modalità di raccolta ed elaborazione non sempre sono chiare e definite, sia che siano *supervised* sia che siano *unsupervised*, ciò da cui

deriva il noto problema della c.d. *black box*). D'altra parte, nelle modalità sta la natura dei sistemi di AI, che sarebbero così considerati intrinsecamente pericolosi. D'altra parte, anche il richiamo all'opportunità di distinguere, nell'identificazione della responsabilità, tra attività concretamente esercitata dal medico e funzionamento del sistema, pone una serie di dubbi, derivanti in primo luogo dallo stretto legame che si instaura tra i due momenti quando l'AI abbia una funzione di "assistenza cognitiva", come si diceva sopra. È immaginabile di poter operare, magari in sede giudiziale, una quantificazione/qualificazione del contributo effettivamente prestato da parte del medico e, per converso, di quello derivante dalla "macchina"? Ha senso chiedersi se l'errore diagnostico sia derivato da un cattivo funzionamento dell'algoritmo, dai dati che ne hanno in parte determinato il risultato, dalla lettura che il medico ha dato del risultato e dall'applicazione che ha inteso farne? Ha senso immaginare di applicare all'AI categorie giuridiche come quelle dell'"oggetto difettoso"? Credo che le risposte a questi interrogativi debbano essere negative. Più logica mi sembra invece la considerazione per cui l'individuazione della responsabilità non sarebbe in fondo diversa da quella che avviene per qualsiasi tipo di decisione medica supportata da strumenti, informazioni, dati, che siano esterni all'autonoma capacità e percezione del medico. Dagli strumenti di diagnostica per immagini fino alle analisi del sangue, il medico si avvale abitualmente di qualcosa che è "altro da sé", ma non per questo è messa in discussione la sua responsabilità diagnostica e di cura; lo stesso potrebbe accadere con l'Intelligenza Artificiale. Il problema però qui è un altro, a mio avviso, e sta nel fatto che con riferimento all'AI risulta particolarmente difficile per il medico verificare

il corretto funzionamento del dispositivo di Intelligenza Artificiale, e dunque il vero rischio sembra essere quello, avanzato anche per l'utilizzo dell'AI nella decisione giudiziaria, di una deresponsabilizzazione di fatto del professionista, che inevitabilmente ricalcherà nelle sue valutazioni i risultati provenienti dal dispositivo intelligente, per la difficoltà di discostarsi da qualcosa che non è in grado di mettere in discussione razionalmente e fondatamente. Anche in questo caso allora, come nelle applicazioni giudiziarie, assume un ruolo centrale quel "diritto alla spiegazione" che si configura come diritto all'intervento umano come diritto in quanto tale, quindi sorretto da una reale comprensione della natura della decisione, medica in questo caso, dei processi che ne stanno a fondamento, del livello di automatizzazione, dei dati che ne hanno orientato il contenuto. Sappiamo che oggi molti *data scientists* stanno lavorando per aprire la *black box*, ma sicuramente il percorso non è compiuto.

La questione allora non riguarderebbe tanto o solo la responsabilità per danni causati nei confronti di terzi, ma la ridefinizione della *soggettività* stessa del medico.

Sotto questo profilo, pur con sfaccettature diverse, la questione della "*spersonalizzazione*" del rapporto medico-paziente si avvicina per certi versi a quella del rapporto giudice-imputato, e non riguarda soltanto la tematica, prettamente civilistica, della responsabilità del medico, rispetto alla quale non sono inedite infatti le fattispecie della responsabilità da "contatto sociale", o il c.d. "contratto di speditività", in cui il rapporto tra medico e paziente sembra concretizzarsi in un obbligo di prestazione sempre e comunque destinata, se esercitata a regola d'arte, a condurre verso la guarigione, come se l'eventualità della morte non venisse, in fondo, neppure presa in considerazione. Riguarda

soprattutto *la relazione*, e come il dispiegarsi della relazione è in grado di incidere sul diritto all'autodeterminazione dei pazienti.

È interessante, da questo punto di vista, la recente sentenza della Cassazione n. 10423 del 2019, in cui, a fronte della richiesta del ricorrente volta ad ottenere un ristoro non soltanto per il danno alla salute conseguente alla mancata informazione, ma per quello determinato dalla «lesione del diritto all'autodeterminazione terapeutica in sé considerato, rispetto al quale il carattere necessitato dell'intervento e la sua corretta esecuzione restano circostanze prive di rilievo», la Corte sottolinea, riprendendo i suoi precedenti, che «in tema di attività medico-chirurgica, è risarcibile il danno cagionato dalla mancata acquisizione del consenso informato del paziente in ordine all'esecuzione di un intervento chirurgico, ancorché esso apparisse, *ex ante*, necessitato sul piano terapeutico e sia pure risultato, *ex post*, integralmente risolutivo della patologia lamentata, integrando comunque tale omissione dell'informazione una privazione della libertà di autodeterminazione del paziente circa la sua persona, in quanto preclusiva della possibilità di esercitare tutte le opzioni relative all'espletamento dell'atto medico e di beneficiare della conseguente diminuzione della sofferenza psichica, senza che detti pregiudizi vengano in alcun modo compensati dall'esito favorevole dell'intervento». E com'è noto quando la Corte costituzionale, nel 2008, afferma la fundamentalità del diritto al consenso informato, ne definisce il contenuto nella «sintesi di due diritti fondamentali della persona: quello all'autodeterminazione e quello alla salute». Dunque se l'attività del medico trova il suo fondamento nel consenso, e il consenso, per come costruito dalla giurisprudenza, è chiaramente definito attraverso la relazione che si instaura tra medico e paziente, non risolvendosi in una

*Forum*

mera attività informativa o un adempimento burocratico, ma il consenso non soltanto è funzionale all'attuazione del diritto all'autodeterminazione, ma quest'ultimo ne definisce il contenuto stesso, possiamo dire che la professione medica abbia una specificità che sta nelle modalità con cui può essere misurato il risultato dell'azione finalizzata alla tutela del diritto alla salute. Come scrive la Cassazione l'assenza del consenso informato rende «irrilevante il fatto che l'intervento medico sia stato eseguito in modo tecnicamente e oggettivamente corretto».

*AdS*

Il pregiudizio arrecato al paziente non informato, non ascoltato, limitato nella sua autodeterminazione, non troverebbe compensazione, in questa visione, nell'esito favorevole dell'intervento chirurgico, nel caso di specie, in generale nella cura. Ciò significa che il compito del medico non è soltanto quello di compiere un'attività diagnostica, di prendere delle decisioni di cura ed applicarle, e, in fin dei conti, non è neppure soltanto quello di guarire, ma si sviluppa, come detto, nella dimensione relazionale.

## 4. Medicina e giustizia: ambiti paradigmatici

### 4.2 La giustizia

#### Massimo Luciani

*Professore ordinario di Istituzioni di diritto pubblico, Università degli Studi di Roma – La Sapienza. Mail: [massimo.luciani@uniroma1.it](mailto:massimo.luciani@uniroma1.it)*

#### 4.2.a Dipende dal suo uso concreto.

La semplice utilizzazione di repertori indicizzati pone meno problemi, anche se si deve considerare che, se non critica e sorvegliata, essa finisce per introdurre uno strisciante vincolo al precedente, per limitare l'adattamento degli indirizzi giurisprudenziali al mutare della realtà sociale e per incidere nelle attitudini culturali del magistrato (che tende ad acquietarsi nella raccolta dei precedenti, trascurando l'approfondimento sulla dottrina e lo studio dei contesti).

Diversa è l'ipotesi dell'affidamento all'AI di una parte del lavoro del giudice, ad esempio quanto alla determinazione della misura di una sanzione in base a paradigmi prefissati. Va da sé che in questo caso sarebbe pregiudicato quel principio di adeguatezza in concreto sul quale tanto ha insistito la nostra Corte costituzionale nella sua giurisprudenza sulle pene fisse. E l'esperienza pratica già maturatasi in alcuni ordinamenti non induce all'ottimismo (si registrano, infatti, casi di paradossale discriminazione).

**4.2.b** In parte vale la risposta al quesito precedente. Quanto alla possibilità di sostituire un giudice umano con un giudice *robot*, fosse pure per controversie bagatellari, sono contrario. Il processo decisionale diventa, così, imperscrutabile, risolvendosi la sua eventuale critica nella difficilissima contestazione (non già della logica di un argomentare giuridico, ma) di un algoritmo. Operazione, questa, che sfugge al giurista e

che meglio compete al tecnico, con conseguenze agevolmente intuibili anche quanto al posizionamento sociale degli operatori del diritto.

## Alessandro Pajno

*Presidente Emerito del Consiglio di Stato e docente di diritto processuale amministrativo, Luiss Guido Carli.*  
 Mail: [a.pajno@giustizia-amministrativa.it](mailto:a.pajno@giustizia-amministrativa.it)

L'AI incontra la giustizia sia come sistema complesso istituzionale, sia come singolo processo, volto ad attribuire il bene della vita ad uno dei contendenti o a stabilire l'innocenza o la colpevolezza dell'imputato.

Sotto il primo profilo, va rilevato che il rapporto fra AI e funzionamento del sistema giudiziario è iniziato da tempo attraverso la progressiva dematerializzazione dei documenti e l'introduzione del processo telematico in grandi settori, come il processo civile e quello amministrativo. L'innovazione tecnologica è stata, all'inizio, vista come elemento decisivo per l'aumento dell'efficienza; successivamente si è realizzato uno spostamento sul potenziale cognitivo dell'innovazione, grazie alla capacità computazionale degli strumenti algoritmici. Una utilizzazione accorta di questi sistemi può consentire di fornire risposte a problemi prima risolti con strumenti soltanto normativi, quali la riduzione del contenzioso attraverso rimedi alternativi automatizzati, la riduzione del margine di errore attraverso la valutazione preventiva del rischio di soccombenza, la riduzione dei tempi attraverso la trattazione automatizzata delle controversie seriali, la riduzione del margine di differenziazione geografica delle risposte alle domande di giustizia. La trattazione attraverso strumenti algoritmici di controversie civili di modesto valore può essere considerata accettabile solo in presenza di una preventiva consapevole accettazione delle parti. Va, peraltro, tenuto presente che, a fronte delle applicazioni algoritmiche a contenuto probabilistico, la decisione non deriva semplicemente dal ragionamento deduttivo, ma è basata sulla considerazione di una pluralità di elementi. Una situa-

zione del genere implica necessariamente l'accettazione del margine di errore.

Sotto il secondo profilo, anche il processo è chiamato a confrontarsi con la rivoluzione digitale. Significativo appare l'apporto fra strumenti algoritmici e formazione della prova, specie nel processo penale. L'impiego di tali strumenti può comportare una opacità del dato probatorio, anche in relazione all'asimmetria conoscitiva delle parti nel processo, legata nella tradizione italiana (ed anche in quella di altri paesi), alla presenza di una parte pubblica. Tale opacità rischia, alla lunga, di sottrarre la prova stessa al confronto dialettico fra le parti. Risulta, pertanto, indispensabile valorizzare, in ogni sua forma, il principio di parità delle armi nel processo.

La prova algoritmica può presentare rischi di innattendibilità legati al piano della teoria scientifica sintetizzata nell'algoritmo e rischi legati alla stessa costruzione di quest'ultimo. Il contraddittorio processuale dovrà, pertanto, riguardare entrambi questi profili.

Con riferimento all'indipendenza del giudice deve essere preso in considerazione l'impatto che gli strumenti di *open access* associati allo strumento algoritmico potrebbero avere sul valore del precedente. È stato segnalato che l'esistenza di una rete neurale che raccolga l'insieme dei precedenti, anche non vincolanti, potrebbe in qualche modo compromettere l'evoluzione giurisprudenziale, scoraggiando il ricorso a posizioni non assistite dal consenso prevalente, così mettendo in discussione l'affermazione, nella stessa giurisprudenza, di nuovi paradigmi interpretativi.

Un ruolo specifico deve essere riconosciuto al processo amministrativo come luogo del sindacato sul potere pubblico quando si serve di algoritmi. Assume, così, rilievo il sindacato giurisdizionale sulle decisioni amministrative algoritmiche, rispetto alle quali la giurisprudenza



amministrativa (Cons. St., 2270/2019; 8472/2019), se da una parte ha segnalato che la relativa utilizzazione va ritenuta una doverosa declinazione dell'art. 97 Cost., ha, dall'altra precisato che:

- 1) l'utilizzo di procedure informatiche non può costituire elusione dei principi che regolano lo svolgimento dell'attività amministrativa;
- 2) non vi sono ragioni per limitare l'uso dell'algoritmo all'attività vincolata;
- 3) l'attività amministrativa algoritmica è soggetta ai principi di piena conoscibilità, di non esclusività della decisione algoritmica (è sempre necessaria la partecipazione umana), di non discriminazione algoritmica.

Anche se la giurisprudenza amministrativa va elaborando principi rilevanti in materia, rimangono moti punti da approfondire, tra cui quello relativo alla conoscibilità dell'algoritmo anche in relazione alle esigenze di riservatezza delle imprese produttrici ed alla stessa possibilità di un sindacato a fronte del *deep learning*.

È stato, altresì, osservato che il diffondersi di strumenti di *quantitative legal prediction* potrebbe provocare anche nei sistemi europeo-continentali il rischio di una rigida centralizzazione del precedente e potrebbe indurre nel singolo giudice il timore di discostarsi dalla decisione preconizzata dallo strumento predittivo, con conseguenze sull'indipendenza del medesimo giudice. Si ritiene, tuttavia, che la tradizione e la cultura dell'indipendenza, ben presenti nell'ordine giudiziario del Paese, costituiscano un antidoto rispetto al rischio prospettato.