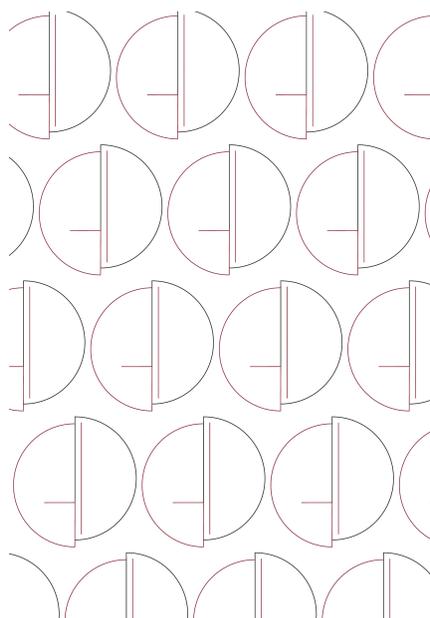


# Annuario 2022 Osservatorio Giuridico sulla Innovazione Digitale

Yearbook 2022  
Juridical Observatory on Digital Innovation

a cura di

Salvatore Orlando e Giuseppina Capaldo



## 6. Diritti fondamentali e ambienti digitali: prime note di una ricerca sul diritto a non essere sottoposto a una decisione interamente automatizzata

*Daniele Imbruglia (Università di Roma La Sapienza)*

“(...) *an algorithm – an opinion formalized in code*”  
Cathy O’Neil

### 6.1. *Onlife*, la quarta rivoluzione

Uno dei più influenti filosofi contemporanei, Luciano Floridi, ha introdotto il termine ‘on-life’ per riferirsi a quelle attività, oggi largamente prevalenti nella vita quotidiana, che si collocano oltre la tradizionale distinzione tra *off-line* e *on-line*. Secondo la definizione offerta dal filosofo italiano, il termine starebbe ad indicare “*the new experience of a hyperconnected reality which it is no longer sensible to ask whether one may be online or offline*”<sup>1</sup>. Si pensi, ad esempio, alla guida di un veicolo secondo le istruzioni che si riceve dal proprio smartphone, alla visione di un film su una smart-tv, e via di seguito. In tutte queste situazioni è difficile, nonché certamente opinabile, rispondere alla domanda se la persona che guida o che guarda il film sia *off-line* oppure *on-line*.

Lo stesso filosofo descrive il più generale impatto delle attuali tecnologie dell’informazione della comunicazione in termini di rivoluzione, la quarta. Dopo quella copernicana che ci ha costretto a smettere

---

Questo articolo costituisce uno sviluppo della relazione tenuta dall’A. alla *Summer School ‘La responsabilità civile nell’era digitale’* del Dipartimento di Giurisprudenza dell’Università di Foggia dal 6 al 10 settembre 2021. Una versione ridotta del testo è pubblicato anche nel volume che raccoglie tutte le relazioni lì svolte, con il titolo *Diritti fondamentali e ambienti digitali. Appunti per una ricerca*. Un lavoro più definito è in corso.

<sup>1</sup> L. FLORIDI, *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Heidelberg – New York – Dordrecht – London, Springer, 2015, p. 1.

di credere alla centralità della terra nell'universo, quella darwiniana che ha smentito l'assunto per cui l'uomo sia al centro del regno biologico e quella freudiana che ha rotto la percezione della coscienza come di un luogo puro e trasparente, vi sarebbe ora una nuova rottura – appunto una quarta rivoluzione – che depone l'uomo “dalla posizione privilegiata ed esclusiva che avevamo nel regno del ragionamento logico, della capacità di processare informazioni per assolvere un determinato compito”<sup>2</sup>.

Ancora. La quantità di attività che sino a qualche tempo fa era di esclusiva competenza dell'essere umano e che oggi è svolta da tecnologie informatiche è in continuo aumento. D'altronde, il vantaggio dell'utilizzo di tali tecniche è lampante. A parità di tempo, il numero di operazioni che si possono così svolgere è infinitamente maggiore rispetto a quello che un essere umano è in grado di processare. Non solo. Affidandosi a questi strumenti, la persona fisica in ipotesi preposta a quella attività potrebbe svolgere compiti diversi, meno ripetitivi e per i quali è necessaria la specifica capacità dell'uomo<sup>3</sup>.

## 6.2. I diritti fondamentali in ambiente digitale

Rispetto a questo contesto in continua evoluzione, si deve registrare un profluvio di dichiarazioni che riconoscono e affermano il rispetto di certi diritti fondamentali in ambienti digitali<sup>4</sup>. Questi testi dimostrano come si stia affermando la consapevolezza circa l'esigenza di una regolazione<sup>5</sup> e che, onde evitare di trovarsi ancora una volta davanti a degli “*unforeseen disbenefits*”<sup>6</sup> dell'innovazione tecnologica, non

---

<sup>2</sup> L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, Raffaele Cortina Editore, 2017, p. 105.

<sup>3</sup> Si pensi al passo di LEIBNIZ, citato in W. JORDAN, *Die Leibniz'sche Rechenmaschine*, in *Zeitschrift für Vertnesaungsweaen*, 1897, p. 307, per cui “*Indignum enim est excellentium virorum horas servili calculandi labore perire, qui Machina adhibita vilissimo cuique secure transcribi posset*”.

<sup>4</sup> Con riferimento all'intelligenza artificiale, ad esempio, si veda il grafico disponibile in [https://wilkins.law.harvard.edu/misc/PrincipledAI\\_FinalGraphic.jpg](https://wilkins.law.harvard.edu/misc/PrincipledAI_FinalGraphic.jpg).

<sup>5</sup> In argomento, classici i riferimenti alla tesi di F.H. EASTERBROOK, *Cyberspace and the Law of the Horse*, in *Univ. Chi. Legal. Forum*, 1996, 207 e alle repliche di L. LESSIG, *The Law of the Horse: What Cyberlaw Might Teach*, in *Harv. Law Rev.*, 1999, p. 501 e, più di recente, R. CALO, *Robotics and the Lessons of Cyberlaw*, in *Cal. Law Rev.*, 2015, p. 513.

<sup>6</sup> M. KRANZBERG, *Technology and History: “Kranzberg's Laws”*, in *Techn. Cult.*, 1986, p. 548.

sia sufficiente un approccio in termini di etica<sup>7</sup>. Il mancato intervento del diritto o la sua articolazione secondo esigenze prettamente capitalistiche e di profitto può determinare un fallimento nell'assicurazione del benessere sociale<sup>8</sup> nonché una mancata tutela di posizioni giuridiche che pure sono al centro del nostro ordinamento e della nostra società (es. i diritti inviolabili della persona)<sup>9</sup>.

Se è vero che si tratta di carte e dichiarazioni che, nella maggior parte dei casi, sono prive di valore giuridico, si deve osservare che tale carattere non ne giustifica una valutazione in termini di irrilevanza e insignificanza, come se questi sforzi non fossero che banali tentativi politici o di mera gestione del consenso. La storia, infatti, ha dimostrato a più riprese come “[p]er molti diritti, anche quelli ritenuti nell’attuale fase di evoluzione pacificamente “giustiziabili”, l’approdo della difesa in giudizio non è stato quasi mai un dato originario, quanto piuttosto un punto di conquista, attraverso la progressiva crescita delle rivendicazioni sociali e della forza culturale che tali diritti sono riusciti col tempo ad esprimere”<sup>10</sup>.

Oltre a rappresentare un importante tentativo di regolazione di un fenomeno in divenire, tali documenti si lasciano apprezzare perché non si esauriscono nella affermazione dei diritti fondamentali esistenti *off-line* anche negli ambienti digitali, ma al loro interno prevedono anche nuove situazioni giuridiche così innovando il catalogo dei diritti fondamentali, i quali, come noto, non vanno intesi come un qualcosa di fuori dalla storia e assoluto: “[a]nche i diritti dell’uomo sono diritti storici, che emergono gradualmente dalle lotte che l’uomo combatte per la propria emancipazione e dalla trasformazione delle condizioni di vita che queste lotte producono”<sup>11</sup>.

Nel prosieguo di questo lavoro, intendo avanzare una possibile ricerca su tali diritti. Il contributo è diviso in due parti. Nei paragrafi immediatamente successivi darò brevemente conto di due recenti dichiarazioni così da rendere il lettore edotto del fenomeno *de qua*. La scelta dei documenti è dettata dal tempo: si tratta di alcuni dei testi più

---

<sup>7</sup> Per tutti, D. TAFANI, *What’s wrong with “AI Ethics” narratives*, in *Boll. tel. fil. pol.*, 2022.

<sup>8</sup> Per tutti, U. PAGANO, *The crisis of intellectual monopoly capitalism*, in *Cambr. Jour. Econ.*, 2014, p. 1409.

<sup>9</sup> Per tutti, S. RODOTÀ, *Il diritto di avere diritti*, Roma-Bari, Laterza, 2012, p. 374.

<sup>10</sup> Così, A. D’ALOIA, *Generazioni future (diritto costituzionale)(voce)*, in *Enc. dir., Ann. IX*, 2016, p. 354.

<sup>11</sup> N. BOBBIO, *L’età dei diritti*, Torino, Einaudi, 2020, p. 57-58.

recenti di quell'ampio insieme di carte digitali ora richiamato<sup>12</sup>. Nella seconda parte del lavoro, invece, cercherò di verificare come nasce e si afferma un diritto fondamentale digitale, discutendo l'ipotesi di decisioni assunte mediante procedimenti automatizzati e dei principi che si stanno via via affermando nelle corti e nella legislazione con riferimento a detto fenomeno.

### 6.3. La Carta derechos digitales

Il 14 luglio 2021, il governo spagnolo ha presentato la *Carta derechos digitales* ('*Carta*'), un documento non normativo che afferma il valore della persona e della dignità umana nella definizione delle regole e delle politiche della nuova realtà digitale<sup>13</sup>. Redatto da un gruppo di esperti di discipline diverse, la *Carta* è il risultato di circa un anno di lavori e di due consultazioni pubbliche. Il testo si compone di ventotto disposizioni suddivise in sei sezioni e anticipate da un preambolo che dà conto delle ragioni dell'intervento, centrale nella c.d. *Plan España Digital 2025*.

La prima sezione della *Carta*, intitolata *Derechos de libertad*, si apre con il riferimento al rispetto negli ambienti digitali dei diritti fondamentali riconosciuti nelle diverse carte e dichiarazioni e prosegue con l'affermazione del diritto all'identità, alla protezione dei dati (con esplicito richiamo al regolamento europeo 2016/679, su cui *infra*) e al diritto all'utilizzo di uno pseudonimo. A tal proposito, la *Carta* prevede che tale pretesa possa essere limitata solo quando l'identificazione personale sia necessaria e che sia comunque possibile la identificazione dell'utente ove richiesto dalla autorità giudiziaria. La medesima sezione prevede poi che il ricorso a sistemi di analisi che impieghino decisioni automatizzate o la profilazione degli individui sia possibile solo quando ammesso dalla normativa nonché il diritto di tutte le persone a strumenti di sicurezza adeguati a un trattamento dei dati sicuro. La sezione si chiude demandando la disciplina del diritto all'eredità digitale (ossia di tutti i beni e e diritti che, nell'ambiente digitale, erano detenuti dalla persona deceduta) al legislatore.

---

<sup>12</sup> In tempi recenti, è altresì intervenuta la *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, adottata lo scorso 26 gennaio 2022 (Com (2022)28) e il cui testo è disponibile in <https://eur-lex.europa.eu>.

<sup>13</sup> Il testo della *Carta* è disponibile in [https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta\\_Derechos\\_Digitales\\_RedEs.pdf](https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf).

La sezione successiva, dal titolo di *Derechos de igualdad*, contiene cinque articoli. Oltre al diritto alla non discriminazione, al diritto all'accesso e al contrasto al divario digitale, la sezione prevede una ricca disposizione in merito alla protezione dei minori. Tale disciplina (art. X) si apre ponendo a carico dei soggetti responsabili (es. i genitori) il compito di assicurare un uso responsabile degli ambienti digitali per garantire il corretto sviluppo del minore. Tra le altre cose, l'articolo X prevede l'introduzione di procedure per la verifica dell'età, il diritto di ricevere una formazione e un'informazione adeguata alle capacità del minore e un generale divieto di trattamento dei dati personali dei minori a fini di profilazione.

La terza sezione della *Carta* contiene disposizioni in tema di partecipazione e di informazione tramite ambienti digitali. Essa si apre con il riferimento alla neutralità della rete Internet e prosegue poi affrontando il tema dell'informazione in ambiente digitale. In particolare, l'art. XV afferma il diritto a ricevere informazioni veritiere e conformi ai protocolli sulla trasparenza (in base ai quali comunicare se l'informazione è stata elaborata mediante processi automatizzati o se ha carattere pubblicitaria o meno). La sezione in parola si conclude con tre articoli che affermano il diritto della cittadinanza alla partecipazione per mezzi digitali, a ricevere una educazione digitale e ad avere rapporti digitali con la pubblica amministrazione.

La quarta e la quinta sezione affrontano precise tematiche degli ambienti digitali. In una, si offrono indicazioni relative al rispetto dei diritti fondamentali dei lavoratori (art. XIX) e alla libertà di impresa in un contesto concorrenziale (art. XX) e nell'altra si affermano articoli in materia di ricerca scientifica, diritto alla salute e all'attività artistica-culturale, nonché all'impiego di programmi di intelligenza artificiale o di neuro-tecnologie. In tale quinta sezione, rubricata *Derechos digitales en entornos e específicos*, la *Carta* fa riferimento anche alla necessità dello sviluppo tecnologico di rispettare la sostenibilità ambientale e le generazioni future. La *Carta* si conclude con una sesta sezione, *Garantias y eficacia*, che riconosce la tutela dei diritti fondamentali anche in ambiente digitale.

Il testo spagnolo ricorda la *Dichiarazione dei diritti in Internet*, approvata dalla Camera dei Deputati nazionale nel 2015 e che conteneva già diverse disposizioni analoghe a quelle ora affermate nella *Carta* (e.g. neutralità della rete, inviolabilità dei sistemi informatici, protezione

dell’anonimato)<sup>14</sup>. Questa parziale coincidenza conferma la diffusione di nuovi diritti fondamentali. Si pensi, ad esempio, al diritto all’utilizzo di uno pseudonimo o alla navigazione anonima. È evidente che un tale diritto non ha senso fuori dagli ambienti digitali, così come è chiaro che esso costituisce uno strumento e una tecnica che consentono la massima libertà di espressione e di pensiero. L’articolo 10 della *Dichiarazione dei diritti in Internet* è esplicito in tal senso, allorché afferma “Ogni persona può accedere alla rete e comunicare elettronicamente usando strumenti anche di natura tecnica che proteggano l’anonimato ed evitino la raccolta di dati personali, in particolare per esercitare le libertà civili e politiche senza subire discriminazioni o censure”. Tutto quanto ora detto sui nuovi diritti fondamentali, ovviamente, non sorprende: “[c]erte richieste nascono, infatti, solo quando nascono certi bisogni e nuovi bisogni nascono in corrispondenza del mutamento delle condizioni della società”<sup>15</sup>.

#### 6.4. Il Blueprint for an AI Bill of Rights statunitense

Lo scorso ottobre l’amministrazione statunitense ha pubblicato un documento teso a “*making automated systems work for the American people*” e noto come *Blueprint for an AI Bill of Rights*<sup>16</sup>. Il testo si compone di una introduzione e di cinque principi, ciascuno dei quali accompagnato da una descrizione e da un commento. L’aspirazione del documento è quella di fornire un kit, una bussola, in grado di orientare le future decisioni politiche in materia di “*automated system*”. Con tale espressione il documento intende riferirsi a quei *software* che usano calcoli computazionali per determinare risultati idonei a rilevare rispetto alla vita di individui o di comunità.

---

<sup>14</sup> Il testo della *Dichiarazione dei diritti in internet* è disponibile in [www.camera.it](http://www.camera.it). Sul c.d. *digital constitutionalism*, si v. i ricchi lavori di E. CELESTE, *Digital constitutionalism: a new systematic theorisation*, in *Int. Rev. Law Comp. Techn.*, 2018, p. 76; T.E. FROSINI, *Il costituzionalismo nella società tecnologica*, in *Dir. informaz. informat.*, 2020, p. 465; G. DE GREGORIO, *The rise of digital constitutionalism in the European Union*, in *Int. Jour. Const. Law*, 2021, p. 41. A conferma della diffusione descritta nel testo è, poi, il dato per cui molti dei temi toccati dal testo spagnolo e prima ancora da quello italiano si rinvencono anche nella citata e recente *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*.

<sup>15</sup> C. FARALLI, *Introduzione*, in EAD. (a cura di), *La storicità del diritto*, Torino, Giappichelli, 2018, p. 8.

<sup>16</sup> Il testo è disponibile in <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.

Il primo principio intende proteggere l'utente da *"unsafe or ineffective systems"*. A tal fine, esso prevede che questi software siano sviluppati con la consultazione di diverse partecipanti, così che tutti i rischi e i diversi potenziali impatti del sistema siano rappresentati. Detti sistemi, poi, devono essere sottoposti a numerose verifiche, volte all'identificazione e alla riduzione dei rischi e a un monitoraggio continuo. Oltre a questi doveri di controllo e di coinvolgimento di rappresentanti sociali, il principio in esame prevede anche il divieto di progettazione di sistemi automatizzati tesi (o con la possibilità ragionevolmente prevedibile) a mettere in pericolo la sicurezza dell'utente o di una certa comunità. Il principio suggerisce di eseguire una valutazione indipendente che confermi la sicurezza e l'efficacia del sistema, compresa la comunicazione delle misure adottate per mitigare i potenziali danni, e i cui risultati devono essere resi pubblici.

Il secondo principio afferma il fondamentale divieto di discriminazione algoritmica (*"algorithmic discrimination"*). Tale patologia si verifica allorché il software determina un trattamento diverso ingiustificato o a un impatto sfavorevole sulle persone in base a razza, colore, etnia, sesso, religione, età, origine nazionale, disabilità, stato di veterano, informazioni genetiche o qualsiasi altra classificazione protetta dalla legge. È interessante notare come nel vocabolo 'sesso' il commento al principio in esame ricomprenda anche gravidanza, parto e condizioni mediche correlate, identità di genere, stato intersessuale e orientamento sessuale. Il divieto di questo genere di discriminazione impone a tutti i soggetti coinvolti nella costruzione del prodotto *software* (i progettisti, gli sviluppatori e i distributori) di adottare misure proattive e continue. Tra queste misure è incluso l'uso di dati rappresentativi e la protezione contro i proxy per le caratteristiche demografiche, i test e la mitigazione delle disparità prima dell'implementazione e in corso, nonché l'adozione di una chiara supervisione organizzativa. Anche qui si suggerisce la predisposizione di una valutazione indipendente dell'impatto dell'algoritmo, che includa i risultati dei test di disparità e le informazioni sulla mitigazione, e che sia reso pubblico e facilmente fruibile.

Il terzo principio prevede il divieto di pratiche abusive di trattamento dei dati e il diritto alla piena autonomia sull'utilizzo dei dati personali. Al fine di proteggere l'utente da violazioni della privacy, il documento insiste sull'importanza di scelte progettuali (*"by design"*) che garantiscano l'inclusione di tali protezioni per impostazione

predefinita, tra cui la garanzia che la raccolta dei dati sia conforme alle aspettative ragionevoli e che vengano raccolti solo i dati strettamente necessari per il contesto specifico. Ancora una volta, tale principio si rivolge *in primis* ai soggetti che implementano detti software, i quali dovrebbero richiedere l'autorizzazione dell'utente e rispettare le sue decisioni in merito alla raccolta, all'uso, all'accesso, al trasferimento e alla cancellazione dei suoi dati nei modi più appropriati e nella misura più ampia possibile. Laddove ciò non sia possibile, si prevede l'obbligo di utilizzo di garanzie alternative di *privacy by design*. Accanto a questo aspetto, il principio in esame prevede il divieto di impiego di meccanismi di progettazione e di esperienza utente che offuschino la scelta dell'utente o che lo gravino con impostazioni predefinite invasive della *privacy*. Il principio insiste altresì sulla raccolta del consenso dell'utente: si afferma che questo possa utilizzato solo per giustificare la raccolta dei dati nei casi in cui può essere dato in modo appropriato e significativo e che le richieste di consenso siano brevi, comprensibili e redatte in un linguaggio semplice. In critica alle attuali pratiche di difficile comprensione, il documento insiste sull'importanza di dare all'utente la possibilità di decidere in merito alla raccolta dei dati e al contesto specifico di utilizzo dei propri dati.

Il penultimo principio concerne il diritto a essere informati sull'impiego di un *software* automatizzato. Tale diritto impone ai progettisti, agli sviluppatori e ai distributori di sistemi automatizzati di fornire una documentazione in un linguaggio semplice e generalmente accessibile, che con descrizioni chiare spieghi il funzionamento generale del sistema e il ruolo svolto dall'automazione, avvisando sull'eventualità che tali sistemi automatizzati siano in uso. Inoltre, l'utente deve sapere come e perché un esito che lo riguarda è stato determinato da un *software* automatizzato, anche quando il sistema automatico non è l'unico elemento che determina l'esito. Anche in questo caso si suggerisce la previsione di rendere pubblici i rapporti che, in un linguaggio semplice, includono informazioni sintetiche e delle valutazioni sulla chiarezza e sulla qualità dell'avviso e delle spiegazioni.

Infine, l'ultimo principio afferma che ciascun utente dovrebbe essere in grado *i)* di rifiutare un procedimento basato sull'impiego di software automatizzanti e *ii)* di avere accesso a una persona umana in grado di valutare e risolvere rapidamente i problemi che incontra. Oltre che nelle ipotesi previste dalla legge, questa facoltà di rifiuto a favore di un'alternativa umana andrà riconosciuta ogni qualvolta sia

opportuno il ricorso a una valutazione da parte di un essere umano. L'accesso a detta valutazione dovrebbe essere affidabile e accessibile. In alcuni settori più sensibili, gli stessi software automatizzati dovrebbero prevedere e incorporare la partecipazione di un essere umano almeno per le decisioni più rilevanti.

## 6.5. Come nasce un diritto fondamentale digitale? Il problema del *machine learning*

Come anticipato, intendo sviluppare la ricerca qui accennata discutendo sul come simili diritti – storicamente dati e quindi tipicamente digitali – sorgano e circolino. In questo contributo, mi concentrerò nell'analisi di quella straordinaria tecnologia rappresentata dai *software* di intelligenza artificiale<sup>17</sup> che fanno applicazione del *machine*

---

<sup>17</sup> Ad oggi, non esiste una definizione comune e pacifica di cosa sia l'intelligenza artificiale (AI; *artificial intelligence*, IA) (in tal senso, tra gli altri, A. SANTOSUOSSO, C. BOSCARATO, F. CAROLEO, *Robot e diritto: una prima ricognizione*, in *Nuova giur. civ. comm.*, 2012, 497 e A. D'ALOIA, *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *Riv. Biodir.*, 2019, p. 8). Un report dell'Unione Europea – *AI Watch Defining Artificial Intelligence* – del 2020 elenca oltre cinquanta definizioni, proposte da Stati, istituzioni internazionali, euro-unitarie, accademici e soggetti privati (*AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, Luxembourg, 2020, doi:10.2760/382730, JRC118163; il testo è disponibile in <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>), ma resta il dato che nessuna delle definizioni lì richiamate sia riuscita ad imporsi e permane una grande eterogeneità e distanza tra le varie concezioni dell'intelligenza artificiale: si è tuttora in una fase in cui è ancora difficile tracciare "a bright-line distinction between what constitutes AI and what does not" (NATIONAL SCIENCE AND TECHNOLOGY COUNCIL COMMITTEE, *Preparing for the future of Artificial Intelligence*, 2016, 7 (in [www.whitehouse.gov](http://www.whitehouse.gov))). Ciò detto, si può qui richiamare la definizione proposta in *National Artificial Intelligence Initiative Act*, adottato dal Congresso degli Stati Uniti d'America e che è entrato in vigore il 1° gennaio 2021 che definisce l'*artificial intelligence* come «a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to (A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action», nonché quella contenuta nella recente proposta di regolamento europeo che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM/2021/206 final, AIA proposal) e che individua un «sistema di IA» nel *software* sviluppato con approcci di apprendimento automatico, basati sulla logica e sulla conoscenza nonché quelli statistici, che possa, per un dato insieme di obiettivi definiti dall'uomo, generare *output*, come contenuti, previsioni, raccomandazioni o decisioni, che influenzano gli ambienti con cui interagiscono.

*learning*, così definito da Tom Mitchell «*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E*»<sup>18</sup>. Per fare ciò, ritengo doveroso premettere una breve, non completa, imperfetta e certamente sommaria descrizione della tecnologia in esame.

### 6.5.1. Esempi di *machine learning*

Si pensi all'attività di riconoscimento di immagini e si prenda l'esempio di un cane. Sino a qualche tempo fa, l'unico modo per determinare se un'immagine raffigurasse quell'animale era sottoporla a una persona fisica che fosse in grado di verificare la presenza di un cane nella figura. Per comprendere come, generalmente, una persona acquista questa capacità si pensi a un genitore che passeggia con il figlio di pochi anni per la città e che, davanti a ciascun cane (*input*) incontrato per strada, si ferma ripetendo il nome dell'animale (etichetta) al figlio. In poco tempo, questi collega la parola all'animale, così da impararne il nome (cane) e quando ne vede un altro esemplare lo chiamerà correttamente (*output*). Orbene, mentre l'essere umano impara a riconoscere un cane vedendone un numero tutto sommato ristretto di esemplari, realizzare una tecnologia in grado di rilevare le immagini che raffigurano un cane richiede tecniche complesse<sup>19</sup>.

---

<sup>18</sup> T. MITCHELL, *Machine Learning*, New York, McGraw-Hill, 1997. Per una definizione più recente e, a parere di chi scrive, preferibile, si v. J.D. KELLEHER, B. MAC NAMEE, A. D'ARCY, *Fundamentals of Machine Learning for Predictive Data Analytics*, Cambr. (Mass.), 2020, p. 5, in base alla quale esso è da intendersi come un "automated process that extracts patterns from data". Il *machine learning* è un sotto-campo dell'intelligenza artificiale, che, a sua volta, è un sotto-campo delle scienze informatiche, insieme all'ingegneria del *software* e al calcolo distribuito.

<sup>19</sup> Come noto, l'origine della espressione intelligenza artificiale risale a uno studio condotto nel 1956 presso il *Dartmouth College* (New Hampshire, USA) e che, per l'appunto, intendeva procedere «on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it» (J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, p. 1, (testo disponibile in <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>)). A partire da questo obiettivo ambizioso si è così parlato di *artificial general intelligence* (AGI) e si è cercato di costruire una macchina che avesse una intelligenza cognitiva generale, paragonabile a quella dell'uomo, e che fosse in grado di svolgere tutto ciò che l'essere umano abitualmente compie. Nonostante gli ingenti investimenti che questa linea di ricerca ha ricevuto sino alla fine degli anni Ottanta del secolo scorso, i risultati attesi

Una prima strada è quella di scrivere una procedura di elaborazione dei dati (programma, *software*) inserendo tutte le istruzioni necessarie all'operazione immaginata (es. riconoscimento cane), tra cui la regola (algoritmo) di soluzione del problema (es. la definizione di cane). Così scritto, questo *software* (c.d. tradizionale) realizza la verifica dell'immagine tramite l'esecuzione degli ordini impartiti. Grazie alle attuali capacità di calcolo automatizzato, tale compito verrà svolto con modalità (velocità, precisione) inconcepibili per l'essere umano.

Una strada diversa è quella che impiega metodi di apprendimento automatico (*machine learning*). Come nel caso del *software* tradizionale, anche in questo caso, a partire da un dato che viene inserito (*input*), l'esecuzione di un algoritmo scritto in linguaggio di programmazione (*software*) restituisce un dato-risultato (*output*). A differenza di quella prima ipotesi, però, i *software* di intelligenza artificiale che fanno applicazione di *machine learning* non sono forniti dell'algoritmo, il quale verrà elaborato automaticamente (*process*) dai dati inseriti (*input*) e poi applicato. Mentre i *software* tradizionali restituiscono un *output* che è il risultato della esecuzione di ordini, funzioni e formule impartitigli dal programmatore, in quelli di *machine learning* l'*output* è il prodotto dall'esecuzione automatica dell'algoritmo elaborato automaticamente (*process*) da certi dati (*input*).

Si tornerà immediatamente sulla descrizione di questa tecnica di *machine learning*, ma è bene comprendere sin d'ora che si tratta di una modalità profondamente e assolutamente diversa di realizzare la tecnologia con cui assolvere il compito immaginato. Mentre il presupposto della prima opzione (*software* tradizionale) è la conoscenza e la scrittura di tutte le istruzioni necessarie all'operazione immaginata (es. riconoscimento cane) tra cui anche la regola (algoritmo) di soluzione del problema, nei *software* di intelligenza artificiale non è necessario scrivere tale conoscenza e ciò in quanto il programma è scritto in modo tale da consentire l'elaborazione automatica di una regola (algoritmo) di soluzione del problema e poi applicarla.

Per elaborare questa regola necessaria ad assolvere il compito (es. riconoscimento delle immagini raffiguranti un cane) si fornisce il programma di un c.d. *training data set*. In una prima ipotesi, i dati forniti

---

non sono stati raggiunti e a tutt'oggi siamo lontani dall'aver una *machine* in grado di emulare tutte le capacità proprie dell'intelligenza umana (e, in primis, l'apprendimento): in argomento, per tutti, S. DEHANE, *Apprendre!, Les talents du cerveau, le défi des machines*, Paris, Odile Jacob, 2018, p. 67.

per l'elaborazione del modello sono etichettati, ossia vi sono *input* con l'indicazione dell'*output* richiesto (= esempi di immagini raffiguranti un cane). A partire da questi dati etichettati (o, *rectius*, esempi delle risposte che si attendono), il *software* elabora una regola di riconoscimento idonea ad assolvere il compito rispetto ad immagini che presentano variabili diverse da quelle degli esempi forniti (perché raffigurano cani in ambienti differenti da quelli degli esempi oppure di una razza diversa). In una seconda ipotesi, i dati forniti per l'elaborazione del modello non sono etichettati e il programma è rifornito dei soli *input*. Rispetto a questo *training data set* non etichettato, l'algoritmo elaborerà (*data mining*), ora per assimilazione, ora per riduzione, etc., un modello idoneo ad assolvere il compito rispetto ad immagini che non erano parte del *training data set* e che presentano variabili diverse da quelle degli esempi forniti (perché raffigurano cani in ambienti differenti da quelli degli esempi oppure di una razza diversa).

In entrambi i casi (dati etichettati e dati non etichettati) e una volta elaborato il modello a partire dai *training data set*, questo viene sottoposto a una verifica di affidabilità. Per fare ciò, si ricorre ad un altro *set* di dati, c.d. *test data set*, e si verifica se la risposta (l'algoritmo, la regola, il *pattern*, il modello) elaborata dal programma conduce ad *output* corretti. In caso di corrispondenza, si può dire che la regola di riconoscimento costruita dal *software* sia efficiente nell'assolvere il compito dattogli.

### 6.5.2. Approcci di *machine learning*

A partire da questa ipotesi relativa all'utilizzo di un programma di riconoscimento immagini per identificare le figure che raffigurano un cane, si possono avanzare delle prime elementari definizioni dei metodi di apprendimento automatico (*machine learning*), c.d. *supervised* (dati etichettati) e *unsupervised* (dati non etichettati).

L'apprendimento automatico supervisionato (*supervised learning*) è un *software* che elabora in via automatica una regola (algoritmo, modello, *pattern*) a partire da esempi (*input* e *output*) forniti da un supervisore che ha il compito di guidare il *software* verso la definizione del criterio (algoritmo) idoneo all'adempimento del compito prefissato.

L'*output* può essere sia un valore numerico (regressione lineare), sia una classe o una categoria (classificazione). Un esempio del primo tipo è un *software* di valutazione di immobili. Si parte da un insieme di dati

relativi a immobili (dimensioni, caratteristiche, servizi, piano, localizzazione, etc.) e dai prezzi delle vendite (che costituiscono le c.d. etichette). A partire da questi dati etichettati, si ricava un algoritmo di apprendimento automatico supervisionato che, sulla base dei dati relativi alle vendite precedenti, è in grado di indicare un valore numerico (prezzo) per una casa nuova (non etichettata). Un esempio di *software* di apprendimento automatico supervisionato il cui *output* è una categoria (classificazione) è, invece, un programma *antispam*. Si parte da un insieme di *e-mails* e si etichettano come *spam* quelle indesiderate e come *no spam* quelle di un qualche interesse. A partire da questi dati etichettati, si ricava un algoritmo di apprendimento automatico supervisionato che, sulla base delle *e-mails* precedenti, è in grado di classificare il nuovo messaggio come *spam* o meno. Un classificatore molto importante è il c.d. *Nearest Neighbor*. A partire dai dati etichettati, il *software* costruisce due o più categorie o classi, così che poi considererà il dato da valutare come appartenente alla categoria con cui presenta le maggiori affinità (così, ad esempio, all'utente di una piattaforma *video on demand* che ha appena finito di vedere un certo film verrà suggerita la pellicola che piace ad altri *users* che hanno già visto quel primo film).

Per contro, l'apprendimento automatico non supervisionato (*unsupervised learning*) è quello in cui un programma è fornito unicamente degli *input* e dovrà elaborare una regola senza avere a disposizione gli *output* corretti. A partire dai dati disponibili, il *software* cerca delle regolarità, delle somiglianze o delle semplificazioni che consentano l'elaborazione automatica di un algoritmo da impiegare per svolgere il compito prefissato. Anche l'apprendimento automatico non supervisionato può essere impiegato per assegnare a ogni nuovo *input* un valore numerico (regressione lineare) o una categoria (classificazione). Oggi, esso è utilizzato, *inter alia*, per raccomandazioni e suggerimenti commerciali. Questi suggerimenti sono determinati dal *pattern* che il *software* ha, automaticamente, rilevato mediante l'analisi dei dati disponibili relativi, ad esempio, agli altri clienti o utenti. I principali approcci di apprendimento automatico non supervisionato sono quelli del *clustering* e dell'associazione. Nel primo caso, il *software* elabora una regola con cui raggruppare i dati secondo le proprietà che questi presentano. Nel secondo approccio, associazione, il programma ricava dai dati degli schemi che si ripetono, delle correlazioni o altri collegamenti che possono fornire una regola. Il vantaggio del ricorso a questo

tipo di metodo è duplice. Per un verso non è necessario avere dei dati etichettati (la cui formazione è necessariamente più lunga) e, per altro verso, dall'osservazione del *software* è possibile comprendere nuovi *pattern* e correlazioni inedite. Il rischio nell'utilizzo di questo tipo di apprendimento automatico *unsupervised* è non meno evidente ed è rappresentato dalla circostanza che, in assenza di etichette, l'affidabilità del modello elaborato automaticamente dal *software* può risultare meno alta e richiedere più dati.

Come è intuibile, i metodi di *machine learning* sono numerosi e spesso si combinano tra di loro. Si pensi al c.d. *semi-supervised learning*, che costituisce uno dei metodi oggi più utilizzati. In tale ipotesi, il *software* è rifornito di *input* solo parzialmente etichettati. Processando questi dati misti (etichettati e non etichettati), il programma elabora automaticamente un algoritmo in grado di valutare un nuovo dato e fornire un certo *output*. Ancora, in tempi recenti si è introdotto un ulteriore metodo c.d. *self-supervised learning*, molto promettente perché consente di svolgere compiti di apprendimento supervisionato (classificazione o regressione lineare), senza avere come *input* dati etichettati dall'uomo, la cui raccolta è necessariamente parziale, lenta, dispendiosa e costosa in termini di risorse. In tale metodo, il *software* riceve dati non etichettati e, per così dire, li etichetta automaticamente, ricercando correlazioni e osservando i dati visibili per ricostruire ciò che manca oppure delle strutture implicite. Svolta questa prima operazione ancillare (*pretext task*), il *software* utilizzerà la rappresentazione elaborata dai dati di *input* per svolgere compiti (*text task* o *downstream task*) di classificazione o di regressione lineare (es. riconoscimento oggetto di cui non era stato fornito esempio). In tempi recenti, diversi *social networks* hanno utilizzato questo metodo per elaborare *software* di *speech recognition* e, più in generale, il *self-supervised learning* è impiegato nel contesto dei programmi di comprensione del linguaggio naturale (*natural language processing*, NLP)<sup>20</sup>.

Oltre ai metodi richiamati *supra*, si trova, poi, quello c.d. rinforzato (*reinforcement learning*), che si distingue dai precedenti per la fase di *input*. Al posto di dati da cui estrapolare un modello con cui assolvere il proprio compito, infatti, il programma riceve come *input* un obiettivo da raggiungere (es. non scottarsi al sole). A partire dall'osservazione dell'ambiente (sole\nuvoloso), il *software* assume una decisione

---

<sup>20</sup> Sul tema, si veda il contributo di S. ORLANDO, *Linguaggi di programmazione e responsabilità*, in questo Annuario (cap. 11).

(prendere \ non prendere la crema di protezione) che riceve un *feedback* positivo (ricompensa) o negativo (penalità) a seconda del successo della decisione assunta rispetto all'obiettivo dato. Considerato che il *software* ha per scopo la massimizzazione del rinforzo, il programma ripeterà nel tempo le decisioni che hanno un *feedback* positivo (sole \ crema). Questa modalità di apprendimento automatico è impiegata, ad esempio, nelle *self-driving cars*.

Peraltro, occorre tenere a mente che il *machine learning* è in continua evoluzione. Se i metodi di *supervised learning* e *unsupervised learning* costituiscono metodi classici di *machine learning*, a partire dagli anni Dieci del secolo attuale si è affermato il c.d. *deep learning*, che costituisce una tecnica di apprendimento automatico che utilizza livelli multipli (reti neurali) per elaborare progressivamente caratteristiche di livello superiore dall'*input* iniziale e così individuare in modo ancora più efficace possibili *pattern* da utilizzare come algoritmi. Oltre che nei menzionati *software* di NLP e in genere di *self-supervised learning*, un esempio molto discusso di rete neurale è la c.d. rete generativa avversaria (*Generative Adversarial Network*, GAN), dove un algoritmo genera e l'altro controlla fino a non rendere possibile per il controllore fare distinzioni tra i diversi risultati<sup>21</sup>.

In pressoché tutti questi metodi, i dati disponibili (*input*) sono trattati (*process*) in modo tale da elaborare automaticamente un algoritmo che restituisce un certo risultato (*output*), variamente usato. Questo meccanismo (*Input, Process, Output*, c.d. *IPO model*) si sviluppa in modo diverso: nel *supervised learning*, ad esempio, il *software* è fornito di *input* e di *output*, mentre in quello *unsupervised* dei soli dati di *input*. Ciò che qui rileva, però, è la comprensione del *machine learning* come di un programma (*software*) che, a partire dai dati disponibili (*input*), elabora automaticamente (*process*) un algoritmo (modello, *pattern*, regola) la cui esecuzione, anche essa automatica, restituisce un certo dato-risultato (*output*). L'elaborazione automatica (apprendimento) dell'algoritmo la cui esecuzione restituisce un certo dato-risultato (*output*) è *data-based-inference* e risponde ai metodi della statistica (che, come visto, passano dalla semplice regressione lineare alle complesse reti neurali).

---

<sup>21</sup> Esempi impressionanti si rinvencono nel campo della generazione di immagini (c.d. *deep fake*): per tutti, si v. T. KARRAS, T. AILA, S. LAINE, J. LEHTINEN, *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, in ICLR, 2018, (disponibile in <https://arxiv.org/abs/1710.10196>).

## 6.6. Il software di machine learning e il diritto: dalla personalità elettronica alle presunzioni

Nella elaborazione automatica di una regola di soluzione (algoritmo) la cui esecuzione automatica restituisce un *output* risiede l'elemento di forza dei *software* di intelligenza artificiale, ossia quella tecnologia che consente il processo automatico (e quindi in modalità inconcepibili per l'uomo per ciò che riguarda la velocità e la precisione) di certe operazioni anche senza disporre in partenza delle istruzioni necessarie (algoritmo) per lo svolgimento automatizzato del compito. Si è tante volte osservato che l'impiego di tale specifica capacità dei *software* di *machine learning* presenti dei rischi, non indifferenti, per la società e ciò, in particolare, allorché di essi se ne fa utilizzo predittivo rispetto a questioni sensibili<sup>22</sup>. Orbene, il necessario inquadramento giuridico di questa specifica capacità del *software* di intelligenza artificiale di elaborare automaticamente l'algoritmo (*pattern recognition*) da applicare al problema non è agevole.

Un primo orientamento ha spinto la discussione verso i concetti della «autonomia» e della «personalità elettronica»<sup>23</sup>. Addirittura, qualche anno fa il Parlamento dell'Unione Europea ha approvato una

---

<sup>22</sup> Letture già divenute classiche: E. PARISER, *The Filter Bubble: What The Internet Is Hiding From You*, New York, 2011; R. MACKINNON, *Consent of the networked: the worldwide struggle for Internet freedom*, New York, 2013; C. O'NEIL, *Weapons of math destruction: How big data increases inequality and threatens democracy*, New York, 2016; S. BAROCAS e A.D. SELBST, *Big Data's Disparate Impact*, in *Calif. Law Rev.*, 2016, p. 671; V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York, 2018; S. LEONELLI, *op. cit.*, p. 43; S.U. NOBLE, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York, 2018; J. BRIDLE, *Nuova era oscura* (trad. it.), Roma, 2019; S. ZUBOFF, *op. cit.*; T. GEBRU, *Race and Gender*, in M.D. DUBBER, F. PASQUALE, S. DAS (eds), *Oxford Handbook of Ethics of AI*, Oxford, 2020, p. 253; K. CRAWFORD, *Né intelligente né artificiale. Il lato oscuro dell'IA* (trad. it.), Bologna, 2021; T. NUMERICO, *Big Data e algoritmi*, cit., p. 91 e p. 217.

<sup>23</sup> Si veda, in particolare, C. LEROUX – R. LABRUTO (eds), *Suggestion for a green paper on legal issues in robotics*, 2012, disponibile in [www.unipv-lawtech.eu](http://www.unipv-lawtech.eu). Già all'inizio degli anni Novanta del secolo scorso, in un contesto in cui la maggioranza delle macchine era priva di connessione a internet, si era avanzata l'idea di attribuire di diritti e doveri a tali *software* e considerarli come persone (L.B. SOLUM, *Legal Personhood for Artificial Intelligences*, in *North Carol. L. Rev.*, 1992, p. 1231). In tempi recenti, S. BAYERN, *The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems*, in *Stan. Tech. Law Rev.*, 2015, p. 93 afferma la possibilità di riconoscere "effective legal personhood for nonhuman systems without wide-ranging legal reform and without resolving, as a precondition, any philosophical questions concerning the mind, personhood, or capabilities of nonhuman systems".

risoluzione in cui invitava a valutare l'ipotesi dell'istituzione di «uno *status* giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi»<sup>24</sup>. Sulla falsa riga di uno studio di poco precedente<sup>25</sup>, con tale formula (“*electronic personhood*”) quella risoluzione intendeva così affrontare e risolvere il rischio di c.d. *liability gap*, ossia di carenze in materia di responsabilità rispetto a robot autonomi, i quali non potrebbero essere “considerati come meri strumenti nelle mani di altri”, in quanto dotati della “capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un’influenza esterna”<sup>26</sup>. Se, da un lato, tale capacità – definita come “autonomia” – dei robot li renderebbe “simili ad agenti che interagiscono con l’ambiente circostante e sono in grado di alterarlo in modo significati”<sup>27</sup>, da altro, la stessa renderebbe le norme tradizionali non sufficienti “per attivare la responsabilità per i danni causati da un robot, in quanto non consentirebbero di determinare qual è il soggetto cui incombe la responsabilità del risarcimento né di esigere da tale soggetto la riparazione dei danni causati”<sup>28</sup>.

---

<sup>24</sup> Art. 59, lett. F, *Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica*, (2015/2103(INL)), 17 febbraio 2017, in *eur-lex.europa.eu*. Su quella risoluzione si vedano i commenti di G. TADDEI ELMI e F. ROMANO, *Il robot tra ius condendum e ius conditum*, in *Inf. Dir.*, 2016, 115; A. ZORNOZA e M. LAUKYTE, *Robotica e diritto: riflessioni critiche sull’ultima iniziativa di regolamentazione in Europa*, in *Contr. impr. Eur.*, 2016, p. 808; N. BUSTO, *La personalità elettronica dei robot: logiche di gestione del rischio tra trasparenza e fiducia*, in *Cyberspazio e dir.*, 2017, 499; G. PASSAGNOLI, *Regolamento giuridico e tutele nell’intelligenza artificiale*, in *Pers. merc.*, 2019, p. 79; G. DI ROSA, *Quali regole per i sistemi automatizzati “intelligenti”?*, in *Riv. dir. civ.*, 2021, p. 823.

<sup>25</sup> C. LEROUX – R. LABRUTO (eds), *Suggestion for a green paper*, cit., p. 60.

<sup>26</sup> Cons. AB, AA, *Risoluzione del Parlamento europeo del 16 febbraio 2017*, cit. Nella letteratura, condividono la preoccupazione per cui “le regole ordinarie di responsabilità sono insufficienti e che sono necessarie nuove norme per aiutare a determinare cosa succede ora che gli strumenti possono produrre danni a causa delle loro “cattive” decisioni”, A. ZORNOZA e M. LAUKYTE, *op. cit.*, p. 810; in senso opposto, al termine di una rilettura delle norme codicistiche, nega l’esistenza di un tale rischio di *liability gap*, U. RUFFOLO, *Intelligenza Artificiale, machine learning e responsabilità da algoritmo*, in *Giur. it.*, 2019, p. 1689.

<sup>27</sup> Cons. Z, *Risoluzione del Parlamento europeo del 16 febbraio 2017*, cit.

<sup>28</sup> Cons. AF, *Risoluzione del Parlamento europeo del 16 febbraio 2017*, cit.

Orbene, ferma la tradizionale distinzione nel termine persona<sup>29</sup> dell'essere umano (persona fisica) e dell'ente creato artificialmente per uno scopo giuridico (persona giuridica)<sup>30</sup> e irrevocata in dubbio la legittimità del legislatore di procedere in tal senso per risolvere problemi di responsabilità, va osservato come la proposta soluzione a questa presunta carenza – i.e. il riconoscimento della personalità elettronica – sia stata fortemente criticata fino ad essere completamente abbandonata dalle istituzioni europee<sup>31</sup>.

A ben vedere, essa poggia su di un errore di valutazione della specifica capacità del *software* di intelligenza artificiale di elaborazione automatica dell'algoritmo la cui esecuzione automatica restituisce un certo *output*. In quell'automatismo, come visto, si è letto una manifestazione di "autonomia"<sup>32</sup>. In effetti, leggere una forma di coscienza

<sup>29</sup> *Ex multis*, J. CHIPMAN GRAY, *The Nature and Sources of the Law* (1909), London, 2020, p. 19; F. FERRARA, *Le persone giuridiche*, *Tratt. Vassalli*, II, 2°, 1958, p. 7.

<sup>30</sup> Di recente, si è soffermato sull'influenza del pensiero di Savigny – cui si deve la definizione di "juristische Person" come, appunto, di "ein des Vermögens fähiges künstlich angenommenes Subjekt" – nel diritto italiano ed europeo, F. RANIERI, *L'invenzione della persona giuridica. Un capitolo nella storia del diritto dell'Europa continentale*, Milano, 2021.

<sup>31</sup> Si veda la lettera con cui centinaia di scienziati hanno criticato la proposta di istituire la personalità elettronica: l'*Open Letter to the European Commission Artificial Intelligence and Robotics*, in <http://www.robotics-openletter.eu/>. Nel senso della posizione del Parlamento europeo militano A. ZORNOZA e M. LAUKYTE, *op. cit.*, p. 8; invece, per una chiara critica alla tesi della personalità elettronica, si veda E. PALMERINI, *Robotica e diritto: suggerimenti, intersezioni, sviluppi a margine di una ricerca europea*, in *Resp. civ. prev.*, 2016, p. 1837, la quale bene evidenzia il carattere fuorviante, prematuro e inadeguato.

<sup>32</sup> A ben vedere, il ricorso ai concetti di autonomia, di *software* cosciente e di personalità elettronica si spiega più con la plurisecolare fantasia della macchina intelligente (in argomento, G. WOOD, *Edison's Eve. A Magical History of the Quest for Mechanical Life*, New York, 2002, nonché A. PUNZI, *L'ordine giuridico delle macchine. La Mettrie Helvétius D'Holbach. L'uomo-macchina verso l'intelligenza collettiva*, Torino, 2003) che con il dato reale. A conferma di questa influenza culturale è la circostanza per cui ogni articolo giuridico che discute la disciplina dell'intelligenza artificiale riprende una qualche immagine che, nel corso della storia, l'uomo si è dato a proposito della macchina animata. Talvolta, si cita l'etimologia del termine 'robot', il cui esordio si vuole risalente al 1923, quale traduzione del vocabolo ceco 'robotnik' (lavoratore forzato), impiegato dallo scrittore Karel Capek nel suo dramma fantascientifico *Rossum's Universal Robots*, talaltra si richiama il mito di Pigmalione o il personaggio di *Frankenstein*. Ancora più diffuso è il richiamo alle tre leggi di Asimov, tratte da *Runaround* (1942), e la cui formulazione originaria così si sviluppa: «A robot may not injure a human being or, through inaction, allow a human being to come to harm; A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law; A robot must protect its own existence as long as such protection does not conflict with the First or Second Law». Tale tributo artistico-letterario, peraltro, non è una esclusiva della dottrina

(necessaria per darsi le regole) nella automatica elaborazione dell'algoritmo appare immotivato. Al più, tale automatismo, come è stato esattamente notato, integra la figura leibniziana dei pensieri ciechi<sup>33</sup>.

Per avere un corretto inquadramento giuridico di questa tecnologia ed evitare fughe non si sa quanto ingenuo verso la personalità elettronica, è sufficiente porre a mente la circostanza per cui il dato-risultato (*output*) ottenuto dall'applicazione dell'algoritmo elaborato (*process*) dai dati disponibili (*input*) non è una regola generale, ma un qualcosa di particolare come, d'altra parte, è particolare anche il dato di partenza (*input*). Tale aspetto (da particolare a particolare) consente di qualificare questi *software* di *machine learning* come un procedimento induttivo o inferenziale e non già come uno di tipo sussuntivo (dal generale al particolare) o per astrazione (dal particolare al generale). In tal senso il *machine learning* richiama la definizione di cui all'art. 2727 c.c. in materia di presunzioni ("le presunzioni sono le conseguenze che la legge o il giudice trae da un fatto noto per risalire a un fatto ignorato").

## 6.7. *Machine learning* e diritto: applicazioni pratiche

Nel campo del diritto, un'applicazione di questa intelligenza artificiale molto diffusa è quella della c.d. giustizia predittiva. Dato conto della capacità dell'intelligenza artificiale quale *software* in grado di produrre risultati elaborando i dati disponibili, non sorprende che in molti Paesi si siano avviati progetti di scrittura di programmi che a partire dai dati (*input*) relativi a procedimenti giudiziari elaborano delle regole che restituiscano risultati (*output*) in merito alla durata del procedimento, all'esito, all'ammontare del risarcimento e via discorrendo. Tali risultati possono essere letti come delle previsioni di un certo risultato processuale (da qui, appunto, l'espressione di giustizia predittiva)<sup>34</sup>.

---

giuridica e la stessa citata risoluzione del Parlamento europeo sulle *Norme di diritto civile sulla robotica* del 16 febbraio 2017 si apre rilevando come «gli esseri umani» abbiano «fantasticato sulla possibilità di costruire macchine intelligenti, spesso androidi con caratteristiche umane».

<sup>33</sup> R. BODEI, *Automatismo del pensiero umano e macchine calcolatrici*, in *Mondo digitale*, 2016, p. 1.

<sup>34</sup> In argomento, v. i saggi raccolti in A. CARLEO (a cura di), *Calcolabilità giuridica*, Bologna, Il Mulino, 2017, nonché A. GARAPON, J. LASSÈGUE, *Justice Digitale. Révolution graphique et rupture anthropologique*, Paris, PUF, 2018; M. LUCIANI, *La decisione*

Tra le tante esperienze fin qui maturate, si può fare riferimento al progetto *DataJust*, avviato dalla Repubblica francese nei primi mesi del 2020 (*Décret* n. 2020-356)<sup>35</sup>. *DataJust* intende raccogliere i dati sui procedimenti giudiziari concernenti il risarcimento del danno conseguente a lesioni fisiche. In particolare, i dati (*input*) forniti al *software* ed estrapolati dalle decisioni di secondo grado rese tra il 2017 e il 2019 sono quelli relativi alla lesione, all'età delle parti, al loro sesso, alla residenza, all'eventuale rapporto familiare, alla vita professionale e finanziaria delle parti, alla lesione, al suo impatto, ai referti medici e alle spese sanitarie sostenute, ai precedenti penali e non. Il nome delle parti non è inviato al ministero. L'obiettivo del Ministero della Giustizia francese perseguito con questa iniziativa è quello di mettere a punto – nel tempo massimo di due anni – un *software* in grado di fornire dati (*output*) attendibili circa l'ammontare del risarcimento (regressione lineare), così che il programma possa rilevare come fonte di informazione utile per le vittime e come ausilio ai giudici per la quantificazione del danno conseguente a lesioni fisiche. L'attendibilità e l'affidabilità del *software* potrebbe giocare un ruolo, infine, molto importante per ridurre il contenzioso giudiziario, consentendo alle parti di avere una stima precisa dell'ammontare del danno.

Differente è poi l'impiego di un *software* di giustizia c.d. predittiva che restituisce come *output* la previsione sulla decisione di un certo organo giudicante. Ad esempio, uno studio americano di qualche anno fa ha costruito un algoritmo in grado di restituire un risultato coincidente, per circa il 70% delle volte, con le decisioni assunte dalla *Supreme Court* degli Stati Uniti in circa due secoli (1816-2015, si tratta di oltre 28000 casi). Per fare ciò, gli studiosi hanno utilizzato i dati disponibili nella banca dati della *Supreme Court* e individuato oltre duecento variabili<sup>36</sup>. Programmi analoghi sono comparsi un po' ovunque negli

---

giudiziaria robotica, in *Rivista AIC*, 2018, p. 22 (disponibile a: [https://www.rivista-aaic.it/images/rivista/pdf/3\\_2018\\_Luciani.pdf](https://www.rivista-aaic.it/images/rivista/pdf/3_2018_Luciani.pdf)); G. TUZET, *L'algoritmo come pastore del giudice? Diritto, tecnologie, prova scientifica*, in *Riv. dir. media*, 2020, 1, p. 45; E. BATELLI, *Giustizia predittiva, decisione robotica e ruolo del giudice*, in *Giust. Civ.*, 2020, p. 281; A. CARRATTA, *Decisione robotica e valori del processo*, in *Riv. dir. proc.*, 2020, p. 491; G. ZACCARIA, *Figure del giudicare: calcolabilità, precedenti, decisione robotica*, in *Riv. dir. civ.*, 2020, p. 277.

<sup>35</sup> Testo disponibile in <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000041763205>.

<sup>36</sup> DM. KATZ, MJ. BOMMARITO II, J. BLACKMAN, *A general approach for predicting the behavior of the Supreme Court of the United States*, in *PLoS ONE* 12(4), 2017: e0174698. <https://doi.org/10.1371/journal.pone.0174698>.

ultimi anni, sfruttando l'enorme mole di dati di cui si dispone, in particolare con riferimento ai giudizi di ultima istanza.

Questo genere di studi è estremamente interessante perché consente di fare supposizioni su quale elemento sia più rilevante nella decisione. Ad esempio, nel lavoro svolto rispetto alla giurisprudenza della Corte Europea dei Diritti dell'Uomo e in cui la percentuale di previsione corretta si è rivelata essere di circa il 79% si è evidenziata l'importanza del fatto, e della sua descrizione<sup>37</sup>. Anche in Italia vi sono degli studi, seri, in questo senso. Ad esempio, quello portato avanti dall'università di Bologna e di Pavia (*Legal Analytics for Italian LAw*, LAILA) e che intende occuparsi dell'applicazione dei metodi di analisi giuridica ad un vasto e diversificato insieme di informazioni giuridiche (legislazione, giurisprudenza e dati giuridici empirici), per estrarre conoscenze giuridiche, dedurre relazioni non ancora scoperte ed effettuare previsioni guidate dai dati<sup>38</sup>. Di recente, la Scuola Universitaria Superiore IUSS di Pavia e il Centro Elettronico di Documentazione della Suprema Corte di Cassazione hanno dato vita a un programma che, attraverso l'uso degli strumenti della c.d. *legal analytics* (*data science*, intelligenza artificiale, *machine learning*, *natural language processing* e statistica) ha l'obiettivo di «estrarre e rappresentare conoscenza giuridica, rinvenire correlazioni implicite, individuare tendenze ed effettuare previsioni relative agli orientamenti giurisprudenziali e/o legislativi in modo che sia meglio consultabile ed elaborabile in sede di attività giudiziaria e di ricerca scientifica»<sup>39</sup>.

## 6.8. *Machine learning* e il diritto: il principio a non essere sottoposto a una decisione interamente automatizzata

Discorso ancora diverso e, di tutta evidenza, più delicato riguarda invece l'impiego di *software* di giustizia predittiva impiegati per

---

<sup>37</sup> N. ALETRAS, D. TSARAPATSANIS, D. PREOȚIUC-PIETRO, V. LAMPOS, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, in *PeerJ Computer Science*, 2016, 2:e93 <https://doi.org/10.7717/peerj-cs.93>.

<sup>38</sup> Sul Progetto LAILA si veda: <https://dsg.unibo.it/it/ricerca/progetti-di-ricerca/progetti-nazionali-e-di-ateneo/prin2017-laila-legal-analytics-for-italian-law>

<sup>39</sup> Il testo dell'accordo tra la Corte di Cassazione e la Scuola Universitaria Superiore siglato lo scorso 5 ottobre 2021 è disponibile in [https://www.cortedicassazione.it/corte-di-cassazione/it/dettaglio\\_ecs.page?contentId=ECS26105](https://www.cortedicassazione.it/corte-di-cassazione/it/dettaglio_ecs.page?contentId=ECS26105)

decidere se rilasciare un detenuto su cauzione, se concedere la libertà vigilata oppure in quale zona della città inviare un maggiore numero di pattuglie di polizia. A tal proposito è interessante una decisione, divenuta celebre e che non a caso si rinviene in pressoché tutte le trattazioni sull'argomento, emessa da un organo giudicante statunitense. Si tratta della sentenza *State of Wisconsin v. Eric L. Loomis* resa dalla *Supreme Court* del Wisconsin il 13 luglio 2016<sup>40</sup>.

Tale decisione aveva ad oggetto l'appello avverso una decisione di un tribunale locale (La Crosse) che aveva condannato il ricorrente, Eric L. Loomis, alla pena reclusiva di sei anni per una serie di reati connessi alla guida di una macchina rubata e che risultava essere stata impiegata per una sparatoria. In primo grado, l'imputato si era dichiarato colpevole rispetto ad alcune delle accuse mossegli e il Tribunale aveva proceduto a una indagine sul suo *background* per determinare gli anni di reclusione (*Pre-sentence Investigation*). La doglianza riguardava la circostanza per cui il Tribunale di La Crosse avesse utilizzato l'*output* del *software Correctional offender management profiling for alternative sanctions* (COMPAS) nel determinare l'entità della pena da irrogare.

COMPAS è un *software* di *risk assessments and recidivism data at sentencing*. Esso ha come compito quello di indicare il rischio di recidiva di un soggetto e i suoi bisogni, a partire dai dati presenti nel fascicolo, dell'interrogatorio che questi ha svolto e dei dati relativi alle persone con un passato simile. Il *software* assegna il profilo del soggetto da valutare al gruppo di persone più vicino (simile, affine), così che l'*output* – uno score relativo alle diverse tipologie di recidiva, processuale, generale e violante - rilevi come previsione circa la probabilità di sua recidiva e funga da ausilio al giudice (*to provide decisional support*). La *pre-sentence investigation* indicava esplicitamente che l'*output* non dovesse essere considerato come l'unico elemento su cui basare il giudizio relativo all'ammontare degli anni di reclusione. Il Tribunale aveva assunto la sua decisione - sei anni di reclusione – anche sulla base dei risultati del *software*, alla luce dei quali aveva anche negato la concessione della libertà vigilata. Contro questo primo provvedimento del Tribunale, la difesa di Loomis aveva presentato una richiesta di revisione della decisione, contestando l'utilizzo del *software* COMPAS, la sua affidabilità nonché la assenza di trasparenza circa il modo in cui veniva raggiunto l'*output*. Contro il provvedimento che respingeva la

---

<sup>40</sup> *Supreme Court of Wisconsin*, 13.7.2016, (881 N.W.2d 749 (2016)): il testo è disponibile in <http://www.scotusblog.com/case-files/cases/loomis-v-wisconsin/>

richiesta di revisione, Loomis decise di proporre ricorso in appello e nell'ambito di questo procedimento di impugnazione la questione viene deferita alla Corte Suprema.

Chiamata a decidere sull'ammissibilità dell'impiego di un *software* e sul diritto a una *individualized sentence based on the charges and the unique character of the defendant*» e ad una pena *based on accurate information*», la Corte Suprema del Wisconsin ha offerto una decisione estremamente chiara. Innanzitutto, la Corte ha osservato come l'*output* del *software* COMPAS non sia stato l'unico elemento utilizzato dai giudici del Tribunale di La Crosse per decidere l'ammontare della pena reclusiva e la mancata concessione della libertà vigilata. Accanto al dato prodotto dall'algoritmo sulla pericolosità di Loomis, infatti, i giudici avevano richiamato anche altri, dati come la gravità dei reati contestatigli e la condotta pregressa. Per quanto riguarda, l'assenza di trasparenza – COMPAS è un *software* di proprietà di un soggetto privato e il suo algoritmo è coperto da brevetto – la Corte ha ritenuto la conoscenza degli *score* che il soggetto riceve dal *software* e che possono essere contestati un elemento sufficiente. Al contempo, essa ha raccomandato che il programma sia correttamente aggiornato, di modo da ridurre il rischio di distorsioni e di *bias*, alcuni dei quali già emersi (gli imputati neri avevano molte più probabilità di essere erroneamente giudicati ad alto rischio di recidiva di quelli bianchi; allo stesso modo, gli imputati bianchi avevano più probabilità degli imputati neri di essere erroneamente segnalati come a basso rischio). Essa ha, infine, ribadito che il *software* non è pensato per essere determinante nella decisione sulla reclusione o sul suo ammontare e non è da intendersi come alternativo alla discrezionalità del giudice. Al contrario, la *Supreme Court* del Wisconsin ha spiegato come COMPAS rilevi quale strumento per fornire il tribunale di un dato attendibile circa la probabilità con la quale un profilo criminale come quello dell'imputato possa reiterare il reato. In particolare, essa chiarito che mentre «*using a risk assessment tool to determine the length and severity of a sentence is a poor fit*», COMPAS può essere utile a) nell'identificare i detenuti che sono a basso rischio di recidiva ai fini del loro dirottamento verso alternative non carcerarie e aiuta a decidere se sospendere tutta o parte della pena di un detenuto; b) per valutare il rischio per la sicurezza pubblica del reo e rilevare nella decisione se il rischio di recidiva presentato dal reo può essere gestito in modo sicuro ed efficacemente ridotto attraverso la supervisione e i servizi della comunità; c) per orientare le decisioni

sui termini e le condizioni della libertà vigilata e della supervisione (le valutazioni del rischio possono essere utili per identificare i criminali a basso rischio che non richiedono una supervisione intensiva e programmi di trattamento). In particolare, può rilevare nelle decisioni quali i requisiti di segnalazione, i test antidroga, il monitoraggio elettronico, il servizio comunitario e le strategie di trattamento più appropriate.

## 6.9. Il principio a non essere sottoposto a una decisione interamente automatizzata e il GDPR

Il diritto affermato dalla *Supreme Court* del Wisconsin e che afferma il principio a non essere sottoposto a una decisione interamente automatizzata circola e attraversa anche altri spazi da quello, evidentemente delicato, del processo penale statunitense. In questa sede, si discuterà della profilazione, ossia di quell'inferenza con cui dalla raccolta di dati (*input*) e dalla loro analisi automatizzata si elabora (*process*) un modello (appunto, un profilo) la cui applicazione porta all'individuazione di caratteristiche (*output*) di comportamento, di capacità, di interessi, etc<sup>41</sup>. A livello europeo (e, quindi, nazionale), la disciplina rilevante è offerta dal regolamento generale sul trattamento dei dati personali (reg. EU/2016/679, *General Data Protection Regulation*, GDPR), adottato il 27 aprile 2016, entrato in vigore il 24 maggio dello stesso anno, operativo a partire dal 25 maggio 2018.

### 6.9.1. La profilazione

Il GDPR definisce la profilazione, come *i*) una qualsiasi forma di trattamento *ii*) automatizzato, *iii*) effettuato su dati personali *iv*) per valutare determinati aspetti personali<sup>42</sup>.

Il primo elemento rilevante della definizione di profilazione offerta dal GDPR rimanda alla nozione di trattamento, definita come «qualsiasi operazione o insieme di operazioni, compiute con o senza l'ausilio

<sup>41</sup> per la definizione di «profilazione» si v. art. 4, lett. 4) GDPR. Sui problemi sottesi alla profilazione, si v. S. RODOTÀ, *Il diritto di avere diritti*, cit., p. 404.

<sup>42</sup> Con riferimento alla profilazione nell'esperienza interna precedente al GDPR, si v. l'art. 37, lett. d), d.lgs. 30 giugno 2003, n. 196 e, in giurisprudenza, Corte Cass., 8 novembre 2021, n. 32411, in <http://www.italgiure.giustizia.it/sncass/> ("il trattamento con modalità automatizzata di dati personali con il fine di definire il profilo o la personalità dell'interessato, o ad analizzare abitudini o scelte di consumo").

di processi automatizzati e applicate a dati personali o insiemi di dati personali, come la raccolta, la registrazione, l'organizzazione, la strutturazione, la conservazione, l'adattamento o la modifica, l'estrazione, la consultazione, l'uso, la comunicazione mediante trasmissione, diffusione o qualsiasi altra forma di messa a disposizione, il raffronto o l'interconnessione, la limitazione, la cancellazione o la distruzione».

Il secondo elemento rilevante della definizione di profilazione offerta dal GDPR attiene all'essere tale trattamento di dati automatizzato. Come si legge nelle *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (nel prosieguo, *Guidelines WP 251/rev01*), la profilazione «is often used to make predictions about people, using data from various sources to infer something about an individual, based on the qualities of others who appear statistically similar»<sup>43</sup>. Il terzo aspetto rilevante della definizione di profilazione contenuta nel GDPR riguarda invece i dati oggetto del trattamento automatizzato: i dati personali, ossia «qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»)»<sup>44</sup>.

Infine, l'ultimo elemento costitutivo della nozione di profilazione adottata dal GDPR attiene alla circostanza per cui essa deve servire «per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica». Giova osservare che è necessario che ricorrano tutti gli elementi ora indicati: per esempio, il trattamento automatizzato di dati personali per finalità diverse da quelle di valutazione della persona fisica non rientra nella nozione di profilazione offerta dal regolamento. Per aversi una profilazione rilevante ai fini del GDPR, quindi, è necessario che il trattamento automatizzato dei dati personali sia finalizzato alla valutazione di aspetti personali dell'interessato (come accade, appunto, nell'esempio proposto relativo alla concessione del prestito). Una volta compresa la nozione di profilazione adottata dal GDPR occorre passare alla disciplina applicabile.

---

<sup>43</sup> *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* adottate dall'Article 29 Data Protection Working Party il 6 febbraio 2018 (testo disponibile in <https://ec.europa.eu/newsroom/article29/items/612053>).

<sup>44</sup> Per la definizione di «dato personale» si v. art. 4, lett. 1, GDPR.

Questa varia a seconda che il processo decisionale si basi sulla profilazione in modo esclusivo o meno.

### 6.9.2. La disciplina della decisione basata anche sulla profilazione

In questo paragrafo, si prenda l'ipotesi di una domanda di prestito formulata *on-line* sul sito internet di un istituto di credito, il quale prevede che la decisione sulla concessione del credito sia assunta da un proprio dipendente anche sulla base di un profilo elaborato con mezzi automatizzati (*software*). Ciò vuole dire che la decisione è assunta sulla base del trattamento automatizzato (profilazione), ma non unicamente: il dipendente della banca, infatti, decide se concedere il prestito considerando anche (= oltre alla profilazione) altri elementi (colloquio con il richiedente, etc.).

All'ipotesi qui descritta di decisione basata anche sulla profilazione (trattamento automatizzato di dati personali per valutare aspetti personali) si applica la disciplina generale del GDPR e ad essa occorre guardare, in modo sintetico, per comprendere le risposte alle domande tipiche del ragionamento presuntivo (quando si può procedere per presunzioni? quali presunzioni si possono utilizzare? come?).

Mentre il secondo comma dell'art. 2729 c.c. definisce in negativo i presupposti per il ricorso da parte del giudice al procedimento induttivo in materia probatoria, l'art. 6 GDPR individua esplicitamente, in positivo, i presupposti («condizioni» il cui ricorrere rende lecito l'attività altrimenti illecite) per il ricorso alla profilazione. In questo senso, l'art. 6 GDPR risponde alla domanda relativa al quando si può procedere per presunzioni con riferimento alla profilazione.

In particolare, il GDPR richiama il consenso<sup>45</sup> dell'interessato al trattamento<sup>46</sup> e la circostanza che il trattamento *de qua* sia necessario –

<sup>45</sup> Per la definizione di «consenso dell'interessato», si v. art. 4, lett. 11) GDPR; per la disciplina, si v. anche l'art. 7, GDPR. Si v. anche le *Guidelines on Consent under Regulation 2016/679*, adottate dall'Article 29 Data Protection Working Party il 28 novembre 2017, [http://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=48849](http://ec.europa.eu/newsroom/just/document.cfm?doc_id=48849).

<sup>46</sup> Limitando i riferimenti alla bibliografia in lingua italiana, sul consenso nel GDPR si vedano, tra gli altri, i contributi di A.M. GAROFALO, *Regolare l'irregolabile: il consenso al trattamento dei dati nel GDPR*, in S. ORLANDO – G. CAPALDO, *Annuario 2021, Osservatorio Giuridico sull'Innovazione Digitale*, Roma, Sapienza Università Editrice, 2021, p. 119; I.A. CAGGIANO, *Il consenso al trattamento dei dati personali nel nuovo Regolamento europeo. Analisi giuridica e studi comportamentali*, in *Oss. dir. civ. comm.*, 2018, p. 67; F. CAGGIA, *Il consenso al trattamento dei dati personali nel diritto europeo*, in *Riv. dir. comm.*, 2019, I, p.

requisito da intendersi restrittivamente – all’esecuzione del contratto o delle misure precontrattuali<sup>47</sup> oppure; all’adempimento di obblighi legali ai quali è tenuto il titolare del trattamento<sup>48</sup>; alla salvaguardia degli interessi vitali di una persona fisica anche diversa dall’interessato<sup>49</sup>; all’esecuzione di un interesse pubblico affidato al titolare del trattamento<sup>50</sup>; al perseguimento di un legittimo interesse anche di terzi, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell’interessato che richiedono la protezione dei dati personali, in particolare se l’interessato è un minore<sup>51</sup>.

L’art. 9 GDPR, invece, stabilisce un generale divieto di trattare – anche in via automatizzata – dati personali che «rivelino l’origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l’appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all’orientamento sessuale della persona». La medesima disposizione individua i casi in cui tale divieto viene meno<sup>52</sup> e, tra questi, spicca il riferimento all’ipotesi in cui l’interessato ha prestato «il proprio consenso esplicito al trattamento di tali dati personali per una o più finalità specifiche, salvo nei casi in cui il diritto dell’Unione o degli Stati membri dispone che l’interessato non possa revocare il divieto»<sup>53</sup>.

In conclusione, in un procedimento decisionale si può ricorrere alle presunzioni e alle inferenze proprie della profilazione autorizzata quando il titolare del procedimento abbia soddisfatto le «condizioni» (es. consenso dell’interessato) previste dal GDPR.

La profilazione è una attività complessa. A una prima fase di raccolta dei dati, segue una seconda di analisi automatizzata degli stessi e, quindi, una terza fase in cui si applica il modello elaborato (profilo) a una persona fisica così «da analizzare o prevedere aspetti riguardanti

---

405; L. GATT, R. MONTANARI, I. A. CAGGIANO, *Consenso al trattamento dei dati personali e analisi giuridico-comportamentale. Spunti di riflessione sull’effettività della tutela dei dati personali*, in *Pol. dir.*, 2017, p. 363; S. THOBANI, *La libertà del consenso al trattamento dei dati personali e lo sfruttamento economico dei diritti della personalità*, in *Eur. dir. priv.*, 2016, p. 513.

<sup>47</sup> Art. 6, para 1, lett. b) GDPR.

<sup>48</sup> Art. 6, para 1, lett. c) GDPR.

<sup>49</sup> Art. 6, para 1, lett. d) GDPR.

<sup>50</sup> Art. 6, para 1, lett. e) GDPR.

<sup>51</sup> Art. 6, para 1, lett. f) GDPR.

<sup>52</sup> Art. 9, para 2, GDPR.

<sup>53</sup> Art. 9, para 2, lett. a) GDPR.

il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica»<sup>54</sup>. Ciascuna delle tre fasi che compongono la profilazione rilevante ai fini GDPR è orientata dai principi affermati ex art. 5, GDPR i quali, appunto, rispondono alla domanda circa il come si possa utilizzare la profilazione in un procedimento decisionale.

Innanzitutto, si deve qui segnalare il principio della trasparenza del trattamento, che si traduce in una pregnante serie di obblighi informativi per il titolare del trattamento<sup>55</sup>. In particolare, questi dovrà spiegare come funziona la profilazione all'interessato<sup>56</sup>, garantendogli altresì il diritto di accesso, ossia il diritto di ottenere informazioni dettagliate sui dati personali utilizzati per la profilazione<sup>57</sup>.

Un ulteriore principio di grande importanza è quello della c.d. esattezza. Come la possibilità di basare una decisione giudiziale su dei collegamenti che non passano per leggi universali e scientifiche, ma per delle inferenze probabilistiche è ammessa nella misura in cui questi siano «gravi, precise e concordanti»<sup>58</sup>, così per il ricorso alla profilazione in un procedimento decisionale è richiesto che i dati siano «esatti e, se necessario, aggiornati»<sup>59</sup>.

Questo principio non sorprende. Il procedimento induttivo che caratterizza i *software* di intelligenza artificiale – e, tra questi, anche quelli che vengono impiegati per la profilazione – presenta un grosso rischio, relativo alla circostanza per cui se i dati di *input* sono inesatti, viziati, superati o parziali, il profilo (algoritmo) che ne deriva sarà distorto e così la decisione (*output*) che su di esso si basa (c.d. *garbage in, garbage out*, GIGO). Connessi a tale principio di esattezza sono, *inter alia*, il diritto di rettifica<sup>60</sup> e quello di cancellazione<sup>61</sup> i quali si applicano sia ai dati *input* sia a quelli *output*<sup>62</sup>. Il primo, in particolare, comporta la possibilità per l'interessato di ottenere «la rettifica dei dati personali inesatti che lo riguardano senza ingiustificato ritardo», nonché, e tenuto

---

<sup>54</sup> Art. 4, lett. 4) GDPR.

<sup>55</sup> Art. 5, para 1, lett. a), GDPR.

<sup>56</sup> Artt. 13, 14, GDPR.

<sup>57</sup> Art. 15, GDPR; *adde Guidelines WP 251/rev01*, cit., p. 17.

<sup>58</sup> Art. 2729, co. 1, c.c.

<sup>59</sup> Art. 5, para 1, lett. d, GDPR.

<sup>60</sup> Art. 16, GDPR.

<sup>61</sup> Art. 17, GDPR.

<sup>62</sup> *Guidelines WP 251/rev01*, cit., p. 17-18.

conto delle finalità del trattamento, «l'integrazione dei dati personali incompleti, anche fornendo una dichiarazione integrativa»<sup>63</sup>.

In conclusione, rispetto ai procedimenti decisionali basati sulle inferenze prodotte dalla profilazione, il GDPR individua i requisiti che rendono lecito tale ricorso e le modalità di utilizzo, stabilendo doveri di informazioni *ex ante* e obblighi di rettifica a cancellazione.

### **6.9.3. La disciplina della decisione basata unicamente sulla profilazione**

Nell'esempio discusso nel paragrafo precedente, la decisione (sulla concessione del credito) è assunta (da un dipendente della banca) anche sulla base di un profilo elaborato con mezzi automatizzati (*software*). Come noto, il GDPR contiene una normativa con riguardo a una distinta ipotesi, in cui la decisione è basata «unicamente» sulla profilazione (trattamento automatizzato di dati personali per valutare aspetti personali). Tale seconda disciplina – che si rinviene nell'art. 22 GDPR – va ad aggiungersi a quella esaminata nel paragrafo precedente.

Per comprendere bene l'ipotesi in cui la decisione è basata «unicamente» sulla profilazione, si immagini che stavolta rispetto alla richiesta di prestito formulata *on-line* l'istituto di credito preveda una procedura in cui la decisione sulla concessione del mutuo si fonda esclusivamente all'*output* elaborato da un *software* di intelligenza artificiale senza che vi sia spazio per altri elementi. Se nell'esempio precedente, l'*output* del *software* era valutato da un dipendente (persona fisica) della banca insieme ad altri fattori (es. colloquio), qui, invece, esso è il solo elemento su cui l'istituto di credito basa la sua decisione.

Nel rispondere alle tipiche domande del procedimento presuntivo (quando si può procedere per presunzioni? quali presunzioni si possono utilizzare? come?), la disciplina dettata a tale riguardo dall'art. 22 GDPR riveste un grande interesse teorico, giacché afferma due principi – quello della non esclusività e quello della conoscibilità – di sicura rilevanza per i *software* di intelligenza artificiale.

Il primo comma dell'art. 22 afferma il c.d. principio di non esclusività della decisione algoritmica, ossia il diritto di non essere sottoposto a una decisione che sia basata unicamente su un trattamento automatizzato (come appunto la profilazione) e che produca effetti giuridici che lo riguardano o incida in modo significativo sulla sua persona,

---

<sup>63</sup> Art. 16, GDPR.

quali il rifiuto automatico di una domanda di credito *online* o di pratiche di assunzione elettronica senza interventi umani<sup>64</sup>. In particolare, la norma prevede che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona»<sup>65</sup>.

Le già citate *Guidelines WP 251/rev01* hanno chiarito i requisiti per il sorgere del diritto, ossia ciò che si deve intendere per decisione basata unicamente su un trattamento automatizzato e per decisione che produca effetti giuridici che riguardano l'interessato o che incida in modo analogo significativamente sulla sua persona. Il primo elemento concerne l'assenza di un fattore su cui basare la decisione diverso dal trattamento automatizzato dei dati<sup>66</sup>. Il secondo requisito, invece, va scomposto in due sub-elementi. Per un verso, rileva la idoneità della decisione assunta sulla base del trattamento interamente automatizzata a produrre effetti giuridici, il che vuole dire, ad esempio, la cancellazione di un contratto, la negazione della cittadinanza e così via. Per altro verso, per essere significativa sulla persona, la decisione, come si legge nelle *Guidelines WP 251/rev01*, deve essere in grado di incidere sulle scelte dell'interessato e sul suo comportamento, come può essere una decisione sull'accesso al credito, ai servizi sanitari o all'università<sup>67</sup>.

Le medesime *Guidelines WP 251/rev01* chiariscono altresì che il termine «diritto» non deve essere inteso immaginando che la non esclusività si applichi solo a seguito di un qualche esercizio del titolare. Ciò sta a significare che l'articolo 22, para. 1, GDPR stabilisce un divieto generale, vigente a prescindere dall'attività dell'interessato<sup>68</sup>. *Mutatis mutandis*, esso va inteso alla stregua del secondo comma dell'art. 2729 c.c. (a mente del quale, «le presunzioni non si possono ammettere nei casi in cui la legge esclude la prova per testimoni») e ciò in quanto anche l'art. 22 para. 1, GDPR indica quando il ricorrere a delle

---

<sup>64</sup> Cons. 71, GDPR.

<sup>65</sup> Art. 22, para. 1, GDPR.

<sup>66</sup> *Guidelines WP 251/rev01*, cit., p. 20-21. Nella letteratura, sull'art. 22 para 1. si veda in particolare G. MALGIERI-G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *Intern. Data Priv. Law*, 2017, p. 251.

<sup>67</sup> *Guidelines WP 251/rev01*, cit., p. 21-22.

<sup>68</sup> *Guidelines WP 251/rev01*, cit., p. 20.

presunzioni (=profilazione) non è possibile (esclusività del trattamento automatizzato come fattore di decisione e rilevanza sul piano personale della decisione).

Si è prima osservato come il divieto di decisioni su base interamente automatizzata e significative per l'interessato sia analogo al divieto di cui al secondo comma dell'art. 2729 c.c. con riferimento alle presunzioni semplici in materia probatoria. In entrambi i casi si impedisce il ricorso a un meccanismo presuntivo. A differenza del divieto di ricorso a presunzioni semplici nei casi in cui la legge esclude la prova per testimoni (art. 2729, secondo comma c.c.), il divieto ex art. 22 GDPR para 1 ammette delle deroghe. Il paragrafo successivo, infatti, stabilisce le ipotesi in cui venga meno il diritto a non essere sottoposto a una decisione interamente automatizzata e significativa<sup>69</sup>.

Questo divieto non si applica allorché la decisione basata esclusivamente su di un trattamento automatizzato (come, appunto, la profilazione) e significativa a livello personale *a*) sia necessaria per la conclusione o per l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; *b*) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; *c*) si basi sul consenso esplicito dell'interessato. Nelle eccezioni di cui all'art. 22, para 2, sub *a*) e *c*) non rientra il processo decisionale interamente automatizzato dei dati personali appartenenti a categorie particolari<sup>70</sup>, il quale non è ammesso salvo che ricorra l'esplicito consenso dell'interessato, l'interesse pubblico al trattamento e che comunque siano in vigore misure adeguate alla tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato<sup>71</sup>.

Per quanto sia affermato un generale divieto a decisioni interamente automatizzate che producono effetti giuridici significativi sulla persona<sup>72</sup>, l'enorme estensione delle deroghe<sup>73</sup> rende di grande importanza la disciplina prevista a proposito dell'utilizzo di queste presunzioni. Il GDPR, infatti, stabilisce un principio fondamentale rispetto a

---

<sup>69</sup> Art. 22, para. 2, GDPR. Per una discussione delle misure adeguate che accompagnano l'utilizzo di decisioni interamente automatizzate, si v. E. FALLETTI, *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, in *Dir. informaz. informat.*, 2020, p. 169.

<sup>70</sup> Art. 9, para 1, GDPR.

<sup>71</sup> Art. 22, para. 4, GDPR.

<sup>72</sup> Art. 22, para. 1, GDPR.

<sup>73</sup> Art. 22, para. 2, GDPR.

due delle ipotesi, ossia quelle ex art. 22, para 2, lett. a) e c) (*i.e.* contratto e consenso), in cui è ammesso il ricorso a decisioni interamente automatizzate che producono effetti giuridici significativi sulla persona.

Nei casi in cui tale processo decisionale è così ammesso, il titolare deve attuare misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, garantendogli, almeno, il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione<sup>74</sup>. Se il diritto di ottenere l'intervento umano sta a significare la possibilità di ripristinare il principio della non esclusività della decisione automatizzata (e quindi il diritto di cui all'art. 22, para. 1, GDPR), il diritto di esprimere la propria opinione e quello di contestare la decisione sono alla base del dibattito sul c.d. diritto alla spiegazione (*explainable artificial intelligence*, XAI) che, almeno secondo certi Autori, ne costituisce, per così dire, il presupposto logico e necessario per l'esercizio<sup>75</sup>.

Ancorché l'esistenza di questo diritto alla spiegazione non sia riconosciuto da tutti gli Autori, le citate *Guidelines WP 251/rev01* dove si legge che l'art. 22, para. 3, GDPR riconoscono in capo al titolare del trattamento un obbligo di «*to provide meaningful information about the logic involved, not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm*»<sup>76</sup>. L'informazione da fornire all'interessato *ex post* dovrà essere resa per iscritto, in forma concisa, trasparente, intelligibile e facilmente accessibile, con un linguaggio semplice e chiaro e in via tempestiva (ex art. 12, GDPR).

---

<sup>74</sup> Art. 22, para. 3, GDPR.

<sup>75</sup> Sul rapporto tra art. 22, para. 3, GDPR e il diritto alla spiegazione va osservato che la dottrina è divisa. A fronte di chi ritiene che la norma ex art. 22, para. 3, GDPR si sostanzia sia nel diritto a comprendere come la tecnologia funzioni sia nel definire chi debba dar conto per come essa funziona (in tal senso, *inter alia*, R. MESSINETTI, *La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata*, in *Contr. impr.*, 2019, p. 861; U. PAGALLO, *Algoritmi e conoscibilità*, cit., p. 96), altri negano tale diritto, osservando che dalla norma in esame si possa ricavare unicamente un dovere di rendere informazioni in merito alle generali funzionalità del *software* (così, S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, p. 76). Una terza posizione, infine, afferma il c.d. diritto alla leggibilità, con ciò intendendo la pretesa del singolo a poter comprendere in via autonoma il processo decisionale automatizzato (G. MALGIERI-G. COMANDE, *Why a Right to Legibility of Automated Decision-Making Exists*, cit., p. 256).

<sup>76</sup> *Guidelines WP 251/rev01*, cit., p. 25.

In sintesi, l'art. 22, terzo comma GDPR offre una risposta alla domanda sull'utilizzo (eccezionale) dell'*output* generato dal trattamento personalmente significativo e interamente automatizzato di dati personali nel senso di prevedere un insieme di diritti per l'interessato e obblighi per il responsabile del trattamento che consentono il rispetto dell'obiettivo fondamentale del GDPR, ossia il controllo sui propri dati personali<sup>77</sup>. Tra questi diritti/obblighi spicca quello, di natura informativa, teso a rendere disponibili, *ex post* e secondo una modalità «sufficientemente comprensiva», i dati che spiegano "*the logic involved*".

## 6.10. Conclusioni

Anche rispetto alla rivoluzione tecnologica in essere il diritto civile assolve il suo compito di regolare la convivenza secondo giustizia e non forza<sup>78</sup>, così da "indirizzare l'intelligenza artificiale verso il bene degli individui e della società"<sup>79</sup>. Tale sforzo passa per l'affermazione di principi nuovi, che, come visto, si affermano nella società, quindi nelle aule dei tribunali e infine circolano tra ordinamenti e sistemi diversi. Di questo sforzo, la ricerca scientifica deve essere partecipe: "*There are all too many people who, in some great period of social change, fail to achieve the new mental outlooks that the new situation demands. There is nothing more tragic than to sleep through a revolution*"<sup>80</sup>.

La strada è lunga.

---

<sup>77</sup> Cons. 71, GDPR.

<sup>78</sup> R. NICOLÒ, *Diritto civile*, in *Enc. dir.*, XII, Milano, Giuffrè, 1964, p. 904.

<sup>79</sup> G. SARTOR, *Introduzione*, in *Riv. fil. dir.*, 2020, p. 69.

<sup>80</sup> MARTIN LUTHER KING, *Commencement Address for Oberlin College*, 1965 (testo disponibile in <https://www2.oberlin.edu>).