Statuto Etico e Giuridico dell'IA

FONDAZIONE LEONARDO
CIVILTÀ delle MACCHINE

Statuto Etico e Giuridico dell'IA

Documento introduttivo

La Conferenza sullo statuto della IA e della robotica è uno spazio aperto agli studiosi, ai tecnici e ai cittadini interessati. L'I.A. non è un terreno destinato a caste ristrette. È la madre delle grandi innovazioni del nostro tempo, coinvolge l'intera umanità e sfida il presente in termini radicali.
La Fondazione Leonardo-Civiltà delle Macchine intende essere parte di questa sfida e intende contribuire alla definizione dei parametri etici e giuridici necessari per favorirne lo sviluppo, orientandolo alla massimizzazione dei benefici e dalla minimizzazione dei rischi.
Per questa finalità, consapevole della complessità della materia, e dell'elevato numero di importanti documenti sul tema, la Fondazione ha individuato l'umanesimo digitale come indirizzo di fondo del lavoro e ha selezionato quattro grandi campi, la finanza, la giustizia, la salute, la sicurezza, come terreni prioritari del proprio impegno anche per il futuro.
Per la partenza ci siamo avvalsi del lavoro di Maria Chiara Carrozza, Alessandro Pajno, Stefano Quintarelli, e dei loro collaboratori, per avere tre documenti autorevoli che pongono a fuoco le questioni generali di carattere tecnico (Carrozza), etico (Quintarelli), giuridico (Pajno). La lezione magistrale di Luciano Floridi, OII's Professor, University of Oxford, a conclusione della prima giornata dei lavori, segnerà l'indirizzo scientifico e ideale della Conferenza.

Le grandi innovazioni tecnologiche, come tutte le grandi scoperte, non si limitano ad aggiungere novità all'esistente, ma lo plasmano, gli danno forma e contenuti nuovi: esigono perciò pensieri nuovi e nuove forme organizzative della società e dei poteri pubblici.
Di fronte a queste innovazioni si aprono due conflitti.
Le società si trovano prigioniere della tensione tra conservazione e innovazione. L'esistente, quando si oppone a ciò che lo minaccia, difende sé stesso; ma allo stesso tempo è affascinato dal cambiamento. Perciò il conflitto tra il vecchio e il nuovo è la forma che assume in ogni tempo il progresso civile e tecnologico delle civiltà. All'interno di questa tensione maturano le esigenze di *governance* con le contrapposte tendenze al freno attraverso misure stringenti o all'incentivazione attraverso regole flessibili.
La forza dirompente della IA e della robotica propone il secondo conflitto, tra tecnocentrismo e antropocentrismo. Il tema si affronta attraverso l'intreccio dei saperi e delle esperienze tanto umanistiche quanto tecnologiche. Un intreccio che non può essere indifferente agli esiti e che perciò va orientato. L'indifferenza, infatti, produrrebbe ineluttabilmente una sorta di totalitarismo tecnologico. Una evoluzione di questo tipo relegherebbe l'uomo in una dimensione marginale rispetto alla tecnologia e a chi lo governa. È necessario invece che i sistemi di intelligenza artificiale si conformino ad un approccio antropocentrico, a beneficio della collettività. Le innovazioni devono essere programmate mettendo al centro l'uomo, il rispetto dei suoi diritti, dei principi e dei valori propri di

una società democratica. Al contempo devono assicurare la propria affidabilità dal punto di vista tecnico sotto il profilo della sicurezza e della capacità di essere utilizzate in modo trasparente. Si avverte la necessità che l'etica e il diritto "guidino" e "orientino" la tecnologia, onde permetterle di percorrere binari rispettosi del sistema di valori oggi riflessi nelle Costituzioni e nei trattati internazionali.

In una società antropocentrica la consapevolezza delle opportunità e dei rischi legati alla tecnologia dipende dall'educazione e dalla formazione tecnologica. È necessario educare la società e le persone ad un uso corretto e ad una coesistenza matura con la tecnologia. La formazione degli utenti deve permettere un uso consapevole delle innovazioni che ne evidenzi le potenzialità responsabilizzando l'individuo di fronte ai rischi. Educare significa saper inquadrare i rapporti tra persona e tecnologia, trasmettendo i saperi che permettono alla prima di conoscere come funziona la tecnologia e quindi di valutarne rischi e potenzialità. La società deve poter accedere a percorsi formativi per qualificarsi e riqualificarsi. L'innovazione, inoltre, ha bisogno di professionisti adeguati alla difficoltà delle sfide che la tecnologia pone a livello tanto etico quanto sociale. In tale contesto, la questione della formazione assume un vero e proprio carattere democratico, dipendendo da essa il concreto accesso delle persone ai vantaggi dell'intelligenza artificiale. Una diversa prospettiva potrebbe introdurre non accettabili situazioni di privilegio. L'uso dell'intelligenza artificiale può avere, infatti, conseguenze significative sull'uguaglianza sociale.
Rivendicare una tecnologia a misura d'uomo esige idee forti e politiche altrettanto forti per tre obbiettivi principali di carattere etico e giuridico: non porre impedimenti irragionevoli allo sviluppo, evitare che esso sia dannoso per l'umanità e per i singoli, favorire la piena utilizzazione da parte di chiunque.
Etica e diritto non sono la macchina, sono i *guard rails* che impediscono alla macchina costruita dalla tecnica di uscire fuori strada.

È in corso, come spiega il *paper* curato dal gruppo di lavoro coordinato da Maria Chiara Carrozza, un processo di *socializzazione* della robotica, determinato dall'ingresso del robot nelle nostre fabbriche e nelle nostre case. Noi deleghiamo alla macchina lo svolgimento di compiti con implicazioni sia fisiche che cognitive in modo che possa sostituirci nel compiere determinate azioni, sul posto di lavoro o in casa. Questo ingresso del robot nella vita di ogni giorno comporta la necessità di utenti addestrati. Il cittadino comune ha il dovere di istruirsi. Negli anni Sessanta del secolo scorso un enorme contributo alla alfabetizzazione del nostro Paese venne da una trasmissione del pomeriggio, "Non è mai troppo tardi", condotta da uno straordinario insegnante, il maestro Manzi che riuscì a far prendere la licenza elementare a circa un milione e mezzo di italiani. Negli anni Venti del nuovo secolo è necessaria una seconda

alfabetizzazione, l'alfabetizzazione digitale che insegni al cittadino comune l'uso della IA e della robotica. Se l'I.A. deve essere sviluppata al servizio dell'uomo, l'uomo dev'essere messo in grado di servirsene consapevolmente: il grado di autonomia di un sistema di IA dev'essere sempre definito con chiarezza e i diversi livelli di controllo umano devono convertirsi in regimi giuridici appropriati ai livelli di utilizzo.
Forse il servizio televisivo pubblico dovrebbe prendere in considerazione questa necessità sociale occupandosi anche dei cd "nativi digitali", che fanno uso intensivo delle nuove tecnologie, ma non sono interessati al loro funzionamento.
L'uomo della società del secolo scorso aveva il dovere di munirsi della istruzione necessaria per vivere in quel sistema di relazioni. L'uomo di questo secolo è condannato alla emarginazione se non sa usare la tecnologia, come il suo predecessore era condannato all'emarginazione se non sapeva leggere e scrivere. D'altra parte, come spiega il *paper* redatto dal gruppo di lavoro coordinato da Stefano Quintarelli, l'operatività dei sistemi autonomi dal controllo umano dev'essere subordinata alle necessità, alle aspirazioni e ai diritti delle persone, quali risultano, per utilizzare le parole della Corte di giustizia dell'Unione Europea, dalle cd *tradizioni costituzionali comuni*. Un approccio del genere è, d'altra parte, necessario perché la tecnocrazia rimanga sempre al servizio dell'uomo. Uno sviluppo della IA e della robotica preoccupata dell'avvenire dell'uomo induce a soffermarsi rigorosamente sui diritti che la tecnica e la politica devono rispettare. Questo approccio può essere integrato facendo riferimento ai doveri di chi deve assicurare il godimento di quei diritti. Accessibilità, trasparenza, affidabilità, non discriminazione, dignità delle persone e delle collettività, riservatezza, identità, coesione sociale e pluralismo sono valori che esprimono altrettanti diritti. Perché questi diritti diventino sostanza nella vita delle persone è necessario che specifici soggetti istituzionali, anche sovranazionali, ne assicurino il godimento. Può trattarsi dello stesso utente che non può usare la tecnologia per danneggiare l'altro, come chi guida l'auto non può usarla per recare danni al prossimo. Lo Stato deve garantire un livello di istruzione digitale di base, l'operatore deve evitare che i suoi pregiudizi diventino criteri discriminatori per l'algoritmo, l'azienda costruttrice deve assicurare la qualità della macchina.

Particolare significato assumono i doveri delle imprese. Alcune di esse possono rivaleggiare con gli Stati per capacità finanziarie, possibilità di orientamento dei privati, capacità di condizionamento dei poteri pubblici. Esse devono adottare scelte incentrate sulla necessità che le innovazioni servano allo sviluppo civile. Tra i doveri delle imprese rientrano:

I. l'obbligo di garantire la sicurezza dei sistemi di IA a diversi livelli di applicabilità: dalla tutela della sicurezza degli individui

alla conservazione dei dati personali, dalla protezione e gestione degli *asset* fisici all'integrità strutturale del sistema.

II. Il dovere di assicurare la fruizione dei sistemi di IA anche a speciali categorie di utilizzatori, come non autosufficienti, portatori di handicap e minori, al fine di garantire loro un utilizzo consapevole della tecnologia in linea con la massima espressione del loro potenziale.

III. La creazione di buone pratiche per la prevenzione del danno relativo all'uso di sistemi di IA, attraverso procedure di *risk assessment* e *management*, per identificare situazioni critiche e prototipare scenari di rischio utilizzabili dall'IA.

È opportuno considerare inoltre l'introduzione di meccanismi di riparazione (*redress*) sulla base di un principio di "*redress by design*". Un sistema di IA non difettoso, perfettamente funzionante, essendo un motore statistico che produce risultati probabilistici, può effettuare predizioni errate. In questi casi, la procedura di ricorso può non esistere o, se esiste, può essere inefficace. Al fine di garantire un'efficace tutela dei diritti, le imprese dovrebbero prevedere, fin dalla fase di progettazione, la creazione di meccanismi atti a garantire procedure alternative per poter individuare efficacemente, verificare e correggere le decisioni sbagliate prese da un sistema non difettoso.

La riparazione (redress) è necessaria, ma non è sufficiente. Un recente importante rapporto del Consiglio d'Europa (Responsability and AI, 2019) sottolinea le necessità che gli Stati si impegnino a garantire i diritti umani nei confronti del crescente potere delle grandi aziende della Big Tech, soprattutto per la preoccupante asimmetria tra queste aziende ed il singolo cittadino. Si tratterebbe di meccanismi istituzionali diretti a prevenire i rischi che da un uso irresponsabile dell'I.A. possono derivare ai diritti fondamentali del singolo e all'intera democrazia.

La diffusione di sistemi di intelligenza artificiale pone quindi questioni di ordine costituzionale, costringendo ad interrogarsi sulla persistenza di idoneità delle norme contenute nella Carta fondamentale, integrate e inserite in un sistema di *governance* di alto livello, a formare un argine rispetto all'esercizio di un potere che si ritrova sempre più concentrato nelle mani di operatori privati e non promana, invece, da autorità pubbliche. Se il rapporto tra Stato e cittadino si forma sul riconoscimento, in capo a quest'ultimo di un sistema di garanzia a fronte dell'illegittimo esercizio del potere pubblico, la dinamica sopra descritta impone un ripensamento alla luce delle caratteristiche del potere privato di chi utilizza tecnologie algoritmiche e sistemi di I.A. per realizzare attività che presentano implicazioni rilevanti per la vita e la libertà delle persone. Questo potere privato, pur libero di svilupparsi, pone una esigenza

di conformazione e di regolazione, da realizzarsi eventualmente con forme di *hard-law* e con l'intervento di appositi soggetti indipendenti. Le garanzie del cittadino vanno d'altra parte affermate anche nei confronti dei poteri pubblici, che a loro volta possono essere fruitori o utilizzatori dei sistemi di I.A. per le finalità più diverse, quali quelle legate alle attività della PA o all'organizzazione del servizio giustizia.

Strettamente connessa al tema dei diritti e dei doveri è la questione della *governance*. Le iniziative finora susseguitesi nell'ambito giuridico-regolatorio, documentano una frammentarietà di linee guida tendenzialmente settoriali e la carenza di interventi di carattere organico. La regolazione si trova ancora in una fase embrionale, caratterizzata da interventi sparsi, in cui sono declinati perlopiù principi a tutela del cittadino, ma spesso privi di una *ratio* sistematica.
Il *paper* redatto dal gruppo di lavoro coordinato da Alessandro Pajno dimostra come codici di condotta e meccanismi di certificazione presuppongano un'autorità, possibilmente indipendente, investita di una funzione pubblicistica, che possa assurgere a garante di uno sviluppo dei sistemi di intelligenza artificiale, in conformità a linee e orientamenti condivisi a livello sovranazionale, in modo da preservare la centralità della persona umana, della sua dignità e dei diritti fondamentali. Tale soggetto, posto all'esterno del circuito politico, indipendente rispetto al governo e caratterizzato dalla tecnicità dei propri comportamenti, potrebbe svolgere un'attività di *public regulation* o comunque volta alla definizione di una disciplina comune – neutrale rispetto alle specifiche sensibilità costituzionali – intervenendo specificamente sugli standard e riconoscendo quali siano idonei a garantire la tutela dell'individuo rispetto all'uso dell'intelligenza artificiale. È opportuna inoltre la costituzione di comitati con la partecipazione dei privati produttori che possono offrire occasioni di consultazione permanente in materie essenziali per aggiornare e per superare la costante difficoltà di inquadrare entro fattispecie giuridiche i progressi della società algoritmica.
Sul piano del metodo si ritiene che un processo di *deregulation* e delegificazione attraverso atti normativi di rango secondario entro un quadro opportunamente definito dal legislatore, possa raggiungere più efficacemente gli obiettivi regolatori. La rapidità dei mutamenti tecnologici e la necessità di disporre di *set* di norme con elevato contenuto tecnico sconsigliano l'adozione delle procedure legislative tradizionali. Spunti non privi di utilità potrebbero derivare dall'esperienza maturata in occasione dell'introduzione di normative riguardanti nuove tecnologie, come, ad esempio, internet e le biotecnologie.
Potrebbe rivelarsi utile la considerazione dell'esperienza di alcuni paesi baltici, più avanzati nelle politiche dell'innovazione e in particolare nello sviluppo del digitale. In questi Paesi sono state avviate politiche all'avanguardia elaborando anche progetti per il

riconoscimento di personalità giuridica alle macchine ed addirittura proposte per l'introduzione di "robot magistrati" destinati a occuparsi di cause civili di minore entità. Al di là della possibile problematicità di alcune scelte, come quella legata al riconoscimento della personalità giuridica delle macchine, allargare lo sguardo a tali ordinamenti può servire ad aggiornare le opzioni e a comparare con continuità tecnologie, utilizzazioni e criteri guida.
Lo spazio fisico, infine, è destinato ad essere affiancato dalla *cybersfera*, apparentemente immateriale, nel quale si muovono algoritmi e bot, ma anche regole, principi, politiche volti a definire e regolare nuovi formidabili poteri, quali quelli connessi con l'uso dell'intelligenza artificiale. Essere consapevoli di quanto avviene attorno a noi è necessario per poter accedere alla *cybersfera* in condizioni di parità.

Paper sui principi tecnici

'Automazione e Autonomia: dalla definizione alle possibili applicazioni dell'Intelligenza Artificiale'

Maria Chiara Carrozza con il contributo di Calogero Oddo, Simona Orvieto, Alberto di Minin, Gherardo Montemagni, Scuola Superiore Sant'Anna

INTRODUZIONE

Questo elaborato intende fornire un quadro di sintesi per comprendere il significato e l'importanza dell'Intelligenza Artificiale, quale tecnologia caratterizzante la quarta rivoluzione industriale e che con alta probabilità rappresenterà, insieme alla robotica e ad altre tecnologie abilitanti ICT, la chiave di lettura per le dinamiche socio-economiche del prossimo futuro.
Da un punto di vista metodologico, il documento si propone di fornire definizioni chiave e comprensibili, da sfruttare nei vari contesti ed utilizzabili nelle molteplici applicazioni di questa tecnologia, tra le altre: medicina, istruzione, sicurezza e difesa.
L'applicazione dell'intelligenza artificiale comporterà cambiamenti anche di natura culturale e comportamentale. Tali ricadute rendono ancora più importante la comprensione delle conseguenze che l'Intelligenza Artificiale (*Artificial Intelligence*, AI) potrà avere in ambito tecnologico, sociologico, politico e giuridico.
Nella redazione dell'elaborato, l'obiettivo ultimo è di definire, nel senso di cui sopra, l'Intelligenza Artificiale, mediante un percorso di analisi che parte dall'elemento alla base di questa tecnologia: il dato, trattando caratteristiche essenziali, potenzialità e limiti di quello che viene definito oggi come "nuovo oro" (Big Data). Seguirà l'analisi delle tecnologie Cloud (Cloud computing), con lo stesso schema logico adottato per il dato. Infine, la trattazione dell'Algoritmo come elemento di congiunzione tra dato e il funzionamento dell'Intelligenza Artificiale. In particolare, verranno discusse le principali tecniche di Machine Learning (ML) e la loro relazione con lo sviluppo dell'AI.
Tali elementi sono propedeutici alla comprensione del concetto di AI, esplicando attraverso il ricorso alle più autorevoli e accreditate definizioni di AI con l'intento di fornire al lettore una guida sintetica e neutrale, per orientarsi nel settore. In conclusione, nel tentativo di fornire concretezza, verrà proposto uno sguardo temporale tra il passato, il presente e il futuro dell'Intelligenza Artificiale, con riguardo ad alcuni casi d'uso e possibili ambiti di applicazione.

IL DATO

Come si definisce il dato? E quali sono gli interrogativi da porsi per gestire al meglio le complessità derivanti dall' acquisizione per ottimizzarne l'utilizzo?

LE CARATTERISTICHE ESSENZIALI DEL DATO

Il dizionario Treccani definisce i dati, con uso più specifico in informatica, come elementi di un'informazione costituiti da simboli (numeri, lettere: *d. numerici*, *alfabetici*, *alfanumerici*) che devono es-

sere elaborati, per lo più elettronicamente, secondo un determinato programma.
Possono esistere diversi tipi di dato, con peculiarità differenti tra loro, espresse in maniera esaustiva nel Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, curato dalla Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale

I tipi di dato includono:
- Dati osservazionali;
- Dati sperimentali di laboratorio;
- Dati relativi a simulazioni al calcolatore;
- Dati relativi ad analisi di testi;
- Oggetti fisici o reperti.

Aspetti chiave da analizzare:
- Completezza e attendibilità dei dati;
- Attenzione alla qualità, relativa anche al controllo della forma (formattazione, unità di misura...) dei dati;
- Distribuzione e accesso ai dati;
- Definizione e progettazione di ontologie condivise;
- Importanza della collocazione e della fisicità del datacenter (dove i dati vengono conservati);
- Attribuzione di un valore economico ai dati;
- Procedure di accesso ai dati;
- Gestione della produzione dei dati;
- Disciplina normativa dell'utilizzo dei dati.

La sfida associata al tema dei dati è principalmente legata alla creazione delle condizioni, anche organizzative, che consentano all'AI di utilizzarli in maniera corretta, garantendo: consistenza, qualità, intelligibilità e quantità. Queste caratteristiche sono fondamentali al fine di rendere il dato fruibile; si pensi ad un form per una transazione online: i campi devono essere ben strutturati, con tipi di dati adeguati. Se, per indicare l'importo di un acquisto, il campo di un dato numerico lasciato come tipo testuale, si potrebbero presentare ambiguità nell'interpretare il dato stesso, perché l'utente potrebbe immettere liberamente virgola o punto come separatore dei decimali, la valuta prima o dopo del numero, e determinare altre ambiguità e conseguenti complessità gestionali: 6.74 € oppure 6,74 € oppure € 6.74 oppure € 6,74 oppure 6.74 Euro, e così via. Risulta quindi fondamentale standardizzare i dati già dal momento della loro prima immissione, in modo da ridurre gli errori e minimizzare i costi di gestione per i livelli successivi, dai database ai sistemi di intelligenza artificiale.
Altri concetti fondamentali nella gestione dei dati sono quelli di latenza e velocità di trasmissione. La latenza (*latency*, misurata tipicamente in s) è il tempo di risposta, che intercorre dal momento che un dato è richiesto fino a quando il dato è reso disponibile all'u-

tente. La velocità di trasmissione (*throughput*, misurato tipicamente in bytes/s) invece è la quantità di dati che si possono trasferire nell'unità di tempo. Nelle reti internet ad alta velocità attualmente disponibili, *latency* tipiche sono nell'ordine della decina di ms (millisecondi), mentre *throughput* tipici sono nell'ordine delle centinaia di Mbytes/s (Megabytes al secondo).
In base all'applicazione, si determinano i requisiti di performance per questi due parametri. Ad esempio, nello stato dell'arte della robotica teleoperata (da operatore oppure da intelligenza artificiale) si ritiene prioritaria la riduzione della latenza al fine di poter controllare il sistema a distanza. A tal fine, è tipicamente tollerabile una latenza inferiore alla decina di ms, che può essere ottenuta dalle reti cablate o che potrà essere in futuro garantita con adeguata continuità e qualità del servizio (Quality of Service, QoS) mediante la tecnologia 5G. Invece, latenze maggiori richiedono necessariamente l'implementazione di tecniche di co-decisione tra controllore remoto e sistema automatico localizzato a bordo del robot.
I più grandi successi conseguiti in ambito AI riguardano applicazioni i cui recenti sviluppi di ricerca sono stati possibili grazie alla disponibilità di dataset ampi e relativamente ben strutturati applicati negli algoritmi di apprendimento automatico. Al contrario, i dati provenienti da una moltitudine di dispositivi possono risultare distribuiti irregolarmente nello spazio e nel tempo e quindi più difficili da gestire e da utilizzare.[1]

COME GENERARE IL DATO

I dati possono essere generati da diverse fonti, in particolare: dagli esseri umani, dalle macchine, da organizzazioni o dal mix di questi attori. Le possibilità di ottenere dati stanno aumentando in maniera esponenziale grazie all'utilizzo di svariati sensori, che danno origine al cosiddetto "internet of things". Questi dispositivi vengono dispiegati nelle smart city, negli smartphone, negli orologi, negli ospedali.
La raccolta massiva dei dati è possibile grazie alla diffusione capillare di reti di dati e infrastrutture ICT come internet. Una connessione ad alta velocità è infatti essenziale per permettere ai dati di fluire dai dispositivi generatori ai database ed essere così elaborati.

COME UTILIZZARE IL DATO

Per essere utilizzabile, la moltitudine di informazioni deve essere raccolta all'interno di un "magazzino", il cosiddetto *datawarehouse*.
Perché un dato possa essere sfruttato efficientemente, l'organizzazione del database può essere di tipo "biologico", organizzando e sistematizzando cioè le informazioni secondo una struttura il più simile possibile al modo di ragionare umano. Effettuando questo passaggio sarà possibile sfruttare al massimo le potenzialità dei dati per effettuare data mining ed estrarre informazioni dai dati, e dunque creare un buon punto di partenza per le successive elaborazioni compiute degli algoritmi di Intelligenza Artificiale. Infatti, come molti esperti sottolineano (e questo vale per un'ampia parte delle

1. Ia.italia.it. (2019). [online] Available at: https://ia.italia.it/assets/libro-bianco.pdf

tecnologie), anche l'AI è fortemente bio-ispirata e bio-orientata. In particolare, nei procedimenti di Machine Learning la componente biologica diventa indispensabile nella realizzazione dei database, i quali, per poter essere funzionali devono essere architettati secondo le strutture tipiche della mente umana.
Il punto di forza della organizzazione dei database è che, contenendo un gran numero di informazioni, 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 infatti – si pensi al campo medico – 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 (vds. cosiddetta *personalized medicine*).
Il valore del dato non è, quindi, generato dal singolo byte di informazione ma dall'aggregazione di tante piccole informazioni che unite tra loro in modo strutturato generano informazioni ad alto valore aggiunto.
Il problema più rilevante che può affliggere i dataset è dato dai bias. Errori di valutazione, nel significato di un'immagine o di un concetto che vanno ad inficiare negativamente tutti gli output dell'analisi. Esempi noti sono riportati dall'utilizzo di algoritmi per prevenire i crimini o per attribuire le pene all'interno di un processo, dove i dati di input erano viziati da una serie storica che enfatizzava differenze etniche. Oppure dataset sbilanciati, che sovrastimano o sottostimano il peso di alcune variabili nella ricostruzione della relazione causa–effetto necessaria per spiegare certi eventi e, soprattutto, per prevederli.
Il valore del dato deriva principalmente dalla coesistenza di due fattori: la buona qualità del dato, e la quantità – maggiore è la quantità di dati a disposizione di sistemi di IA e di Ml, migliore sarà la loro affidabilità.

IL DATO COME "NUOVO ORO": CARATTERISTICHE DEI BIG DATA

Il dato è descritto in letteratura come "nuovo oro", principalmente con riferimento ai Big Data i quali possiedono le seguenti caratteristiche:

- Volume: Elemento quantitativo/massivo dei dati;
- Velocità: Capacità di essere raccolti ed estratti con modalità vicina al tempo reale (latency e throughput);
- Varietà: differenziazione nella tipologia dei dati estratti;
- Variabilità: il contenuto dei dati muta di significato a seconda dell'analisi a cui è sottoposto;
- Valore. Idoneità ad estrarre un valore/significato dai dati o dalla loro analisi.

Queste "5 V", sono spiegate nel presente documento relativamente al settore Health Care[2]. La scelta di questa analisi non è casuale, ma è connessa all'ultimo paragrafo di questo elaborato, che avrà ad oggetto proprio alcune possibili applicazioni dell'Intelligenza Arti-

2. Jain, A., Health, W., Health, W. and O'Brien, P. (2019). The 5 V's of big data - Watson Health Perspectives. [online] Watson Health Perspectives. Available at: https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

ficiale in tale ambito.
Alle precedenti caratteristiche alcuni studiosi ne aggiungono altre due:

- Visualizzazione, che consiste nella necessità di specifiche competenze nella realizzazione di strumenti in grado di presentare in maniera comprensibile i risultati dell'estrazione/analisi;
- Veridicità, corrispondente all'accuratezza dei dati in relazione alla sempre maggiore complessità dei data set.

Anche queste ultime due V sono importanti per avere un database che sia il più fruibile possibile. I dati sono considerati il nuovo oro perché negli ultimi anni stanno permettendo alle aziende di scoprire nuovi fattori interessanti su clienti e personalizzare sempre di più i servizi, aprendo nuovi scenari e modelli di business.
Alla luce di questa breve sintesi circa le caratteristiche principali del dato, possiamo affermare che quella dei dati rappresenta una vera e propria sfida dell'AI. Che si gioca sulla necessità che siano di buona qualità, il più possibile esenti da *bias* derivanti da errori nella generazione del dato e dalle annotazioni degli esseri umani. In particolare, questi limiti condizionano tipicamente tutti quei dati prodotti dai sistemi IoT che, pur essendo collegati gli uni agli altri, sono frammentati, eterogenei e poco interoperabili. Lo stesso vale per i cosiddetti Linked Open Data" (LOD) che, per poter essere efficacemente utilizzati, devono essere recuperati e filtrati per mezzo di metodologie semantiche ed ontologiche.[1]

IL CLOUD

LA DEFINIZIONE DI CLOUD[3,4]

Letteralmente "nuvola informatica", termine con cui ci si riferisce alla tecnologia che permette di elaborare e archiviare dati in rete.[2]
In altre parole, attraverso internet il *cloud computing* consente l'accesso a dati memorizzati ed applicazioni su un hardware remoto invece che sulla workstation locale. Per le aziende questo modello comporta dunque una radicale trasformazione della struttura organizzativa e di costi; i potenti sistemi hardware di elaborazione delle informazioni possono essere tenuti in remoto, mediante infrastrutture proprie o di terze parti, con macchine a più bassa prestazione localizzate nei siti dove i dati vengono generati e trattati.
Sono possibili diverse architetture di gestione dell'infrastruttura cloud, in funzione della localizzazione dei sistemi di memorizzazione ed elaborazione dell'informazione: soluzioni con cloud privati, con cloud pubblici e con cloud ibridi.
Sono inoltre possibili diversi modelli di business per la distribuzione di soluzioni di cloud computing, tra cui: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS).
L'introduzione di soluzioni in cloud è un cambio di paradigma rispetto all'Intelligenza Artificiale, perché rappresenta una tecnologia abilitante chiave per mettere a disposizione notevoli moli di dati, generati in modo distribuito nello spazio e nel tempo. Solu-

1. Per una definizione: https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_1.2_introduction_to_linked_data_en_edp.pdf
2. Dizionario Treccani
3. SearchCloudComputing. (2019). What is cloud computing?
4. European Commission Cloud Strategy (2019). Cloud as enabler for the European Commission Digital

zioni in cloud possono consentire la condivisione delle esperienze (dataset) generate da reti di sensori, da macchine, da dispositivi, da persone, e dell'apprendimento generato a partire da tali basi di dati. Di converso, le soluzioni in cloud hanno notevoli implicazioni relativamente alla sicurezza dei dati (sensibili).

IL RUOLO DEL DATA CENTER NELLA GESTIONE DEL DATO

Il cloud è implementato mediante infrastrutture fisiche, detti data center. Quindi, anche se solitamente si associa a questa tecnologia la caratteristica dell'immaterialità, è necessario sottolineare come, al contrario, si tratti di uno strumento materiale, ben localizzato in un determinato luogo fisico (importante per il regime giuridico applicabile). La disponibilità di data center e dei dati in esso contenuto rappresenta un asset patrimoniale, con possibili implicazioni di sicurezza nazionale, che deve essere tutelato con apposite strategie aziendali e previsioni normative nazionali e sovranazionali, come dimostrato dall'attenzione del GDPR al tema della localizzazione fisica dei dati.

L' ALGORITMO

LE CARATTERISTICHE DELL'ALGORITMO

Seguendo l'ottima definizione che si trova sul dizionario Treccani, in informatica 'l'algoritmo viene definito come sequenza finita di operazioni elementari, eseguibili facilmente da un elaboratore che, a partire da un insieme di dati (input), produce un altro insieme di dati (output) che soddisfano un preassegnato insieme di requisiti".[5] La definizione dei requisiti è dunque l'elemento essenziale in cui entra l'operatore umano che deve tradurre le specifiche di progetto in vincoli ovvero 'requisiti che devono essere soddisfatti in ogni caso', e in obiettivi, ossia 'requisiti che devono essere soddisfatti il meglio possibile secondo un qualche criterio specificato'. Nel definire i vincoli e gli obiettivi si verificano i passi ingegneristici che traducono il contesto in cui l'algoritmo deve operare e le sue finalità in operazioni matematiche che realizzano la cornice in cui opera l'algoritmo. È chiaro che l'efficacia e l'attendibilità di un algoritmo dipendono dalla qualità con cui un operatore umano realizza questo trasferimento. L'algoritmo è una operazione matematica ed è predicibile, mentre la traduzione del contesto e delle specifiche in vincoli e obiettivi rispecchia fortemente le caratteristiche e lo scenario dell'operatore e può avere limiti, bias e contenuti anche fortemente ispirati al contesto culturale, per questo ha risvolti anche etici o giuridici.

Sempre partendo dalla definizione tecnica che troviamo nel Dizionario, da un punto di vista ingegneristico, e quindi delle prestazioni che può avere in quanto macchina, 'l'algoritmo è caratterizzato essenzialmente da due elementi: la complessità computazionale, relativa al numero di operazioni elementari necessarie per produrre l'output (direttamente legato al tempo di calcolo necessario per eseguire l'algoritmo), e l'approssimazione, relativa al grado di soddisfazione degli obiettivi secondo il criterio specificato'. Da queste due specifiche grandezze dipendono due fattori importanti e determi-

5. Per una definizione di algoritmo: http://www.treccani.it/vocabolario/algoritmo/

nanti, il tempo necessario per dare una risposta e l'accuratezza della risposta stessa, che sono anche due grandezze fortemente correlate.

GLI ALGORITMI E MACHINE LEARNING (ML)

Gli algoritmi e gli insiemi di dati sono alla base dei metodi di apprendimento automatico, o di machine learning, e in questo ambito, possono essere sostanzialmente suddivisi in due diverse tipologie, che sono legate all'intervento dell'operatore umano. Gli algoritmi *supervised* sono 'annotati' dagli operatori umani o dalle macchine già addestrate che, supervisionando l'apprendimento, permettono la classificazione dei dati. Anche in questo caso il fattore umano entra in causa nella supervisione fornendo una interpretazione che può essere anche culturale e quindi con bias, nei dati o nelle regole dell'algoritmo di addestramento, che influenzano l'output finale. Algoritmi di apprendimento non supervisionato (*unsupervised*) applicano strumenti matematici per definire regolarità, correlazioni, clustering e quindi estraggono dai dati informazioni in maniera automatica che possono mettere in evidenza tendenze, fenomeni e complessità ma che poi richiedono comunque un modello interpretativo che spieghi il fenomeno nel contesto di riferimento con strumenti scientifici e culturali appropriati secondo il settore di applicazione. Nel caso di algoritmi *unsupervised*, il numero di classi di una base di dati può essere non noto a priori, e l'algoritmo di addestramento può creare categorie in modo automatico.

È proprio a questo livello che la responsabilità diventa evidente, basta pensare al caso in cui si debba esprimere una diagnosi medica o somministrare una terapia, il tempo di calcolo che si dedica ai sistemi di apprendimento automatico e l'accuratezza della risposta attesa sono due elementi fondamentali in cui entra la decisione umana, che influenza requisiti, obiettivi e tempi di risposta desiderati. In effetti in medicina i movimenti scientifici che propongono la medicina basata sull'evidenza scientifica cercano proprio di definire metodologicamente come riuscire a procedere nelle decisioni, utilizzando gli strumenti matematici, statistici e informatici ma preservando l'accuratezza del risultato finale ed evitando i bias che in questo campo possono essere letali.

Abbiamo quindi visto come da una definizione puramente tecnica e informatica intervengono i fattori umani e la disponibilità dei dati, fattori che influenzano la messa in pratica dell'algoritmo e quindi possono determinarne l'efficacia, la velocità, l'accuratezza.

DAL MACHINE LEARNING ALL'ARTIFICIAL INTELLIGENCE

LE DEFINIZIONI DI ARTIFICIAL INTELLIGENCE

"Il principale pericolo dell'Intelligenza Artificiale è senza dubbio rappresentato dalla possibilità che le persone arrivino troppo presto alla conclusione di averla capita" Eliezer Yudkowsky

Il problema della definizione di intelligenza artificiale è che molti di noi pensano di sapere di cosa si tratta perché la letteratura o la fantascienza hanno posto l'attenzione su di essa in modo problematico e da molti anni, e ci hanno prospettato la possibilità di avere macchine in mezzo a noi, in grado di agire, quindi pensare, come un essere umano. Come è accaduto per la robotica, anche l'intelligenza artificiale e le sue implicazioni sociali ed etiche sono state largamente anticipate dalla fantascienza. Oggi alcune delle proprietà dei robot e dell'intelligenza artificiale sono rese possibili dalla tecnologia e dalla scienza, e la rivoluzione industriale che stiamo vivendo tratta proprio questo inserimento delle macchine intelligenti nella società ed in mezzo a noi. Sembra quasi che alcuni problemi si stiano materializzando e ci troviamo di fronte ad una necessità di chiavi interpretative che possano regolare l'accesso delle macchine alla società, definendo la cornice regolatoria, etica e giuridica di riferimento. L'altro elemento importante è l'aspetto di emulazione e antagonismo con l'uomo. In effetti quando parliamo di Intelligenza Artificiale, l'emulazione del pensiero umano è alla sua base, e piuttosto che cercare di trovarne una definizione filosofica o comunque astratta, possiamo fare riferimento all'opera e al contributo di Alan Turing ed al suo contributo fondamentale in cui fa un riferimento specifico alla emulazione del pensiero umano. Turing si pose l'interrogativo: "Possono pensare le macchine?" e per rispondere a questo interrogativo partì della definizione di "macchina" e "pensare" e partendo da questa considerazione introdusse, come soluzione al problema sollevato dalla domanda, il "gioco dell'imitazione"[1].
Turing trasformò la domanda se una macchina possa pensare in un problema basato su una prova di tipo dialogico: se non si riesce a distinguere il comportamento verbale di un computer da quello di un essere umano, che per definizione pensa, allora il computer pensa.
Ma cos'è l'Intelligenza Artificiale? Vi sono molteplici definizioni, tra le più accreditate si annovera quella dell'Università di Stanford, che la identifica come "una scienza e un insieme di tecniche computazionali che vengono ispirate - pur operando tipicamente in maniera diversa – dal modo in cui gli esseri umani utilizzano il proprio sistema nervoso e il proprio corpo per sentire, imparare, ragionare e agire." [2]
L'intelligenza artificiale è in continua evoluzione, ma generalmente:

- Coinvolge le macchine per trovare informazioni rilevanti in grandi quantità di dati;
- È la capacità di eseguire compiti ripetitivi con i dati senza la necessità di una costante guida umana. In questo caso si tratta di emulazione di compiti cognitivi.

1. A.M. Turing, Computing machinery and intelligence, in Mind, 59 (1950) 433-460. Traduzione italiana in: V. Somenzi, R. Cordeschi, La filosofia degli automi. Origini dell'intelligenza artificiale, Paolo Boringhieri, Torino, 1986, pp. 157-183.
2. "Artificial Intelligence and life in 2030, One hundred year study on Artificial Intelligence", Stanford University, 2016, p. 5.

ROBOT VS BOT

Nel linguaggio corrente a volte si rileva una certa imprecisione nel riferirsi ai *robot*, che sono dotati di strutture fisiche in grado di interagire con l'ambiente circostante, rispetto ad altri agenti quali assistenti virtuali o sistemi automatici per l'apprendimento e l'elaborazione delle informazioni (detti anche *bot*).

I Robot sono sistemi intelligenti dotati di un corpo con attuatori e sensori, e di un sistema di controllo che utilizza intelligenza artificiale per prendere decisioni utili per il suo funzionamento e per la sua operatività nello spazio di lavoro. Mentre il dizionario Garzanti definisce bot come:

'*Software, utilizzato per cercare informazioni in Internet, che esplora automaticamente numerosissime pagine web per trovare, registrare e catalogare dati; software che è in grado di simulare, in un videogioco, in una chat ecc., il comportamento di un utente umano: chattare con un bot*.'

La Robotica permette di estendere le potenzialità dell'intelligenza artificiale rendendo possibili compiti motori nello spazio fisico, e quindi convertendo energia ed impartendo azioni. La Robotica è nata come automazione del lavoro manifatturiero e di servizio, per produrre beni e servizi. La rivoluzione industriale della Robotica ha portato un cambiamento già negli anni '80, stravolgendo l'impianto industriale fordista e le linee di produzione e incrementando i livelli di produttività.

Gli effetti della Quarta Rivoluzione Industriale potranno riguardare il modo di produrre beni e servizi, che verrà trasformato mediante le cosiddette tecnologie abilitanti come la robotica, l'Intelligenza Artificiale, il Machine Learning, il Cloud. La vera novità del prossimo futuro sarà soprattutto la trasformazione della società, attraverso l'ingresso nelle nostre vite dei Robot. Nei prossimi anni si giocherà la possibilità per la Robotica per diventare una tecnologia 'di consumo', non più per applicazioni manufatturiere, ma per un utilizzo a contatto con i consumatori e non più con personale addestrato e formato per utilizzarla.

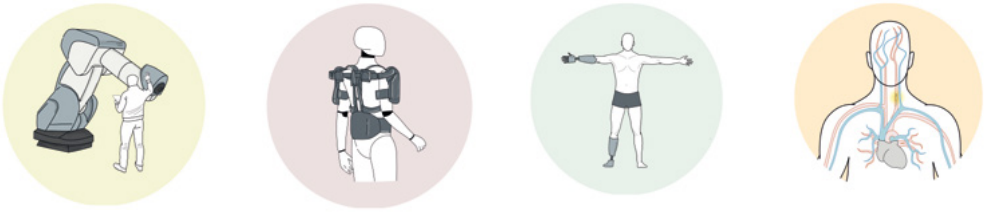
Dobbiamo considerare che il nostro Paese è tra i primi produttori al mondo e rappresenta un'area di grande competitività per la Robotica, e quindi l'Italia potrà essere protagonista di questa trasformazione.

L'impiego della robotica nell'industria e nella catena di montaggio ha caratterizzato la scorsa rivoluzione industriale. Quello che però sta accadendo oggi è diverso: il robot smette di essere distante dall'operaio, si avvicina a lui, sino a divenirne collaboratore come nel caso della robotica collaborativa che dal 2016 sta cambiando la robotica industriale (Figura 1a). Il passaggio successivo vede la robotica indossabile dall'operatore umano (Figura 1b), come per esempio avviene con esoscheletri industriali impiegati per supportare l'operaio nel gesto lavorativo, limitare i danni alla salute causati dallo sforzo e dalla ripetitività del movimento, sostenerlo nello spostamento dei carichi alleggerendone il peso. La robotica, arriva a giocare un ruolo dirimente nella possibilità di ottimizzare processi

produttivi, limitare il rischio connesso ad alcune tipologie di lavoro e sicuramente per la riduzione dell'impatto ambientale di certi processi. L'integrazione dei robot collaborativi o indossabili con strategie di intelligenza artificiale permette inoltre ai robot di generalizzare, e potenzialmente di rispondere a circostanze inizialmente non previste, superando con l'apprendimento i confini limitati delle regole e dell'esperienza impartite durante la programmazione iniziale. Anche questa caratteristica tecnica, conseguenza della fusione di robotica e intelligenza artificiale, pone però questioni critiche di tipo etico e giuridico, perché rende il comportamento del robot meno predicibile, e quindi più difficilmente ricostruibile la catena di responsabilità in caso di situazioni critiche come danni a persone o cose.

Ancora più estremo è il caso delle protesi cibernetiche e della bionica (Figura 1c-d), dove l'integrazione nell'uomo del dispositivo tecnologico è ancora più forte, e la robotica e l'intelligenza artificiale entrano all'interno del corpo umano e permettono di supportarne il funzionamento fisiologico mediante sistemi impiantabili.

Figura 1. I robot oggi sono in grado di grado di interagire con le persone: dalla robotica collaborativa e indossabile alle protesi di arto ai sistemi impiantabili. Illustrazioni di Andrea Aliperta.



a b c d

AUTOMAZIONE VS NARROW VS GENERAL AI[1,2]

Il concetto di agente autonomo afferisce sia ai sistemi fisici che no. Il primo caso fa riferimento agli agenti dotati di corpo, come i robot, cioè gli agenti che hanno apparati che emulano sia il cervello (*brainware*) che il corpo (*bodyware*) e che quindi sono direttamente capaci di effettuare azioni fisiche, mentre il secondo caso fa riferimento ad agenti che non sono dotati di presenza fisica evidente, come nel caso di operatori non umani nelle transazioni finanziarie (ad esempio, nei mercati azionari oppure in piattaforme business-to-business che gestiscono i sistemi di approvvigionamento con soluzioni Industria 4.0).

Gli agenti autonomi presentano sia sfide Scientifiche e Tecnologiche (S&T) che implicazioni Etiche, Legali e Sociali (ELS) collegate, con particolare riferimento ai profili di responsabilità associati al loro impatto nella società in relazione agli eventuali danni alla proprietà e alla persona conseguenti alla loro attività. Il concetto di autonomia è intrinsecamente multi-scala a seconda del livello della

1. Fondazione Leonardo, Draft Comments AI Public Consultation.
2. Questo capitolo è liberamente tradotto e adattato da Stradella, E., Salvini, P., Pirni, A., Carlo, A. D., Oddo, C. M., Dario, P., & Palmerini, E. (2012). Robot Companions as Case-Scenario for Assessing the "Subjectivity" of Autonomous Agents. Some Philosophical and Legal Remarks. In ECAI Workshop on Rights and Duties of Autonomous Agents (RDA2) (pp. 24-31).

gerarchia di controllo che è dotato di un livello di autonomia, o che coinvolge co-decisione ambientale o umana.
L'autonomia può andare dal controllo di basso livello (ad esempio, nel seguire una traiettoria di riferimento nello spazio dei giunti di un robot), alla pianificazione ed esecuzione del *task* coerentemente con un obiettivo specifico (ad esempio, in definire percorsi ottimali per spostarsi tra due luoghi, come con gli strumenti di navigazione come *Google Maps*), alla definizione di obiettivi specifici coerentemente con un obiettivo generale (ad esempio, la sequenza delle fermate intermedie nella catena di distribuzione dei prodotti, come nel caso dei servizi postali, di Amazon e simili), alla gestione di risorse energetiche (ad esempio, politiche di risparmio energetico e di ricarica delle batterie), alla *cloud robotics* (ad esempio, agenti che condividono decisioni ed esperienze su infrastrutture ICT), all'interazione e comunicazione (ad esempio, il caso del "Chinese room thought experiment"[1], che è collegato al "gioco dell'imitazione" di Turing richiamato in precedenza), alla definizione di obiettivi strategici in forma astratta, alla capacità di apprendere relazioni e collegamenti (tipica della cosiddetta "narrow artificial intelligence") fino all'obiettivo di emulare in futuro le capacità intellettive umane di apprendimento e creatività (la cosiddetta "general artificial intelligence").
Tutti tali livelli di autonomia, da quelli di basso livello gerarchico a quelli astratti e generali, presentano aspetti subdoli nel tentare di definire il concetto di autonomia o di differenziare un controllo automatico da un livello di autonomia. Ovviamente il concetto di autonomia è collegato direttamente al controllo automatico, sebbene l'autonomia sia notevolmente più controversa. L'influenza dell'esperienza passata sui comportamenti futuri non è sufficiente per caratterizzare l'autonomia da un controllo automatico: il semplice operatore matematico di integrale è influenzato dall'esperienza passata, ma nessuno affermerebbe che un integratore è una macchina autonoma (piuttosto, è una macchina automatica, ed infatti è un blocco fondamentale della teoria tradizionale di controllo e automazione).

LE APPLICAZIONI DELL'INTELLIGENZA ARTIFICIALE TRA POTENZIALITÀ E RISCHI

La novità, che forse rappresenta la parte più critica, riguarderà tutti quei processi tecnologici, frutto di interazione tra intelligenza artificiale, machine learning, cloud, basi di dati che permetteranno al robot di migliorare e potenziare le proprie prestazioni.
Il miglioramento delle performance dovrà infatti correttamente essere bilanciato con la garanzia di affidabilità e sicurezza del robot nella coesistenza con l'essere umano. In questo senso per il robot sarà un salto di qualità diventare generalista e non essere specialistico, cioè industriale o di servizio e quindi operato solo da utenti addestrati e in un contesto certificato. Possiamo dunque parlare di un vero e proprio processo di "socializzazione" della robotica. Se il robot entra nelle nostre case, interagiamo con esso, deleghiamo ad

1. Per un maggiore approfondimento: https://plato.stanford.edu/entries/chinese-room/

esso lo svolgimento di una serie di compiti con implicazioni, come dicevamo, sia fisiche che cognitive sino al punto di trasformarlo in un nostro "alias" cioè in grado di agire al nostro posto nell'esecuzione di determinate azioni.
In conclusione, Intelligenza Artificiale, robotica e la cybersecurity costituiscono oggi le chiavi di lettura della Quarta Rivoluzione Industriale e verosimilmente, rappresenteranno anche gli elementi centrali delle rivoluzioni scientifico – industriali del prossimo futuro. Essi forniscono importanti opportunità di sviluppo e dispongono di un elevato potenziale innovativo in grado di impattare su svariati ambiti della vita economica del paese, nelle dinamiche sociali e nel miglioramento delle qualità della vita.
Vi sono però una serie di aspetti, legati alla diffusione di queste nuove tecnologie, che non possono essere sottovalutati. Tra di essi certamente la sostenibilità ambientale ed energetica, lo sviluppo di nuovi materiali impiegabili in robotica, si pensi alle ricerche connesse ai materiali impiegabili nella robotica così detta "soft".
Altro aspetto è certamente quello connesso all'interazione tra uomo e robot che muterà inevitabilmente i comportamenti dell'individuo nel suo essere "sociale", dagli smartphone, ai robot domestici che sono già ampiamente entrati nelle nostre vite mutando visibilmente il nostro approccio a certe dinamiche comportamentali e sociologiche. I requisiti di sicurezza, controllo, capacità di prevedere il comportamento del robot e regolamentazione dello stesso, sono aspetti essenziali per la competitività del Paese nei settori dell'IA, della Robotica che si concretizzano in valore economico per il nostro sistema.

UN ESEMPIO DI APPLICAZIONE DELL'IA: LE ESPERIENZE DELLE STARTUP NEL CAMPO HEALTHCARE

Le nuove tecnologie, in particolare l'intelligenza artificiale, rendono migliore l'interazione con i clienti e con gli utenti del prodotto/servizio offerto da un'azienda. Attraverso l'integrazione e l'implementazione di tecnologie come Internet of Things, Machine Learning e AI è possibile ottenere benefici, non soltanto nella relazione con i clienti, ma anche internamente al perimetro aziendale, ad esempio si assiste solitamente ad un aumento della produttività e di conseguenza al miglioramento delle performance economiche.
Questi motivi stanno spingendo sempre un numero maggiore di aziende ad investire su questo tipo di innovazioni. Nella varietà degli scenari potenziali entro cui operano aziende che sfruttano tecnologie di AI la scelta, nell'ambito di questo contributo, è stata orientata principalmente allo studio dell'innovazione in campo healthcare. In questo settore altamente innovativo e dalle conseguenze dirette sulla salute delle persone si è scelto di analizzare, in modo necessariamente non esaustivo, il mondo europeo delle startup, le quali sono solitamente un valido indicatore della direzione verso cui si sta muovendo la ricerca applicata al mondo reale.

La prima azienda oggetto di analisi è eB^2, una startup spagnola che

si è posta l'obiettivo di migliorare le condizioni di vita dei pazienti affetti da patologie psichiatriche. Nello specifico l'azienda vuole cambiare il modo in cui questi disagi vengono monitorati. Ad oggi infatti i pazienti si sottopongono a visite mediche per monitorare le proprie condizioni. Chiaramente questo modello di controllo porta con sé dei bias dovuti all'essere "fuori dalla vita quotidiana", i pazienti tendono infatti a tenere sotto controllo il proprio comportamento durante questi incontri. La startup vuole innovare radicalmente questo modello tramite l'utilizzo di un sistema composto da smartphone, cloud e algoritmi. Sul primo device è possibile installare un'app appositamente creata. Questa è capace di "catturare" moltissime informazioni sul paziente e sullo stato della patologia con il vantaggio di raccogliere questi dati nel corso di tutta la giornata. Una volta raccolti i dati essi vengono inviati su un cloud e successivamente elaborati in "tempo reale" (in tempi molto brevi) attraverso l'utilizzo di specifici algoritmi (l'azienda li definisce come "Personalized Artificial Intelligence Models"). I risultati di questa analisi saranno inviati ai medici di ciascun paziente in modo che essi possano elaborare la cura ad hoc e scoprire comportamenti insoliti del paziente. Il sistema permette anche di individuare situazioni critiche ed avvisare gli operatori sanitari. L'azienda ritiene di essere riuscita a creare un modello di AI con la capacità di applicare modelli e metodi scientifici psicologici.

Diabeloop è una startup francese che cerca di migliorare la qualità della vita dei pazienti con diabete di tipo 1. In particolare, oggi tali pazienti sono soliti assumere dosi più o meno costanti di insulina quando ne sentono il bisogno. Questo comportamento non è ottimale, infatti il dosaggio potrebbe essere troppo o insufficiente. L'azienda francese sta cercando di creare un sistema per permettere la personalizzazione e l'ottimizzazione delle dosi di insulina in base all'effettivo bisogno del paziente. Per far questo la startup si avvale di un sistema composto da 3 strumenti: un sensore per misurare il livello di zuccheri nel sangue, capace di inviare anche i dati raccolti; un erogatore di insulina anch'esso con caratteristiche IoT e infine il sistema software in cui una serie di algoritmi di AI elabora l'enorme quantità di informazioni raccolte dal paziente e calcola in anticipo il momento in cui sarà necessario assumere la dose di insulina e la quantità esatta della stessa.
Dal punto di vista business ci sono alcune caratteristiche chiave che stanno aiutando l'azienda a raggiungere il pieno di sviluppo del potenziale: esperienze pregresse dei fondatori e le molte collaborazioni. Il team si compone di medici, ingegneri biomedici e un investitore venture capital. Per quanto riguarda le partnership, Diabeloop vanta sia partnership tecniche (con CEA-Leti), Mediche (CERITD) e centri medici (12 centri medici hanno iniziato a sperimentare la tecnologia).

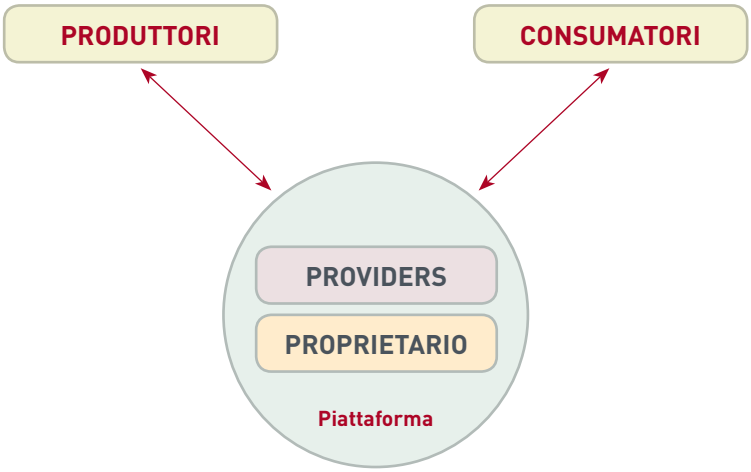
Dermosafe viene dalla Svizzera ed opera nello studio e nell'individuazione dei melanomi. Secondo la ricerca condotta dai fondatori il numero di questo tipo di tumore è in crescita negli ultimi anni e tale trend è dovuto all'invecchiamento della popolazione mondiale. A questa variabile se ne aggiungono altre identificate, tra cui alcuni geni del DNA, ed altre non ancora scoperte che portano un paziente allo sviluppo di questo tipo di tumore. Una cosa è certa: l'individuazione precoce è fondamentale per aumentare esponenzialmente le chance di salvezza del paziente.
In questo caso, il sistema attualmente utilizzato nella maggior parte degli ospedali ha buone probabilità di individuare le situazioni critiche ed intervenire per tempo per arginare lo sviluppo del tumore. Questo metodo però comporta costi molto alti, infatti la diagnosi viene effettuata da medici esperti, e tempi di attesa molto lunghi, in particolare è necessario fare più visite prima di accertare la pericolosità/rischiosità del tumore. Il team di Dermosafe ha ideato una soluzione che si propone di identificare rapidamente un melanoma nella fase iniziale di sviluppo. Questa caratteristica può rappresentare un ulteriore passo in avanti rispetto all'attuale scienza medica che non è capace di individuare il tumore in tale fase.
La soluzione è composta da una Dermoclic, un apparecchio che fa semplice una foto/scan della pelle del paziente, inviando poi questa immagine dermoscopica a Dermoview, un'applicazione web dove sono raccolte tutte le informazioni ottenute dai pazienti. Infine, le informazioni vengono rielaborate, giungendo infine ad una diagnosi (dal sito sembra che ci sia anche il contributo di un medico specializzato in melanomi che controlla le diagnosi fatte dal sistema per verificarne la correttezza) che viene comunicata al paziente e medici.
Inoltre, la soluzione può essere utilizzata anche da personale infermieristico e non esperti, riducendo quindi il costo. Infine, gli algoritmi tramite la comparazione e l'utilizzo di big data saranno in grado di dare giudizi molto affidabili sulla diagnosi. Anche in questo caso le caratteristiche chiave per il successo dell'azienda sono rappresentate dal team e dalle collaborazioni.

INTELLIGENZA ARTIFICIALE E PLATFORM ECONOMY

Con il termine piattaforma si definisce uno strumento digitale capace di mettere in contatto più utenti, i quali entrano a far parte della piattaforma per sfruttare i benefici da essa offerta.
L'interazione stessa di questi utenti è ciò che crea il valore dello strumento, che svolge il ruolo di intermediario nello scambio di beni e servizi.
Negli ultimi anni intorno a questo strumento si è andato a sviluppare un nuovo concetto di modello di business, la cosiddetta "platform economy", ossia l'economia di piattaforma. Con questo termine si fa riferimento a una larga fetta di business che oggi funziona in maniera digitale.
Lo sviluppo e il miglioramento delle tecnologie, ad esempio quelle

informatiche, internet e la proliferazione di device mobile ha permesso lo sviluppo esponenziale delle piattaforme, riducendone i costi di sviluppo e rendendo la collaborazione e l'interazione facile, rapida e sicura. Queste caratteristiche hanno permesso la creazione e il rafforzamento dei network tra i partecipanti, andando a creare ancora maggior valore per la piattaforma stessa.
Il sistema di piattaforma si compone di diversi attori (Figura 2): il proprietario della piattaforma che ha la possibilità di accettare i partecipanti e decidere le funzioni a loro concesse; i Providers che si interfacciano direttamente con la piattaforma, rendendola disponibile agli utenti; i Creators (detti anche Complementors), ossia coloro che effettivamente creano i servizi e i contenuti offerti sulla piattaforma; ed infine gli utenti che usufruiscono dei servizi ed interagiscono con i Creators.

Figura 2. Il sistema di piattaforma.



Gli ecosistemi di piattaforma vengono solitamente suddivisi in tre gruppi: Le piattaforme di transizione, dette anche *digital matchmachers*, hanno lo scopo di mettere in contatto diversi utenti, funzionando come un mercato virtuale; Le piattaforme di innovazione hanno lo scopo di fornire diverse tecnologie di base adattabili ai bisogni specifici di ciascun utente; Le piattaforme di integrazione nascono dalla combinazione delle due tipologie precedenti ed hanno funzionalità miste.
In tutti e tre i casi l'obiettivo del proprietario della piattaforma è di creare un ecosistema di complementors che costruiscano il proprio business in questo ecosistema digitale. Maggiore sarà l'impegno e l'investimento di questi soggetti nel creare contenuti attrattivi per i clienti, maggiori saranno gli utenti che entreranno a far parte della piattaforma. Questo andamento rappresenta un circolo vizioso, infatti le nuove realtà imprenditoriali cercheranno di entrare nelle piattaforme che offrono il bacino più ampio di utenti, i quali, allo

stesso tempo, vorranno partecipare alla piattaforma migliore, ossia quella che offre migliori servizi, ergo quella con più Creators.
Molto spesso i proprietari di una piattaforma tendono ad offrire diversi benefit e risorse (*application program interface*, *software development kits* ed altro) per aumentare il grado di appetibilità dell'ambiente virtuale.
Una dinamica a cui si assiste solitamente in questo ecosistema vede un'azienda affermarsi nel ruolo di "platform leader", a cui si aggiungono varie aziende complementari nel ruolo di innovatori (tra questi vengono identificati anche i clienti e gli utenti della piattaforma).

Gli imprenditori traggono diversi benefici di varia natura derivanti dalla partecipazione ad una platform. I vantaggi immediati e comuni ai tre gruppi sono:

- Sharing knowledge con gli altri complementor, caratteristica molto importante ed intrinseca in base al grado di openess di una piattaforma;
- Sharing risks, lavorando in maniera complementare è possibile minimizzare il rischio di adozione della piattaforma;
- Ampia community di utenti con cui interagire, scambiare e trarre importanti informazioni;
- Innovation collaboration, la collaborazione è ritenuto elemento caratterizzante dell'ecosistema di piattaforma.

La facilitazione offerta dai sistemi virtuali sta creando sempre maggior dipendenza per gli imprenditori che partecipano alla piattaforma. Questo vincolo ha determinato la nascita dei cosiddetti dependent-entrepreneur, ossia imprenditori soggetti al volere del proprietario della platform. Questo assoggettamento è dovuto all'incidenza sempre maggiore di questi strumenti sul business degli imprenditori. 'Se non risulti su Google, vuol dire che non esisti': questa frase esprime pienamente il concetto, le aziende sono infatti obbligate a partecipare al network di piattaforma per mantenere il proprio business. Al tempo stesso, si viene a creare un effetto di *lock-in* che assoggetta l'azienda al volere del proprietario della piattaforma, limitando la "libertà" dell'imprenditore.
Vi sono anche alcuni rischi a cui l'imprenditore è esposto direttamente, essi sono legati alla possibilità di "aiutare" indirettamente un competitor a sviluppare il proprio business, attraverso lo sfruttamento di informazioni chiave raccolte sulla piattaforma. A questo se ne aggiunge un altro, forse ancora di maggior peso, non va infatti dimenticato che il proprietario della piattaforma ha accesso a tutte le informazioni che vengono scambiate sulla stessa. Questo ricco database contiene dati chiave riguardo clienti, strategie e prodotti ed esso è a completa disposizione del creatore della piattaforma (nel momento stesso in cui ci si iscrive alla piattaforma si accetta questa condizione).

Già oggi ci sono aziende, ad esempio Amazon, che utilizzano queste informazioni per capire e anticipare i bisogni dei consumatori. L'azienda di retail è solita utilizzare algoritmi di intelligenza artificiale per analizzare alcune informazioni chiave sulle ricerche, gli oggetti salvati e acquistati dai clienti. Tramite questi dati, Amazon riesce a capire le preferenze dei consumatori (e ad anticiparle), offrendo poi un prodotto ad hoc per surclassare le aziende presenti nella piattaforma.
Nonostante questi rischi insiti nelle caratteristiche stesse di questi strumenti, essi sono ritenuti da molti studiosi come la miglior soluzione per lo sviluppo dell'imprenditorialità, in un contesto dinamico e globale come quello odierno. Le piattaforme hanno permesso e stanno permettendo a moltissime startup di costruire il proprio business rapidamente e sostenendo costi molto inferiori rispetto a quelli che avrebbero dovuto affrontare qualche decade fa.

Per avere successo in questo ambiente digitale è però necessario essere pronti ad un nuovo modello di business, in cui:

- Gli assets più importanti diventano la community e i contenuti da essa creati;
- Il valore viene creato dalla facilitazione dell'interazione tra produttori esterni e consumatori per cui la governance dell'ecosistema diventa una skill chiave;
- Le relazioni di piattaforma permettono ad un'azienda di sviluppare la propria tecnologia molto rapidamente, andandosi prima a specializzare in un ambito particolare, e poi diversificando rapidamente, e a basso costo, l'ambito target della tecnologia;
- Le collaborazioni tra imprese e utenti aumentano qualitativamente e quantitativamente, non si basano più su rapporti formali e la contaminazione avviene di continuo;
- La piattaforma ha l'obiettivo di massimizzare il valore totale dell'ecosistema tramite un processo circolare, interattivo e composto di feedback continui.

Queste caratteristiche rendono la competizione molto più complicata rispetto ai modelli di business tradizionali. Gli executives devono fare attenzione a nuove metriche di valutazione dei rischi per mantenere la posizione competitiva di un'azienda operante nel mondo delle piattaforme.
Nel contesto dell'intelligenza artificiale è evidente come esso sia già uno strumento molto importante per i proprietari delle piattaforme, che lo utilizzano per analizzare e migliorare (e non solo vedi caso Amazon sopra) l'ecosistema digitale. Al tempo stesso, dato che molti strumenti e varie tecnologie sono offerte dai proprietari stessi è prospettabile che in futuro le aziende avranno la possibilità di utilizzare e sfruttare questa tecnologia (AI), come servizio aggiuntivo. Sembra quindi sensato considerare in futuro le aziende come fruitori delle tecnologie e non più proprietarie delle stesse, come

già sta avvenendo in alcuni settori come quello delle stampanti. Gli imprenditori creeranno in futuro moduli che sfruttano la tecnologia altrui per un determinato scopo e al tempo stesso saranno sfruttati da altre aziende nel realizzare il proprio obiettivo, si delinea quindi una catena di valore integrata in cui ciascuna azienda diventa modulo tecnologico e commerciale di un altro modulo (ossia di un'altra azienda).

BIBLIOGRAFIA

- AA.VV (2016). Artificial Intelligence and life in 2030, One hundred year study on Artificial Intelligence, Stanford University Press.
- AGID, Task Force IA, (2019). L'Intelligenza Artificiale al servizio del Cittadino. [online] Available at: https://ia.italia.it/assets/librobianco.pdf
- Carrozza, M.C. (2017). I Robot e Noi, Il Mulino – AREL.
- Cennamo, C. and Santalo, J. (2019). "Platform competition: Strategic trade-offs in platform markets". In Strategic Management Journal, 34:11, pp. 1331-1350.
- Consumer Value Creation. (2016). Pipelines, Platforms, and the New Rules of Strategy. [online] Available at: https://consumervaluecreation.com/2018/03/11/pipelines-platforms-and-the-new-rules-of-strategy
- Cutolo, D. and Kenney, M. (2019). Current Publications | The Berkeley Roundtable on the International Economy. [online] Brie.berkeley.edu. Available at: https://brie.berkeley.edu/recent-publications/current-publications
- Fondazione Leonardo, Draft Comments AI Public Consultation. [unpublished]
- Kenney, M. and Zysman, J. (2016). The Rise of the Platform Economy | Issues in Science and Technology. [online] Issues in Science and Technology. Available at: https://issues.org/the-rise-of-the-platform-economy
- Kenney, M., Rouvinen,, P., Seppälä, T. and Zysman, J. (2019). [online] Available at: https://www.researchgate.net/publication/332036703_Platforms_and_Industrial_Change
- Nambisan, S., Siegel, D. and Kenney, M. (2018). On open innovation, platforms, and entrepreneurship. In Strategic Entrepreneurship Journal, 12:3.
- Stradella, E., Salvini, P., Pirni, A., Carlo, A. D., Oddo, C. M., Dario, P., & Palmerini, E. (2012). Robot Companions as Case-Scenario for Assessing the "Subjectivity" of Autonomous Agents. Some Philosophical and Legal Remarks. In ECAI Workshop on Rights and Duties of Autonomous Agents (RDA2) (pp. 24-31).
- Stradella, E. (2019). La regolazione della Robotica e dell'Intelligenza artificiale il dibattito, le proposte, le prospettive. Some points for reflection, in MediaLaws - Journal of media law, 1.
- Turing, A.M. (1950). Computing machinery and intelligence, in Mind, 59, 433-460. Italian translation in: V. Somenzi, R. Cordeschi, La filosofia degli automi. Origini dell'intelligenza artificiale, Paolo Boringhieri, Torino, 1986, pp. 157-183.
- White Paper on Artificial Intelligence at the service of the citizen, Version 1.0 March 2018, edited by the Task Force on Artificial Intelligence of the Agency for Digital Italy (ia.italia.it).
- Zysman, John & Kenney, Martin (17.10.2016). "The Next Phase in the Digital Revolution: Platforms, Abundant Computing, Growth and Employment". ETLA Reports No 61. https://pub.etla.fi/ETLA-Raportit-Reports-61.pdf

Paper sui principi etici

Nota metodologica

Il presente documento si prefigge lo scopo di definire un insieme di obblighi e raccomandazioni pratiche per lo sviluppo di applicazioni e sistemi basati su tecniche di Intelligenza Artificiale (IA). Gli stessi sono derivati a partire da una definizione di diritti conseguenti a principi e valori etici radicati nei documenti fondamentali della nostra organizzazione sociale.
È stato redatto da un gruppo multidisciplinare di ricercatori costituito da Francesco Corea, Fabio Fossa, Andrea Loreggia, Salvatore Sapienza con la supervisione di Stefano Quintarelli. Il manoscritto è stato quindi sottoposto alla revisione delle Proff. Maria Chiara Carrozza, Monica Palmirani, Francesca Rossi e del Prof. Carlo Casonato ed il testo è stato riesaminato alla luce dei commenti ricevuti.
Il presente documento è stato elaborato dagli autori in piena autonomia, libertà ed indipendenza e riflette unicamente ed esclusivamente le opinioni del gruppo di lavoro.
Il documento è stato redatto con il metodo del consenso ("consent"), ovvero con l'assenza di obiezioni significative per ogni suo enunciato. Pertanto, sebbene i membri del gruppo sostengano il documento nel suo complesso, non necessariamente essi condividono ogni singola affermazione contenuta nel documento stesso.

RIFERIMENTI
LAVORI PRECEDENTI

Nella elaborazione si è tenuta in particolare considerazione quanto definito dai seguenti lavori:

- Partnership on AI (2016): lista di principi centrata sulla necessità di sviluppare una cultura di cooperazione fra ricercatori in IA, di garantire una distribuzione il più possibile equa dei benefici delle nuove tecnologie e il coinvolgimento di stakeholders pubblici e aziendali. Promosso dai principali Over The Top.
- Principi di Asilomar (2017): manifesto di principi di roboetica e linee guida per lo sviluppo delle nuove tecnologie, definiti da accademici e professionisti del settore.
- AI in the UK (2018): studio realizzato dalla Camera dei Lord a supporto della condivisione sociale dei benefici derivanti dall'uso di una IA trasparente e sicura.
- Villani (2018): report dell'esecutivo francese che definisce la propria strategia sull'Intelligenza Artificiale, elencando i principi fondamentali per il suo sviluppo.
- AI4People (2018): il documento, elaborato da Atomium-EISMD, chiarisce rischi e opportunità che l'IA presenta nei confronti della società contemporanea e delinea princìpi etici a cui adeguare ricerca e utilizzo dell'IA.
- CEPEJ (2018): il documento, redatto dalla European Commission for the Efficiency of Justice, ha lo scopo di valutare impatti etici e potenzialità dell'uso dell'IA in contesti giudiziari.
- HLEG_AI Ethics Guidelines (2019): linee guida definite dal

gruppo di esperti della Commissione Europea per la creazione di una intelligenza artificiale affidabile ed attendibile.
- IEEE Ethically aligned design (2019): stabilisce che le tecnologie devono incorporare, attraverso pratiche da attuare già in sede di progettazione, i valori fondamentali a cui associare provvedimenti di policy e inquadramenti legali corrispondenti.

FONDAMENTI

Si è scelto di ancorare il lavoro ad alcune carte fondamentali del nostro tessuto sociale, partendo da quelle di carattere globale per arrivare al livello nazionale. Si fa quindi riferimento a:

- Sustainable Development Goals: formulati dalle Nazioni Unite nel 2015, con l'obiettivo di promuovere lo sviluppo sostenibile attraverso la soluzione di alcuni fra i maggiori problemi economico sociali dell'umanità; lo sviluppo dell'IA si lega a doppio filo con il loro raggiungimento.
- I guadagni di produttività offerti dell'IA devono anche favorire un'industrializzazione inclusiva rispettosa del lavoro umano, come stabilito negli obiettivi n.8 e n.9. Allo stesso tempo, bisogna assicurare un'ampia diffusione dei suoi benefici al fine di ridurre le disuguaglianze, in accordo con l'obiettivo n. 10.
- La promozione della parità di genere fissato nell'obiettivo 5, richiede di garantire l'eliminazione di bias dal design degli algoritmi; la tutela di un'educazione paritaria e di qualità definita nell'obiettivo n. 4, si necessita della diffusione di cultura digitale a tutti i livelli d'istruzione. Sul piano politico, la promozione di pace, giustizia ed istituzioni forti prevista nell'obiettivo 16 dipende in misura crescente da un utilizzo eticamente corretto dell'IA nella personalizzazione della comunicazione di massa, nella prevenzione e repressione del crimine e nell'amministrazione della giustizia, senza scadere in forme di manipolazione e di controllo statale autoritario. Per quanto riguarda la promozione della pace, in particolare, l'IA non deve in alcun modo sostituire il giudizio umano nei sistemi di arma.
- Dichiarazione Universale dei Diritti Umani: costituisce uno dei testi essenziali per ogni riflessione etico-giuridica da cui iniziare anche un dialogo sul quadro etico dell'IA. Tale dialogo è fondato sul riconoscimento della dignità umana quale fondamento del rispetto e della promozione dei diritti, indipendentemente dal sesso, dall'appartenenza etnica e religiosa, dalle opinioni politiche o da altri fattori che possono dar luogo a discriminazione. Il carattere universale dei diritti umani riconosciuti nella Dichiarazione la rende idonea a favorire una discussione globale e inclusiva sui temi e sulle sfide lanciate dall'IA nei confronti della società pluralista e multiculturale contemporanea. Di particolare rilevanza, per gli scopi di questo documento, sono la dignità (art. 1), la tutela della riservatezza (art. 12), la libertà di informazione (art. 19), e l'attenzione verso i temi di uguaglianza e non discriminazione, fondamentale nella riflessione sui bias

algoritmici. Sono soggetti agli obblighi che scaturiscono dalla Dichiarazione gli Stati membri delle Nazioni Unite, che devono trasporne i suoi principi negli ordinamenti nazionali riguardanti i soggetti da loro regolati, non direttamente vincolati dalla Dichiarazione.

- Carta dei diritti fondamentali dell'Unione Europea: stabilisce valori e obiettivi fondamentali dell'Unione Europea e rappresenta un rilevante quadro di riferimento per lo sviluppo e l'uso di sistemi di IA. I valori delineati si incardinano sull'inclusione, la tolleranza, la giustizia, la solidarietà e la non discriminazione e si declinano nel rispetto della dignità e dei diritti umani, delle libertà individuali, degli ideali democratici, dell'uguaglianza dei cittadini davanti alla legge e dello stato di diritto. Alla luce di tali valori, L'UE si impegna a promuovere la pace e il benessere, la libertà e la sicurezza dei propri cittadini; a garantire giustizia e libertà di spostamento; a favorire lo sviluppo sostenibile basato su una crescita economica equilibrata e sulla stabilità dei prezzi, su un'economia di mercato altamente competitiva, con la piena occupazione e il progresso sociale, e la protezione dell'ambiente; a lottare contro l'esclusione sociale e la discriminazioni; a promuovere il progresso scientifico e tecnologico; a rafforzare la coesione economica, sociale e territoriale e la solidarietà tra gli Stati membri nel rispetto della diversità culturale e linguistica che ne contraddistingue la natura.
- Costituzione della Repubblica Italiana: tra i suoi principi fondamentali, l'inviolabilità dei diritti dell'uomo e il riconoscimento dell'uguaglianza formale e sostanziale devono essere garantiti nella ricerca e nell'utilizzo di sistemi IA. La centralità del lavoro, inoltre, impone una riflessione sulle conseguenze economiche e sociali dello sviluppo di tali sistemi. Nell'applicazione dei sistemi d'arma, sottolinea il rifiuto della guerra come mezzo di risoluzione di controversie internazionali. Il rispetto dei diritti dei cittadini alla libertà personale, alla riservatezza e alla libera manifestazione del pensiero deve guidare l'impiego da parte di soggetti pubblici e privati, dei sistemi IA in grado di minacciare queste libertà fondamentali. Per questi soggetti, va raccomandato il rispetto della dignità umana nell'iniziativa economica. Nell'attività di tutela della sicurezza e dell'ordine pubblico, nonché sulle attività di prevenzione e repressione dei crimini, occorre tanto che l'attribuzione di responsabilità civili e penali per un utilizzo improprio di sistemi IA sia personale, quanto rafforzare il rispetto del principio di legalità e del giusto processo.

PRINCIPI E VALORI ETICI

Sulla base dell'analisi svolta nei paragrafi precedenti, il gruppo ha identificato un insieme di principi e valori etico-sociali, organizzati in tre macro livelli con una stratificazione di raggio crescente, da quello individuale a quello globale.
Le tre macro categorie, ed i relativi principi, sono da intendersi secondo una prospettiva integrata, non di rilevanza [Fig. 1: cerchi concentrici]. La scelta di un tale approccio non riflette solamente necessità formali, come la compilazione di un aggregato di enunciati, ma sottende una volontà sostanziale: proporre una visione organica.

LIVELLO INDIVIDUALE

Il livello individuale si propone di identificare i valori che pertengono alla persona e si incardinano sul fondamento monolitico della dignità umana. Da quest'ultima derivano i diritti civili e il principio della non discriminazione, i quali tutti si esplicano sia nella dimensione materiale che nella dimensione immateriale delle attività umane.

DIGNITÀ UMANA

Per quanto sia estremamente difficile accordarsi sulla sua definizione, il principio della dignità umana è largamente diffuso e comunemente riconosciuto. In quanto tale, esso ricopre un ruolo fondamentale nella Dichiarazione universale dei diritti umani, nella Carta dei diritti fondamentali dell'Unione Europea e nella Costituzione della Repubblica Italiana.
Nel suo senso più fondamentale, per dignità si intende il valore intrinseco che pertiene ad ogni individuo in quanto essere umano—valore che, per rifarsi ad una nota formula kantiana, impone di non trattare mai un altro essere umano solo come un mezzo per i propri scopi, ma anche e sempre come un fine in sé, cioè come un soggetto in grado di determinare sé stesso in maniera autonoma. Il principio della dignità umana costituisce una limitazione del potere di autodeterminazione e di azione del singolo, per cui il valore intrinseco del suo simile funge da confine della propria libertà.
L'IA, data la sua pervasività sociale e l'impatto profondo che si ritiene eserciterà su ogni aspetto della vita, potrà avere effetti rilevanti sul rispetto della dignità umana. Applicazioni in campo industriale, sanitario, educativo, assistenziale e sociale potranno offrire nuovi potenti mezzi per la produzione, il mantenimento o il rafforzamento delle condizioni associate alla vita dignitosa. Tuttavia, le stesse tecnologie che possono essere indicate come mezzi per il rispetto e l'affermazione della dignità umana potrebbero anche minacciarne l'integrità sia morale che fisica, intersecando i temi bioetici della transumanità. Contrarie alla dignità umana sembrano essere tecnologie che manipolano l'utente - anche a fine di bene - o a cui sono delegate decisioni di grande importanza sociale o esistenziale senza che sia possibile comprenderne le dinamiche. Ancora, la dignità umana è messa significativamente a rischio da tecnologie che non colgono il valore intrinseco di ogni individuo dissolvendo la sua

particolarità nella generalità di modelli statistici.
In conclusione, il principio di dignità è ampiamente riconosciuto come un'istanza fondamentale per lo sviluppo e l'uso etico dell'IA.

LIBERTÀ E DIRITTI CIVILI

Il principio su cui si basa il modello etico dei diritti civili è la dignità della persona. Il modo più affidabile per assicurare la felicità e la giustizia è l'affermazione del valore dell'essere umano che lo differenziano dagli altri esseri naturali e gli conferiscono anche la propria dignità.
Si può dire che la Dichiarazione Universale dei Diritti Umani è considerata l'origine e il nucleo fondamentale di una costruzione etica come base della soluzione ai conflitti della convivenza umana.
La Dichiarazione universale dei diritti umani costituisce un'etica materiale che stabilisce valori, contiene norme che devono essere rispettate, diritti che devono essere garantiti e libertà che devono essere protette.
Storicamente, i primi diritti che si sono sviluppati sono stati i cosiddetti diritti di prima generazione, i diritti di libertà, che limitano il potere dello Stato, come la libertà di pensiero, di coscienza e di opinione, in risposta contro monarchie assolute e regimi dittatoriali. Successivamente i diritti di seconda generazione hanno riguardato i diritti di uguaglianza ed i diritti politici, che assicurano una parità di condizioni nella partecipazione al potere politico
La terza generazione, il cui valore fondamentale è la solidarietà, comprende i diritti sociali, il diritto alla sicurezza sociale, il diritto al lavoro.
Come si vede, i diritti umani vanno inquadrati in una prospettiva dinamica: si sono evoluti nel corso dell'esperienza storica, ed è ragionevole ritenere che possano continuare a farlo.
I diritti civili si radicano nella Dichiarazione universale dei diritti dell'uomo che riconosce il diritto di tutte le persone alla libertà (di circolazione, di pensiero, di opinione, di associazione, ecc.), alla giustizia, un livello di vita adeguato, alla salute e al benessere, in particolare alle cure mediche e ai servizi sociali. Tutti questi sono ambiti in cui le tecnologie assumono un ruolo preponderante. Grazie alla IA, la dimensione immateriale, è infatti divenuta (o sta divenendo) la principale interfaccia utente per le relazioni sociali ed economiche delle persone, la sede prima in cui tali diritti vanno assicurati (Quintarelli, 2019). Tale assicurazione deve essere sostanziale, prima ancora che formale, bilanciando pertanto gli squilibri esistenti nella dimensione materiale tra diversi individui ed includendo una cautela particolare per le persone più deboli che statisticamente sarebbero relegati ad *outlier* nei modelli statistici.

NON DISCRIMINAZIONE

La Carta dei diritti fondamentali dell'Unione Europea afferma il principio di uguaglianza: riconoscere a tutti i cittadini gli stessi diritti davanti alla legge. Il principio della parità tra uomo e donna è alla base di tutte le politiche continentali, ed è l'elemento su cui si

fonda l'integrazione europea. Si applica in tutti i settori.
La Costituzione italiana recita:
"*Tutti i cittadini hanno pari dignità sociale e sono eguali davanti alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali. È compito della Repubblica rimuovere gli ostacoli di ordine economico e sociale, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese.*"
Ed anche:
"*La Repubblica riconosce a tutti i cittadini il diritto al lavoro e promuove le condizioni che rendano effettivo questo diritto.*"
Da queste previsioni consegue la lotta contro la discriminazione, la tutela dei diritti delle minoranze e dei settori più fragili della popolazione in relazione alla loro situazione oggettiva.
I dati rilevati ed utilizzati nei sistemi di *machine learning* dipingono i tessuti sociali incorporandone i relativi pregiudizi. In assenza di specifiche cautele e previsioni, i modelli statistici prodotti cristallizzano e possono amplificare tali bias.

CAPIRE I BIAS E COME GESTIRLI

I sistemi di IA, inevitabilmente, ereditano dagli esseri umani molti dei bias di cui soffrono. Le modalità in cui questi errori vengono trasmessi possono essere molteplici, e difatti possiamo facilmente identificarne almeno cinque (Hammond, 2016): bias nei dati stessi; bias nati da interazioni; bias di similarità; bias che scaturiscono da obiettivi contrastanti; bias emergenti. Nel primo caso, l'errore è contenuto direttamente nei dati che alimentano il sistema, ed è spesso associato a banche dati incomplete, errate, o mal classificate. Nel caso di "bias da interazioni", gli errori sorgono nel momento in cui il programma interagisce con utenti esterni e apprende attraverso tali interazioni (si pensi al caso del chatbot Tay, tristemente noto esempio di un *chatbot* che sviluppò ideologie naziste attraverso scambi di opinioni con utenti su Twitter). I bias di similarità, invece, nascono da sistemi che implementano correttamente le azioni per le quali sono stati programmati, ma che involontariamente in questo modo restringono le possibilità del sistema stesso (un esempio è il *news feed* di Google o Facebook, che mostra agli utenti notizie simili). Per quanto riguarda i bias che derivano da obiettivi contrastanti, esistono situazioni in cui un sistema disegnato per assolvere una specifica funzione in un ambito ben delineato crea conseguenze negative in applicazioni laterali e secondarie. Infine, i bias emergenti sono quelli che rafforzano pregiudizi e comportamenti umani che potrebbero essere opinabili.

LIVELLO SOCIALE

I sistemi di IA, essendo non deterministici, tendono a generare alcune predizioni errate. Nella valutazione delle loro conseguenze si determina una tensione tra il livello individuale ed il livello sociale. Questa sezione definisce i valori rilevanti da rispettare e promuovere in un'ottica di bilanciamento tra bene comune e individuale.
Dal punto di vista dell'equità, l'adozione dell'IA deve garantire a

INCLUSIVITÀ

tutti, incluso categorie deboli o svantaggiate, accesso equo a opportunità, servizi e del lavoro prodotto, evitare concentrazioni di risorse e potere.
Gli effetti dello sviluppo dell'IA sulle diseguaglianze riguardano infatti non soltanto l'aspetto monetario, ma anche quello relativo a questioni socio-culturali, dove a condizioni di disagio si sovrappongono la mancanza di accesso all'educazione digitale e discriminazioni su base etnica o di genere. Un uso pregiudizievole delle nuove tecnologie rischia quindi di minare la posizione di categorie in posizione già precaria, innescando un circolo vizioso di marginalizzazione e ulteriore aumento delle diseguaglianze.
Un'adozione equa delle tecnologie intelligenti richiede che i relativi vantaggi e le opportunità connesse - finanziarie, educative, giuridiche, sanitarie, assistenziali, e così via - siano inclusive di un numero più ampio possibile di cittadini, a prescindere dalla loro condizione sociale, classe di reddito. ubicazione geografica e da altri fattori analoghi [*Sustainable Development Goals*].

RIDUZIONE DISUGUAGLIANZA

L'IA dovrebbe essere sviluppata in modo da prevenire e ridurre attivamente le diseguaglianze, garantendo la massima condivisione dei benefici socio-economici delle nuove tecnologie e vigilando affinché i guadagni di produttività garantiti dalla sua implementazione non diventino monopolio di una ristretta cerchia di soggetti, ma siano invece distribuiti equamente attraverso diverse categorie e classi sociali.
L'IA può diventare una forza attiva per la riduzione delle diseguaglianze, incorporando un concetto di giustizia distributiva che guardi alle categorie marginali come soggetti di intervento prioritario: strumenti basati sull'IA, ad esempio, possono risultare utili, nell'ambito del sistema educativo, per colmare divari di apprendimento, mentre in ambito sanitario, possono essere utilizzati per stimolare il *social empowerment* ed i servizi destinati agli individui affetti da disabilità.
L'impatto sul mondo del lavoro, particolarmente rilevante in relazione alla Costituzione Italiana, richiede non solo particolare attenzione affinché siano adeguatamente ammortizzati fenomeni negativi quali esuberi di massa, disoccupazione generalizzata, *de-skilling* e deprezzamento del lavoro umano, ma anche la definizione di nuove *policy* del lavoro che affrontino la relazione tra uomo e macchina nonché l'elaborazione di nuovi dispositivi sociali capaci di mitigare le esternalità negative dell'automazione e favorire generalizzate condizioni di esistenza dignitose.

COESIONE SOCIALE

Lo sviluppo dell'IA deve favorire la coesione sociale e garantire la robustezza del procedimento democratico. La ricerca (Sirbu et al., 2019) mostra come il design dei sistemi di IA usati in una comunità possa facilitare la formazione di un consenso in un numero ridotto di interazioni ma sia anche in grado, viceversa, di rafforzare e

radicare le divisioni nel tempo. Questo effetto, noto e sfruttato da tempo nei mass media, assume una rilevanza estremamente significativa nell'era dei *personalized* media.
In tale contesto, obiettivi socialmente desiderabili e legittimi interessi aziendali possono divergere ed entrare in conflitto. Da una parte, l'interesse sociale nel rafforzamento della coesione sociale suggerisce l'adozione di tecnologie capaci di favorire la composizione di opinioni differenti e promuovere confronti tolleranti. Al contrario, per massimizzare l'*engagement* degli utenti, la quantità di interazioni, e quindi lo *screen time* e i ricavi collegati, le aziende sono portate a adottare tecnologie IA tese ad amplificare le divisioni ed esacerbare gli animi, caratteristiche sfruttate per la diffusione delle cosidette *fake news* e *deep fakes*.
Un'altra modalità in cui si esplica questa divergenza di obiettivi è l'uso di algoritmi in grado di sfruttare il *confirmation bias* degli utenti effettuando una iper-personalizzazione dei messaggi. La riproposizione mirata di contenuti affini (c.d. *echo chamber*), giustificata dall'esigenza di migliorare l'esperienza dell'utente, rischia di compromettere il pluralismo informativo.
Questi fenomeni, se non gestiti, influenzano negativamente i processi democratici e minano la coesione sociale con effetti socioeconomici profondi e di lungo termine.

LIVELLO GLOBALE

In continuità con i principi discussi a livello individuale e sociale, prevenzione del danno, ricerca di pace e giustizia e sostenibilità si configurano infine come cardini globali dello sviluppo etico dell'IA.

PREVENZIONE DEL DANNO

I sistemi informatici consentono di affrontare problemi con una scalabilità sostanzialmente illimitata, ben superiore a quella possibile agli esseri umani. I sistemi di IA consentono di affrontare problemi di natura differente rispetto ai tradizionali domini applicativi dei sistemi algoritmici deterministici, quali ad esempio i problemi di percezione e classificazione, precedentemente riservati alla attività umana, che possono così essere oggetto di una scalabilità sostanzialmente infinita a livello globale. Ciò aumenta le possibilità umane ma ne espande i possibili rischi: così come è globale l'utilizzo delle tecnologie, è globale la propagazione di eventuali errori e dei danni relativi.
La prevenzione del danno si concretizza in una valutazione dei rischi finalizzata ad adottare o applicare misure che ne prevengano la manifestazione o ne mitighino l'esposizione o gli effetti. La prevenzione diventa quindi un processo dinamico che periodicamente valuta i sistemi attraverso procedure di *risk assessment*, promuovendo procedure e protocolli per il *risk management*. La creazione di buone pratiche come quelle introdotte nel campo della sicurezza informatica (ISO/IEC 27001) permette di identificare e descrivere situazioni critiche. Le pratiche di *risk management* adottano e promuovono "gli scenari di rischio" come metodologia utile per l'analisi del rischio.

Diventa di attuale interesse raccogliere e prototipare scenari che possano essere utilizzati per valutare i sistemi autonomi in modo tale da poter definire diverse classi di rischio. Tra gli scenari ormai conosciuti che possono essere censiti e raccolti per una periodica valutazione rientrano la trasmissione di bias ai dati utilizzati per il *training*, il *data poisoning*, lo *adversarial attack*.

PACE E GIUSTIZIA

L'articolo 3 del Trattato Fondamentale dell'Unione Europea sancisce come obiettivi condivisi dagli Stati membri la promozione della pace e della giustizia, in aderenza alla Convenzione Europea sui diritti umani ed alla Carta delle Nazioni Unite.
La rivoluzione tecnologica legata all'IA corre il rischio, da un lato, di accentuare le disuguaglianze interne ai paesi avanzati; dall'altro, di scavare un solco ancora più profondo tra questi ed i paesi in via di sviluppo.
Uno sviluppo etico dell'IA deve quindi garantire la protezione dei valori della concordia e della fratellanza. Inoltre, l'innovazione tecnologica deve essere condotta nel rispetto dei principi di giustizia e contribuire a prevenire l'erompere di conflitti e tensioni internazionali.
Nella prevenzione e nella repressione dei crimini, i principi di legalità e di "giusto processo", garantiti dal diritto internazionale, dalla Costituzione e riconosciuti dagli ordinamenti, devono collocarsi come presupposti irrinunciabili al dispiegamento di sistemi IA in questi settori.

AWS: AUTHONOMOUS WEAPON SYSTEMS

L'applicazione di IA e sistemi autonomi promette di rivoluzionare il contesto dei conflitti militari e della *governance* di sicurezza degli attori statali. Un approccio etico in questo contesto deve preservare l'autonomia e la determinazione umana al fine di garantire sempre il controllo dei meccanismi decisionali autonomi e la responsabilità soggettiva, quantomeno in quegli ambiti ritenuti critici dal punto di vista strategico e, soprattutto, morale, rispecchiando l'esigenza di proteggere valori di rilevanza costituzionale quali la vita, l'incolumità fisica e la dignità umana.
La garanzia del controllo umano significativo sui sistemi autonomi è necessaria al fine di preservare la natura benefica della tecnologia, in linea con i principi prevalenti nel campo della bioetica e dell'etica della tecnologia.
Ciò è valido in particolare nel caso delle *Lethal Authonomous Weapon Systems* (LAWS), i sistemi d'arma autonomi. Il divieto di produzione ed utilizzo di LAWS rispetta gli standard prevalenti di *human-on-the-loop* (HOTL) e *human-in-the-loop* (HITL), secondo cui l'applicazione di sistemi autonomi in ambiti safety-critical deve avvenire sempre sotto la supervisione ed il controllo di un operatore umano.

In ogni caso, i sistemi autonomi progettati per arrecare un danno fisico ad infrastrutture e persone militari o militarizzate hanno speciali e inconsueti risvolti etici se paragonati con le armi di tipo tradizionale o con i sistemi non armati. Per questo motivo è necessario da un punto di vista etico poter garantire almeno i seguenti requisiti:

- La finalità difensiva dell'esercizio dei sistemi autonomi (Ovvero la natura di **Defensive Authonomous Weapon Systems**)
- Assicurare il loro controllo e la validazione finale della decisione esecutiva da parte dell'uomo (**controllo**).
- Progettarli in modo che abbiano sistemi di tracciamento che garantiscano l'attribuzione delle responsabilità nel loro uso (**responsabilità**).
- I loro sistemi di apprendimento e adattamento devono documentare ed essere in grado di spiegare in modo comprensibile all'operatore umano le loro determinazioni (**trasparenza e spiegabilità**).
- Sulla scorta del precedente, si deve fare in modo che l'operatore umano preveda il comportamento delle loro funzioni autonome (**fiducia**), tanto in ambito direzionale quanto in quello operativo.
- Bisogna sviluppare codici etici professionali ed addestrare operatori umani che siano responsabili del loro uso e siano chiaramente identificabili (**formazione**)

SOSTENIBILITÀ

Il più grande singolo fattore tecnologico che favorirà il raggiungimento dei 17 obiettivi di sviluppo sostenibile (SDG) nei prossimi anni sarà la rivoluzione digitale, determinata dai continui progressi nel campo dell'informatica ed in particolare del *machine learning* e della robotica. La rivoluzione digitale rivaleggia con il motore a vapore, il motore a combustione interna e l'elettrificazione per gli effetti pervasivi su tutti i settori dell'economia e della società e pertanto impattante molti ambiti indirizzati dai 17 SDG.

Ad esempio la IA permeerà in misura crescente quasi tutti i settori dell'economia, dall'agricoltura (agricoltura di precisione), all'industria mineraria (veicoli autonomi), alla produzione (robotica), alla commercializzazione (profilazione), alla finanza (modelli comportamentali), ai media (*targeting* individuale), alla salute (diagnostica), ecc.

In generale, questi contributi della tecnologia possono aumentare la produttività, ridurre i costi di produzione, espandere l'accesso a beni e servizi, dematerializzare la produzione riducendo l'impatto sull'ambiente, migliorare il funzionamento dei mercati, migliorare la ricerca e le terapie farmacologiche, semplificare l'accesso ai servizi pubblici, ecc.

Tuttavia, vi sono anche evidenti rischi e svantaggi della rivoluzione che devono essere identificati e affrontati. Forse il più temuto è la perdita di posti di lavoro e lo spostamento della distribuzione del reddito dal lavoro al capitale.

I processi di automazione sono in corso da decenni e una conseguenza importante, a quanto pare, è la riduzione della domanda di

lavoratori meno qualificati. Con i progressi nell'IA e nella robotica, molti più lavoratori possono ora vedere minacciati il loro lavoro e i loro redditi.
Mentre i nuovi posti di lavoro potrebbero sostituire quelli vecchi, i nuovi posti di lavoro potrebbero avere redditi reali e condizioni di lavoro più bassi.
Ci sono molte altre minacce percepite dalla rivoluzione digitale. Le identità digitali possono essere rubate. I governi e le imprese private possono invadere la privacy e monitorare gli individui contro la loro volontà o a loro insaputa. Alcune aziende possono sfruttare i loro vantaggi nell'accumulare grandi dati per conquistare una posizione dominante di monopolio nei rispettivi mercati, consentendo loro di porsi sostanzialmente al riparo rispetto alla concorrenza da parte di nuovi entranti inficiando il funzionamento del mercato.
I *social media* possono essere manipolati e *cyber* attacchi possono paralizzare una società interrompendo i flussi di informazioni o colpendo i dispositivi collegati a Internet.
Il problema dell'impatto ambientale delle tecnologie di IA richiede particolare attenzione in ragione dell'elevato consumo energetico.
Per i progetti che coinvolgono Paesi meno avanzati o in via di sviluppo, è fondamentale tenere in considerazione se e in che modo i sistemi IA possano integrarsi con le soluzioni già adottate in questi contesti e quali risorse siano necessarie per la loro effettiva implementazione. In particolare, la scarsa quantità e qualità delle informazioni digitali raccolte presso le zone di intervento può ostacolare l'adozione di sistemi IA. Occorre riflettere sull'eventualità di mettere a disposizione dei Paesi meno avanzati o in via di sviluppo dati semanticamente interoperabili raccolti o elaborati da Stati tecnologicamente più avanzati.

DIRITTI

In questa sezione si elencano i diritti che discendono dai principi e valori etici precedentemente analizzati e che devono informare le raccomandazioni circa lo sviluppo etico delle tecnologie IA.

INFORMAZIONE

Tutti i sistemi autonomi prevedono che l'individuo abbia uno scambio attivo di informazioni, utili ad elaborare lo scenario e suggerire una soluzione al problema affrontato.
È un diritto dell'utenza conoscere ed essere informato sull'intero processo: dalla raccolta dei dati e delle informazioni, dal procedimento di elaborazione ai rischi, e sulla stessa natura dell'interazione con il sistema (dove questa sia con un sistema autonomo in grado di elaborare le informazioni o meno). Il consenso informato dovrebbe essere presentato in modo chiaro e succinto permettendo una scelta consapevole ed evitando adesioni irriflesse.
Nel caso di decisioni che possono impattare significativamente sulla vita degli utenti o sulla società nel suo complesso, è necessario tutelare il diritto di scelta circa il livello di autonomia/intelligenza del sistema durante l'interazione, specificando le conseguenze della

scelta, nonché di poter richiedere un intervento totalmente umano.

EDUCAZIONE

La consapevolezza riguardante potenzialità e rischi legati alla tecnologia va di pari passo con l'educazione e la formazione tecnologica.
È desiderabile educare, istruire e formare la società e le persone ad un uso corretto e ad una coesistenza matura con la tecnologia. Differenziare educazione, istruzione e formazione permette di considerare separatamente aspetti importanti del rapporto uomo-macchina.
In questo contesto, educare significa saper inquadrare i rapporti tra persone e tecnologia, in particolare come l'individuo dovrebbe rapportarsi e interagire in modo consapevole con gli strumenti forniti. Istruire significa saper trasmettere i saperi che permettono alla persona di conoscere (anche in modo sommario o generale) come funziona la tecnologia e di conseguenza di valutarne rischi e potenzialità. Formare si riferisce ad un processo di apprendimento attraverso il quale l'utenza (conscia delle proprie conoscenze e lacune) migliora e incrementa la propria istruzione.
In quest'ottica l'abuso della tecnologia diventa quindi una carenza di educazione, mentre il proliferare di allarmismi catastrofici una mancanza di istruzione della tecnologia.
Per questo, alzare il livello di *information literacy* risulterebbe in una maggiore adeguatezza e consapevolezza degli individui rendendoli maggiormente adeguati ad affrontare la rapidità di evoluzione del mondo.

AUTODETERMINAZIONE DELL'IDENTITÀ

Proponiamo di usare come definizione di identità di una persona l'insieme di attributi materiali ed immateriali che la definiscono descrivendone unicità e diversità. L'autodeterminazione dell'identità è così un diritto imprescindibile e inalienabile radicato nel diritto alla dignità di ciascun individuo.
La natura sociale del vivere umano rende l'identità personale una caratteristica anche sociale, ovvero plasmata attraverso un percorso dove libere interazioni con l'ambiente e altri individui permettono di costruire una narrativa che cambia con il tempo e l'ambiente stesso.
La raccolta di dati e l'interazione con la tecnologia nelle pratiche quotidiane rende l'intelligenza artificiale uno strumento potente in grado di svolgere mansioni e soddisfare bisogni migliorando la qualità della vita, ma allo stesso tempo può divenire un mezzo per manipolare le decisioni degli individui minandone l'autodeterminazione (Taddeo e Floridi, 2018). Appartengono, inoltre, al medesimo ambito aspetti più pragmatici della tutela dell'identità quali il diritto alla portabilità, alla rettifica, all'oblio ed altri diritti connessi, giuridicamente tutelati.
Per rendere azionabili queste forme di tutela, l'individuo deve essere sempre messo in condizione di sapere e conoscere la natura del

suo interlocutore (artificiale o meno), le sue finalità e potenzialità, al fine di consentirgli di scegliere come interagire con l'agente e quali facoltà accordargli sui propri dati.

RISERVATEZZA

La tutela della riservatezza si declina nel riconoscimento di una sfera privata all'interno della quale l'individuo deve essere immune da intromissioni di terze parti, siano esse pubbliche o private.
L'aumento della capacità di calcolo, di *storage* e di connessione determinato dall'evoluzione dell'elettronica porta all'accumulo di dati nel tempo, provenienti da dispositivi, sensori e sonde di ogni tipo, che si fondono con l'ambiente materiale in cui le persone vivono e l'ambiente immateriale in cui esse operano; l'aumento della capacità di elaborazione, grazie ai modelli statistici che ne vengono distillati, permette di passare da applicazioni algoritmiche deterministiche ad applicazioni probabilistiche.
Ci sono quindi tre dimensioni la cui rilevanza cresce esponenzialmente rispetto alle applicazioni informatiche tradizionali: la compenetrazione spaziale, l'accumulo temporale e la modellazione statistica.
La prima dimensione impone una re-ingegnerizzazione concettuale dello spazio all'interno del quale l'individuo si muove. Con la progressiva erosione della barriera che separa immateriale e materiale, occorre valutare l'impatto dei sistemi IA nella sfera privata dell'individuo considerando l'interconnessione tra l'ambiente fisico in cui hanno luogo i suoi movimenti e la percezione digitalizzata che di essi hanno i sistemi stessi.
Una seconda prospettiva ruota intorno al tempo come elemento distintivo dell'analisi. L'utilizzo di dati riferiti al passato per lo sviluppo automatico di modelli predittivi richiede una riflessione sull'effetto di cristallizzazione sociale che si potrebbe determinare se le valutazioni più invasive della riservatezza fossero prese senza analizzare la correzione, l'integrazione o l'eliminazione di bias o dati non più rilevanti. Nei casi di accumulo permanente e indiscriminato di dati (*always on*), occorre riflettere quali concrete possibilità di sottrarsi, anche temporaneamente, alla raccolta di informazioni siano da garantire agli individui per ragioni di riservatezza.
Una terza prospettiva è legata ai profili cognitivi dell'IA e all'estrazione di nuova conoscenza a partire dai dati. Nella fase di valutazione della pervasività dei sistemi IA nella sfera privata dell'individuo, occorre porre l'accento non solo sulla gestione del dato personale osservato, ma anche sull'impatto delle inferenze che tali sistemi sono in grado di generare e sull'ottica super-individuale attraverso cui essi consentono di osservare la realtà.

ANONIMATO PROTETTO (DIFFERENTIAL PRIVACY)

Sempre più spesso emerge la delicatezza dell'equilibrio tra la necessità di poter perseguire alcuni reati e la garanzia della libertà di espressione rispetto al rischio di censura. Il concetto di reputazione, ovvero di fiducia attribuita ad un soggetto, è uno dei pilastri centrali della società in tutti i suoi aspetti relazionali, dall'economia alla politica.
La Costituzione italiana recita:
"Tutti hanno diritto di manifestare liberamente il proprio pensiero con la parola, lo scritto e ogni altro mezzo di diffusione."
Ed anche:
"La responsabilità penale è personale."
Rispetto al diritto alla libertà di espressione, le valutazioni circa la rimozione di contenuti sono affidate a moderatori che operano globalmente secondo principi definiti dalle piattaforme, non necessariamente allineati con le disposizioni dei governi locali, e con quanto stabilito dagli ordinamenti.
La libertà di espressione non deve esporre a possibili ritorsioni personali da parte di gruppi di pressione. Nel contempo, deve essere altresì garantita la possibilità, nel caso in cui il fatto costituisca reato, che l'autorità giudiziaria possa risalire alla identità di chi lo ha commesso.
Queste esigenze sono contrastanti e difficilmente conciliabili in quanto gli attuali servizi digitali e l'infrastruttura giuridica non sono state disegnate in modo coerente per gestirle. Il rapporto tra anonimato e conoscibilità è oggetto è riferibile al paradigma di differential privacy o anonimato protetto.

TUTELA DEI DIRITTI

Una efficace tutela dei propri diritti, è esso stesso un diritto fondamentale delle persone, in linea con l'obiettivo 16 degli SDG e la necessità di assicurare effettività alle condizioni essenziali della democrazia (Rodotà, 2012 p.63)
Lo sviluppo dell'elettronica determina un aumento delle capacità di calcolo, di archiviazione e di comunicazione dei dispositivi elettronici ovvero una riduzione del loro costo rendendoli pervasivi ed in grado di catturare dati da ogni azione ed interazione. Dal computer su ogni scrivania, si passa ad un computer pervasivo in cui le funzioni ed i dati delle persone sono distribuiti e l'accesso a tali funzioni e dati è assicurato tramite interfacce uomo-macchina semplificate e basato sul riconoscimento dell'identità.
Grazie a questa disponibilità di dati e capacità di calcolo, le tecniche di IA consentono di realizzare software non più algoritmicamente deterministici ma basati sull'applicazione di modelli statistici distillati dai dati. La natura delle applicazioni realizzabili cambia, consentendo di applicare la scalabilità tipica delle applicazioni informatiche ad ambiti differenti non indirizzabili con precedenti algoritmi deterministici, tra cui, ad esempio, casi tradizionalmente vincolati alla percezione umana e la loro classificazione.
In questo genere di usi dell'IA, il fattore di scala muta la natura stessa dell'applicazione. Si consideri ad esempio il passaggio dell'e-

same delle foto segnaletiche, effettuato da persone e quindi limitato a poche decine di migliaia di individui, alla sua effettuazione con strumenti IA, indefinitamente scalabile, e quindi potenzialmente riguardante molti milioni di persone. Nel definire le riserve di legge, i padri costituenti potevano prevedere esclusivamente l'intervento umano che implicitamente incorpora frizioni e limiti alla scalabilità. Con la scalabilità offerta dall'IA la natura del controllo, può mutare da eccezione a regola sociale che non può essere considerata alla stregua di un semplice *upgrade* tecnologico ma che pone nuovi interrogativi in merito alla tutela dei diritti delle persone.

DIRITTI DEI SOGGETTI DEBOLI

Alcune categorie di soggetti, quali minori, anziani, e non autosufficienti si possono trovare nell'incapacità di prendere decisioni autonome, incapacità legata all'età o a condizioni psico-fisiche.
La Convenzione ONU sui diritti dell'infanzia e dell'adolescenza stabilisce in particolare come l'individuo di minore età sia titolare di diritti imprescindibili oltre che un elemento al quale garantire particolare tutela in quanto oggetto di cure e assistenza speciali [11]. L'art. 29 sancisce l'importanza dello sviluppo della personalità, identità e attitudini con speciali riferimenti all'ambiente e alle relazioni interpersonali descritte attraverso diversi livelli di interscambio culturale.
Oggi si osserva come, generalmente, la tecnologia, ed in particolare l'IA, sia invece progettata e costruita per catturare l'attenzione (soprattutto dei più giovani) e mantenere quest'ultima il più a lungo possibile al fine di raccogliere o produrre più materiale informazionale possibile.
La tecnologia dovrebbe invece considerare queste categorie un elemento di forte tutela, promuovendo valori e metodologie che lo inquadrino in un ambiente di assistenza e crescita, favorendone lo sviluppo cognitivo e che ne permetta la libera determinazione, senza vincolarne scelte, preferenze ed attitudini, in particolare se strumentali ad uno sfruttamento commerciale o con finalità manipolative (SDG 3).

OBBLIGHI E RACCOMANDAZIONI

Alla luce dei principi e dei diritti identificati in precedenza, questa sezione individua obblighi e propone raccomandazioni a cui attenersi nello sviluppo etico e nell'utilizzo regolamentato di sistemi di IA.

FIDUCIA

La fiducia gioca un ruolo cruciale in ogni processo di innovazione: solo tutelando e promuovendo il capitale di fiducia sociale relativo all'IA sarà possibile coglierne appieno le potenzialità. Di conseguenza, è necessario individuare i fattori che potrebbero minare la fiducia in questa tecnologia e mettere in campo misure efficaci che ne limitino o eliminino le esternalità negative.
La fiducia è un collante sociale primario e sostiene l'organizzazione del lavoro, la divisione dei compiti e la delegazione delle mansioni,

rendendo disponibili a livello comunitario opportunità e prospettive altrimenti irraggiungibili.
Data la rilevanza della fiducia in ogni interazione umana, è necessario prendere precauzioni affinché anche le tecnologie basate sull'IA — in quanto mediatori di relazioni sociali — risultino affidabili, degne della fiducia dei diversi attori coinvolti nel loro utilizzo (SDG 9).
Le tecnologie realizzate devono rispecchiare i valori degli utenti e della società nel cui contesto vengono utilizzate, conformemente a obiettivi socialmente condivisi.
È necessario che siano definiti *framework* di fiducia, ovvero corpus di metodologie, regole, certificazioni, controlli, sanzioni, *benchmark* finalizzati al raggiungimento di target socialmente desiderabili e politicamente determinati.

ACCESSIBILITÀ

Essendo assai rilevante l'impatto sociale delle decisioni prese per mezzo di algoritmi, è fonte di grande preoccupazione il fatto che le modalità tramite cui tecniche di apprendimento automatico processano informazioni non siano facilmente accessibili, ovvero non siano trasparenti e spiegabili. Questa è la ragione per cui si raccomanda a più voci che i sistemi decisionali autonomi siano i più trasparenti possibile e che le loro determinazioni possano essere spiegabili.
L'opacità dei modi in cui questi sistemi elaborano i dati su cui sono basate le decisioni prodotte — che fa di essi delle *black box* (Pasquale, 2015) — è problematica sia da un punto di vista tecnico che da un punto di vista sociale: l'assenza di accessibilità può ingenerare il sospetto che alcune correlazioni su cui la decisione è basata possano incorporare pregiudizi eticamente condannabili e causare discriminazioni, trattamenti non equi e ingiustizie.
Sul piano tecnico è necessaria la trasparenza per rendere conoscibile come funzioni il processo decisionale del sistema, la sua logica interna, e per poterlo validare da un punto di vista tecnologico.
Sul piano sociale la spiegabilità attiene alla traduzione della funzione dell'algoritmo in termini comprensibili agli utenti, per poter fornire le motivazioni dell'output.
Infatti, si può avere spiegabilità senza trasparenza ad esempio quando una decisione venga presa sulla base di un criterio noto ma senza accesso all'algoritmo che lo elabora.
Si può avere trasparenza senza spiegabilità laddove venga fornito un pieno accesso ai dati ed agli algoritmi ma la determinazione dell'output non sia motivabile dal sistema. Il valore della trasparenza richiede lo sviluppo di nuove tecniche di spiegazione capaci di aprire le black box e rendere conto dei loro processi interni.
La garanzia dell'accessibilità è inscindibile dall'elaborazione di policy che, includendo trasparenza e spiegabilità, rafforzi la fiducia e protegga il diritto degli utenti – cittadini, autorità e comunità scientifica - ad essere informati in modo semplice e chiaro circa l'uso di intelligenze artificiali e dei limiti collegati a tale uso.

SICUREZZA

La tutela della sicurezza della persona deriva direttamente dai valori fondamentali della dignità personale e come tale richiede di essere rispettata in ogni fase del processo di design e utilizzo delle tecnologie IA. Di conseguenza, concentriamo l'attenzione sulla sicurezza intesa come integrità del sistema.

La pervasività della tecnologia rende la sicurezza un obbligo imprescindibile che il fornitore, di servizi o della tecnologia, deve assicurare a diversi livelli di applicabilità: dalla tutela della sicurezza degli individui alla conservazione dei dati personali, dalla protezione e gestione degli asset fisici all'integrità strutturale del sistema.

Per quanto riguarda i dati personali, ricordiamo le indicazioni riportate dal Regolamento Ue 2016/679 [7]. Tra gli altri diritti garantiti dal testo normativo, è necessario predisporre misure idonee a limitare il trattamento di dati personali. Tale facoltà assume un particolare significato di tutela della sicurezza dei dati nei confronti del loro impiego da parte di sistemi informatici (Rodotà, 2012 p. 399) e, a maggior ragione, di sistemi di IA.

Un sistema sicuro dovrebbe poi escludere ed evitare possibili esternalità negative, come anche essere strutturato in modo tale da non incentivare il sistema stesso o terze parti a raggiungere l'obiettivo prefissato con strumenti o azioni non idonee. Allo stesso tempo, dovrebbe essere consentito al sistema di apprendere nuove strategie per completare un'azione senza che queste abbiano ripercussioni inattese, e garantirgli un grado di flessibilità tale da potersi adattare a diverse situazioni senza dover essere controllato ad ogni passo, specie qualora i meccanismi di monitoraggio siano estremamente complicati o dispendiosi.

Identificare accuratamente il livello di sicurezza di un sistema è di capitale importanza tanto quanto fare in modo di comunicare tale livello agli utenti in modo intuitivo e chiaro. Da questo punto di vista, dovrebbe essere elaborato un linguaggio funzionale – ad esempio, tramite il ricorso a certificazioni o etichette – che semplifichi la comprensione del livello di sicurezza ed affidabilità di una tecnologia IA.

USABILITÀ

È risaputo che alcuni sistemi IA si fondano su complessi modelli computazionali, talvolta poco interpretabili e opachi (*black box*), che possono ingenerare una percezione di mancato controllo nella fase di esecuzione.

Se, da un lato, la rilevanza dell'intermediazione tra sistema ed essere umano trova una sua giustificazione nell'esigenza di assicurare trasparenza, controllo e trust, dall'altro essa acquista ancor maggiore significato quando speciali categorie di utilizzatori usufruiscono del sistema. I benefici offerti dai sistemi IA devono essere accessibili a persone non autosufficienti o portatori di handicap, per garantire la massima espressione del loro potenziale e una vita significativa (art. 3 Cost., Art. 26 Carta dei Diritti Fondamentali

dell'UE). Speciale attenzione, inoltre, va posta sull'utilizzo consapevole di sistemi IA da parte dei minori.
Occorre porre in bilanciamento, da una parte, l'esigenza di adattare i sistemi alle diverse fasi di sviluppo del minore (ad esempio, sviluppo del linguaggio o del pensiero matematico) e alle sue condizioni soggettive (diversità culturali e di linguaggio o disturbi di apprendimento) e, dall'altra, i limiti da porre allo sviluppo di i nterfacce manipolative o in grado di ingenerare confusione circa la natura artificiale del sistema (SDG 4) .
Risulta quindi necessario, per una corretta interazione tra umani e sistemi di IA, che le interfacce pongano il fruitore umano del sistema al centro del progetto di sviluppo.

CONTROLLO

Una partecipazione attiva di un essere umano nelle decisioni prese da sistemi IA è necessaria affinché l'operatore non rivesta il ruolo di esecutore passivo esente da responsabilità morali e giuridiche connesse all'utilizzo di tali sistemi. Tale supervisione previene la riduzione dei destinatari delle decisioni a mere variabili di un calcolo probabilistico, condizione inaccettabile nel caso di situazioni critiche in cui è presente un rischio per valori etici di rilevanza costituzionale o in cui sia necessaria una valutazione morale delle conseguenze della decisione. La discussione sul controllo dei sistemi IA può articolarsi in due direzioni: una descrittiva, in cui si definisce il grado di autonomia di un sistema IA (ad esempio, tramite una grandezza ordinale in cui a ciascun valore corrisponde un certo grado di autodeterminazione della macchina e, per converso, il grado di controllo umano (SAE International, 2018)); una normativa, in cui si trasla ogni livello di controllo in un determinato regime giuridico attribuendo responsabilità differenziate e appropriate al contesto di utilizzo.

RESPONSABILITÀ

Il dilemma della responsabilità è sicuramente uno degli aspetti più problematici nello sviluppo di nuovi sistemi di IA. Non è tuttavia chiaro se la responsabilità per alcune decisioni prese da un sistema intelligente debbano essere attribuite al suo sviluppatore, al venditore del software, all'utilizzatore o a terze parti. Inoltre, è interessante notare come diverse persone abbiano sviluppato una naturale e irrazionale "avversione verso gli algoritmi", che introduce un elemento aggiuntivo di responsabilità: se un medico decide di non seguire la raccomandazione di un sistema di IA non reputato affidabile, ma sbaglia, può essere ritenuto responsabile per gli esiti di tale decisione? Fino a che punto, quindi, possiamo ignorare un sistema di IA? É chiaro che il problema non sia di facile risoluzione, e che probabilmente sia difficile se non controproducente creare e utilizzare un unico sistema di riferimento per gestire la responsabilità di attori diversi in circostanze differenti. Potrebbero essere richiesti, infatti, molteplici framework che considerino contemporaneamente non solo la responsabilità come tale, ma anche la trasparenza,

equità, accuratezza e il grado di controllo di un algoritmo.
Per le applicazioni che possono avere un impatto significativo su società, persone e cose occorre attribuire *ex ante* le responsabilità connesse all'impiego di tali sistemi invece di attendere la valutazione *ex post* di un soggetto chiamato ad allocare responsabilità oggettive.
I termini contrattuali che dettagliano diritti, facoltà, immunità e privilegi devono essere chiari ed accessibili, soprattutto nelle applicazioni che sono destinate o che incidono su un ampio numero di persone.
Inoltre, è necessario elaborare meccanismi di *accountability* che impediscano strategie di deresponsabilizzazione o di attribuzione di responsabilità a soggetti non umani.

RIPARAZIONE

Come abbiamo il principio della "privacy by design" per i sistemi di gestione dei dati personali, è opportuno considerare, per i sistemi basati sull'IA che prendono decisioni che possono influenzare la vita delle persone, l'introduzione di meccanismi di riparazione (*redress*) sulla base di un principio di "*redress by design*".
La considerazione di base è che un sistema di IA non difettoso, perfettamente funzionante, effettuerà predizioni errate.
Ciò accade, sia a causa del bias presente nei dati di addestramento, sia perché un sistema di questo tipo è inerentemente non deterministico come, per esempio, un autovelox può essere: se si supera il limite di velocità con la propria auto, l'autovelox lo rileva e si ottiene una multa. È possibile fare ricorso, ma l'automobilista è ritenuto colpevole fino a prova di innocenza perché un sistema deterministico, non difettoso (correttamente configurato, certificato e controllato) stabilisce la sua colpevolezza.
Ma con un sistema di intelligenza artificiale non difettoso, perfettamente funzionante, essendo un motore statistico che produce necessariamente risultati probabilistici, questa decisione potrebbe essere giusta nel 98% delle volte e sbagliata nel 2% delle volte (sarebbe inappropriato classificare queste predizioni errate come sbagli), il che significa che in questo 2%, la persona viene riconosciuta colpevole anche quando non lo è (o non può ottenere un servizio, anche se ha pieno diritto di ottenerlo).
Per la persona, la decisione sbagliata può generare ricadute, superando la portata della decisione stessa, ad esempio generando biasimo sociale, feedback negativi online e altre conseguenze che possono diffondersi nella Rete e diventare impossibili da rimuovere.
Questo 2% di errore tollerato non é da intendersi negativamente, poiché garantisce una flessibilità strutturale all'algoritmo e la capacità di adattarsi e includere nuovi elementi emergenti.
In questi casi errati (possono essere falsi positivi o falsi negativi), la procedura di ricorso può non esistere o, se esiste, può essere inefficace, il suo costo può essere eccessivo, può risultare concretamente non accessibile a tutti, può richiedere un tempo eccessivo o può non rettificare le suddette ricadute.

Una più efficace tutela dei diritti dovrebbe prevedere il principio di *redress by design*, ovvero la previsione, fin dalla fase di progettazione, di meccanismi atti a garantire la ridondanza, sistemi alternativi, procedure alternative, ombudsman, ecc. per poter individuare efficacemente, verificare, correggere le decisioni sbagliate prese da un sistema non difettoso, perfettamente funzionante ed eventualmente, affinare le capacità predittive del sistema.

UN ESEMPIO DI REDRESS BY DESIGN

Le norme di prossima vigenza in materia di enforcement del copyright implicano la predisposizione di filtri dei contenuti caricati dagli utenti sulle piattaforme di condivisione. Tali filtri sfruttano l'IA per confrontare i video caricati con un database di firme di materiale protetto da copyright. La presenza di sanzioni elevate spingerà i gestori delle piattaforme a configurare i loro sistemi massimizzando il recupero, a scapito della precisione. Video di persone innocenti verranno ritenuti in violazione del copyright a causa di predizioni errate legate all'errore statistico inerente al modello, intaccando la libertà di manifestazione del pensiero degli individui i cui contenuti saranno erroneamente censurati. Sebbene ciò possa essere ritenuto un danno marginale quando esaminato a livello sociale complessivo, per la singola persona sarà una violazione totalizzante di un proprio diritto. Con questo meccanismo si istituisce un primo grado di giudizio in cui, un sistema non deterministico, che è noto commettere errori, stabilisce la illiceità di un comportamento di una persona, fino a prova contraria. La procedura di ricorso avverso le decisioni prevede l'intervento di una autorità nazionale, come ad esempio AGCOM. È prevedibile che la procedura di ricorso sarà effettivamente non accessibile a tutti, non sufficientemente tempestiva o valutata in molti casi eccessivamente complessa.

Un'applicazione del principio di *redress by design* potrebbe prevedere una procedura alternativa di trattamento dei casi in cui il sistema IA predice la violazione: il contenuto potrebbe essere immediatamente pubblicato qualora l'utente accetti di associare la propria identità al contenuto stesso (ad esempio in modo analogo a quanto avviene per gli accessi al Wi-Fi con credenziali via SMS) assumendosi quindi la responsabilità di eventuali illeciti che, con riti e garanzie abituali, verrebbero stabiliti da una corte in grado così di conoscere l'identità del presunto colpevole.

TITOLARITÀ DEI DATI

I dati pertengono all'individuo che li genera. Diverse sono le realtà che utilizzano il dato al fine di estrarne informazione. L'informazione migliora il livello dei sevizi ottenuti in termini di accuratezza ed affidabilità. Basti pensare per esempio a come, nei moderni sistemi di navigazione, le informazioni aggregate permettano di conoscere lo stato del traffico in un determinato tratto stradale informando gli utenti e abilitandoli ad un eventuale cambio di tragitto, migliorando in questo senso il livello di servizio erogato.
Tale informazione è fondamentale per la definizione di meccanismi

in grado di fornire servizi utili alle persone, al fine di migliorarne la qualità di vita. La tutela del dato grezzo, della sua titolarità, del mantenimento e possesso è solo un primo passo non ancora sufficiente a garantire la tutela dell'individuo e della collettività. Con l'aumentare dei dati raccolti ed elaborati aumenta la necessità di definire dei meccanismi tecnici, contrattuali e regolamentari che disciplinino l'estrazione e la gestione dell'informazione, identificandone livelli di accessibilità, modalità di impiego e divulgazione. Anche data la natura non rivale e solo parzialmente escludibile dei dati, agli utenti devono essere tecnicamente forniti piena trasparenza e controllo dei dati raccolti ed elaborati da sistemi di IA, garanzie che devono essere assicurate a livello contrattuale. Lo scenario può radicalmente mutare con l'introduzione di sistemi crittografici o sistemi di calcolo distribuiti (overlay network crittografici, crittografia omomorfica o sistemi di IA distribuita). Vanno perciò seguiti con attenzione i loro sviluppi, al fine di consentire agli utenti una piena titolarità in merito alla disponibilità dei propri dati.

GOVERNANCE

La questione riguardante la governance dell'IA non riflette, tecnicamente, un singolo problema, bensì una moltitudine di aspetti differenti. É infatti una problematica omnicomprensiva che tocca tematiche legate tanto alla giustizia, alla responsabilità, dalle strategie e politiche nazionali che concernono l'IA fino alla sorveglianza "intelligente".

- La considerazione che l'IA determina una scalabilità che trascende i naturali limiti alle attività di percezione e classificazione umane, ponendo pressioni su processi sociali consolidati, suggerisce la opportunità di considerare di disporre l'introduzione di frizioni per limitare o rallentare tale scalabilità nei casi in cui ciò possa determinare esternalità negative (si pensi al sopracitato esempio relativo alle foto segnaletiche o alla diffusione di fake news che minano coesione sociale e processi democratici).
- Dalla natura intrinsecamente statistica dell'IA consegue che singole istanze di un sistema non difettoso possano determinare eventi avversi a causa di predizioni errate, nel contempo determinando effetti benefici complessivi assai maggiori rispetto alla situazione precedente. Istanze di un sistema di ausilio alla guida possono errare causando qualche incidente, anche fatale, ma nel complesso del suo utilizzo riducendo grandemente il numero di incidenti e vittime. Per certi versi si tratta di una situazione analoga a quella dei prodotti farmaceutici.
- La governance dell'IA dovrebbe assicurare un'adeguata individuazione, misurazione e classificazione delle previsioni errate causate da sistemi non difettosi, per garantire che rientrino nei valori obiettivo socialmente desiderabili stabiliti grazie a processi democratici (SDG 16). In determinati casi potrebbe essere necessario condurre test di validazione dei sistemi e di misura dei loro effetti prima della loro commercializzazione. In taluni

altri dovrebbero essere definite procedure di validazione e valutazione di conformità, almeno per i sistemi che possono avere rischi di impatti significativi su cose, persone e società, consentendo di escludere o attribuire correttamente dolo e negligenza.

- Per affrontare tali situazioni andrebbe considerata la realizzazione di una autorità o agenzia incaricata di monitorare la diffusione dell'IA e rilevare le sfide emergenti, fornendo informazioni ai decisori politici ed applicandone le determinazioni. Dovrebbe assicurare il soddisfacimento degli obiettivi a livello di sistema fissati dai responsabili politici in relazione alle classi di applicazioni di IA, sulla base degli effetti sugli individui e sulle organizzazioni sociali. A tal fine potrebbe emettere linee guida e raccomandazioni per gli sviluppatori di sistemi IA ed assistere le aziende nell'applicazione di un approccio basato su valutazioni di rischio ed impatto.
- Tale organismo potrebbe infine assicurare un coordinamento con l'Unione Europea ed altri organismi internazionali di standardizzazione. Inoltre, potrebbe giovarsi della collaborazione dei corpi intermedi quali sindacati ed associazioni dei consumatori.

FORMAZIONE

L'IA detiene una notevole rilevanza anche nel settore dell'educazione (SDG 4). Si rende necessario, pertanto, identificare in che misura essa generi opportunità e rischi in settori attinenti.

Il corpo sociale tutto deve poter accedere a percorsi formativi per qualificarsi e riqualificarsi, in modo che l'impatto dell'IA sul mondo del lavoro possa incontrare professionisti preparati a coglierne le opportunità e all'altezza delle sfide che tale tecnologia pone a livello tanto etico quanto sociale, evitando così la formazione di gruppi marginalizzati e incapaci di trovare il proprio ruolo nel nuovo panorama lavorativo.

La formazione degli utenti deve permettere un uso consapevole della tecnologia che non ne demonizzi la natura ma ne evidenzi le potenzialità responsabilizzando l'individuo di fronte a rischi e pericoli.

Il mondo aziendale deve promuovere percorsi educativi volti a facilitare l'integrazione di considerazioni di carattere etico legate alle tecnologie IA in via di sviluppo, sia in sede di design – sostenendo percorsi interdisciplinari e critici – sia in tutti gli altri momenti relativi alla presentazione e alla pubblicizzazione del prodotto.

In ultimo, i *decision maker* di ogni livello devono acquisire piena consapevolezza della natura e del funzionamento dei sistemi di IA al fine di redigere regole adeguate al loro utilizzo.

SITOGRAFIA

- European commission for the efficiency of justice (CEPEJ) (2018). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment [https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c]
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Schafer, B. (2018). 'AI4People—An ethical framework for a good AIsociety: Opportunities, risks, principles, and recommendations'. Minds and Machines, 28(4): 689-707.
- Future of Life Institute (2017). AI Principles – Future of Life Institute [https://futureoflife.org/ai-principles/]
- Hammond, K. (2016). 5 unexpected sources of bias in artificial intelligence. [https://techcrunch.com/2016/12/10/5-unexpected-sources-of-bias-in-artificial-intelligence/]
- House Of Lords (2018). AI in the UK: ready, willing and able? [https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf]
- IEEE (2019). Ethically Aligned Design. [https://ethicsinaction.ieee.org]
- Independent High-Level Expert Group on AI set up by the European Commission (2019). Ethics guidelines for trustworthy AI [https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai]
- ISO/IEC 27000 family – Information security management systems – [https://www.iso.org/isoiec-27001-information-security.html]
- Partnership on AI. (2016). Tenets – The Partnership on AI [https://www.partnershiponai.org/tenets/]
- Pasquale, F. (2015). The black box society. Harvard University Press.
- Quintarelli, S. (2019). Capitalismo immateriale. Bollati e Boringhieri
- Regulation 2016/679 of the European Parliament and of the European Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L 119 (GDPR)
- Rodotà, S. (2015). Il diritto di avere diritti. Gius. Laterza & Figli Spa.
- SAE International (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles [https://www.sae.org/standards/content/j3016_201806/]
- Sîrbu, A., Pedreschi, D., Giannotti, F., Kertész, J. (2019). 'Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model'. PLOS ONE, 14(3), p.e 0213246.
- Taddeo, M., Floridi, L. (2018). 'How AI can be a force for good'. Science, 361(6404): 751-752.
- Villani, C., Bonnet, Y., Rondepierre, B. (2018). For a meaningful artificial intelligence: towards a French and European strategy. [https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf]

Paper sui principi giuridici

Il Presente documento è stato redatto da Marco Bassini, Giovanni De Gregorio, Marco Macchia, Alessandro Pajno, Francesco Paolo Patti, Oreste Pollicino, Serena Quattrocolo, Dario Simeoli, e Pietro Sirena, con il coordinamento di Alessandro Pajno.

PREMESSA

Il presente contributo nasce dal desiderio della Fondazione Leonardo - Civiltà delle Macchine di approfondire lo studio dei moderni sistemi di intelligenza artificiale e delle relative implicazioni anche da una prospettiva giuridica. L'impegno che si intende profondere mira a stimolare la comunità scientifica e, più in generale, gli *stakeholders* e gli attori istituzionali, a una meditata presa di coscienza sulle conseguenze che derivano dalla diffusione di nuove tecnologie su larga scala che offrono inedite opportunità ai cittadini, alle imprese e alle pubbliche amministrazioni ma al contempo, ove non adeguatamente governate, possono dare origine a rischi per i diritti e le libertà fondamentali.

Si impone, altresì, la necessità di assicurare che il progresso tecnologico si svolga in armonia con le esigenze di tutela individuali e collettive, nel rispetto di una dimensione antropocentrica. Occorre così delineare le modalità di intervento che minimizzino le esternalità negative sullo sviluppo delle nuove tecnologie garantendo al contempo la salvaguardia del nucleo duro dei diritti fondamentali.

Con il presente documento si è cercato di effettuare una mappatura dei tentativi sinora condotti a titolo sia di *soft law*, sia di *hard law*, in campo giurisprudenziale e in ambito dottrinale, per offrire un congruo inquadramento giuridico dell'intelligenza artificiale. Si è cercato altresì di illustrare quali principi dovrebbero orientare e guidare i prossimi sviluppi. Svolta una mappatura preliminare, infatti, il documento procede secondo una logica *de jure condendo,* effettuando una rassegna di problematiche attraverso ideali carotaggi entro i vari settori del diritto che appaiono attinti dalle innovazioni tecnologiche di cui si discute.

DALLA SOCIETÀ DELL'INFORMAZIONE ALLA SOCIETÀ DELL'ALGORITMO

L'emersione di tecnologie caratterizzate dall'impiego di sistemi di intelligenza artificiale ha inaugurato una nuova stagione di dibattito tra soggetti pubblici e privati in merito alle principali questioni etiche, sociali e giuridiche attorno all'impiego e alle conseguenze

relative all'impiego di tali tecnologie. Università, governi e imprese sono soltanto tre dei centri nevralgici che hanno dimostrato uno spiccato interesse per il tema dell'automazione con il fine di presentarsi preparati al nuovo cambio di paradigma che può essere descritto come un'evoluzione della società dell'informazione in quella dell'algoritmo. L'intelligenza artificiale è ormai pronta per fare il suo ingresso massivo nella società, provocando un conseguente bisogno di riadattare concetti e categorie tradizionali al nuovo modello sociale governato dall'automazione. Una tale diffusa attenzione al tema dell'intelligenza artificiale deriva non solo dai numerosi risvolti concernenti l'impiego di tali tecnologie in una pluralità di settori distinti ma anche da un'evidente, e più generale, dicotomia che descrive la relazione tra uomo e macchina.

Volendo ricorrere a un parallelismo con la dimensione digitale, l'intelligenza artificiale sembra esprimere la stessa capacità di mettere in discussione categorie preesistenti (*disruptiveness*) specialmente se si osservano le pretese anarchiche sviluppate all'indomani della diffusione della rete negli anni '90 (Barlow, 1996; Johnson and Post, 1997). La pretesa anarchica può, infatti, collegarsi in senso lato alla non prevedibilità degli esiti dell'utilizzo di sistemi di intelligenza artificiale in funzione o di *bias* in fase di elaborazione di dati alla fonte (*garbage in, garbage out*) ovvero in funzione di capacità di apprendimento autonomo (*machine learning*). In questo caso, tuttavia, non si assiste all'impossibilità di applicare alla dimensione digitale le regole del mondo atomico, ma semplicemente all'esigenza di ripartire in maniera ottimale anche sotto un profilo economico i rischi connessi alla produzione di un evento lesivo o di decisioni che possano avere un effetto sui diritti degli individui.

La lezione di Lessig in merito alla possibilità di regolamentare lo spazio digitale ha mostrato che il diritto fatica a imporre il rispetto delle sue regole nella tecnologia e dunque le sue finalità non paiono trovare soddisfazione "in autonomia". La stessa lezione di Lessing ha mostrato, però, che è sempre possibile ricorrere a una regolazione che incida sull'architettura della tecnologia, dato che il *code* si nutre e si alimenta di fattori anche diversi dal diritto, tra cui il suo design, le norme sociali e le regole del mercato (Lessig, 1996). Si deve dunque prendere atto che le norme giuridiche necessitano di essere calate sulla peculiarità delle tecnologie, al fine di preservare la propria funzione prescrittiva. Il diritto può quindi rimarcare i suoi fini tramite una regolazione dell'architettura della tecnologia (Reidenberg, 1997). Applicando queste considerazioni nel campo dell'intelligenza artificiale, ne conseguirebbe una differenziazione tra le varie forme di intelligenza artificiale basata su una selezione di quelle "più virtuose" in funzione di un minor livello di rischio per l'individuo.

Tali considerazioni, tuttavia, non devono spingere a cadere nell'errore di ritenere, per analogia con le caratteristiche di una *disruptive technology* quale Internet, che l'intelligenza artificiale abbia rag-

giunto uno stadio di maturità tecnologica. Al contrario, la situazione odierna è solo l'inizio, se non una timida anticipazione, di un fenomeno le cui implicazioni non sono ancora del tutto misurabili e considerabili in modo esaustivo.
Gli obiettivi del documento sono diversi. Oltre a sollecitare un dibattito sul codice etico dell'intelligenza artificiale e passare in rassegna i casi più critici in Italia e all'estero, il manifesto ambisce a individuare un primo nucleo di principi giuridici che devono essere rispettati nella regolamentazione delle nuove tecnologie. Detti principi traducono alcune scelte etiche in disposizioni dell'ordinamento, offrendo spunti per la configurazione di uno statuto giuridico della produzione e dell'utilizzazione dell'intelligenza artificiale nel rispetto dei valori espressi dalla Costituzione, nonché dalle fonti internazionali ed europee.

GLI STRUMENTI DI SOFT LAW

Il carattere ancora acerbo del dibattito, e allo stesso tempo la sua rilevanza, vengono sottolineate dalla elaborazione di carte e strumenti a livello sovranazionale che si sono susseguiti nel corso degli ultimi due anni al fine di fornire le prime linee guida sull'intelligenza artificiale da almeno due punti di vista: la *governance* e la tutela dei diritti.
All'interno della prima categoria si possono certamente menzionare i principi sull'intelligenza artificiale emanati dall'Organizzazione per la cooperazione e lo sviluppo economico (OCSE) nel marzo 2019. La raccomandazione costituisce il primo *standard* intergovernativo sull'intelligenza artificiale e mira a perseguire diversi scopi: tra di essi, la promozione della ricerca e nello sviluppo dell'automazione, di un ecosistema digitale per l'intelligenza artificiale nonché la cooperazione internazionale in questo campo. Il punto centrale del documento consiste nel bilanciamento tra l'innovazione, che riveste un ruolo cruciale per lo sviluppo di nuovi servizi intelligenti, e la gestione responsabile di tali tecnologie che si concretizza nel rispetto dei diritti umani e dei valori democratici. Più in particolare, tale documento identifica cinque linee guida complementari ai fini di una gestione responsabile dell'intelligenza artificiale: crescita inclusiva, lo sviluppo sostenibile e il benessere; la salvaguardia dell'equità e di valori parametrati sull'uomo; la trasparenza e comprensibilità; la robustezza, sicurezza e affidabilità; la responsabilità.
A livello regionale, il rapporto pubblicato nel giugno 2018 dal Gruppo di esperti europeo sull'intelligenza artificiale costituisce un punto di riferimento per i prossimi passi. In particolare, il documento in questione sottolinea il ruolo di queste nuove tecnologie nel plasmare il tessuto sociale e costituire un fattore cruciale per la crescita dell'Europa. Allo stesso tempo, gli effetti positivi derivanti dallo sviluppo e diffusione di tali tecnologie vengono mitigati dai rischi che possono derivare dall'adozione di un approccio tecno-centrico, nel quale l'uomo viene relegato in una dimensione marginale rispetto alla tecnologia. Ed è qui che il documento sotto-

linea come le tecnologie basate su sistemi di intelligenza artificiale debbano conformarsi a un approccio antropocentrico. Tali tecnologie dovrebbero, infatti, costituire non solo un obiettivo dettato dall'innovazione ma anche un mezzo per aumentare il benessere umano. Da una tale ricostruzione deriva che siffatte tecnologie devono, da un lato, essere programmate nel rispetto dei diritti fondamentali, della normativa applicabile e dei principi e dei valori di base, dall'altro, assicurare affidabilità dal punto di vista tecnico, dal punto di vista della sicurezza e per capacità di essere utilizzate in modo trasparente dalla società.

Sotto un altro punto di vista, la Carta etica sull'uso dell'intelligenze artificiale nei sistemi giudiziari adottata nel contesto del Consiglio d'Europa dall'*European Commission for the Efficiency of Justice* (CEPEJ) nel dicembre del 2018 si occupa di definire i principi che i responsabili politici, i legislatori e i professionisti dovrebbero adottare nell'affrontare il rapido sviluppo dell'intelligenza artificiale nel campo della giustizia e, più in particolare, nei sistemi giudiziari nazionali. Similmente al *report* del gruppo di esperti europei, il CEPEJ sottolinea l'importanza dell'applicazione di tali tecnologie nel campo della giustizia al fine del miglioramento dell'efficienza e della qualità dei processi. Allo stesso tempo, viene messo in luce come l'implementazione di tali tecnologie debba essere attuata in modo responsabile, nel rispetto dei diritti fondamentali garantiti, in particolare nella Convenzione europea sul Diritti umani (CEDU) e nella Convenzione del Consiglio d'Europa sulla protezione dei dati personali. Tra i suddetti principi occorre richiamare *in primis* il principio del rispetto dei diritti fondamentali, che mira ad assicurare che strumenti e servizi di intelligenza artificiale siano progettati e attuati secondo modalità idonee a salvaguardare la tutela di tali diritti. Da un tale approccio costituzionale derivano il principio di non discriminazione che mira a prevenire specificamente lo sviluppo o l'intensificazione di qualsiasi discriminazione tra individui o gruppi di individui; il principio di qualità e sicurezza che richiede di ricorrere a fonti certificate e dati immateriali con modelli concepiti in modo multidisciplinare in un ambiente tecnologico sicuro; il principio di trasparenza, imparzialità ed equità atto a rendere i metodi di trattamento dei dati accessibili e comprensibili, autorizzando *audit* esterni; e il principio di controllo finalizzato a garantire che gli individui siano attori informati e in controllo delle loro scelte.

A livello dell'Unione europea, il Parlamento europeo si è occupato di fornire le prime linee guida in merito alla responsabilità civile nel settore della robotica già nel febbraio 2017. In particolare, la risoluzione del Parlamento fornisce diverse proposte in materia di responsabilità per danno causato da un robot come l'applicazione degli istituti della responsabilità oggettiva, la gestione dei rischi, l'istituzione di un regime di assicurazione obbligatorio nonché l'istituzione di uno *status* giuridico *ad hoc* consistente in una perso-

nalità elettronica, che permetta di ritenere i robot più sofisticati responsabili delle proprie azioni dannose. Più di recente, la Commissione europea ha adottato una Comunicazione su "*L'intelligenza artificiale per l'Europa*", documento non vincolante che sintetizza la strategia europea rispetto al fenomeno dell'automazione. Tale comunicazione mira a dare impulso alla capacità tecnologica e industriale dell'UE e all'adozione dell'intelligenza artificiale in tutti i settori economici, sia privati che pubblici, attraverso investimenti in ricerca e innovazione e un migliore accesso ai dati. Essa è volta, altresì, a facilitare la preparazione ai cambiamenti socio-economici apportati dall'intelligenza artificiale, incoraggiando la modernizzazione dell'istruzione e dei sistemi di formazione, sostenendo il talento, anticipando i cambiamenti nel mercato del lavoro e fornendo appoggio alle transizioni nel mercato del lavoro e all'adeguamento dei sistemi di protezione sociale. Tale strategia viene radicata in un quadro etico e giuridico adeguato, basato sui valori dell'Unione e coerente con la Carta dei diritti fondamentali dell'UE.

Le iniziative di *soft law* prese in considerazione rivelano alcune tendenze comuni.

Sottesa a larga parte dei documenti in parola è una generale preoccupazione affinché lo sviluppo dei moderni sistemi di intelligenza artificiale e delle tecnologie algoritmiche si svolga in armonia con la tutela dei diritti e delle libertà individuali. Vi è una diffusa percezione circa gli innumerevoli vantaggi che l'implementazione di queste tecnologie può recare entro una varietà di settori (dalla giustizia alla pubblica amministrazione, passando per l'esecuzione degli impegni contrattuali). Ad essa si accompagna la ferma consapevolezza in ordine alla necessità di conservare una funzionalizzazione dei loro utilizzi alla salvaguardia della dignità umana e dei diritti.

Questa preoccupazione si percepisce soprattutto nell'insistenza di alcuni documenti nel declinare principi che possano offrire un adeguato margine di manovra a legislatori e regolatori, costituendo al contempo importanti indicazioni di *policy*.

Dal punto di vista della *governance*, sembra proporsi un approccio *multistakeholder*, fondato sul coinvolgimento dei diversi soggetti che rivestono un ruolo cruciale nel funzionamento della tecnologia. Non possono essere trascurati i dubbi derivanti dall'applicazione di un tale approccio i cui effetti hanno rivelato in altri settori risultati insoddisfacenti specialmente quando si concretizzano in forme di concertazione quali, ad esempio, la *co-regulation* o la *self-regulation*. Alcuni esempi possono essere riscontrati nei codici di condotta relativi a *hate speech* e alla disinformazione *online*, dove la provenienza da parte degli attori non istituzionali delle regole segna il confine della loro efficacia. Esistono forme di concertazione tra parti private e soggetti pubblici che permettono a questi ultimi di definire con cognizione di causa e in maniera rispondente alla sensibilità dei primi standard, regole di condotta, principi e linee guida. L'adozione di un approccio *bottom-up* non esclude la validità di possibili di-

verse iniziative. Tanto più tecnico sarà il loro contenuto, tanto più potranno rivelarsi efficaci. Viceversa, se il contenuto corrisponde a scelte di *policy*, sarà più difficile immaginarne l'efficacia e pretenderne la condivisione e il supporto da parte di attori pubblici-istituzionali.

Un ulteriore profilo che emerge chiaramente dai primi tentativi di inquadrare il tema dell'intelligenza artificiale riguarda la potenziale codificazione dei diritti digitali. Tale tendenza verso la positivizzazione di nuovi interessi trova spazio alla luce dell'esigenza di coniare un nuovo statuto della persona a fronte dell'erompere di poteri privati. Tracce di questo processo emergono già dalla definizione di un diritto alla spiegazione in seno al Regolamento Generale sulla Protezione dei Dati Personali ("GDPR") nell'ambito della decisione algoritmica automatizzata. Eppure, l'art. 22 coglie soltanto una porzione, se non una fase embrionale, di sviluppo di queste situazioni giuridiche. Giova osservare che lo stesso GDPR potrebbe offrire una chiave di lettura su come approntare strumenti a protezione degli individui nel rapporto non tra pubblico e privato, ma tra soggetti che agiscono meramente come controparti di un rapporto orizzontale tramite una amministrativizzazione della protezione dati, ossia attraverso la definizione di alcune misure che rispondono all'obiettivo di assicurare maggiore *accountability* e trasparenza da parte di chi utilizzi IA.

Si pone dunque il tema della possibile cristallizzazione di nuovi diritti sostanziali e nuovi diritti procedurali, che però sono accomunati dalla "orizzontalità", ossia dal riconoscimento in capo all'individuo nel rapporto con l'utilizzatore di intelligenza artificiale che agisce come soggetto in posizione non necessariamente autoritativa, non escludendo la possibilità che ciò avvenga anche da parte di autorità pubbliche.

Tali diritti, tuttavia, necessitano di un sistema affidabile di *enforcement* in assenza del quale resterebbero soltanto delle prescrizioni senza alcuna pretesa di vincolatività. Sotto tale profilo, meno indagata rispetto alle precedenti questioni, ma allo stesso tempo cruciale per assicurare la tutela dei diritti, è la questione riguardante quali soggetti possano assicurare un *enforcement* effettivo dei nuovi diritti digitali. In particolare, l'approccio basato su una pluralità di *stakeholder* sembra essere la strada intrapresa. Si può osservare come tra i canali privilegiati da parte del GDPR per garantire il rispetto dei principi sulla *data protection* vi siano codici di condotta e meccanismi di certificazione, che naturalmente per un verso sono frutto di forme di *multistakholderism*, ma presuppongono al contempo il ruolo di un soggetto pubblico o comunque investito di una funzione pubblicistica. Questo soggetto potrebbe essere congegnato alla stregua di un ente certificatore che a livello sovranazionale, onde definire uno *standard* comune, neutrale alle specifiche sensibilità costituzionali, intervenga specificamente sugli *standard*, riconoscendo quali siano quelli idonei a garantire la tutela dell'individuo

rispetto all'uso dell'intelligenza artificiale.
Nella medesima direzione, ma sul piano degli attori pubblici, merita senz'altro attenzione l'esempio pionieristico del Codice dell'Amministrazione Digitale (d.lgs. n. 82/2005), che ha costituito un primo esempio di sistematizzazione delle norme concernenti la digitalizzazione della pubblica amministrazione nei rapporti sia con i cittadini sia con le imprese e che da ultimo è parso offrire spazi sempre maggiori al concetto di "cittadinanza digitale".
In tale quadro d'insieme, le iniziative finora susseguitesi nei vari ambiti, e in modo particolare in quello giuridico-regolatorio, documentano una frammentarietà nella declinazione di linee guida tendenzialmente settoriali e la carenza di interventi di carattere organico sotto forma di *hard law*. Si è ancora in una fase embrionale caratterizzata da interventi sparsi che declinano perlopiù principi a tutela del cittadino, ma che non sembrano ispirati da una *ratio* omogenea e unificante.
Tale frammentazione porta a riflettere sul ruolo del giurista nella società dell'algoritmo, funzione che assume una rilevanza fondamentale come avviene in tutte le fasi di transizione tra nuovi paradigmi che comportano cambiamenti radicali dello *status quo*. Tuttavia, la trasformazione del giurista non è solo dovuta a un adattamento delle capacità di affrontare le sfide dell'intelligenza artificiale utilizzando strumenti tradizionali operando nel rispetto del principio di *rule of law*, ma è anche legata al supporto che proprio le tecnologie di intelligenza artificiale promettono di fornire come ausilio alle attività del giurista. Ne consegue che le nuove tecnologie dell'automazione costituiscono non solo una sfida per il giurista, ma anche un'opportunità nella gestione efficiente dei propri compiti confermando il duplice binario (non necessariamente "oppositivo") che si sta delineando tra uomo e macchina.
In chiave di tecnica normativa, si deve porre un'ulteriore questione metodologica in ordine alla modalità che consenta di raggiungere più efficacemente gli obiettivi perseguiti da legislatori e regolatori. Sul piano delle fonti, infatti, un processo di *deregulation* e delegificazione in favore dell'intervento di atti normativi di rango secondario, entro un quadro opportunamente definito dal legislatore, appare opzione preferibile alla luce della rapidità dei mutamenti tecnologici e della necessità di disporre di *set* di norme talvolta connaturate anche da un elevato contenuto tecnico che più difficilmente potrebbero derivare dal procedimento legislativo ordinario.
Da ultimo, occorre che le autorità nazionali si confrontino con l'esperienza di alcuni paesi baltici, notoriamente più avanzati nelle politiche dell'innovazione e in particolare nello sviluppo del digitale. L'Estonia, per esempio, ha da tempo intrapreso politiche all'avanguardia in materia, elaborando sia progetti per il riconoscimento di personalità giuridica alle macchine sia proposte per l'introduzione di "robot magistrati" destinati a occuparsi di cause di minore entità. Occorre, pertanto, allargare lo sguardo a questi ordinamenti onde

comprendere se le opzioni prescelte anche in campo normativo possano prestarsi a una circolazione e a un eventuale condivisione su più larga scala, costituendo la "cartina di tornasole" della varietà dei profili al centro del dibattito.

L'ELABORAZIONE DOTTRINALE DI FRONTE ALLA SOCIETÀ ALGORITMICA

Lo sviluppo dei primi strumenti di *soft law* non ha frenato le riflessioni in campo dottrinale, destinate a confrontarsi con la difficoltà (comune alla giurisprudenza) di sussumere entro fattispecie pensate prima dell'avvento della società algoritmica fenomeni che paiono discostarsi da paradigmi consolidati. La letteratura giuridica, pur incline da decenni a confrontarsi con l'emersione del fattore tecnologico e con le sue conseguenze (Costanzo, 2012) sotto una moltitudine di prospettive – filosofia del diritto, diritto pubblico, diritto privato, diritto penale, solo per citarne alcune – appare attraversare una fase ancora embrionale. Il formante giurisprudenziale non manca di scontare analoghe criticità Tuttavia presenta una propria specificità: la necessità di risolvere un caso concreto in ossequio al principio che esige la soggezione dei giudici soltanto alla legge, e dunque entro il perimetro dello strumentario giuridico esistente.

Il contributo tanto della dottrina quanto della giurisprudenza ha senz'altro risentito, in parte, delle analoghe problematiche che si sono poste già in passato con riferimento all'ambito della robotica, pur nella consapevolezza della non piena sovrapponibilità delle categorie in parola. Le caratterizzazioni dell'intelligenza artificiale, in particolare le proprietà che ne descrivono le capacità di *machine learning*, suggeriscono tuttavia una considerazione peculiare e specifica di questo fenomeno.

Sul piano dottrinale, uno dei principali interrogativi concerne la possibilità di ipotizzare un riconoscimento di soggettività giuridica in capo ai sistemi di intelligenza artificiale (Sartor, 2009; Pagallo, 2013). Si tratta del presupposto per imputare loro non solo il riconoscimento di diritti (e doveri) ma altresì la responsabilità per possibili eventi lesivi derivanti dal loro funzionamento. Tale problematica, tra le più cogenti all'attenzione dei commentatori (Sartor, 2009), si trova al centro di un diffuso dibattito alimentato soprattutto dalla possibilità che i sistemi di intelligenza artificiale elaborino e attuino comportamenti che deviano dagli *input* ricevuti da programmatori e/o utilizzatori. Come si segnalerà oltre, le riflessioni dei commentatori si sono appuntate sull'esame dei vari paradigmi di responsabilità civile codificati dall'ordinamento giuridico, al fine di appurare la possibilità di estenderne analogicamente l'applicazione anche all'intelligenza artificiale. In ambito nordamericano, si è addirittura assistito alla teorizzazione, da parte di un autore (Hallevy, 2010), di possibili forme di responsabilità penale della macchina fondate sull'implicita attribuzione di personalità giuridica, anticamera di una sostanziale assimilazione tra agenti umani e agenti robotici. Questo modello postula la differenziazione tra tre possibili scenari: *(i)* la *perpetration through another,* ove il

sistema di intelligenza artificiale è mero esecutore materiale, privo di capacità cognitiva e dunque volitiva, di una condotta istigata e voluta del programmatore o dell'utente; in tali circostanze, l'intelligenza artificiale sarebbe paragonabile a un soggetto incapace e pertanto non potrebbe subire alcuna imputazione di responsabilità; *(ii)* la *natural probable consequence*, ove programmatori e utenti sono considerati penalmente responsabili di un reato commesso dall'intelligenza artificiale come conseguenza naturale e probabile di un loro comportamento doloso o colposo, per esempio di un errore nella programmazione o nell'uso; *(iii)* la *direct liability*, ove si presuppone che l'intelligenza artificiale sia dotata di *mens rea* e dunque compatibile con un'attribuzione di responsabilità per una condotta da essa materialmente eseguita. A tale scenario viene equiparata l'ipotesi di *aberratio delicti* in cui, programmata per commettere un determinato reato, l'intelligenza artificiale, deviando dagli input di programmatori o utenti, ne commetta uno di diverso tipo.

Nel terzo dei paradigmi descritti, Hallevy ipotizza una possibile applicazione di pene costruite secondo un principio di equivalenza tra macchina e umano. Si tratterebbe della cancellazione del *software* volta a neutralizzare il sistema oppure della sua disattivazione per un periodo di tempo prestabilito onde favorirne una rieducazione.

All'esame della dottrina si pongono ulteriori profili che indirettamente riflettono la difficoltà di conciliare le categorie giuridiche esistenti e il carattere innovativo della tecnologia. Tra le altre, la possibilità di realizzare una perfetta sostituzione tra automi o sistemi di intelligenza artificiale e agenti umani nell'espletamento di attività lavorative, con le evidenti ripercussioni che questa opzione potrebbe comportare nell'ambito di svariate professioni (Estlund, 2018). Nel campo della giustizia e dell'assistenza legale, le prime applicazioni in via sperimentale di tecniche algoritmiche sollevano diversi interrogativi, non circoscritti alla effettiva funzionalità di questi sistemi.

Da ultimo, la dottrina ha esplorato approfonditamente, soprattutto nel corso degli anni più recenti, i profili inerenti alla protezione dei dati personali, al cospetto di una importante riforma del quadro legislativo vigente a livello europeo, che ha visto l'entrata in vigore del già ricordato GDPR (Pizzetti, 2018). Tale atto delinea un primo sistema di protezioni rispetto all'individuo i cui dati costituiscono oggetto di un processo decisionale automatizzato (Wachter-Mittelstadt-Floridi, 2017). Non è un caso che i principi di *privacy by design* e *by default* siano stati richiamati come guida per il titolare del trattamento al fine di assicurare il rispetto dei dati personali attraverso la predisposizione di misure tecniche e organizzative adeguate, volte ad attuare in modo efficace i principi di protezione dei dati (*by design*), nonché a garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento (*by default*).

Come osservato, però, queste ricerche dovrebbero emanciparsi

dall'ambito specifico della protezione dati. Lo studio sulle dinamiche di controllo della persona sui processi automatizzati andrebbe collocato entro un contesto di più ampio respiro, che permetta di apprezzarne le implicazioni (per esempio, nel contesto relativo a decisioni amministrative o provvedimenti giurisdizionali automatizzati).

LA GIURISPRUDENZA DI FRONTE ALLA SOCIETÀ ALGORITMICA

La giurisprudenza non ha fatto mancare alcune prime, embrionali indicazioni al confronto con casi concreti che hanno riguardato l'utilizzo di sistemi di intelligenza artificiale.
L'attitudine al confronto con l'innovazione non costituisce di certo una novità per i giudici., Essi, infatti sono gli attori istituzionali più prossimi "per vocazione" alle situazioni in cui si danno nuove esigenze e istanze di tutela per effetto del mutato contesto tecnologico e sociale.
Le prime indicazioni giurisprudenziali si sono avute negli Stati Uniti a fronte della diffusione dei primi dispositivi di natura robotica e del verificarsi di alcuni incidenti produttivi di eventi dannosi e talvolta letali (su cui Bassini-Pollicino-Liguori, 2018).
Pur in carenza di un vero e proprio *leading case* da parte della Corte suprema, i temi al centro della giurisprudenza più risalente in materia concernono in particolare la definizione dello standard di diligenza richiesto ai produttori di componenti robotiche e ai relativi utilizzatori. Alcune delle pronunce si sono focalizzate sull'esistenza di un nesso di causalità che rendessero addebitabili all'utilizzatore ovvero al produttore le conseguenze dannose degli incidenti. In tale prospettiva era analizzato il contributo causale offerto da eventuali comportamenti negligenti da parte delle vittime. Altri precedenti in questa direzione hanno affrontato il tema della separabilità tra le componenti difettose e i macchinari nei quali queste ultime venivano incorporate.
Indicazioni di significativa importanza si sono intraviste anche nella giurisprudenza della Corte di giustizia dell'Unione europea, che si è confrontata con l'applicazione di sistemi di filtraggio automatici sui flussi di comunicazione da parte delle piattaforme digitali e sull'esistenza di una responsabilità di questi operatori per eventuali illeciti commessi dagli utenti terzi a mezzo dei servizi (specialmente nelle cause C-70/10, *Scarlet c. Sabam* e C-360/10, *Sabam c. Netlog*; oltre che in cause riunite C-236, 237 e 238/08, *Google France*).
A differenza delle decisioni incentrate sulla robotica, che hanno spesso fatto ricorso alle categorie della responsabilità da prodotto difettoso in relazione a eventi dannosi, un filone giurisprudenziale più recente cerca di conciliare l'utilizzo di sistemi algoritmici con, da un lato, l'amministrazione della giustizia e, dall'altro, gli eventuali provvedimenti della pubblica amministrazione.
Le pronunce paradigmatiche in questo ambito risalgono, per un verso, al Consiglio di Stato italiano (sez. VI, 8 aprile 2019, n. 2270) e al *Conseil Constitutionnel* francese (sentenza n. 2018-765 DC del 12

giugno 2018) e, per altro, alla Corte Suprema del Wisconsin (*Loomis v. Wysconsin,* 881 N.W.2d 749 (Wis. 2016), *certiorari* negato con sentenza 137 S.Ct. 2290, 2017).
Da un lato, il Consiglio di Stato si è espresso, nello scorso aprile, in una vicenda originata dall'utilizzo da parte del MIUR di un sistema algoritmico per la definizione delle assegnazioni del personale docente della scuola secondaria. Il supremo organo di giustizia amministrativa ha chiarito che la regola tecnica che governa ciascun algoritmo resta pur sempre una regola amministrativa generale, costruita dall'uomo e non dalla macchina, per essere poi (solo) applicata da quest'ultima. Pertanto, la "regola algoritmica" deve rispettare alcuni requisiti: *(i)* ancorché declinata in forma matematica, ha piena valenza giuridica e amministrativa, e come tale deve soggiacere ai principi di pubblicità e trasparenza, di ragionevolezza e di proporzionalità; *(ii)* non può lasciare spazi applicativi discrezionali ma deve prevedere con ragionevolezza una soluzione definita per tutti i casi possibili, cosicché la discrezionalità amministrativa, non demandabile al software, possa rintracciarsi al momento dell'elaborazione dello strumento digitale; *(iii)* presuppone che sia l'amministrazione a compiere un ruolo *ex ante* di mediazione e composizione di interessi, anche per mezzo di costanti test, aggiornamenti e modalità di perfezionamento dell'algoritmo; *(iv)* deve contemplare la possibilità che sia il giudice a svolgere, sul piano "umano", valutazioni e accertamenti fatti direttamente in via automatica, per permettergli di apprezzare la correttezza del processo automatizzato in tutte le sue componenti.
La decisione del Consiglio di Stato ha fatto seguito a una serie di ricorsi sui quali si era precedentemente pronunciato il Tribunale Amministrativo Regionale del Lazio, con sentenze non sempre lineari, frutto della difficoltà di inquadrare univocamente una materia da ricollocarsi entro le categorie esistenti (tra le altre, Tar Lazio – Roma, sez. III bis, 22 marzo 2017, n. 3769; 10 settembre 2018, n. 9227; 12 marzo 2019, n. 3238; 27 maggio 2019 n. 6606). In particolare, ha suscitato dibattito la possibilità per un algoritmo, anche laddove "istruito" da un agente umano e dunque pre-impostato in base a un sistema assiologico, di assicurare il rispetto delle garanzie procedimentali contemplate dalla legge sul procedimento amministrativo.
In Francia, una decisione del *Conseil Constitutionnel* del 2018 si è pronunciata sulla legittimità di una norma che ampliava la possibilità per la pubblica amministrazione di ricorrere (seppure a titolo di eccezione) a decisioni in grado di produrre effetti giuridici sugli individui fondate su un trattamento automatico di dati personali. La stessa disposizione legittimava decisioni automatizzate nel caso in cui *(i)* l'attività algoritmica non riguardasse dati sensibili, *(ii)* fosse possibile una via di ricorso amministrativa e *(iii)* fossero fornite adeguate informazioni in relazione all'utilizzo di algoritmi. Di tale norma veniva dedotto un possibile conflitto con la distribuzione dei poteri esecutivi prevista dall'art. 21 della Costituzione, soprattutto

in relazione alle capacità di autoapprendimento degli algoritmi che avrebbero potuto determinare l'applicazione di regole differenti da quelle pre-impostate. Il *Conseil* ha però escluso l'esistenza di profili di incostituzionalità, ritenendo che fossero state osservate tutte le garanzie necessarie alla salvaguardia dei diritti e delle libertà degli individui. Tra queste, rientravano fattori come: la limitazione dell'utilizzo a specifiche tipologie di decisioni, la previsione delle ricordate condizioni legittimanti e la possibilità per l'individuo destinatario ultimo di una decisione di ottenere una spiegazione in modalità intellegibili e dettagliate del funzionamento del processo algoritmico.

Dall'altro lato dell'Oceano, invece, nel caso *State v. Loomis*, la *Supreme Court* del Wisconsin ha confermato nel 2016 la decisione d'appello in un procedimento penale conclusosi con la condanna dell'imputato in cui era stato tenuto in conto, ai fini della recidiva, il risultato di un *Presentence Investigation Report* prodotto attraverso l'uso di un *software* proprietario (COMPAS), il cui funzionamento risulta coperto da segreto industriale. Tale *software* restituisce *risk assessment* sulla base sia delle informazioni raccolte sulla base di un colloquio con l'imputato sia delle informazioni relative al suo storico criminale. In questa pronuncia, la Corte ha ritenuto che l'utilizzo del *software* non implicasse una violazione del principio del giusto processo (essendo i suoi risultati peraltro rilevanti solo per la valutazione della recidiva e non per la decisione circa la condanna); tuttavia, ha enunciato alcune cautele di cui tenere conto onde garantire che il risultato ultimo sia sempre il frutto di un apprezzamento da parte di un agente umano, che potrà eventualmente rivedere gli esiti del processo algoritmico.

PARS CONSTRUENS: PER UNA GOVERNANCE DELL'INTELLIGENZA ARTIFICIALE

L'utilizzo di sistemi di intelligenza artificiale pone in discussione la funzionalità degli istituti giuridici esistenti all'interno di branche disparate del diritto. In ambiti diversi, infatti, la capacità delle norme vigenti di soddisfare le finalità di interesse pubblico affidate loro da legislatori e regolatori, oltre che dalla pubblica amministrazione, sembra vacillare.

Come si è ricordato nella premessa, Lessig, già descrivendo le criticità connesse alla regolazione della rete Internet, osservava come il web rendesse precaria la capacità delle norme giuridiche di assolvere gli obiettivi "politici" ivi sottesi. Lo studioso evocava la necessità che legislatori e regolatori utilizzassero il diritto per regolare indirettamente l'architettura del cyberspazio, condizionata nella sua fisionomia da vari fattori (le norme sociali, le regole del mercato e il contesto infrastrutturale) di cui l'ideale sintesi veniva definita come il *code* (Lessig, 1999). Infatti, a differenza di altre posizioni che sostenevano l'impossibilità di una regolamentazione della rete, la posizione di Lessig è stata supportata da diversi interventi normativi atti invece a regolarla. Ne consegue che, dal punto di vista della *governance*, la domanda principale non è costituita dall'*an* ma

dal *quantum* di regolamentazione ossia da quel principio di proporzionalità che esige di ridimensionare la portata delle regole al fine di trovare un punto di equilibrio tra interessi confliggenti.
Contestualizzando tali riflessioni nel campo dell'intelligenza artificiale e astraendole a un livello metagiuridico, il punto di equilibrio deve essere trovato tra l'uomo e la macchina al fine di scongiurare che lo sviluppo dell'automazione porti l'uomo in una posizione di subordinazione a una sua stessa creazione., Appare dunque necessario che, a fronte dell'erompere dei sistemi di intelligenza artificiale, i giuristi profondano gli sforzi per restituire un ecosistema antropocentrico, nel quale il fattore tecnologico non sovrasti e non si imponga sui valori, primo fra tutti la dignità, su cui si incentra la tutela dell'individuo. In tal senso, si avverte la necessità che il diritto "guidi" e "orienti" la tecnologia, onde permetterle di percorrere binari rispettosi di un sistema assiologico di valori riflesso nelle costituzioni degli stati moderni e nei trattati internazionali.
Affinché tale missione sia portata a compimento, però, non è sufficiente (né forse, a rigore, è necessaria) un'opera di rivisitazione delle garanzie costituzionali dei diritti fondamentali; occorre, invece, una puntuale verifica circa le norme dell'ordinamento positivo e la loro perdurante abilità a riflettere alcuni principi ispiratori che ne giustificano la collocazione all'interno del sistema delle fonti.
Il diritto deve pertanto regolare la tecnologia, così da condizionarne virtuosamente il funzionamento, declinandolo al perseguimento delle finalità sottese all'azione di legislatori e regolatori. Occorre tuttavia assicurare che l'intervento delle regole giuridiche, cruciale onde scongiurare derive tecnocratiche che preludano all'affermazione del dominio della tecnica, si concretizzi mediante modalità appropriate che non assumano i contorni di un controllo pubblico di carattere "censorio". Al contrario, tali modalità dovranno saper attuare l'esigenza di certezza del diritto secondo regole resistenti alla natura "*disruptive*" dell'intelligenza artificiale. La certezza del diritto costituisce, in ultima analisi, l'obiettivo cui deve tendere l'ordinamento giuridico al cospetto delle novità immesse dalle evoluzioni tecnologiche. La conoscenza e l'"accessibilità" delle implicazioni derivanti dall'utilizzo di tecniche e sistemi algoritmici, infatti, costituiscono fattori dirimenti per scelte imprenditoriali come l'adozione di un determinato modello di *business* o l'implementazione di sistemi innovativi da parte delle pubbliche amministrazioni.
Queste tecnologie consegnano inedite opportunità di sviluppo e di esercizio da parte degli individui dei diritti di cittadinanza (una cittadinanza sempre più "digitale"). Esse recano, tuttavia, rischi e possibili insidie derivanti dalla intrinseca difficoltà di decodificarne la fisionomia e il funzionamento secondo i paradigmi esistenti.
In questo specifico versante si avverte l'interesse dei giuristi e la rilevanza del compito di definire regole appropriate. Con una fondamentale premessa: la declinazione di un *set* di regole funzionale all'efficace ricorso a sistemi di intelligenza artificiale non presup-

pone necessariamente la codificazione di una *lex specialis* che ne governi il funzionamento, nutrendosi invece della valutazione a mo' di "*fitness test*" della perdurante capacità delle norme giuridiche di rispondere alla relativa *ratio* anche nella loro applicazione ai sistemi di intelligenza artificiale.
Inoltre, il ruolo della *rule of law* costituisce un punto di riferimento cruciale per l'attività giurisprudenziale che, in assenza di regole chiare, come si è assistito in diverse occasioni nel caso di Internet, è naturalmente spinta a interpretare e, alle volte, manipolare il diritto attraverso forme di *judicial activism* che ricalcano l'attività del legislatore (Pollicino-Bassini, 2014). Pertanto, al fine di garantire il principio di separazione dei poteri evitando non solo derive tecnocratiche ma anche eccessivi interventi giurisprudenziali giustificati dalla necessità di colmare le lacune normative, la certezza del diritto costituisce il punto di equilibrio a cui l'intero sistema giuridico deve tendere anche, e soprattutto, nella società dell'algoritmo.
Entro questo quadro, si impone di effettuare una rassegna dei vari ambiti nei quali è opportuno rivalutare la tenuta delle categorie esistenti, individuandoli nel diritto costituzionale e amministrativo, nel diritto civile, nel diritto penale e nel diritto processuale. Naturalmente le riflessioni che riguardano ciascuno dei segmenti non potranno che avere un impatto anche sugli altri, stante la necessità che le norme di diritto positivo si conformino a una serie di principi comuni al di là della specifica branca cui appartengono.

INTELLIGENZA ARTIFICIALE E DIRITTO PUBBLICO

La diffusione dei sistemi di intelligenza artificiale pone anzitutto questioni di **ordine costituzionale**. Al costituzionalista, infatti, compete individuare le forme idonee per assicurare che i valori e i diritti tutelati dalla carta fondamentale e della Carta di Nizza possano trovare adeguata protezione anche al cospetto dei sistemi di intelligenza artificiale. Occorre inoltre porsi alla ricerca di norme che agiscano come argine rispetto all'esercizio di un potere che non promana più esclusivamente dalle autorità pubbliche, ma si ritrova sempre più concentrato nelle mani di operatori privati (Teubner, 2004). Il rapporto tra Stato e cittadini si fonda, come noto, sul riconoscimento in capo a questi ultimi di un sistema di garanzie. Queste ultime intendono porre al riparo gli individui da possibili abusi e arbìtri da parte dei temporanei detentori del potere pubblico. Questa dinamica impone oggi un ripensamento alla luce delle caratteristiche diffuse del potere "privato". Si tratta di un potere detenuto da chi utilizza, spesso alimentandosi di ingenti quantità di dati (*big data*), tecnologie algoritmiche e sistemi di intelligenza artificiale per realizzare attività che presentano implicazioni rilevanti per i diritti e le libertà delle persone (De Gregorio, 2019).
Questa necessaria estensione delle garanzie non comporta però l'esclusione dei poteri pubblici, che a loro volta possono essere fruitori e utilizzatori di sistemi di intelligenza artificiale per finalità disparate, senz'altro consentite dall'ordinamento, tra cui l'efficien-

tamento dell'attività amministrativa. Infatti, l'intelligenza artificiale non coinvolge soltanto la sfera dei dritti dell'individuo ma anche altri principi costituzionali quali quelli formanti l'organizzazione della giustizia e l'attività della pubblica amministrazione.
La circostanza che criticità emergano anche nel rapporto tra cittadini e poteri pubblici, e non soltanto nell'ambito dei rapporti orizzontali inter-privati, illustra efficacemente come sia necessario un ripensamento di alcuni aspetti del rapporto di cittadinanza (sempre più digitale). Le norme costituzionali, pensate esclusivamente per attagliarsi al rapporto tra potere pubblico e individuo, paiono vacillare al cospetto di questo nuovo scenario.
La tutela dei diritti appare, invero, soltanto uno dei versanti rispetto ai quali è auspicabile che si registri un'evoluzione in grado di restituire un elevato livello di controllo da parte degli individui. Nella prospettiva del diritto pubblico si palesano altresì esigenze di *governance* della tecnologia, che implicano la necessità per i regolatori di realizzare opportune forme di concertazione a livello nazionale e sovranazionale.
L'esigenza di rafforzare la tutela dei diritti fondamentali si coglie guardando alle embrionali forme di codificazione di nuovi diritti, di cui si ha traccia, per esempio, nel già ricordato GDPR, al Considerando 71 e all'art. 22. Tali referenti paiono dare fondamento a un diritto "alla spiegazione" nell'ambito dei processi decisionali automatizzati, così da costituire al contempo una barriera a tutela degli individui rispetto a decisioni del tutto estranee a un intervento umano che incidano significativamente sui loro diritti. La norma appare veicolare un messaggio evidente: restituire centralità al fattore umano, pur senza che la necessaria precedenza logica e giuridica dell'agente umano possa ergersi a ostacolo all'innovazione tecnologica e al miglioramento dei processi che parti pubbliche e private pongono in essere. Il diritto alla spiegazione cela evidentemente un'esigenza primitiva di trasparenza, che soprattutto (ma non esclusivamente) nel campo dei provvedimenti amministrativi riflette un principio cruciale per conferire al cittadino la possibilità di verificare le modalità di esercizio del potere ed esercitare su di esso un controllo, compreso il sindacato giurisdizionale. Questa esigenza non manca di essere avvertita anche nell'ambito delle applicazioni nel settore privato, dove sono invece carenti specifici meccanismi che consentano analogo controllo da parte di chi subisce gli effetti di una decisione automatizzata (in dottrina, Wachter-Mittelstadt,-Floridi, 2017; Kaminski, 2019). Accanto al diritto alla spiegazione si potrebbe valutare la codificazione di un ulteriore principio, ossia il diritto a conoscere il proprio interlocutore e la sua natura, onde assicurarsi non soltanto della trasparenza del processo decisionale automatizzato ma anche della sua effettiva corrispondenza a una logica umana ovvero algoritmica. Il diritto a non essere sottoposti a un trattamento interamente automatizzato si iscrive nella disciplina specifica sulla tutela dei dati personali, ma ambisce

ovviamente a un riconoscimento più ampio, che lo emancipi dalla dimensione della *data protection*. Sotto questo profilo, è importante osservare il *trend* che ha condotto negli ultimi anni la Corte di giustizia a interpretare il diritto alla privacy e alla protezione dei dati con effetti orizzontali, in particolare nella decisione *Google Spain* del 2014 (causa C-131/12), che a sua volta concerneva l'attività di un motore di ricerca consistente nella indicizzazione automatica di dati personali contenuti in pagine web di siti sorgente. In questa e altre pronunce la Corte di giustizia ha offerto importanti indicazioni a sostegno di un'applicazione dei diritti fondamentali anche nei rapporti orizzontali inter-soggettivi, quale per esempio quello che si instaura tra motore di ricerca e individuo, i cui dati sono oggetto di memorizzazione tra i risultati delle ricerche (Pollicino, 2018). Proprio questo *trend* parrebbe sostenere l'opportunità di codificare o comunque riconoscere nuove situazioni giuridiche che non trovano immediata tutela nelle carte dei diritti in quanto si mostrano sguarnite di protezione e dunque di giustiziabilità soprattutto nei rapporti inter-privati. Infatti, di fronte all'attività dei poteri pubblici, princìpi come la trasparenza impongono una possibilità di accesso e di controllo sull'operato dell'amministrazione. Invece, nei rapporti privati tali tutele non sarebbero direttamente applicabili in assenza di una codificazione normativa. Il GDPR ha tentato di realizzare una tale codificazione sotto il profilo della tutela dei dati personali, ma essa inevitabilmente aspira a un confezionamento entro sedi normative più generali, financo, per alcuni tratti, di rango costituzionale.

Le implicazioni di carattere costituzionale sul fronte dei diritti e dell'eguaglianza non sono circoscritte soltanto a questo ambito, ma conoscono manifestazioni in forme variegate (Casonato, 2019; Simoncini, 2019).

Tensioni si pongono anzitutto rispetto all'eguaglianza dei cittadini, da intendersi in particolare come parità di *chances* e sotto il profilo "sostanziale", non meramente ancorata a un vincolo formalistico. La diffusione di sistemi di intelligenza artificiale è destinata, per esempio, a determinare conseguenze di grande momento in ambito occupazionale: per un verso, sostituendo gli individui nell'esecuzione di mansioni che possono essere svolte analogamente da una macchina, per altro verso sottraendo però opportunità lavorative. L'impatto di questi sistemi riguarderà tanto le professioni intellettuali quanto le attività diverse da queste ultime; è ragionevole pronosticare, tuttavia, che saranno soprattutto quelle artigiane e operaie (normalmente esercitate da personale con livello meno elevato di istruzione), a risentire del rapporto di succedaneità tra uomo e macchina, con ricadute sociali che potrebbero attingere l'eguaglianza sostanziale dei cittadini. Un altro versante esemplificativo riguarda l'ambito della salute, ove è ragionevole pronosticare un più diffuso ricorso a tecniche e dispositivi medicali di nuova generazione, che però potrebbero, per gli elevati costi di sviluppo e di messa

a regime, causare una disparità nell'accesso alle cure somministrate tra abbienti e meno abbienti.

Un secondo profilo meritevole di attenzione si ricollega alla sfera dei diritti politici. L'impiego di tecniche algoritmiche favorisce dinamiche di profilazione degli utenti basate sulla raccolta delle loro preferenze, in grado di rifletterne convinzioni e orientamenti politici, morali, filosofici, religiosi, etc. Come dimostra anche la vicenda *Cambridge Analytica*, la raccolta di queste informazioni a carattere personale può alimentare una *bubble democracy* nella quale gli utenti sono destinatari di messaggi personalizzati, che in quanto selezionati in base all'analisi algoritmica delle loro preferenze, si presume riscuoteranno maggiore successo (Sunstein, 2009). Questa logica di "customizzazione" rischia tuttavia di deformare la reale estensione del quadro delle opinioni e visioni politiche. Si condiziona il rafforzamento delle posizioni di partenza (*confirmation bias*) anziché favorire una loro ponderazione ed eventuale rivisitazione tramite il confronto critico con altre di segno opposto, che esistono nella realtà ma divengono "invisibili" sul *web* e sui *social network* in particolare (Pariser, 2011).

Sul versante della *governance* è avvertita l'esigenza di individuare o costituire un'autorità indipendente che possa assurgere a garante di uno sviluppo dei sistemi di intelligenza artificiale in conformità a linee e orientamenti condivisi a livello sovranazionale, in modo da preservare la centralità della persona umana, della sua dignità e dei diritti fondamentali. L'amministrazione, del resto, non può astenersi dall'intervenire a fronte di fatti economici rilevanti in cui viene sviluppata la capacità delle macchine di riprodurre o attuare operazioni tipiche delle funzioni cognitive umane, compiendo azioni con un certo grado di autonomia. La *public regulation* potrebbe essere affidata ad un apparato estraneo al circuito politico, indipendente rispetto al governo e a singoli ministri, e caratterizzato dalla tecnicità dei componenti. La specializzazione del personale è necessaria dal momento che l'attività richiede conoscenze specifiche che non si ritrovano ordinariamente nell'ambito della funzione pubblica, ed è utile a prevenire conflitti di interessi. In via alternativa, potrebbero essere istituiti uffici di regolazione sul modello statunitense, vigilati o incardinati presso i dipartimenti dell'esecutivo e con competenze settoriali. La collocazione ideale di un'autorità di questo tipo è sovranazionale, perché di rilevanza internazionale sono i temi, le tecniche, i caratteri delle questioni affrontate. Ma la stessa è suscettibile di essere collocata anche a livello domestico, pur sempre entro un *network* europeo sul modello di altre regolazioni economiche. A una simile Autorità potrebbero essere attribuiti poteri sia amministrativi che "quasi legislativi" e "quasi giudiziali", ricorrendo a tecniche di *hard* e *soft law*, e cumulando diverse funzioni che allo stato non appaiono univocamente ascrivibili agli attori pubblici esistenti. Potrebbe svolgere funzioni di carattere regolatorio dell'iniziativa privata, provvedendo anche alla formulazione di indirizzi generali

nel settore. Ovvero funzioni di certificazione pubblica a garanzia dell'affidabilità e della trasparenza per i produttori e per gli utenti. Ovvero ancora funzioni di vigilanza o di accertamento rispetto a possibili rischi di manipolazione e profilazione. A questa Autorità a potrebbe essere rimesso il compito di disciplinare i requisiti di trasparenza, sicurezza, affidabilità, sostenibilità, *accountability* di chi opera nel settore – anche mediante sistemi di accreditamento –, nonché le modalità di tutela dei soggetti vulnerabili. L'Autorità dovrebbe regolarmente riferire in Parlamento sull'attività svolta. La scelta delle politiche pubbliche necessarie per l'attuazione dell'intelligenza artificiale, dovrebbe, invece, rimanere affidata agli organismi politico-rappresentativi.

A questo sistema potrebbe validamente aggiungersi la costituzione di un organo consultivo permanente che replichi il modello *multi-stakeholder* – attuato, per esempio, con l'*Internet Governance Forum* – aprendosi ai vari attori del settore pubblico e privato per promuovere il dibattito sulle implicazioni etiche, giuridiche ed economiche dei sistemi di intelligenza artificiale. Nel complesso, la prospettiva proposta sembra confermare la necessità di un sistema di governo pubblico in grado di incidere dall'alto sulle posizioni economiche degli operatori privati, secondo una filosofia differente da quella che ispira la *Internet Corporation for Assigned Names and Numbers* (ICANN). In qualità di ente privato che opera senza fini di lucro sul sistema dei nomi a dominio, degli indirizzi e dei protocolli internet, esso ha dato vita ad un modello di *governance* contrattuale, basato sulla necessità della stipulazione di accordi privatistici tra l'ente e gli operatori. Eppure, in tale articolato sistema di governo, le istituzioni pubbliche e gli elementi di autoritatività non scompaiono affatto, poiché la rete necessita di un sistema unitario in cui singole parti possono differenziarsi, ma a condizione che ciò avvenga secondo regole tali da assicurare l'uniformazione e l'integrazione funzionale.

Sempre sul versante pubblicistico, come si è evidenziato, l'intelligenza artificiale promette un impatto di grande momento rispetto all'attività della pubblica amministrazione, e dunque nell'ambito precipuo del **diritto e** del **processo amministrativo**. La già ricordata vicenda giurisprudenziale che ha trovato seguito nella pronuncia del Consiglio di Stato ha descritto importanti indicazioni sul rapporto tra regola algoritmica e regola giuridica (Luciani, 2018). L'utilizzo da parte della pubblica amministrazione di sistemi algoritmici impone pertanto di costituire una serie di vincoli che possano conformare la tecnologia ai principi sottesi all'attività amministrativa.

Naturalmente, al fine di assicurare un utilizzo virtuoso e rispondente all'interesse pubblico dell'intelligenza artificiale potrebbe essere utile stabilire alcuni presìdi mediante legge ordinaria, per esempio nella seguente formula: "*Le pubbliche amministrazioni istituiscono una rete di comitati tecnici per il controllo e il monitoraggio costante dei processi di intelligenza artificiale composti da scienziati, economisti e*

giuristi, con il compito di fornire al Governo gli elementi da illustrare al Parlamento sulla verifica annuale dei processi di sviluppo e applicazione dell'intelligenza artificiale. I comitati tecnici valutano anche gli impatti sociali e giuridici dell'uso dell'intelligenza artificiale da parte del settore pubblico al fine di modulare opportunamente gli obiettivi settoriali".
Questi interventi potrebbero permettere di arginare le insidie insite nell'esistenza di un ineliminabile *bias* dovuto alla programmazione degli algoritmi da parte di agenti umani, sensibili, , a una serie di fattori "esterni". Si mira così ad assicurare che in questo ambito i parametri di funzionamento dell'algoritmo costituiscano il risultato di un processo decisionale pubblico di cui sia garantita la massima trasparenza. Come contraltare, ricorre però l'esigenza di un controllo democratico sui pubblici poteri da parte dei cittadini, che presuppone un sufficiente grado di alfabetizzazione corrispondente all'elevato contenuto tecnico degli algoritmi, che trova eco anche normativamente nel principio di totale accessibilità degli atti e dei documenti amministrativi (cosiddetto "FOIA", d.lgs. n. 97/2016). Tali interventi potrebbero trovare attuazione, per esempio, a partire da disposizioni come l'art. 8 del Codice dell'amministrazione digitale. Analoghe misure dovrebbero mirare a consentire alla stessa pubblica amministrazione e ai funzionari che vi operano di mantenere un effettivo potere di dominio sul prodotto di una decisione automatizzata, loro imputabile, di cui agli effetti giuridici essi continuano a rispondere.
Questi interventi paiono rappresentare le condizioni ottimali affinché i processi automatizzati, già diffusi in misura apprezzabile nell'ambito di atti amministrativi vincolati o a bassa discrezionalità amministrativa, possano conoscere un'estensione della loro applicazione anche nell'ambito della formazione delle decisioni amministrative contraddistinte da un più elevato livello di discrezionalità. Sotto questo profilo, particolari criticità andranno affrontate rispetto alla motivazione dell'atto amministrativo, adempimento la cui complessità può variare significativamente a seconda del grado di discrezionalità richiesto. Sul versante dell'intensità del controllo giurisdizionale, le proprietà dell'algoritmo non andranno considerate come un ambito riservato alla pubblica amministrazione, non attingibile dal sindacato giurisdizionale, se non in termini di ragionevolezza e proporzionalità, bensì dovranno essere oggetto di una piena e diretta verifica istruttoria.

INTELLIGENZA ARTIFICIALE E DIRITTO CIVILE

L'intelligenza artificiale e le nuove tecnologie avranno un significativo impatto anche sulle regole del **diritto privato**. In primo luogo, nell'area concernente "le persone", studiosi quali Teubner, hanno discusso in merito alla esigenza di creare una nuova figura giuridica dotata della capacità di agire e della capacità di essere titolare di situazioni giuridiche soggettive: il robot. Il nuovo soggetto dovrebbe affiancarsi alle persone naturali e agli enti nel quadro dei rapporti privatistici (Teubner, 2018). I vantaggi risiederebbero nella possibi-

lità di limitare la responsabilità dei programmatori e degli utilizzatori dei servizi del nuovo potenziale soggetto al fine di promuovere lo sviluppo tecnologico. Nelle trattazioni più approfondite si osserva che il futuro soggetto, dotato di intelligenza artificiale, potrebbe essere sottoposto a regole che allo stato vigono per gli enti. In questo modo, ad esempio, sarebbe possibile assicurare l'esistenza di una garanzia patrimoniale utile a far fronte a eventuali pretese risarcitorie di terzi.

In proposito, è stato opportunamente messo in luce che, diversamente rispetto ai soggetti che attualmente intrattengono rapporti nell'ambito del diritto privato, il robot non persegue un proprio interesse. L'interesse è pur sempre riconducibile a un diverso soggetto, inteso in senso tradizionale, che si avvale degli strumenti messi a disposizione dell'intelligenza artificiale. Questo fondamentale rilievo è sufficiente per affermare che non sussiste la necessità di creare nuovi soggetti giuridici. La tecnologia e l'intelligenza artificiale sono (e devono permanere) al servizio dei soggetti giuridici intesi in senso tradizionale. Peraltro, la netta presa di posizione non deve indurre il giurista a distogliere l'attenzione dai procedimenti decisionali, in parte imprevedibili, che caratterizzano le nuove tecnologie. Il profilo è di primaria importanza per comprendere come, nei diversi settori, il diritto privato debba disciplinare atti posti in essere con l'ausilio della tecnologia, che in alcuni settori si sostituirà alle condotte umane.

Alcune innovazioni sono già attuali e coinvolgono la materia cardine dei rapporti privatistici, ossia il contratto. Il pensiero è subito rivolto allo *smart contract* che, nel contesto del c.d. "decreto Semplificazioni" all'art. 8-*ter* (aggiunto dalla l. 12/2019 in sede di conversione del d.l. 135/2018), è stato di recente definito come "un programma per elaboratore che opera su tecnologie basate su registri distribuiti e la cui esecuzione vincola automaticamente due o più parti sulla base di effetti predefiniti dalle stesse". In questo ambito, di indubbia rilevanza da un punto di vista economico, l'ambizione dei programmatori è quella di escludere completamente l'operatività del diritto. Secondo la visione più estrema, il codice dovrebbe prendere il posto del diritto, poiché è proprio il codice computerizzato – in queste ricostruzioni – a fungere da "diritto". In modo provocatorio Savelyev ha osservato inoltre che gli *smart contracts* rappresentano l'inizio della fine del diritto contrattuale (Savelyev, 2016).

Di fronte ad affermazioni di questo genere, i giuristi devono rispondere con cautela, venire a conoscenza degli aspetti tecnici e affrontarne i profili problematici. Si tratta di una tematica spesso intrisa di luoghi comuni, da superare facendo uso delle tradizionali categorie civilistiche. In questa prospettiva, è necessario mettere in luce che, nonostante il seducente appellativo, coniato negli anni novanta da Szabo, gli *smart contracts* non sono affatto intelligenti (Szabo, 1997). Essi eseguono semplicemente una prestazione al ricorrere di un determinato evento (*trigger event*), che opera come una sorta di

condizione: una volta verificatosi l'evento, l'esecuzione del contratto diviene inesorabile, alla stregua delle operazioni di un distributore automatico che, in seguito all'immissione della moneta, dispensa il prodotto e l'eventuale resto. Lo *smart contract* in quanto tale non è dunque dotato di intelligenza artificiale e il procedimento automatizzato, allo stato, riguarda soltanto l'esecuzione e non la conclusione del contratto. Ne deriva che, a rigore, lo *smart contract* non potrebbe neppure considerarsi un contratto (Durovic-Janssen, 2018). Le questioni più significative si pongono per gli *smart contracts* connessi alla tecnologia *blockchain*. Le transazioni avvengono su piattaforme informatiche che prendono il posto degli intermediari e assicurano che la corretta esecuzione del contratto, concluso attraverso l'incontro tra la proposta e l'accettazione, non possa più essere messa in discussione. Pertanto, in linea teorica, gli *smart contracts* non conoscono inadempimenti. Inoltre, sono in grado di generare vantaggi in termini economici, in virtù della eliminazione dei costi di transazione, principalmente connessi alle negoziazioni e alle attività di intermediazione. In questo quadro, il grande cambiamento rispetto al passato riguarda l'affidamento nelle capacità di adempiere dell'altro contraente. Chi compie una transazione su una piattaforma *blockchain* generalmente non conosce l'altro contraente, il suo affidamento concerne soltanto il funzionamento della piattaforma. In questo senso, si afferma che gli *smart contracts* sono *self-executing* e *self-enforcing*. In realtà, il procedimento non sempre è pienamente automatizzato poiché spesso le informazioni rilevanti per l'esecuzione del contratto vengono fornite da un terzo che, nel gergo tecnico, viene denominato *oracle*. L'affidamento dei contraenti deve pertanto estendersi anche a quest'ultima figura (Durovic-Janssen, 2018).
Nonostante le incomprensioni e le incongruenze, gli *smart contracts* avranno un impatto significativo nel mercato. Tuttavia, non corrisponde al vero che il codice si sostituirà al diritto; il diritto continuerà certamente ad essere applicabile. Eventuali falle nei sistemi operativi potrebbero agevolare fraudolente manomissioni da parte di soggetti terzi, che richiederebbero l'applicazione delle norme sulla responsabilità. Inoltre, non può escludersi che uno *smart contract* violi norme imperative o che venga concluso da un soggetto incapace di agire. Molto rilevante, anche in virtù dei risvolti per il mercato concorrenziale, si profila l'applicazione del diritto dei consumatori, che contiene norme idonee a tutelare la parte debole del rapporto contrattuale, limitando l'autonomia dei contraenti. Le riflessioni non dovrebbero riguardare soltanto l'astratta questione dell'applicabilità delle norme imperative, bensì essere volte ad individuare in che modo tali norme possano adattarsi a protocolli informatici. In questo senso, sarà necessario valutare l'applicabilità della direttiva 93/13/CEE, concernente le clausole abusive nei contratti stipulati con i consumatori. Non è escluso che la normativa europea possa applicarsi a codici informatici e che la tecnologia possa rendere più efficace il controllo sostanziale dei contratti nei rapporti B2C.

Le procedure di esecuzione automatizzata possono comportare un aumento dell'effettività del diritto. Oltre a quanto indicato con riguardo alle clausole abusive, un esempio sovente menzionato in dottrina riguarda la compensazione pecuniaria alla quale è tenuto il vettore aereo, in caso di ritardo o cancellazione del volo, ai sensi dell'art. 5, par. 3, del Regolamento CE n. 261/2004. Rendendo la tecnologia degli smart contracts obbligatoria per i vettori aerei sarebbe possibile assicurare ai consumatori gli indennizzi in via automatica. Attualmente, in media solo il 5 per cento dei passeggeri esercita i propri diritti in caso di ritardo o cancellazione del volo. Gli *smart contracts* potrebbero avere un impatto significativo sui rapporti tra vettori aerei e passeggeri e incidere sulle modalità di erogazione dei servizi. Da questo punto di vista, si teme che l'aumento di effettività del diritto possa rendere quest'ultimo insostenibile da un punto di vista economico. Se i vettori aerei fossero sistematicamente obbligati al pagamento dell'indennizzo in caso di ritardo o cancellazione del volo, si assisterebbe con tutta probabilità a un incremento significativo del prezzo dei biglietti aerei. In questo senso, in maniera paradossale, Basedow ha affermato che il diritto non sempre si presta ad essere pienamente effettivo.

Al di là di questa problematica, occorre rilevare che le nuove tecnologie potrebbero apportare un significativo miglioramento delle norme giuridiche, rendendole più adeguate alla soluzione del caso concreto. In base alla normativa europea relativa alla responsabilità del vettore aereo pochi minuti di differenza nella durata del ritardo possono assumere un'importanza decisiva ai fini del riconoscimento dell'indennizzo. Con la tecnologia degli *smart contracts* sarebbe possibile quantificare un indennizzo in proporzione al ritardo accumulato. Il problema specifico si collega a quello di carattere più generale concernente la personalizzazione delle norme giuridiche. Con l'avvento dei *big data*, i software potrebbero modellare le norme giuridiche sulla base delle peculiarità di ogni singolo consociato (c.d. "*granular norms*"). Secondo Casey e Niblett, ciò potrebbe determinare il superamento dell'utilizzazione di standard e clausole generali nel diritto privato (Casey-Niblett, 2017).

I processi decisionali algoritmici assumono rilievo nel contesto dei mercati attraverso la pratica del c.d. '*High-frequency trading*' (HFT). Si tratta di una modalità di intervento sui mercati compiuta con sofisticati strumenti software, e talvolta anche hardware, che permettono di porre in essere negoziazioni ad alta frequenza, guidate da algoritmi matematici, che agiscono su mercati di azioni, opzioni, obbligazioni, strumenti derivati, *commodities* (Balp-Strampelli, 2018). La durata di queste transazioni può essere brevissima, con posizioni di investimento fatte proprie per periodi di tempo variabili, da poche ore fino a frazioni di secondo. Lo scopo di questo approccio è quello di lucrare su margini estremamente esigui, anche pochi centesimi. Per ottenere significativi ricavi mediante margini minimi, la strategia HFT deve necessariamente operare su grandi

quantità di transazioni giornaliere. L'HFT può causare distorsioni nel mercato, in quanto gli scambi ad alta velocità conferiscono un vantaggio a chi li mette in atto rispetto a chi si serve di strumenti tradizionali. Inoltre, la continua azione degli operatori HFT tra mercati diversi sullo stesso titolo conferisce ai mercati un'estrema volatilità, senza tuttavia che i movimenti speculativi si consolidino su una posizione di investimento sul capitale azionario di specifiche aziende. Ciò dimostra che le operazioni prescindono da valutazioni sui fondamenti economici esibiti dalle aziende. La potenziale pericolosità dell'HFT per la stabilità dei mercati è confermata dal caso Goldman Sachs, uno dei più importanti operatori a far ampio uso della suddetta tecnologia. Un suo dipendente, Sergey Aleynikov, si è impossessato dei codici sorgente con cui la compagnia accedeva a simili operazioni di trading: la vicenda di Alenikov si concluse con l'arresto da parte dell'FBI e l'applicazione di una pena detentiva.
La pericolosità dell'HFT è ulteriormente testimoniata da interventi normativi volti a porre un freno alle relative attività speculative e a garantire una certa trasparenza con riferimento alle modalità operative. Sotto questi profili, deve essere valutata con attenzione la revisione della direttiva europea, c.d. MiFID II (*Markets in Financial Instruments Directive*) e ulteriori regolamenti attuativi, che obbligano gli high frequency traders a registrarsi come imprese di investimento e a rendere pubblici i loro algoritmi, fornendo garanzie sull'attendibilità dei loro software. In ambito nazionale, la CONSOB segue con particolare concentrazione gli sviluppi tecnologici e ha pubblicato *discussion papers* e altri contributi nel tentativo di pianificare interventi regolatori idonei a fronteggiare le principali problematiche a livello interno e sovranazionale.
Da un altro angolo di visuale, *i* problemi relativi alla responsabilità civile investono soprattutto l'applicabilità delle norme relative all'illecito in ipotesi in cui una scelta, che si rivela dannosa, sia stata compiuta in base alle elaborazioni di un algoritmo. Si pongono problemi soprattutto in ordine alla prova della sussistenza degli elementi della responsabilità civile. Più in generale, sono in discussione le funzioni della responsabilità civile, che dovrebbe continuare a garantire una certa carica deterrente nei confronti dei possibili *tortfeasors* e, al contempo, rispondere all'esigenza di indennizzare adeguatamente i soggetti danneggiati. Le nuove tecnologie rendono necessaria l'elaborazione di nuovi parametri per attribuire la responsabilità e, posta la diversa gestione del rischio, impongono ai giuristi di ripensare alcune norme concernenti la responsabilità oggettiva.
Per quanto attiene ai parametri di riferimento per attribuire la responsabilità, si tratta di valutare la condotta di sistemi in grado di apprendere autonomamente e di decidere sulla base dei dati raccolti. Un primo parametro utilizzabile è quello relativo alle capacità umane. In presenza di un danno causato da un sistema operativo autonomo, ci si dovrebbe chiedere se un uomo nella stessa posizio-

ne sarebbe stato in grado di evitare il danno. La soluzione è criticata da parte della dottrina, poiché il *software* dovrebbe avere capacità superiori all'uomo e, in molti casi, le due condotte non sarebbero paragonabili, posto che le nuove tecnologie possono generare rischi diversi rispetto a quelli connessi a una condotta umana. In ogni caso (come osservato da Teubner, 2018), le capacità dell'uomo potrebbero costituire – almeno nel primo periodo – una soglia minima di "diligenza" richiesta al sistema operativo. In prospettiva, Chagal-Feferkorn propone di sviluppare un parametro autonomo per i sistemi intelligenti, denominato "algoritmo ragionevole", diverso a seconda del settore preso in esame (Chagal-Feferkorn, 2018). Altri autori, come Wagner, concentrano invece l'attenzione sui dati statistici concernenti le attività esercitate da software dotati di intelligenza artificiale (Wagner, 2017).

Non mancano voci critiche in merito alla possibilità di sviluppare parametri adeguati a valutare la condotta dei sistemi operativi autonomi. Il diritto potrebbe assicurare la tutela dei danneggiati mediante normative speciali che prevedono ipotesi di responsabilità oggettiva. In questo contesto, emerge il problema della responsabilità del produttore con riguardo a sistemi operativi che apprendono autonomamente. Rispetto alle nuove tecnologie, le norme di conio europeo, contenute nella direttiva 85/374/CEE sulla responsabilità del produttore, appaiono del tutto inadeguate poiché si riferiscono a tecnologie obsolete e non chiariscono con precisione in quali casi un sistema operativo in grado di apprendere autonomamente è da considerare difettoso. Allo stato, un gruppo di lavoro istituito dalla Commissione si sta occupando dello sviluppo di *guidelines*, che dovrebbero permettere di orientare l'interprete chiamato ad applicare le norme della direttiva.

Un ambito peculiare in cui si presentano le descritte problematiche è quello della responsabilità da circolazione di veicoli a guida autonoma, in cui è sotto osservazione l'apparato di regole relative alla responsabilità del proprietario del veicolo, alla responsabilità del produttore e alla assicurazione obbligatoria (v. Comunicazione della Commissione al Parlamento Europeo, al Consiglio Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, "Verso la mobilità automatizzata: una strategia dell'UE per la mobilità del futuro", del 17 maggio 2018). Da più parti si discute di questioni che attengono a diversi soggetti interessati: i produttori, gli utenti, le compagnie assicurative. Pur afferendo all'area del diritto privato, il settore dei veicoli a guida autonoma incide certamente su interessi generali poiché, allo stato, riguarda le norme sulla responsabilità civile che trovano più sovente applicazione e concerne interessi economici di notevole rilevanza.

In un primo periodo, come è suggerito dalla suddetta Comunicazione, le attuali norme relative alla responsabilità da circolazione di veicoli, assistite dalla previsione dell'assicurazione obbligatoria, potrebbero continuare a fornire una regolamentazione soddisfacente.

In prospettiva futura, occorrerà riflettere su nuove forme di responsabilità che coinvolgano maggiormente i produttori delle vetture. Questi ultimi sono infatti i soggetti che immettono i sistemi operativi nel mercato e gestiscono i rischi connessi al funzionamento del software. Il tema è affrontato a livello globale e, tenuto conto delle principali tesi, si tratterà di decidere se modificare o interpretare diversamente le norme in materia di responsabilità del produttore (come suggerito da Geistfeld, 2017), oppure se creare un nuovo sistema di responsabilità *no-fault*, in cui i danneggiati dovrebbero essere risarciti da un apposito fondo istituito con contributi erogati dai produttori di veicoli a guida autonoma (come proposto da Abraham-Rabin, 2019).

INTELLIGENZA ARTIFICIALE E DIRITTO PENALE

L'impatto dell'intelligenza artificiale. e delle questioni ad essa connesse, appare rilevante anche con riferimento al diritto penale.
Sul piano della **responsabilità per condotte penalmente rilevanti** realizzate da soggetti non umani, sembra opportuno avviare al più presto una discussione obiettiva, priva di preconcetti, nella consapevolezza del rischio rappresentato da una deriva verso forme di responsabilità penale oggettiva in capo a coloro che hanno progettato e attivato il soggetto agente non umano.
Si tratta di una discussione che sembra spingersi al di là della realtà contingente. Tuttavia occorre liberarsi di ogni condizionamento distopico per accettare che numerosissime sono già oggi le fonti decisionali automatizzate le cui conseguenze possono violare beni giuridici tutelati dalla legge penale. Ferma la granitica previsione dell'art. 27, comma 1, Cost. – che ha di fatto impedito, ad oggi, il riconoscimento di responsabilità penale in capo alle persone giuridiche (con il conseguente ricorso al concetto di responsabilità amministrativa dell'ente per il reato commesso nel suo interesse) – il problema dell'imputazione di responsabilità per scelte riferibili a decisioni automatizzate non fa parte, appunto, di un futuro distopico, ma della realtà. Indubbiamente la questione pone in discussione l'intero concetto di pena e coinvolge nella discussione prospettive filosofico-culturali che trascendono la dogmatica penalistica, innervandosi ancora una volta nel dibattito sul riconoscimento di una soggettività giuridica. Il rischio da evitare, *in primis*, è quello di rivolgersi a paradigmi di tipo oggettivo che, formalmente estranei al diritto penale, talvolta si affermano attraverso l'intervento giurisprudenziale, per mezzo dell'amplificazione di posizioni di garanzia pur previste dalla legge.

INTELLIGENZA ARTIFICIALE E PROCESSO

Non va sottaciuto il rilevante contributo che l'intelligenza artificiale può fornire nel campo dell'organizzazione del servizio giustizia, del funzionamento dei sistemi giudiziari nazionali nonché del diritto processuale. Ivi l'applicazione delle tecnologie algoritmiche può indubbiamente rappresentare un fattore di miglioramento dell'efficienza e della qualità dei processi, sebbene l'implementazione della

tecnologia debba avvenire in modo responsabile e nel rispetto dei diritti fondamentali garantiti.
L'intelligenza artificiale applicata al settore dell'azione pubblica nel contesto della produzione, diffusione, valutazione e miglioramento dei servizi pubblici e dei beni collettivi, è intesa come un insieme di dispositivi di carattere tecnologico, tecnico e matematico la cui regolazione va distinta per le fasi di sviluppo, utilizzo e monitoraggio e per i livelli di conoscenza fattuale, di architettura e di *agency* che esso implica. Nel settore dell'azione giudiziaria, in particolare, le promesse della giustizia predittiva si sono riverberate nella espansione delle cosiddette *legal tech*, nell'aumento della attenzione degli attori internazionali per la elaborazione di standard di *accountability* e *responsiveness* (di sviluppo ed uso), nella sperimentazioni in taluni ordinamenti di strumenti di aiuto alla decisione del giudice, di elaborazione in via preliminare del potenziale di costo/beneficio di un contenzioso e della sua soluzione giurisdizionale.
Un utilizzo avanzato di questo insieme di strumenti ad alta intensità tecnologica, articolati sulla applicazione della scienza matematica e della scienza informatica, è in essere solo in alcuni Paesi europei. Eppure già nella distribuzione automatizzata dei casi, nella analisi di banche dati giurisprudenziali (anche se non realizzata attraverso la traduzione del linguaggio naturale in una rappresentazione digitale) e nella gestione ponderata dei carichi di lavoro e dei flussi, lo strumento dell'automazione ha fatto la sua entrata nell'organizzazione del servizio giustizia ormai da tempo.
Sul piano della qualità della giustizia le promesse e le aspettative legate all'applicazione di strumenti di intelligenza artificiale vanno nella direzione di: *(i)* riduzione del contenzioso attraverso rimedi alternativi automatizzati; *(ii)* riduzione dei margini di errore nella valutazione preventiva del rischio di soccombenza (tale profilo riguarda in particolare le professioni ordinistiche); *(iii)* riduzione dei tempi attraverso la trattazione in via automatizzata delle controversie seriali e standardizzabili; *(iv)* riduzione dei margini di differenziazione distrettuale e circondariale per tipologie di risposta a simili tipologie di contenzioso.
Muovendo dai risultati conseguiti da società già operanti nel settore è possibile affermare che, attualmente, esistono *software* in grado di anticipare correttamente l'esito di un giudizio. Il dato assume rilevanza nell'ambito delle professioni forensi e della gestione del contenzioso, posto che i *software* potrebbero essere in grado – quantomeno – di influire sulla scelta del privato di agire in giudizio. Una prospettiva maggiormente di frontiera auspica l'utilizzazione di software predittivi per la soluzione di questioni di natura bagatellare.
La massima aspirazione di alcuni fautori dell'intelligenza artificiale. è quella di eliminare del tutto il ruolo del diritto nell'espletamento di significative attività dell'uomo. Ciò potrebbe implicare l'adozione di *meccanismi alternativi di risoluzione delle controversie* completamente automatizzati, eventualmente collegati a piattaforme

blockchain (Ortolani, 2019). L'obiettivo, da molti considerato utopistico, è quello di creare un "giudice *robot*" (il c.d. "*robo-judge*"), in grado di decidere controversie sulla base dell'elaborazione statistica di dati. Sebbene siano state messe in evidenza le potenzialità della tecnologia negli accertamenti del fatto compiuti in ambito civile, la ricerca scientifica si presenta ancora limitata con riferimento al possibile ruolo dell'intelligenza artificiale.

Con la transizione da applicazione algoritmiche *deterministiche* ad applicazioni *probabilistiche*, la decisione non deriva semplicemente dalla pedissequa applicazione del ragionamento deduttivo, ma è basata sul contemporaneo apprezzamento di una pluralità di interessi ed istanze che vengono in rilievo nel caso concreto. Si realizza in tal senso un passaggio sintetizzabile nella formula dalla *delega di processo* alla *delega di decisione*, che implica necessariamente l'accettazione dell'esistenza del margine di errore. La tecnologia potrebbe assicurare maggiore velocità e un significativo risparmio di risorse. Come in altri campi, le innovative soluzioni che la tecnologia è in grado di offrire per la soluzione delle controversie civili dovranno necessariamente conseguire un elevato livello di accettazione sociale prima di essere implementate. Merita attenzione, altresì, l'impatto che i sistemi di *open access*, associati a strumenti algoritmici predittivi, possono produrre sul piano del valore del precedente e dell'indipendenza degli organi giudicanti.

La regolazione etica e giuridica di tali profili deve tenere conto degli snodi organizzativi entro cui domanda ed offerta di giustizia si incontrano. Quattro appaiono le dimensioni funzionali che necessitano di essere poste alla attenzione della norma etica e giuridica in questo contesto: conoscenza, rito del processo, *status* dell'organo terzo dirimente le controversie, dati personali. I principi che vanno richiamati in tale senso sono: *(i)* responsabilità professionale e trasparenza della produzione della conoscenza; *(ii)* applicazione delle garanzie processuali di difesa; *(iii)* autonomia della giurisdizione e indipendenza del giudice; *(iv) privacy* e sicurezza dei dati. Essi possono essere declinati in norme etiche e in norme giuridiche.

Sul piano etico un codice deontologico che assicuri la responsabilità professionale degli sviluppatori e di coloro che intervengono nel processo di integrazione e applicazione dei dispositivi di automazione appare necessario e in linea con le forme di regolazione delle *expertise* peritali e di consulenza di cui ci si avvale in via endoprocedimentale. Ancora la estensione delle garanzie processuali massime – in tal senso il caso ordinamentale italiano appare in una prospettiva comparata di sicura ispirazione – per i riti nei quali le parti possono avvalersi (con riconoscimento nel processo) di strumenti di giustizia predittiva deve essere assicurata con il più alto livello di tutela costituzionale.

L'autonomia del giudice a fronte della disponibilità di conoscenza induttiva derivata dalla analisi di banche dati giurisprudenziali va assicurata sia con norme che obblighino le giurisdizioni supreme

a un *set* di standard comuni per il trattamento di tale conoscenza da parte delle giurisdizioni di primo e secondo grado, sia con norme che integrino nella argomentazione del giudice la esplicitazione della fonte da cui tale conoscenza viene desunta. Sul piano giuridico ordinamentale dovrebbero essere previste delle unità specializzate a livello di giurisdizioni di secondo grado – dove il giudizio di fatto trova la sua sintesi – per la valutazione dei dispositivi di analisi delle banche dati. Infine, la qualità dei dati e la loro tutela va prevista sul piano giuridico. *Sub condicione* di anonimato del giudice, le banche dati devono rispondere a sistemi di regolazione pubblica responsabili della sicurezza. In tal senso, andrebbe previsto in via legislativa la disposizione nei ministeri di uffici preposti allo sviluppo reti e politiche di mantenimento della sicurezza. Le garanzie di *privacy* dovranno essere conformi alle norme sovranazionali in materia (GDPR).

Anche la sfera del processo – civile, penale, amministrativo, contabile – è chiamata a confrontarsi con le implicazioni del *digital turn*, non soltanto per via del senso, assai diffuso, di ineluttabile transizione verso un quadro in cui tutte le attività umane debbono necessariamente offrirsi alla rivoluzione basata sull'inedito connubio tra incommensurabile potenza computazionale e *big data*. É, infatti, incontestabile la condizione di grave inefficienza e inefficacia della giustizia, che pone tutti gli operatori nella posizione di dover studiare come le potenzialità dell'era digitale e, in particolare, l'intelligenza artificiale, possano eventualmente migliorare l'attuale situazione.

In particolare, con riferimento al problema della prova, con considerazioni mirate soprattutto sul processo penale ma estensibili, con opportune distinzioni, al processo in generale, può essere osservato quanto segue.

L'impiego di strumenti algoritmici nella ricerca e nella formazione della prova implica una inevitabile opacità del processo di creazione del dato probatorio.

Occorre, pertanto, per bilanciare l'asimmetria conoscitiva tra le parti del processo, valorizzare in pieno il principio di parità delle armi.

Questa asimmetria conoscitiva, fortemente radicata nella tradizione, post-inquisitoria, italiana e di molti altri Stati europei continentali, è inevitabilmente legata alla presenza di una parte pubblica nell'indagine e nel processo penale. Essa rischia di essere oltremodo amplificata dal ricorso a mezzi di ricerca della prova e mezzi di prova la cui opacità, accentuata dall'impiego di soluzioni di *machine learning* o *deep learning*, può sottrare completamente la prova stessa al confronto dialettico delle parti sulla sua attendibilità. Ciò trasformerebbe la "prova algoritmica" in un mezzo apoditticamente attendibile e, quindi, il giudice, in mero annotatore di un processo di valutazione della prova interamente assorbito dalla natura algoritmica della medesima.

Allo scopo di attuare più efficacemente le suddette considerazioni, occorre inoltre identificare e distinguere, all'interno della "prova algoritmica", i profili di rischio di inattendibilità legati al piano della teoria scientifica sintetizzata nell'algoritmo e quelli legati, invece, alla costruzione e al funzionamento dell'algoritmo stesso. Risulta evidente dalla esperienza di alcune giurisdizioni locali statunitensi che l'impiego processuale di strumenti di natura predittiva nasconde plurimi livelli di opacità e di rischio, dipendenti non soltanto dalla natura digitale e automatizzata dello strumento probatorio. Infatti, il *software* che confeziona il dato utilizzato con valore probatorio all'interno del processo, innanzitutto, si basa su una teoria scientifico-probabilistica che potrebbe non rispondere ai parametri "Daubert" i quali, trascendendo la sfera della Corte Suprema statunitense, sono divenuti, nel tempo, una sorta di decalogo per l'ammissibilità della prova scientifica nella maggior parte degli ordinamenti occidentali. Il primo profilo di contraddittorio processuale deve riguardare, allora, la validità e la validazione della teoria scientifica sottesa all'algoritmo da parte della comunità scientifica di riferimento. Il profilo dell'opacità dell'algoritmo in cui essa viene sintetizzata rappresenta un secondo, ineluttabile, tema di confronto tra le parti, in contraddittorio, davanti al giudice imparziale.
Un altro passaggio consiste nel rivalutare la tradizionale distinzione tra fase cognitiva ed esecutiva, in relazione all'impiego di valutazioni predittive del comportamento violento e del rischio di recidivanza. In diretta connessione con quanto sottolineato nel punto precedente, si osserva un crescente ricorso a strumenti algoritmici di predizione del rischio di recidivanza e/o di comportamenti violenti (come anche nel caso *State v. Loomis* già ricordato), basati su teorie bayesiane che sfruttano correlazioni ottenute dall'analisi automatizzata di dati di varia provenienza. Come accennato, l'area geografica di riferimento, al momento, è per lo più quella nordamericana e australiana, ma con progressive incursioni anche in Europa (in particolare, Inghilterra e Galles). Poiché tali strumenti predittivi si basano su teorie psico-criminologiche più o meno accettate dalla comunità scientifica, il vaglio di cui si accennava anche *supra* rimane essenziale.
Altrettanto essenziale è la rivalutazione del ruolo che la prova dell'attitudine caratteriale e psicologica può avere nel processo penale. Si tratta di un elemento certamente rilevante nella quantificazione della pena e nella modulazione del trattamento sanzionatorio in concreto, ma non certo nell'accertamento dei fatti oggetto del singolo procedimento penale. In questa prospettiva, la tradizionale distinzione che vede l'Italia e altri Paesi continentali opporsi all'idea che la valutazione caratteriale possa incidere sull'accertamento della responsabilità per i fatti accaduti torna ad assumere grande significato. La prova (predittiva, algoritmica, automatizzata) caratteriale può acquisire elementi conoscitivi su come l'imputato potrebbe essere portato a comportarsi in futuro, ma nulla dice ri-

spetto alla responsabilità di questi rispetto ai fatti di cui è accusato. Sul diverso piano dell'indipendenza del giudice merita di essere preso in considerazione attentamente l'impatto che i sistemi di *open access*, associati a strumenti algoritmici predittivi, possono produrre sul piano del valore del precedente e dell'indipendenza degli organi giudicanti. L'accesso completo e automatizzato alle decisioni giudiziarie rappresenta un tassello del più ampio movimento per l'*openness* delle pubbliche amministrazioni che in alcuni ordinamenti, come ad esempio, la Francia, è già stato assicurato da tempo. Certamente tale strumento agevola molto i professionisti legali, favorendo una migliore conoscenza e circolazione anche delle decisioni di merito. Pertanto, tale esteso bacino di sentenze presenta fortissima appetibilità per chi sia interessato a costruire, a beneficio – principale ma non esclusivo – delle *law firms*, strumenti predittivi della decisione giurisdizionale. Alimentando una rete neurale con l'insieme di tutti i precedenti, pur non vincolanti, si possono stabilire, correlazioni utili per immaginare come il giudice si pronuncerà rispetto a una determinata questione e all'interno del collegio che concorre a comporre.

Ciò può rappresentare, come accennato, un grande vantaggio per la strategia difensiva e per l'allocazione delle risorse all'interno dello studio legale. Esso ha diverse ricadute assai perniciose: 1) le correlazioni stabilite possono non essere corrette; 2) l'evoluzione dell'interpretazione giurisprudenziale sarebbe fortemente compromessa. Il ricorso a simili strumenti scoraggerebbe, infatti, il patrocinio di posizioni potenzialmente 'non vincenti'. Rischierebbe così di essere depotenziato l'essenziale fattore di evoluzione che deriva dal tradizionale impegno dell'avvocatura nel promuovere nuovi paradigmi interpretativi che, scalando la struttura delle corti, arrivano ad affermarsi come giurisprudenza dominante, di legittimità, e talvolta, a trasformarsi in modifiche normative; 3) anche nei sistemi continentali che non sono basati sul valore del precedente, il diffondersi di strumenti di *quantitative legal prediction* può provocare un effetto di centralizzazione del valore del precedente. Ciò può indurre il singolo giudice persona fisica a temere conseguenze di vario genere (disciplinare, civile) per essersi discostato, con la sua decisione, dall'esito preconizzato dallo strumento predittivo. L'effetto potrebbe inizialmente indurre un apparentemente neutro onere suppletivo di motivazione, per distaccarsi dal precedente, come accade negli ordinamenti di *common law*, che potrebbe tuttavia progressivamente incidere sul senso di indipendenza del giudice. Questi, infatti, potrebbe sentirsi, più o meno consciamente, indotto ad aderire sistematicamente al risultato della predizione.

Alla luce delle suddette osservazioni, bisogna distinguere attentamente il concetto di conoscibilità dei precetti e di prevedibilità della sanzione, essenziali a soddisfare il moderno concetto di legalità, da quello di "prevedibilità" della decisione del giudice.

Infatti, se è vero che la più moderna concezione del principio di

legalità passa per il riconoscimento dell'essenzialità della garanzia di conoscibilità del precetto normativo da seguire e di prevedibilità della sanzione che sarà applicata in caso di accertata trasgressione, si deve riconoscere che lo scenario evocato dalla *quantitative legal prediction* non è ispirato all'attuazione della predetta garanzia o che, quantomeno, può sovrapporvi effetti contrari e perniciosi.
Sempre con riferimento alla giurisdizione, centrale appare il ruolo del processo amministrativo, destinato a divenire, ad un tempo, il luogo del sindacato sul potere dell'amministrazione che si serve di algoritmi e il luogo dell'esame del corretto esercizio dell'eventuale potere del sistema pubblico di regolare e conformare l'uso, da parte dei soggetti privati, dell'intelligenza artificiale.
Sotto il primo profilo, assume particolare rilievo il sindacato sulle c.d decisioni algoritmiche dell'amministrazione. Pur nelle comprensibili oscillazioni, la giurisprudenza amministrativa, come sopra visto, ha già avuto modo di precisare le condizioni fondamentali per assicurare la qualità di tali decisioni. Quel che val la pena di sottolineare è che, ai fini della legittimità del provvedimento basato sull'utilizzazione degli algoritmi, deve essere assicurata una piena accessibilità di questi ultimi – assimilabili ad un concetto tecnico e non ad un concetto giuridico indeterminato – alla loro formazione e alla loro evoluzione, anche con riferimento alla loro provenienza.
Sotto il secondo profilo, deve essere osservato che esso riguarda uno scenario ancora lontano dall'essere realizzato. Quel che forse, in questa sede, può essere anticipato è che potrà costituire un utile ausilio la giurisprudenza del giudice amministrativo sull'attività di regolazione, che richiede un supplemento di partecipazione, anche in funzione di controllo dei soggetti interessati (c.d legalità procedurale) e sul sindacato delle valutazioni tecniche. Mediante tale sindacato, difatti, potrà essere valutata con attenzione la maggiore attendibilità delle scelte operate dall'autorità amministrativa.

BIBLIOGRAFIA

- Abraham, K.S.-Rabin, R.L., Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era, in 105 Virginia Law Review, 2019, p. 127 ss.
- Balp, G.-Strampelli, G., Preserving Capital Markets Efficiency in the High-Frequency Trading Era, in 1 University of Illinois Journal of Law, Technology & Policy, 2018, p. 349 ss.
- Barlow, J.P., A Declaration of Independence of the Cyberspace, Electronic Frontier Foundation, 1996
- Bassini, M.-Pollicino, O., Liguori, L., Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?, in Pizzetti, F. (a cura di), Intelligenza artificiale, protezione dei dati personali e regolazione, Giappichelli, 2018, p. 333 ss.
- Casey, A.J.-Niblett, A., The Death of Rules and Standards, in 92 Indiana Law Journal, 2017, p. 1401 ss.
- Casonato, C., Intelligenza artificiale e diritto costituzionale: prime considerazioni, in Diritto pubblico comparato ed europeo, fascicolo Speciale, maggio 2019, p. 101 ss.
- Chagal-Feferkorn, K.A., The Reasonable Algorithm, in 1 University of Illinois Journal of Law, Technology & Policy, 2018, p. 111 ss.
- Costanzo, P., Il fattore tecnologico e le trasformazioni del costituzionalismo, in Rassegna Parlamentare, 4, 2012, p. 811 ss.
- De Gregorio, G., From Constitutional Freedoms to the Power of the Platforms: Protecting Fundamental Rights Online in the Algorithmic Society, in 11(2) European Journal of Legal Studies, 2019, p. 65 ss.
- Durovic, M.-Janssen, A., The Formation of Blockchain-based Smart Contracts in the Light of Contract Law, in 26 European Review of Private Law, 2018, p. 753 ss.
- Estlund, C., What Should We Do After Work? Automation and Employment Law, in 128 Yale Law Journal, 2018, p. 254 ss.
- Geistfeld, M.A., A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation, in 105 California Law Review, 2017, p. 1611 ss.
- Hallevy, G., The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control, in 4(2) Akron Intellectual Property Journal, 2016, p. 171 ss.
- Johnson D.-Post. D., Law and borders: The rise of law in cyberspace, in 48(5) Stanford Law Review, 1996, p. 1367 ss.
- Kaminski, M., The Right to Explanation, Explained, in 34(1) Berkeley Technology Law Journal, 2019, p. 143 ss.
- Lessig, L., Code and other Laws of Cyberspace, Basic Books, 1996
- Luciani, M., La decisione giudiziaria robotica, in Rivista AIC, 3, 2018, p. 872 ss.
- Ortolani, P., The impact of blockchain technologies and smart contracts on dispute resolution: arbitration and court litigation at the crossroads, in Uniform Law Review, 2019, p. 1 ss.
- Pagallo, U., The Laws of Robots, Springer, 2013
- Pariser, E., The Filter Bubble: What the Internet Is Hiding from You, Penguin, 2011
- Pizzetti, F. (a cura di), Intelligenza artificiale, protezione dei dati personali e regolazione, Giappichelli, 2018
- Pollicino, O., L'efficacia orizzontale dei diritti fondamentali previsti dalla Carta La giurisprudenza della Corte di giustizia in materia di digital privacy come osservatorio privilegiato, in Rivista di diritto dei media, 3, 2018, p. 138 ss.
- Pollicino, O.-Bassini, M., The Law of the Internet between Globalisation

and Localisation, in Maduro, M.P.-Tuori, K.-Sankari, S. (eds.), Transnational Law: Rethinking European Law and Legal Thinking, Cambridge, 2014, p. 346 ss.
- Reidenberg, J. R., Lex Informatica: The Formulation of Information Policy Rules through Technology, in 76 Texas Law Review 1997, p. 553 ss.
- Sartor, G., Cognitive automata and the law: Electronic contracting and the intentionality of software agents, in 17(4) Artificial Intelligence and Law 2009, p. 253 ss.
- Savelyev, A., Contract Law 2.0: «Smart» Contracts as the Beginning of the End of Classic Contract Law, Higher School of Economics Research Paper No. WP BRP 71/LAW/2016
- Simoncini, A., L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà, in BioLaw Journal – Rivista di BioDiritto, 1, 2019, p. 63 ss.
- Sunstein, C., Republic.com 2.0, Princeton University Press, 2009
- Szabo, N., Formalizing and Securing Relationships on Public Networks, in 2(9) First Monday, 1 September 1997
- Teubner, G., Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten, in 218 Archiv für die civilistische Praxis, 2018, p. 155 ss.
- Teubner, G., Societal Constitutionalism: Alternative to State-Centred Constitutional Theory? in Joerges, C.-Sand, I.-Teubner, G., Transnational Governance and Constitutionalism, Hart, 2004
- Wachter S., Mittelstadt B. and Floridi L., Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, in 7(2) International Data Privacy Law, 2017, p. 76 ss.
- Wagner, G., Produkthaftung für autonome Systeme, in Archiv für die civilistische Praxis 217, 2017, p. 708 ss.

"PRINCÌPI – GUIDA"

1. DIGNITÀ – La dignità dell'individuo costituisce il caposaldo su cui si fonda la società dell'algoritmo. La relazione tra uomo e macchina è incardinata sul principio di dignità che impedisce alla tecnologia di marginalizzare l'individuo ponendolo in una situazione di soggezione.

2. DIVIETO DI DISCRIMINAZIONE – In nessun caso, l'utilizzo dell'intelligenza artificiale deve condurre a forme ingiustificate di discriminazione diretta e indiretta, specialmente quando basata su caratteristiche quali la razza, il colore, la lingua, la religione o le convinzioni, la nazionalità o l'origine nazionale o etnica, nonché l'ascendenza, l'età, la disabilità, il sesso, l'identità di genere, l'orientamento sessuale.

3. TRASPARENZA – I sistemi di intelligenza artificiale devono assicurare la trasparenza dei propri processi intesa come comprensibilità per l'individuo e non soltanto come spiegazione dettagliata del processo algoritmico. In tal senso, le tecnologie devono essere sviluppate in modo che i soggetti pubblici e privati siano in grado di fornire all'individuo quegli elementi di comprensione tali da renderlo consapevole dell'impatto dell'utilizzo di tali tecnologie sulle proprie libertà e diritti.

4. ACCOUNTABILITY – La complessità dell'intelligenza artificiale non può costituire una giustificazione che conduca a una deresponsabilizzazione in capo agli attori che fanno uso di tali tecnologie. È opportuno che sia individuato un centro d'imputazione sul quale l'individuo può fare affidamento quando l'impiego di decisioni automatizzate influenza i propri diritti e libertà fondamentali.

5. LIBERTÀ D'ESPRESSIONE – La libertà di espressione costituisce uno dei fondamenti sui cui si basano le società democratiche. L'impiego di sistemi di intelligenza artificiale per la gestione delle informazioni e dei contenuti deve evitare forme di polarizzazione e esclusione che possano minare la possibilità per l'individuo di comprendere e partecipare in una società democratica. L'utilizzo di tali sistemi deve tendere al rispetto della libertà di ogni individuo di esprimere il proprio pensiero e il suo diritto a informare e informarsi.

6. PRIVACY E PROTEZIONE DEI DATI PERSONALI – L'utilizzo di tecnologie ubiquitarie e l'impiego massivo di dati mettono a rischio la tutela della privacy e dei dati personali. I principi di *privacy by design* e *by default* costituiscono la guida per i soggetti coinvolti nell'attività di trattamento dati al fine di assicurare il rispetto dei dati personali attraverso la predisposizione di misure tecniche e organizzative adeguate, volte ad attuare in modo efficace i principi di protezione dei dati (*by design*), nonché a garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento (*by default*).

7. GIUSTIZIABILITÀ – Avverso le decisioni automatizzate che riguardano diritti e libertà fondamentali dell'individuo devono essere previsti meccanismi di ricorso pubblici e/o privati attraverso cui l'individuo possa chiedere e, eventualmente, ottenere la revisione o la temporanea sospensione di un'eventuale decisione automatizzata.

8. PUBBLICA AMMINISTRAZIONE – L'utilizzo di sistemi di intelligenza artificiale da parte della Pubblica amministrazione non deve mirare soltanto a esigenze di efficienza dell'attività amministrativa, ma essere strumento attraverso cui migliorare i processi amministrativi e interfacciarsi con l'individuo nel rispetto dei principi di legalità, trasparenza e correttezza.

9. REGOLAMENTAZIONE – La regolamentazione dei sistemi di intelligenza artificiale deve mirare alla tutela dell'individuo al fine di evitare qualsiasi deriva tecnocratica o mercantilista. La tutela di altre libertà e diritti, quali la libertà d'impresa, devono essere parametrate rispetto alla necessità che l'individuo non sia soggiogato alla tecnica ma sia parte attiva del progresso umano e scientifico.

10. SOGGETTIVITÀ - Diversamente rispetto ai soggetti che attualmente intrattengono rapporti nell'ambito del diritto privato, il robot dotato di intelligenza artificiale non persegue un proprio interesse. L'interesse è pur sempre riconducibile a un diverso soggetto, inteso in senso tradizionale, che si avvale degli strumenti messi a disposizione dell'intelligenza artificiale. Non sussiste dunque la necessità di creare nuovi soggetti giuridici. La tecnologia e l'intelligenza artificiale sono (e devono permanere) al servizio dei soggetti giuridici intesi in senso tradizionale.

11. SMART CONTRACTS E TECNOLOGIA BLOCKCHAIN - Non corrisponde al vero che il codice si sostituirà al diritto; il diritto continuerà ad essere applicabile. Eventuali falle nei sistemi operativi potrebbero agevolare fraudolente manomissioni da parte di soggetti terzi, che richiederebbero l'applicazione delle norme sulla responsabilità. Inoltre, non può escludersi che uno *smart contract* violi norme imperative o che venga concluso da un soggetto incapace di agire.

12. EFFETTIVITÀ - Le procedure di esecuzione automatizzata dei contratti e delle norme possono comportare un aumento dell'effettività del diritto. Il fenomeno riguarda, ad esempio, la compensazione pecuniaria alla quale è tenuto il vettore aereo, in caso di ritardo o cancellazione del volo, ai sensi dell'art. 5, par. 3, del Regolamento CE n. *261/2004. Rendendo la tecnologia degli smart contracts obbligatoria per i vettori aerei, sarebbe possibile assicurare ai consumatori gli indennizzi in via automatica.*

13. GRANULARITÀ - Le nuove tecnologie potrebbero apportare un significativo miglioramento delle norme giuridiche, rendendole più adeguate alla soluzione del caso concreto. Con l'avvento dei big data, i software potrebbero modellare le norme giuridiche sulla base delle peculiarità di ogni singolo consociato (c.d. "granular norms"). Ciò potrebbe determinare il superamento dell'utilizzazione di standard e clausole generali nel diritto privato.

14. MERCATI - I processi decisionali algoritmici assumono rilievo nel contesto dei mercati attraverso la pratica del c.d. "high-frequency trading" (HFT). La pericolosità dell'HFT è testimoniata da interventi normativi volti a porre un freno alle relative attività speculative e a garantire una certa trasparenza con riferimento alle modalità operative. Sotto questi profili, deve essere valutata con attenzione la revisione della direttiva europea, c.d. MiFID II (Markets in Financial Instruments Directive) e ulteriori regolamenti attuativi, che obbligano gli high frequency traders a registrarsi come imprese di investimento e a rendere pubblici i loro algoritmi, fornendo garanzie sull'attendibilità dei loro software.

15. RESPONSABILITÀ CIVILE - *La tutela risarcitoria* dovrebbe continuare a garantire una certa carica deterrente nei confronti dei possibili tortfeasors e, al contempo, rispondere all'esigenza di indennizzare adeguatamente i soggetti danneggiati. Le nuove tecnologie rendono necessaria l'elaborazione di parametri per attribuire la responsabilità in capo al danneggiante e, posta la diversa gestione del rischio, impongono ai giuristi di ripensare alcune norme concernenti la responsabilità oggettiva.

16. COLPEVOLEZZA - Per quanto attiene ai parametri di riferimento per attribuire la responsabilità, si tratta di valutare la condotta di sistemi in grado di apprendere autonomamente e di decidere sulla base dei dati raccolti. Un primo parametro utilizzabile è quello relativo alle capacità umane. Le capacità dell'uomo potrebbero costituire – almeno nel primo periodo – una soglia minima di "diligenza" richiesta al sistema operativo. In prospettiva, si potrebbe sviluppare un parametro autonomo per i sistemi intelligenti, diverso a seconda del settore preso in esame. Anche i dati statistici concernenti le attività esercitate da *software* dotati di intelligenza artificiale potrebbero giocare un ruolo significativo.

17. PRODUCT LIABILITY - Rispetto alle nuove tecnologie, le norme di conio europeo, contenute nella direttiva 85/374/CEE sulla responsabilità del produttore, appaiono inadeguate poiché si riferiscono a tecnologie obsolete e non chiariscono con precisione in quali casi un sistema operativo in grado di apprendere autonomamente è da considerare difettoso.

18. VEICOLI A GUIDA AUTONOMA - Un ambito peculiare è quello della responsabilità da circolazione di veicoli a guida autonoma. In un primo periodo, le attuali norme relative alla responsabilità da circolazione di veicoli, assistite dalla previsione dell'assicurazione obbligatoria, potrebbero continuare a fornire una regolamentazione soddisfacente. In prospettiva futura, occorrerà riflettere su nuove forme di responsabilità che coinvolgano maggiormente i produttori delle vetture. Questi ultimi sono infatti i soggetti che immettono i sistemi operativi.

19. PROCESSO – L'opacità che ancora contraddistingue i sistemi di intelligenza artificiale rischia di determinare, anche nelle applicazioni a carattere predittivo, conseguenze pregiudizievoli in ambito processuale, dove il principio della parità delle armi e il rispetto dei diritti fondamentali devono continuare a operare come nucleo inalienabile.

20. SINDACABILITÀ – Per assicurare un sindacato sulle decisioni algoritmiche della pubblica amministrazione occorre assicurare piena accessibilità degli algoritmi (da intendersi come concetto tecnico, e non solo giuridico), in particolare in relazione alle fasi di formazione, evoluzione e anche con riferimento alla loro provenienza.