

# Il diritto di comprendere, il dovere di spiegare.

## *Explainability* e intelligenza artificiale costituzionalmente orientata

Elisa Spiller\*

THE RIGHT TO UNDERSTAND, THE DUTY TO EXPLAIN. EXPLAINABILITY AND CONSTITUTIONAL ORIENTED ARTIFICIAL INTELLIGENCE

ABSTRACT: In the last years, explainable algorithms and AI raise significant interest (and concerns) in the legal debate. However, adequate recognition of this elusive right to explanation remains ambiguous and uncertain, mainly because of the current legal framework's limits. This paper addresses the problem from the perspective of constitutional law, examining if and how the so-called law in action can consolidate rights and duties to explainable and comprehensible algorithms. Focusing on EU law, the paper considers the achievements gained in personal data protection, focusing on the innovations introduced by art. 22, reg. EU 679/2016 (GDPR). In light of the shortcomings that emerged in this field, the reflection thus recalls the constitutional *rationale* of data protection. It, therefore, examines how national case-law – using the same method – fosters a more solid recognition of an effective right to understand and a duty to explain algorithms.

KEYWORDS: Algorithmic Decision-Making; right to explanation; Explainable AI (XAI); AI society; Constitutional rights and AI

SOMMARIO: 1. Considerazioni introduttive – 2. Interoperabilità terminologica tra tecnologia e diritto. Alcune brevi premesse – 3. *Explainability* e data protection: gli approdi attorno all'art. 22 GDPR – 4. Fondamenti costituzionali per l'*explainability* nella tradizione europea – 5. *Explainability* e IA costituzionalmente orientata: la via italiana e le conferme europee – 6. Conclusioni.

### 1. Considerazioni introduttive

La diffusione dei *big data* e dell'intelligenza artificiale (IA) sta portando ad una progressiva automatizzazione della società. Si tratta di una trasformazione che interessa i settori più vari<sup>1</sup> incluso il mondo del diritto, in cui da più parti si tenta di orientare i futuri sviluppi di questa nuova rivoluzione tecnologica.

\* Dottoressa di ricerca di in Giurisprudenza (diritto costituzionale) e docente a contratto di Law & Data presso l'Università degli Studi di Padova. Mail: [elisa.spiller@unipd.it](mailto:elisa.spiller@unipd.it). Contributo sottoposto a doppio referaggio anonimo.

<sup>1</sup> L'espressione si ispira a *Automating society. Taking Stock of Automated Decision-Making in the EU*, a cura di Bertelsmann Stiftung e Algorithm Watch, ed. 2019, (disponibile a questo link: <https://automatingsociety.algorithmwatch.org>, ultimo accesso: 18.05.2021).



In particolare, se osservata dal punto di vista del diritto costituzionale, questa congiuntura desta non pochi interrogativi. Si assiste non solo ad una redistribuzione del potere ma ad una sua radicale riconfigurazione, sempre più incentrata sul paradigma *data-driven*. L'utilizzo di tecnologie "intelligenti" implica una delega di responsabilità, dall'uomo alla macchina, che implicitamente porta con sé una progressiva de-umanizzazione delle persone coinvolte. Quel che si osserva è dunque un autentico mutamento strutturale dell'interazione, dovuto essenzialmente al fatto che la tecnica da *protesica* (cioè volta a compensare un deficit) diventa sempre più *mimetica* (imitante) fino a sostituirsi o a sovrastare la componente umana<sup>2</sup>.

Il presente contributo si concentra sull'impatto di questa trasformazione sul fronte delle garanzie e dei diritti fondamentali, soffermandosi sul controverso tema dell'*explainability*: la c.d. "spiegabilità" degli algoritmi e dell'IA.

Il presupposto di partenza, affrontando l'argomento, è abbastanza immediato. Perché si possa (ri)stabilire un rapporto collaborativo tra intelligenze di diversa natura, non solo è imprescindibile confidare nell'affidabilità delle macchine. È inoltre necessario che si possa raggiungere, almeno in ultima istanza, un'adeguata comprensione (anche almeno relativa) delle logiche di funzionamento (e ciò, possibilmente, in ogni singolo caso).

Da un punto di vista operativo, tuttavia, ci si trova di fronte ad una sfida epocale. In questo nuovo contesto non solo si rinvengono forti interessi in contrasto con principi democratici di trasparenza e partecipazione<sup>3</sup>. A ciò si aggiunge l'esigenza di dover fare i conti anche con i condizionamenti tecnici che si possono interporre ad una migliore comprensibilità dei sistemi. E se il diritto può offrire soluzioni variabili a seconda delle peculiarità dell'ordinamento in esame, il dato tecnico invece presenta dei tratti di universalità. Diventa perciò fondamentale comprendere come il *design* tecnologico influisca sulla portata dei principi e sull'efficacia dei rimedi offerti dalla normativa, e come la perimetrazione giuridica della programmazione *ex ante* possa migliorare gli standards di tutela e i livelli di trasparenza. Lo scopo di questo scritto è di esaminare se e come la disciplina vigente fornisca simili criteri di progettazione e il grado di prescrittività di questi ultimi, ponendo l'accento sugli approdi che consentono di consolidare il riconoscimento di un diritto di comprendere e un dovere di spiegare gli algoritmi. Focalizzando l'attenzione sull'ordinamento giuridico europeo, si esaminerà lo stato dell'arte sulla regolamentazione dei processi decisionali automatizzati (ADMs, *automated decision-making systems*). Si partirà dai riferimenti offerti dalla disciplina sulla protezione dei dati personali (reg. UE 679/2016, o GDPR, *General Data Protection Regulation*), ambito in cui da principio si è cominciato ad affrontare le questioni accennate. Individuati i limiti e le lacune che emergono in questo campo, si passerà quindi ad esaminare l'apporto delle prime pronunce che a livello nazionale stanno contribuendo a chiarire la portata delle garanzie e i limiti legati alla spiegabilità degli algoritmi. L'obiettivo dell'itinerario proposto è di estrapolare da questa giurisprudenza d'avanguardia le premesse per un'effettiva affermazione di una garanzia di comprensibilità e trasparenza dei processi decisionali automatizzati, soprattutto nell'ottica di promuovere uno sviluppo costituzionalmente orientato di questi sistemi.

<sup>2</sup> A. SORO, *Democrazia e potere dei dati. Libertà, algoritmi e umanesimo digitale*, Milano, 2019, 21 (enfasi aggiunta).

<sup>3</sup> Per tutti, F. PASQUALE, *The Black Box Society. The Secret Algorithm that Control Money and Information*, Cambridge, 2015.



## 2. Interoperabilità tra tecnologia e diritto. Alcune premesse

Prima di addentrarsi nel vivo del problema giuridico dell'*explainability*, è necessario soffermarsi brevemente su alcune premesse circa l'interoperabilità terminologica tra tecnologia e diritto.

Innanzitutto, è bene precisare come, generalmente, l'espressione "intelligenza artificiale" identifichi i «sistemi che dimostrano un comportamento intelligente, analizzando il proprio ambiente e compiendo azioni con un certo grado di autonomia, per raggiungere obiettivi specifici»<sup>4</sup>. Il campo d'azione di queste tecnologie, ovviamente, si estende ben oltre questa definizione, riconducendo a tale concetto una pluralità di sistemi<sup>5</sup>.

Esaminando però come tali soluzioni abbiano e stiano trovando spazio nel mondo del diritto, non si può fare a meno di notare come l'attenzione spesso converga su un tipo specifico di IA, ossia il *machine learning* (soprattutto nelle forme del *supervised learning*). Per cogliere le peculiarità di questo tipo di algoritmi – seppur in modo assai elementare – conviene pensare alla conformazione e al funzionamento di un neurone (da qui la definizione *reti neurali*). Dalle articolazioni più periferiche della struttura passa un *input* che, se correttamente elaborato, produce una reazione attesa. Nel caso dell'IA, l'*input* è costituito dai contenuti di un certo *dataset*, mentre la struttura dell'algoritmo assume questa forma reticolare che, a seconda dei casi, può essere più o meno complessa. Il sistema viene prima allenato con dei dati-esempio (c.d. *labeled data*) verificando che, forniti certi *input*, il programma arrivi al risultato per cui è progettato. Un operatore, di volta in volta, controlla e corregge eventuali storture, fintanto che il *software* non raggiunga un livello di performance ottimale. A questo punto, il sistema può lavorare in modo autonomo, cioè elaborare altri set di dati (originali rispetto a quelli di rodaggio) orientandosi verso la soluzione che teoricamente gli consente di raggiungere al meglio l'esito per cui è stato programmato.

Il ricorso a simili soluzioni ad integrazione (o in sostituzione) di procedure finora svolte da sistemi più "semplici" – ossia connotati da un minor grado di discrezionalità – certo comporta qualche margine di rischio. Il fatto che un *software* sia in grado di calibrare i propri risultati sulle performance e sui riscontri in precedenza ottenuti comporta infatti che la macchina possa esercitare funzioni più complesse con maggior autonomia e minor supervisione. Queste preoccupazioni, in particolare, valgono soprattutto quando l'esito di tali operazioni sia destinato a produrre conseguenze giuridicamente rilevanti, in grado di incidere sullo status e sulle prerogative dell'interessato. Lo scenario che si appalesa in questi casi tende a considerare la persona coinvolta non tanto come soggetto coinvolto nel procedimento, bensì la vede ridotta quasi ad oggetto del medesimo.

<sup>4</sup> Gruppo indipendente di esperti ad alto livello sull'Intelligenza artificiale istituito dalla Commissione europea nel giugno 2018, *Una definizione di IA: principi, capacità e discipline (definizione elaborata ai fini dei documenti del gruppo)*, aprile 2019, 1.

<sup>5</sup> Basti pensare a quelli agilmente censiti dall'Agenzia per l'Italia Digitale nel *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino*, pubblicato nel marzo 2018, in cui si considerano il *machine learning*, il *deep* e il *reinforce learning*, la *computer vision*, il *natural language process*, la robotica e molti altri (26 ss.). Lo stesso approccio è confermato dalla proposta di Regolamento per l'IA della Commissione UE del 21 aprile 2021 (Commissione europea, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, Bruxelles, 21 aprile 2021, COM(2021) 206 final).



Sulla base della correttezza<sup>6</sup> dei dati con cui è stato addestrato e dei *datasets* sui quali poi è chiamato ad operare, il sistema infatti elabora e lavora su delle rappresentazioni della realtà inedite, ed è in funzione di queste “datizzazioni” che si determina rispetto al suo obiettivo.

Come facilmente intuibile, se la soluzione proposta è affidabile, a tutta prima, non si incontrano particolari criticità. Se però i risultati ottenuti dimostrano delle alterazioni o sollevano qualche dubbio, ecco che si pone il problema di come mettere in discussione e validare gli esiti conseguiti. È a questo punto che si innesta il problema giuridico dell’*explainability*, e così delle garanzie e degli obblighi ricollegabili alla spiegabilità degli algoritmi utilizzati.

Innanzitutto, va detto che inquadrare i processi digitali appena descritti nelle categorie tradizionali del diritto è impresa assai ardua. Spesso infatti non si dispone di definizioni e discipline in grado di cogliere con la debita puntualità le caratteristiche e i problemi che connotano l’uso di questi sistemi. Prendendo in esame come normativa di frontiera il GDPR, ad esempio, i riferimenti più espliciti a questi trattamenti *data intensive* poggiano soprattutto sui concetti di “profilazione”<sup>7</sup> e di “processo decisionale automatizzato”<sup>8</sup>; due nozioni che risultano spesso problematiche se calate nella complessa realtà dell’IA e delle sue diverse applicazioni.

In secondo luogo, si apre il problema rimediabile, ossia quali garanzie – tecniche e giuridiche – debbano presidiare questi processi. È qui si intrecciano due diverse prospettive circa il rapporto tra tecnologia e diritto. È innegabile che la tecnologia abbia acquisito nel tempo una sempre più spiccata attitudine normativa; tant’è che il *design* e il funzionamento di uno strumento incide profondamente sulle possibilità di azione dei soggetti che con questo interagiscono, finanche a produrre importanti conseguenze sui diritti degli interessati. Al contempo, si avverte chiara l’esigenza che le tecnologie siano allineate ai principi cui l’ordinamento si ispira (non solo nell’uso ma anche nella loro logica di funzionamento) caldeggiando così una convergenza verso i fondamenti di diritto ereditati dalla tradizione costituzionale nazionale ed europea<sup>9</sup>.

Perché si verifichi questa confluenza tra principi di diritto e *design* tecnologico, è però necessario individuare puntualmente quali siano i parametri cui la programmazione deve ispirarsi e quali siano le

<sup>6</sup> Qui con parlando di «correttezza» si allude alla qualità dei *dataset* considerati, ossia l’adeguatezza dei medesimi rispetto alle diverse finalità di utilizzo in termini di quantità, variabilità, accuratezza, descrittività, rappresentatività (solo per citare alcuni dei principali criteri di valutazione).

<sup>7</sup> Reg. UE 679/2016, art. 4, n. (4). In particolare, la definizione identifica «qualsiasi forma di trattamento automatizzato di dati personali consistente nell’uso di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica».

<sup>8</sup> Nozione con cui di prassi si tenta di circoscrivere i casi in cui una determinazione che innesca effetti giuridici (e così la valutazione) sia effettuata in tutto o in parte da un algoritmo sulla base dell’elaborazione di dati direttamente forniti dall’utente o indirettamente osservati, derivati o desunti da altri trattamenti (inclusa la profilazione). Gruppo di lavoro Articolo 29, *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, Bruxelles, 6 febbraio 2018, 7 ss. Ulteriori spunti di riflessione circa le definizioni qui esaminate sono trattati anche da O. SASSO SARTI, *Profilazione e trattamenti dei dati personali*, in L. CALIFANO, C. COLAPIETRO (a cura di), *Innovazione tecnologica e valore della persona. Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Napoli, 585 ss.

<sup>9</sup> Per ulteriori approfondimenti, si rimanda a K. YEUNG, *Design for the Value of Regulation*, in J. VAN DEN HOVEN, P.E. VERMAAS, I. VAN DE POEL (a cura di), *Handbook of Ethics, Values, and Technological Design, Sources, Theory and Application Domains*, Dordrecht, 2015, 447-472; K. YEUNG, *Towards an understanding of regulation by design*, in R. BROWNSWORD, K. YEUNG (a cura di), *Regulating Technologies*, Oxford, 2008, 79-107.



caratteristiche irrinunciabili (e dunque cogenti) che una certa soluzione deve avere per essere conforme alle aspettative giuridiche. Per potersi individuare l'ipotesi che meglio soddisfa le aspettative sociali, serve chiarire in funzione di quali criteri *normativi* questa possa essere ritenuta più soddisfacente di altre. E la chiarezza con cui possono essere reperite queste indicazioni, da un lato, può facilitare la contestazione dei sistemi non conformi e, dall'altro, può incoraggiare ulteriori sviluppi della ricerca verso gli obiettivi indicati. Tutto ciò induce quindi a verificare se e come la normativa vigente fornisca tali parametri e come questi siano stati recepiti e sviluppati nei diversi contesti.

### 3. *Explainability e data protection: gli approdi attorno all'art. 22 GDPR*

Affrontando il problema della *explainability*, in un primo momento, l'attenzione della dottrina internazionale ed europea si è concentrata soprattutto su quanto previsto dal GDPR e, in particolare, dall'art. 22 di questa normativa.

Tale disposizione ha il primo merito di ribadire l'importanza di un assunto storico, ossia un «diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato compresa la profilazione». La nuova formulazione, peraltro, specifica anche l'opportunità di ricondurre a questa categoria *tutti* i trattamenti e i processi in grado di produrre effetti giuridici che riguardino o incidano in modo significativo nella sfera giuridica dell'interessato, dettagliando così ulteriormente i requisiti per una corretta qualificazione del dato tecnologico<sup>10</sup>.

Sancito questo limite, tuttavia, l'articolo 22 GDPR ammette il ricorso a queste soluzioni in molteplici situazioni<sup>11</sup>, rendendo così la regola quella che formalmente avrebbe dovuto essere soltanto un'eccezione. In questi casi, a tutela dell'interessato, sono previste delle garanzie particolari, prescrivendo al titolare del trattamento l'obbligo di attuare misure appropriate per tutelare i diritti, le libertà e i legittimi interessi del destinatario (fatti salvi alcuni casi in cui è ammesso un regime particolare<sup>12</sup>). Nonostante la complessità di questo intreccio normativo, tra le tutele ivi previste una in particolare ha catturato più delle altre l'attenzione. Si legge infatti che, qualora sottoposto ad un procedimento decisionale automatizzato, l'interessato ha «il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione»<sup>13</sup>. Ed è proprio quest'ultimo enunciato ad aver generato un fervente dibattito circa l'esistenza e la portata di un diritto alla spiegazione degli algoritmi (*c.d. right to explanation*) e sul ruolo che dovrebbe (o potrebbe) toccare agli operatori umani nell'esercizio di questo diritto.

<sup>10</sup> Reg. UE 679/2016, art. 22, § 1.

<sup>11</sup> *Ibidem*, art. 22, § 2, ossia i casi in cui la decisione sia «(a) necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; (b) 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 o (c) si basi sul consenso esplicito dell'interessato».

<sup>12</sup> *Ibidem*, art. 22, § 3.

<sup>13</sup> *Ibidem*, art. 22, § 3. Tali garanzie, peraltro, non sono cogenti nelle ipotesi contemplata al punto (b) del paragrafo precedente (art. 22, § 2), ossia nei casi in cui la decisione 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.



In tal senso, le indicazioni ermeneutiche che precedono il testo normativo (*considerando* 71) prevedono che in ogni caso in cui sia ammesso l'utilizzo di un ADM, tale trattamento debba essere subordinato a garanzie adeguate, che dovrebbero sempre comprendere il dovere di fornire all'interessato una specifica informativa. A questi fini, però, qui si rinviengono due indicazioni distinte: da un lato, il diritto di ottenere l'intervento umano e di esprimere la propria opinione; dall'altro, il diritto *di ottenere una spiegazione della decisione conseguita* dopo tale valutazione e di contestare la decisione<sup>14</sup> medesima. Sebbene l'articolato non si spinga tanto oltre, lette unitamente ai contenuti dell'art. 22 GDPR, tali indicazioni sono state accolte da alcuni<sup>15</sup> come un'esplicita scelta del legislatore europeo di istituire una nuova ed essenziale garanzia di spiegabilità dei procedimenti decisionali automatizzati; un'ipotesi particolarmente apprezzata, soprattutto da quanti sono preoccupati che la diffusione dell'IA possa portare ad un maggior e definitivo determinismo tecnologico a discapito della persona.

Le linee guida europee<sup>16</sup>, tuttavia, si sono dimostrate più caute nell'interpretazione di tali riferimenti. Come si legge, infatti, l'interessato certo gode di un diritto a conoscere l'esistenza di questo tipo di procedimenti e deve poter ottenere «informazioni significative sulla logica utilizzata», così da poter comprendere la *ratio* di tali trattamenti laddove le conseguenze possano essere più severe<sup>17</sup>. Cionondimeno – si sottolinea – all'interno dell'articolato, di fatto, non si rinviene alcun riferimento puntuale all'*explainability* in quanto tale.

Se il regolamento quindi non contempla esplicitamente un diritto ad ottenere una spiegazione, esaminando la portata dei possibili rimedi contro questo tipo di decisioni, è cruciale chiarire in che cosa consistano effettivamente gli oneri posti in capo al titolare del trattamento (ragionando sia *ex ante* che *ex post*).

Va detto che la possibilità di riconoscere un vero e proprio *dovere* di fornire una spiegazione degli algoritmi utilizzati (o, in aggiunta o in alternativa, l'accesso al codice) potrebbe in alcuni casi risultare inconciliabile con altri interessi, primo tra tutti la tutela della proprietà intellettuale. Specificando che il titolare del trattamento ha l'onere di «fornire informazioni *significative* sulla logica del procedimento»<sup>18</sup>, il regolamento però promette comunque di dare piena espansione alle garanzie legate alla possibilità di «esprimere la propria opinione e di contestare la decisione»<sup>19</sup>. Ciò per dire che all'interessato dovrebbero essere fornite delle informazioni sufficientemente complete affinché questi possa comprendere i motivi della risoluzione che gli viene in capo, e così usufruire di rimedi effettivi per potersi eventualmente opporre<sup>20</sup>.

Anche con questi chiarimenti, però, la portata concreta di queste prescrizioni è rimasta per molti versi ambigua. Le diverse soluzioni interpretative nel tempo proposte, infatti – seppur pregevoli – sono

<sup>14</sup> *Ibidem*, *considerando* 71, secondo periodo (corsivo aggiunto dell'Autore).

<sup>15</sup> B. GOODMAN, S. FLAXMAN, *European Union regulations on algorithmic decision-making and a "right to explanation"*, in *AI Magazine*, 38, 3, 2017, 50-57; V. MAYER-SCHÖNBERGER, Y. PADOVA, *Regime Change? Enabling Big Data through Europe's New Data Protection Regulation*, in *Columbia Science and Technology Law Review*, 17, 2, 2016, 188 ss.

<sup>16</sup> *Linee guida sul processo decisionale automatizzato*, cit.

<sup>17</sup> Reg. UE 679/2016, art. 13, § 2, lett. (f) e art. 14, § 2, lett. (g).

<sup>18</sup> *Ibidem*.

<sup>19</sup> *Ibidem*, art. 22, § 3, ultimo periodo.

<sup>20</sup> *Linee guida sul processo decisionale automatizzato*, cit.



andate via via rimodulando i contenuti dell'art. 22 GDPR in termini sempre più modesti, senza peraltro fornire criteri utili circa le tecnologie o il *design* più affini alle garanzie accennate.

Rimanendo confinati dentro il recinto del GDPR, il diritto ad una spiegazione è stato più volte rideclinato in termini assai variegati. Da una "semplice" interpretazione estensiva del diritto di essere informati<sup>21</sup> (sintesi degli obblighi di informativa e del diritto di accesso<sup>22</sup>) si è passati ad un *alias* del diritto ad ottenere informazioni significative sulla logica del procedimento<sup>23</sup>, considerando al contempo le ipotesi di rinvenire in via ermeneutica un diritto alla leggibilità degli algoritmi attraverso procedure di *auditing*<sup>24</sup> o un diritto a sole *reasonable inferences*<sup>25</sup>. Tutto questo, però, è avvenuto senza effettivamente arrivare a sancire un diritto o specifici oneri circa i requisiti minimi di spiegabilità delle tecnologie adottate.

Le lacune messe in luce da queste riflessioni hanno così portato a concludere che la normativa del GDPR possa rappresentare senz'altro un valido punto di partenza per la regolamentazione dell'IA ma certamente non il suo di approdo definitivo<sup>26</sup>. Così come la letteratura più recente suggerisce<sup>27</sup>, si deve anzi cominciare a prendere il largo dall'art. 22 e dal relativo regolamento, guardando invece con maggior favore le soluzioni più attente all'equo bilanciamento della molteplicità di diritti coinvolti negli ADMs. Ed è in questa direzione in cui senz'altro si muove la proposta di regolamento sull'intelligenza artificiale presentata il 21 aprile 2021<sup>28</sup>, di cui ora non rimane che monitorare gli sviluppi e gli esiti.

In attesa di queste evoluzioni del quadro normativo, nei paragrafi che seguono, si andranno ad individuare i fondamenti che, da un punto di vista costituzionale, supportano le pretese di spiegabilità nell'alveo della tradizione giuridica europea, approfondendo poi gli spunti offerti dalla giurisprudenza nazionale più recente per un *design* tecnologico più affine ai principi costituzionali comuni.

#### 4. Fondamenti dell'*explainability* nella tradizione costituzionale europea

Rispetto agli approdi descritti nel paragrafo precedente, conviene tentare un ribaltamento di prospettiva. Il problema dell'*explainability*, infatti, offre l'occasione di apprezzare la disciplina sulla protezione

<sup>21</sup> S. WATCHER, 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*, 7, 2, 2017, 76–99.

<sup>22</sup> Reg. UE 679/2016, art. 15.

<sup>23</sup> A.D. SELBST, J. POWELS, *Meaningful information and the right to explanation*, in *International Data Privacy Law*, 7, 4, 2017, 233–242.

<sup>24</sup> G. MALGIERI, G. COMANDÉ, *Why a Right to Legibility of Automated decision-making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 4, 2017, 243–265.

<sup>25</sup> S. WATCHER, B. MITTELSTADT, *A right to reasonable inferences: re-thinking data protection law in the age of big data and AI*, in *Columbia Business Law Review*, 2, 2019, 494–620.

<sup>26</sup> A questa conclusione s'invita a guardare, in particolare ed *ex multis*, in F. PIZZETTI, *Intelligenza artificiale, protezione dei dati e regolamentazione*, Torino, 2018.

<sup>27</sup> G. GONZÁLEZ FUSTER, *Artificial Intelligence and Law Enforcement. Impact on Fundamental Rights*, Studio richiesto dal Comitato per le libertà civili, la giustizia e gli affari interni per il Parlamento europeo, Bruxelles, luglio 2020, 16-17.

<sup>28</sup> Commissione europea, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, Bruxelles, 21 aprile 2021, COM(2021) 206 final.





dei dati non tanto (e non solo) per i suoi attuali contenuti, quanto piuttosto per gli spunti di metodo che quest'ultima propone.

Tale normativa, infatti, così come il diritto consacrato dall'art. 8 della Carta di Nizza, rappresenta per certi versi un precipitato dei principi comuni alla tradizione costituzionale europea; qui però appositamente rideclinati per fronteggiare le sfide poste dalle tecnologie informatiche. A ciò si aggiunge che la materia in questione storicamente non era stata pensata per tutelare "soltanto" la riservatezza delle informazioni personali. Al contrario, quest'ultima è stata concepita per tutelare una pluralità di interessi, ora ridefiniti in modo inedito nelle nuove dinamiche imposte dalla c.d. società dell'informazione. Da un lato, quindi, vi sono una serie di norme volte a promuovere la libera circolazione e il corretto utilizzo dei dati (definendo così le responsabilità e le modalità con cui questi possono essere lecitamente trattati). Dall'altro, invece, si è andati a tessere una fitta maglia di garanzie e diritti legati all'utilizzo dei contenuti, così da tutelare la posizione dei soggetti interessati e fornire loro degli strumenti per contestare eventuali abusi<sup>29</sup>.

Si possono cogliere importanti parallelismi tra la *ratio* sottesa a questa disciplina e la logica propria del costituzionalismo. Il legislatore, infatti, in questo «crescendo normativo»<sup>30</sup> si concentra su due aspetti tra loro complementari: la definizione di specifici limiti a quanti siano nella posizione di esercitare il potere tecnologico e la predisposizione di particolari garanzie per far valere gli interessi individuali e sociali, secondo una dialettica che ha dato spazio alla creazione di una prima *rule of law* tecnologica<sup>31</sup>. Guardando al passato, la disciplina sui dati può quindi essere letta come il prodotto di una serie di stratificazioni normative che, a partire dagli obblighi relativi alla conformità dei sistemi, gradualmente ha introdotto una serie di nuove garanzie volte a far valere i diritti fondamentali non solo nei confronti delle pubbliche autorità, ma anche dei poteri informativi privati, attraverso speciali profili di responsabilità<sup>32</sup>. Continuando poi ad esaminare la complessa struttura della *data governance* europea con questa chiave di lettura, si coglie come ancor oggi in quest'ambito si avvicendino molte delle sfide che oggi interessano il costituzionalismo post-moderno. In un sistema che vede l'aperta concorrenza tra

<sup>29</sup> Rispetto a questo punto, la dottrina ha analizzato l'impianto di questi sistemi normativi in chiave comparata in molteplici occasioni. Giova peraltro ricordare che, anche all'intero dello stesso panorama europeo si registrano tutt'oggi importanti variazioni nel modo in cui vengono interpretate queste garanzie; diversità essenzialmente dovute alle differenze che caratterizzano la tradizione giuridica e costituzionale di ciascuno Stato. A riguardo, si rimanda alle diverse analisi proposte, *ex multis*, in F.W. HONDIUS, *Emerging Data Protection in Europe*, Amsterdam, 1975; D.H. FLAHERTY, *Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada and the United States*, Chapel Hill, 1989; C.J. BENNETT, *Regulating Privacy. Data Protection and Public Policy in Europe and the United States*, Ithaca-Londra, 1992; G. BUTTARELLI, *Banche dati e tutela della riservatezza. La privacy nella Società dell'Informazione*, Milano, 1997; D. VANNI, *La protezione dei dati personali in prospettiva comparatistica*, Roma, 2012; G. GONZÁLEZ FUSTER, *The emergence of data protection as a fundamental right of EU*, Cham-Heidelberg-New York-Dordrecht-Londra, 2014.

<sup>30</sup> A. SIMONCINI, *The Constitutional dimension of the Internet: some research path*, Working Paper, EUI LAW, 16, 2016, 4.

<sup>31</sup> Quanto alla scelta della chiave di lettura, ci si è rifatti alla prospettiva proposta in A. SIMONCINI, *The Constitutional dimension of the Internet: some research path*, cit. Quanto alle sue rideclinazioni del concetto di *rule of law* in chiave digitale e informativa, sia permesso poi rimandare a E. SPILLER, *La sentenza Tele2 Sverige: verso una digital rule of law europea?*, in *Ianus*, 2017.

<sup>32</sup> V. MAYER-SCHÖNBERGER, *Generational Development of Data Protection in Europe*, in P.E. AGREE, M. ROTENBERG (a cura di), *Technology and Privacy: The New Landscape*, Cambridge (MA), 1997, 219 ss.



attori pubblici e privati, le partite che si giocano nella regolamentazione del potere informativo, infatti, non possono prescindere dalle categorie ereditate dal passato (e così dalla *rule of law*) ma devono al tempo stesso rivolgere le previsioni a tutti i soggetti coinvolti in questo ecosistema, andando così ad assumere tratti sempre più liquidi e trasversali<sup>33</sup>.

È in quest'ottica, dunque, che conviene affrontare le questioni relative alla regolamentazione dell'IA, così come quelle inerenti alla *ratio* e ai confini della protezione dei dati personali. Pertanto, fermi restando gli spunti che si possono trarre dall'interpretazione del GDPR, per accreditare i principi etici che spingono verso lo sviluppo di una IA spiegabile è utile guardare ai passi in avanti fatti dalla giurisprudenza di alcuni Stati membri, ossia al loro tentativo di aggiornare principi costituzionali condivisi in chiave tecnologica.

### 5. Explainability e IA costituzionalmente orientata: la via italiana e le conferme europee

Alla luce di quanto emerso finora, non sorprende che di fronte alle sfide poste dall'IA sia stata proprio una giurisprudenza "costituzionalmente orientata" ad offrire nuovi spunti di riflessione sul tema dell'*explainability*.

Anche a livello nazionale, infatti, i primi tentativi di armonizzazione rispetto alle novità introdotte dal GDPR hanno lasciato non poche zone d'ombra. Le normative statali solo raramente forniscono indicazioni orientative circa le soluzioni tecnologiche da preferire per i processi di decisione automatizzata e, nel dare attuazione al regolamento, la maggior parte degli Stati membri ha adottato per lo più un'impostazione negativa (o comunque abbastanza neutra)<sup>34</sup>.

Ragionando sulle garanzie legate agli ADMs, assumono quindi un valore particolare le decisioni che, pur attingendo da altri ambiti, mirano ad uno sviluppo *constitutional-friendly* di tali soluzioni. Pur considerando tecnologie tra loro diverse, gradualmente, in diversi Paesi europei la giurisprudenza sta

<sup>33</sup> Quanto alle riflessioni che nello specifico riguardano il rapporto tra diritto costituzionale, protezione dei dati e rapporto con i soggetti privati, si rimanda soprattutto a L.A. BYGRAVE, *Data Protection Law*, cit., 125 ss. Quanto, invece, ad alcune delle criticità che il diritto costituzionale incontra nella c.d. postmodernità, si rinvia, *ex multis*, a G. ZAGREBELSKY, *Diritti e Doveri*, in R.E. KOSTORIS (a cura di), *Percorsi giuridici della postmodernità*, Bologna, 2016, 71 ss.; M.R. FERRARESE, *Le istituzioni della globalizzazione*, Bologna, 2000; C. CROUCH, *Postdemocrazie*, Roma, 2015.

<sup>34</sup> G. MALGIERI, *Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislation*, in *Computer Law and Security Review*, 35, 5, 2019, 2 ss. La maggior parte degli Stati membri, infatti, pur avendo provveduto ad introdurre le debite misure di coordinamento dopo l'entrata in vigore del reg. 679/2016, non ha comunque previsto *ex lege* alcuna nuova eccezione oltre a quelle previste a livello europeo (l'Italia, la Romania, la Svezia, la Danimarca, la Polonia, la Finlandia, Cipro, la Grecia, la Repubblica Ceca, l'Estonia, la Lituania, la Bulgaria, la Lettonia, il Portogallo, la Croazia, la Slovacchia, il Lussemburgo, Malta, e la Spagna). Germania e (in parte) Austria hanno invece provveduto all'implementazione dell'art. 22 reg. 679/2016 nella propria legislazione nazionale, ma senza prevedere alcuna specifica e ulteriore garanzia a tutela della persona interessata.

Soltanto alcuni ordinamenti hanno colto l'occasione per rinnovare il proprio sistema di tutele con nuove garanzie. In tal senso, fan scuola le esperienze di Regno Unito, Slovacchia e Irlanda, in cui sono stati introdotti, ad esempio, particolari obblighi informativi (includendo anche accurate descrizioni dei meccanismi decisionali) o alcune forme di valutazione di impatto sugli algoritmi. A questi si aggiungono gli esempi di Francia e Ungheria, dove è stato introdotto un diritto a conoscere il peso che i diversi parametri assumono nei singoli passaggi del processo decisionale.



cominciando a mettere in chiaro nuovi criteri, con l'intento di favorire un miglior accreditamento sociale delle tecnologie più affini ai diritti fondamentali e ai principi di partecipazione e trasparenza. Certo, le pronunce finora disponibili non presentano un grado di omogeneità sufficiente a procedere ad una valida comparazione. Tuttavia, questi pioneristici approdi hanno il merito di documentare come stia evolvendo la *law in action* in materia di *explainability*, dando conto delle tecnologie prese in esame, della loro qualificazione e delle soluzioni ermeneutiche che cominciano a formare una comune coscienza a livello europeo verso una *lawful AI*.

### 5.1. La via italiana...

Assecondando il metodo accennato nelle pagine precedenti, la giurisprudenza italiana, tra le prime, ha avuto l'occasione di cimentarsi in un apprezzamento costituzionalmente conforme dei sistemi utilizzati dal Ministero dell'Istruzione nell'ambito della riforma *La buona scuola* (L. 13 luglio 2015, n. 107) per l'assegnazione delle cattedre.

È bene premettere che qui sarebbe improprio parlare di IA, poiché l'algoritmo utilizzato fondamentalmente mancava dei requisiti di autonomia che invece caratterizzano queste tecnologie. Cionondimeno, esaminando un procedimento automatizzato con le lenti del diritto<sup>35</sup>, i giudici amministrativi hanno avuto la possibilità di far chiarezza sulla portata di due principi essenziali – i principi di accessibilità e non-esclusività – prescrivendo la doverosità di un sindacato tecnico e giurisdizionale sugli algoritmi usati dalla p.a.<sup>36</sup>.

A tal proposito, si è osservato come il fatto di affidare attività automatiche o classificatorie ad una macchina non sia di per sé pregiudizievole, anzi. Dove ciò contribuisca a rendere più efficienti i procedimenti, la scelta appare una doverosa declinazione dei principi che informano il buon andamento della pubblica amministrazione (art. 97 Cost.)<sup>37</sup>. Tuttavia – si è ribadito – laddove si decida di adottare

<sup>35</sup> La metafora relativa a lenti e occhiali si ispira alle parole di Paolo Grossi, così come si ritrovano in Id., *Prima lezione di diritto*, Roma, 2003, *passim*.

<sup>36</sup> I riferimenti sono soprattutto alle pronunce del TAR Lazio, sez. III bis, sentt. n. 3769/2017 e 9224-9230/2018, nonché dalla recentissima decisione del Consiglio di Stato, sez. VI, sent. n. 2270/2019. In tutti e tre i casi, oggetto di censura era quanto previsto dalla legge n. 107/2015 che autorizzò il Ministero dell'Istruzione, dell'Università e della Ricerca ad attuare un piano straordinario di assunzioni a tempo indeterminato di personale docente per le istituzioni scolastiche di ogni ordine e grado. Tale normativa, nello specifico, aveva previsto che l'intera procedura di assegnazione dovesse essere gestita da un sistema informatico che avrebbe provveduto automaticamente all'assegnazione delle sedi, secondo i criteri stabili dalla legislazione vigente. A queste peraltro sono seguite una lunga serie di pronunce, anche di fronte ai giudici ordinari, che in linea con quanto si dirà a breve hanno progressivamente consolidato un orientamento garantista.

<sup>37</sup> Principi sanciti nel cotesto italiano principalmente dall'art. 97 Cost. ed analogamente confermati in sede europea con riferimento al diritto ad una buona amministrazione (art. 42 Carta di Nizza) come constatato, peraltro, in A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in R. CAVALLO PERIN, D.U. GALETTA (a cura di), *Il diritto dell'amministrazione pubblica digitale*, Torino, 2020, 17; Id., *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019, 1163; M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018, 879; C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE Online*, 3, 2020, 3369 ss.; A. PAJNO et al, *AI: profili giuridici Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *Biolaw Journal – Rivista di BioDiritto*, 3, 2019, 205-235; A. CARLEO (a cura di), *Decisione robotica*, Bologna, 2019; Idem (a cura di), *Calcolabilità giuridica*, Bologna, 2017; F. PATRONI GRIFFI, *La decisione robotica e il giudice amministrativo*, in A. CARLEO (a cura di), *Decisione robotica*, cit., 170; A. D'ALOIA, *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019.





tali soluzioni, l'utilizzo di procedure "robotizzate" non può in alcun modo giustificare l'elusione dei principi posti a fondamento del sistema costituzionale e dell'attività amministrativa<sup>38</sup>. L'algoritmo, infatti, in questi casi, può arrivare a gestire in modo automatico parte del procedimento, dando implicitamente corpo ad una «diretta espressione dell'attività svolta dalla pubblica amministrazione»<sup>39</sup>. Trattandosi di un'attività che indubbiamente intercetta un pubblico interesse, è necessario capire quindi come rideclinare in chiave tecnologica non solo il procedimento ma anche le garanzie.

Quanto alla conoscibilità della logica di funzionamento, si è così sancito che «la regola algoritmica deve essere *non solo conoscibile in sé, ma anche soggetta alla piena cognizione*, e al pieno sindacato, del giudice amministrativo»<sup>40</sup>, poiché «solo in questo modo è possibile svolgere, anche in sede giurisdizionale, una valutazione piena della legittimità della decisione»<sup>41</sup> vagliando il rispetto dei principi di imparzialità, pubblicità e trasparenza.

Quanto invece al rapporto uomo-macchina, si è chiarito che le procedure informatiche, anche dove consentano di conseguire un maggior grado di precisione (o addirittura la perfezione) comunque «non possano mai soppiantare, sostituendola davvero appieno, l'attività *cognitiva, acquisitiva e di giudizio che solo un'istruttoria affidata ad un funzionario persona fisica è in grado di svolgere*»<sup>42</sup>, ponendo così un primo limite alle possibili derive di un eccessivo determinismo tecnico<sup>43</sup>.

Esaminando i passaggi appena accennati, subito si percepiscono alcune assonanze tra le conclusioni qui raggiunte e i suggerimenti sottesi alla disciplina europea sulla protezione dei dati personali. A differenza di quest'ultima, però, gli argomenti proposti dalla giurisprudenza in esame muovono da differenti presupposti. Affermando l'illegittimità dell'utilizzo di algoritmi predittivi in vece dell'attività amministrativa a contenuto discrezionale, infatti, si viene a delineare una regola di confine che sembra tracciare una linea di indirizzo per il futuro. Ragionando a contrario, il ricorso al *machine learning*, qualora considerabile, comunque non può prescindere dal dovere di fornire informazioni sulla logica dell'algoritmo e il diritto ad ottenere l'intervento umano, e questo in virtù delle solide coordinate giuridiche e valoriali segnate dal quadro costituzionale di riferimento. Implicitamente, i principi di accessibilità e non-esclusività<sup>44</sup> si accreditano così come effettivi parametri di scelta e sviluppo delle

<sup>38</sup> Cons. di Stato, sez. VI, sent. n. 2270/2019, § 8.1. Una conclusione, questa, che porterebbe ad escludere la possibilità – anche solo in astratto – di ipotizzare l'utilizzo di forme di *unsupervised machine learning* di apprendimento progressivo e di *deep learning* nell'ambito delle attività della pubblica amministrazione (*ibidem*, § 8.2).

<sup>39</sup> TAR Lazio, sez. III bis, sent. n. 3769/2017, § 2.

<sup>40</sup> *Ibidem*, § 8.4. Corsivo aggiunto.

<sup>41</sup> *Ibidem*.

<sup>42</sup> *Ibidem*, § 5.1. Corsivo aggiunto

<sup>43</sup> A tal proposito – continua la riflessione del TAR – si porrebbe «in contrasto ostando alla deleteria prospettiva orwelliana di dismissione delle redini della funzione istruttoria e di abdicazione a quella provvedimentale, il presidio costituito dal baluardo dei valori costituzionali scolpiti negli artt. 3, 24, 97 della Costituzione oltre che all'art. 6 della Convenzione europea dei diritti dell'uomo» (*ibidem*, § 5.1).

<sup>44</sup> Chiaramente, le questioni legate alla tutela della proprietà intellettuale dell'algoritmo, qui sono stati agilmente ovviate. Probabilmente, in altri frangenti – come, ad esempio, nei rapporti tra privati o nel caso in cui gli algoritmi fossero semplicemente in dotazione della p.a. – gli esiti potrebbero essere diversi. Il completo disvelamento delle logiche utilizzate dai sistemi, infatti, potrebbero comportare un importante pregiudizio in capo a quanti hanno investito nel loro sviluppo; ragion per cui sarebbe più opportuno valutare altri rimedi. Cfr. A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e futuro delle libertà*, in *BioLaw Journal – Rivista di Biodiritto*, 1, 2019, 76 ss. confrontando il caso *Loomis v. Wisconsin* con la giurisprudenza qui in esame.



tecnologie in esame, criteri giuridici destinati ad avere un valore dirimente nella scelta delle soluzioni tecnologiche e così in uno sviluppo *constitutional-friendly* della società dell'algorithm.

## 5.2. ... e le conferme europee

Questi pionieristici approdi hanno trovato conferma anche in altre sedi europee, consolidando così le prime fondamenta di quello che, con il tempo, potrebbe costituire un terreno di riflessione comune per l'Unione e gli Stati membri.

In un primo momento, facendo eco a quanto sancito dai colleghi italiani, il *Conseil constitutionnel* francese nel 2018 si è pronunciato su un tema analogo. A fronte di un controllo di legittimità costituzionale preventivo, il collegio, soffermandosi proprio sulla corretta interpretazione delle garanzie ex art. 22 GDPR, ha concordato che, nei casi in cui sia possibile utilizzare sistemi completamente automatizzati, rimane comunque indispensabile poter spiegare in modo comprensibile e dettagliato il procedimento operato dal sistema, così come per le procedure ordinarie. Attingendo dunque dai principi fondativi dell'ordinamento e dalla pregressa tradizione costituzionale in materia di diritti e tecnologie, si è così giunti a dare nuovo compimento alle prerogative proprie della *liberté informatique*<sup>45</sup>.

Questi casi, tuttavia, ancora non avevano affrontato puntualmente le sfide poste dall'IA e dal *machine learning*, trattandosi sempre di tecnologie dotate di un grado di autonomia sostanzialmente ridotto.

Un significativo passo in avanti circa la conformità e la legittimità di un'IA è stata possibile soltanto nel marzo 2020, grazie ad una decisione (ormai già "storica") scritta dai giudici del tribunale distrettuale de L'Aia<sup>46</sup>.

Qui si trattava di vagliare la legittimità (e l'accettabilità) di un sistema *predittivo* per il monitoraggio delle frodi nel settore previdenziale (*Systeem Risico Indicatie*, SyRI). La questione verteva quindi sulla validità di un sistema propriamente "intelligente", ossia composto di algoritmi che non si limitano a svolgere un compito meramente descrittivo-organizzativo, bensì forniscono indicazioni circa la possibilità che un certo fatto avvenga (e non invece il suo effettivo verificarsi)<sup>47</sup>.

Analizzando attentamente il dato tecnologico, si è potuto apprezzare come i sistemi di *deep-learning* e *self-learning* rischiano di essere spesso fraintesi e sopravvalutati. L'algorithm – è stato ribadito – non è in grado di capire ed interpretare la realtà ma di fare delle "semplici" previsioni proponendo un possibile esito per una certa vicenda; il tutto, senza vagliarne il merito. Per far questo, però, il sistema segue logiche complesse; logiche che pongono problemi analoghi a quelli cui si accennava in principio circa il funzionamento delle reti neurali e i rischi legati al c.d. «*garbage in-garbage out*» (ossia il pericolo di risultati distorti a causa di dataset imperfetti, incompleti, o imprecisi).

Vi è stata quindi una prima occasione di constatare di fronte ad un tribunale come in questi frangenti gli operatori umani, pur servendosi di queste rappresentazioni, non siano in grado di comprendere quanto ottengono né di verificare il procedimento eseguito dalla macchina per proporre un certo

<sup>45</sup> Conseil constitutionnel francese, sentenza n. 2018-765 DC del 12 giugno 2018, § 70.

<sup>46</sup> Tribunale distrettuale de L'Aia (*Rechtbank Den Haag*), sentenza n. 1878 del 5 febbraio 2020.

<sup>47</sup> *Ibidem*, § 6.45. Si legge, infatti: «the application of SyRI constitutes a dragnet, untargeted approach in which personal data are collected for investigation purposes. It argues that SyRI is a digital tracking system with which citizens are categorised in risk profiles and in the context of which the State uses 'deep learning' and data mining. According to NJCM et al., SyRI is a proactive system with a large-scale, unstructured and random automated linking of files of large groups of citizens and secret processing of personal data».

risultato. Di conseguenza, si è concluso che il risultato proposto non possa costituire una decisione valida ai fini del diritto, e si è anzi reputato inopportuno l'utilizzo di tali sistemi per simili scopi. Da un lato, infatti, la risoluzione proposta dal programma non è fondata sulle specificità del singolo caso (quanto su correlazioni altamente probabili). Dall'altro, gli incaricati non sono in grado di spiegare e giustificare gli esiti in modo sufficientemente esaustivo, rendendo tali soluzioni inadeguate rispetto alle regole sulla trasparenza<sup>48</sup>.

In virtù dei principi fondamentali che informano l'ordinamento costituzionale olandese, si è stabilito dunque che un sistema che non permetta un adeguato controllo della logica e degli esiti proposti da un'IA non può essere considerato "legittimo" (quanto meno, non *tout court*) poiché imporrebbe un sacrificio sproporzionato degli interessi delle persone coinvolte<sup>49</sup>. Non solo un simile trattamento avrebbe come prima conseguenza una seria compromissione della libertà personale – e così del diritto di autodeterminazione, rispetto all'amministrazione così come nei confronti della macchina – ma potrebbe anche portare ad indebite discriminazioni, consolidando storici motivi di stigma sociale sulla base di possibili *bias* confermativi<sup>50</sup>.

Si osserva quindi come, gradualmente, sulla base dei principi di trasparenza e correttezza su cui poggiano le pretese sancite dall'art. 22 GDPR, la giurisprudenza nazionale abbia cominciato a porre solide premesse, da un lato, per un riconoscimento di un diritto a comprendere gli algoritmi e, dall'altro, di un dovere di preferire algoritmi spiegabili. Simili conclusioni, però, non poggiano direttamente sulla protezione dei dati, bensì sulla tradizione e sui principi che hanno ispirato questa stessa materia, attingendo dall'eredità del diritto costituzionale. Si è così inaugurato un *trend* che conferma la correttezza del metodo, ma spinge ad andare oltre, e il tempo dirà se questi orientamenti potranno trovare conferma o smentita.

## 6. Conclusioni

Come acutamente osservato,

---

<sup>48</sup> *Ibidem*, § 6.46.

<sup>49</sup> *Ibidem*, § 6.87 e 6.89, in cui si legge: «The principle of transparency is the leading main principle of data protection that underlies and is laid down in the Charter and the GDPR (see 6.27- 6.34 for the principles of data protection). The court is of the opinion that in view of Article 8 paragraph 2 ECHR this principle is insufficiently observed in the SyRI legislation. The court finds that the SyRI legislation in no way provides information on the factual data that can demonstrate the presence of a certain circumstance, in other words which objective factual data can justifiably lead to the conclusion that there is an increased risk».

<sup>50</sup> *Ibidem*, § 6.90-95. Sono considerati *bias* o errori confermativi, le distorsioni rappresentative che derivano dal consolidarsi di una convinzione o di un pregiudizio. In questo caso, va inteso in questi termini: i dati raccolti spesso ripropongono (volontariamente o meno) alcuni modelli di discriminazione consolidati (età, genere, reddito, propensione al rischio, etc.). Il sistema elaborando questi dati senza poterne vagliare la qualità finisce per consolidare le cause di stigma sociale (o addirittura di introdurne di nuove). In questi casi, tuttavia, queste disparità possono essere più difficili da identificare e confutare: da un lato, infatti, nel momento in cui si adotta un certo strumento si confida nel suo corretto funzionamento e, dall'altro, i procedimenti automatizzati sono normalmente considerati meno discrezionali della scelta umana, e dunque meno sindacabili.

«la tecnologia digitale [mai come oggi] concorre alla definizione di criteri valoriali e orienta sempre più le decisioni private e pubbliche, rischiando di perdere la sua funzione di “strumento” a servizio dell’uomo per assurgere a fine autonomo cui, paradossalmente la persona finirebbe per divenire strumentale»<sup>51</sup>.

Le riflessioni qui proposte hanno messo in luce come di fronte alle sfide poste dall’IA sia necessario un cambio di paradigma<sup>52</sup>; e questo non solo in campo tecnologico ma anche in ambito giuridico.

La disciplina sulla protezione dei dati personali, cui tradizionalmente era stato assegnato il compito di assicurare un’adeguata protezione dei diritti fondamentali, nella transizione vero la società dell’automazione dimostra di non poter più mantenere efficacemente le sue promesse. Come dimostrato, pur intercettando i problemi, di fronte ai rischi che si avvertono di fronte al continuo sviluppo dell’IA e dell’*automated decision making*, questa materia non è sempre in grado di dare risposte adeguate.

Così se è ormai chiaro che non si può prescindere dal dotarsi di un nuovo complesso normativo per la regolamentazione dell’IA, allo stesso tempo, vi è l’urgenza di elaborare nuovi metodi per risolvere le antinomie valoriali che questa tecnologia porta con sé. La parabola storica della protezione dei dati e la cultura giuridica che ne ha accompagnato l’evoluzione vanno dunque osservate da una diversa prospettiva, così da poterne cogliere gli spunti di metodo senza spingere il dato normativo oltre le sue possibilità.

Come dimostrano anche le proposte di regolamentazione più recenti<sup>53</sup>, sulla base dell’esperienza maturata, la *data protection* ha consentito di individuare i problemi: le criticità dei trattamenti data-intensivi, l’opacità dei procedimenti decisionali automatizzati, l’importanza di promuovere la collaborazione uomo-macchina (e dunque la comprensibilità del *software*), l’esigenza di garantire il confronto tra persone laddove si dubiti dei risultati computazionali. Il diritto costituzionale, invece, ha il compito di offrire il quadro di riferimento valoriale: identificando gli interessi in gioco, stabilendo se e come prenderli in considerazione e come procedere nel contemperare le diverse esigenze, avendo a mente le implicazioni ultime di ognuna di queste scelte sulla tenuta giuridica e ideale dell’ordinamento.

Le soluzioni qui esaminate, nel promuovere una lettura novativa della disciplina di dettaglio e un’interpretazione costantemente aggiornata delle garanzie costituzionali, dimostrano una convergenza tra diverse sensibilità europee, in linea con la tradizione costituzionale comune su cui si edifica il progetto comunitario. Assecondando questi orientamenti, quindi, non resta che attendere gli ulteriori sviluppi della tecnologia e della normativa, avendo a mente il mutuo condizionamento che ciascuna di queste componenti esercita sull’altra, e così il compito assegnato al diritto di promuovere anche in questi percorsi i principi che possono rendere socialmente sostenibile lo sviluppo tecnologico e la rivoluzione 4.0.

<sup>51</sup> A. SORO, *Democrazia e potere dei dati. Libertà, algoritmi e umanesimo digitale*, cit.

<sup>52</sup> A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell’intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1, 2019, 87 ss.

<sup>53</sup> Commissione europea, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, Bruxelles, cit.