INTELLIGENZA ARTIFICIALE E DIRITTO: UNA RIVOLUZIONE?

A CURA DI
ALESSANDRO PAJNO, FILIPPO DONATI E ANTONIO PERRUCCI

VOLUME I
DIRITTI FONDAMENTALI, DATI PERSONALI E REGOLAZIONE

SOCIETÀ EDITRICE IL MULINO

Alla pubblicazione di questa ricerca ha contribuito il Gruppo AlmavivA, che Astrid vivamente ringrazia

ISBN 978-88-15-29967-3

Copyright © 2022 by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati. Nessuna parte di questa pubblicazione può essere fotocopiata, riprodotta, archiviata, memorizzata o trasmessa in qualsiasi forma o mezzo – elettronico, meccanico, reprografico, digitale – se non nei termini previsti dalla legge che tutela il Diritto d'Autore. Per altre informazioni si veda il sito **www.mulino.it/fotocopie**

Redazione e produzione: Edimill srl - www.edimill.it

Capitolo dodicesimo

I PROCESSI DECISIONALI AUTOMATIZZATI E IL DIRITTO ALLA SPIEGAZIONE

1. *Introduzione*

Quando si parla del «diritto alla spiegazione» gli studiosi di diritto e tecnologie assumono spesso una posizione scettica. È noto il dibattito esistente in dottrina tra coloro che ritengono questo diritto esistente sulla base di un Considerando del Regolamento europeo sulla protezione dei dati personali (GDPR) e coloro che invece escludono tale diritto perché non è presente nel testo degli articoli dello stesso GDPR[1]. La questione è altamente discussa e non appare ragionevole che si possa trovare a essa una soluzione con le norme attualmente in vigore. Malgrado gli importanti ostacoli giuridici, sono le caratteristiche stesse dei sistemi di intelligenza artificiale che rendono alquanto difficile prevedere il diritto a una spiegazione *ex post* in capo agli interessati e, correlativamente, il dovere per il titolare del trattamento di fornire le specifiche ragioni del perché l'IA ha deciso in un certo modo.

Se questo è veramente in estrema sintesi lo scenario in cui ci muoviamo, in questo lavoro piuttosto che cercare di individuare una risposta filosofica a tali problemi, ci occu-

Questo capitolo è di Erik Longo.

[1] Prima dell'entrata in vigore del GDPR nel 2018 è sorto in dottrina un notevole dibattito sul rilievo della lettura del Considerando n. 71 in combinato disposto con il testo normativo degli artt. 13, 14 e 22 tra chi riconoscerebbe solo un diritto di accesso e di informazione (S. Wachter, B. Mittelstadt e 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, n. 2) e chi invece sosterrebbe l'esistenza di un «diritto alla spiegazione» (A.D. Selbst e J. Powles, *Meaningful information and the right to explanation*, in «International Data Privacy Law», 2017, n. 4).

peremo prima del problema delle decisioni automatizzate e del diritto alla spiegazione e poi della recente Proposta di Regolamento europeo sulla intelligenza artificiale[2]. Proveremo a mettere in evidenza le principali questioni che tale documento solleva con riguardo alle decisioni automatizzate.

2. *I processi decisionali automatizzati*

Il termine processo decisionale automatizzato è entrato a far parte del gergo giuridico solo di recente. È a partire dalla Convenzione di Strasburgo n. 108/1981 (ratificata in Italia con la l. 21/2/1989, n. 98) che anche in ambito legislativo si inizia a parlare di «elaborazioni automatizzate» per indicare la protezione dei dati personali nei confronti di «operazioni effettuate nel loro insieme o in parte grazie a procedimenti automatizzati». La nozione è stata ripresa e precisata all'interno della Direttiva 95/46/CE che prevedeva la tutela nei confronti di qualsiasi trattamento, manuale o automatizzato, dei dati personali e che stabiliva all'art. 15 il divieto, salvo alcune eccezioni, di procedimenti interamente automatizzati che possano determinare effetti giuridici o che abbiano conseguenze significative nei confronti dell'interessato al trattamento[3]. Il divieto era stato contornato da numerose deroghe, segno del difficile bilanciamento tra la protezione dei dati e l'uso delle decisioni automatizzate, soprattutto nei rapporti commerciali tra privati.

Il divieto e le deroghe sono stati riprodotti con alcuni cambiamenti nell'art. 22 del GDPR, il quale stabilisce il diritto dell'interessato a «non essere sottoposto ad una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo

[2] European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM(2017) 85 final, Bruxelles.

[3] Su tale disposizione si veda L.A. Bygrave, *Minding the machine: Article 15 of the EC data protection directive and automated profiling*, in «Computer Law & Security Review», 2001, n. 1.

riguardano o che incida in modo analogo significativamente sulla sua persona»[4].

La quantità e qualità delle eccezioni al principio sancito nel primo comma dell'art. 22 GDPR, compreso la non banale individuazione del termine «unicamente» a indicare la natura di tali decisioni[5], ha fatto parlare di questa norma come una «barriera difensiva simbolica»[6], frutto di un compromesso, insufficiente ad impedire decisioni formate senza l'intervento umano che incidano su diritti individuali umani[7].

Il concetto a cui le norme europee citate si riferiscono è la *(fully) Automated Decision Making* (ADM)[8], in base alla quale un sistema informatico debitamente programmato può produrre una decisione rilevante per i soggetti coinvolti nel trattamento senza che vi sia alcun ausilio dell'intervento umano,

[4] Sul punto si veda O. Sesso Sarti, *Profilazione e trattamento dei dati personali*, in L. Califano e C. Colapietro (a cura di), *Innovazione tecnologica e valore della persona: il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, Napoli, 2017; A. Pierucci, *Elaborazione dei dati e profilazione delle persone*, in V. Cuffaro, R. D'Orazio e V. Ricciuto (a cura di), *I dati personali nel diritto europeo*, Torino, 2019.

[5] Con riferimento al termine «unicamente», il Gruppo di lavoro articolo 29 ha sostenuto che, per aversi il «livello minimo di interazione umana» richiesto dalla norma, non potrebbe valere un intervento meramente passivo, ma si presuppone una significativa revisione umana «condotta da qualcuno che abbia l'autorità e la competenza per cambiare la decisione», Article 29 - Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (amended version)*, 6/2/2019, Bruxelles.

[6] E. Falletti, *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, in «Il Diritto dell'informazione e dell'informatica», 2019, n. 2.

[7] A. Simoncini, *Art. 22*, in R. D'Orazio, G. Finocchiaro, O. Pollicino e G. Resta (a cura di), *Codice della privacy e data protection*, Milano, 2021, p. 387.

[8] In linea generale, le ADM possono essere definite in diversi modi. Si tratta essenzialmente di quei processi attraverso i quali raccogliere, elaborare, modellare e utilizzare dati (personali e non) per prendere decisioni automatizzate; cfr. S. Newell e M. Marabelli, *Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of «datification»*, in «The Journal of Strategic Information Systems», 2015, n. 1. La nozione di ADM coinvolge quindi una serie ampia di processi, dai sistemi che aiutano i decisori umani fino ai processi decisionali completamente automatizzati, in un'ampia varietà di contesti.

basandosi esclusivamente sulla valutazione algoritmica dei dati personali dei soggetti/utenti[9]. Un esempio tipico di decisione automatizzata è la «profilazione», che è definita nell'art. 4 del GDPR come «qualsiasi forma di trattamento automatizzato».

Come già ricordato, l'art. 22 del GDPR non menziona il diritto ad ottenere una spiegazione della decisione che invece prevede il Considerando n. 71. La mancata riproduzione di tale espressione solleva una duplice questione: da un lato, capire se esiste un tale diritto e, dall'altro, immaginare se è possibile richiedere al titolare del trattamento un obbligo giuridico di fornire una spiegazione.

Secondo molti, la tesi affermativa andrebbe al di là di quanto indicato dal legislatore. Oltre ai motivi giuridici, vi sarebbero anche ragioni di tipo tecnico legate al fatto che i sistemi automatizzati lascerebbero poco spazio a un tale obbligo.

Bisognerebbe, infatti, intendersi sul significato del termine «spiegazione»[10]. Come rilevano le linee guida del Gruppo di lavoro articolo 29, il concetto di spiegazione può assumere diversi connotati, potendosi distinguere tra una *explanation* relativa al funzionamento dello stesso sistema, quindi al meccanismo decisionale considerato nel suo complesso, e la spiegazione relativa alla singola decisione. In base al criterio temporale, inoltre, si distingue tra spiegazione *ex ante*, che quindi è relativa a un sistema decisionale che effettua valutazioni, e spiegazione *ex post*, nel caso di decisioni relative a casi specifici. La precisazione di questo tipo di «spiegazioni», per quanto riguarda la profilazione, si trova oggi nell'art. 7, par. 4, lett. *a*, della Direttiva sui diritti dei consumatori così come modificata dalla Direttiva (UE) 2019/2161.

Da un punto di vista tecnico il problema più consistente quando si parla di spiegazione riguarda la trasparenza che l'intelligenza artificiale, specie quella più evoluta, dovrebbe

[9] La norma si riferisce soprattutto al caso della profilazione, che è un tipico esempio di processo decisionale automatizzato. Sul punto si veda A. Pajno, M. Bassini, G. De Gregorio *et al.*, *AI: profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in «BioLaw Journal - Rivista di BioDiritto», 2019, n. 3.

[10] Simoncini, *Art. 22*, cit., pp. 385 ss.

garantire e che non riesce quasi mai a permettere per via della «naturale» opacità di tali sistemi[11]. Le macchine, infatti, sono estremamente più efficienti di noi umani nell'effettuare calcoli complessi ma non hanno una «intelligenza situazionale» e la capacità di analogie nel caso in cui non siano state allenate con dati specifici. Sebbene l'IA abbia superato gli esperti umani in una gamma sempre crescente di compiti di riconoscimento, previsione e processo decisionale, è molto difficile generare modelli causali e spiegazioni per le percezioni, decisioni, raccomandazioni e azioni di tali macchine. Ciò le rende molto performanti per certe attività ma estremamente inutili o dannose in altre[12].

Quando ci approcciamo al rapporto con le macchine non possiamo tutelarci contro gli eventuali errori computazionali allo stesso modo in cui procediamo con l'errore umano, né possiamo pretendere una spiegazione, allo stato della conoscenza e dell'evoluzione tecnologica, così come la pretendiamo nella realtà analogica. Occorre infatti comprendere molto bene cosa accade in tali situazioni e, semmai, ideare «nuove» forme di tutela per quelle circostanze nelle quali le macchine sono delegate a produrre una decisione valida, fino anche ad immaginare nuovi principi giuridici che servano nei rapporti uomo-macchine[13].

Se quanto detto è vero, un mero diritto alla spiegazione che si basi solo sul rendere trasparente la «logica» utilizzata dagli algoritmi di intelligenza artificiale potrebbe non bastare. Per realizzarsi nella sua pienezza, in molte circostanze l'in-

[11] F. Pasquale, *The black box society: The secret algorithms that control money and information*, Boston, 2015.

[12] Diversamente – ed è uno scarto immenso – l'esperienza umana si caratterizza per il necessario sforzo di completamento attraverso l'ingegno umano che raggiunge soluzioni operando attraverso informazioni necessariamente incomplete, ipotetiche e spesso implicite; cfr. A. Condello, *Il non-dato e il dato. Riflessioni su uno «scarto» fra esperienza giuridica e intelligenza artificiale*, in «Ars interpretandi», 2021, n. 1.

[13] Si pensi al problema della «cattura del decisore» in quei casi in cui si delega alla macchina anche solo una parte della decisione; cfr. A. Simoncini, *Il diritto alla tecnologia e le nuove diseguaglianze*, in F.S. Marini e G. Scaccia (a cura di), *Emergenza Covid-19 e ordinamento costituzionale*, Torino, 2020.

formazione fornita dovrebbe includere anche la conoscibilità dei dati che hanno dato vita alla decisione automatizzata e soprattutto si dovrebbe consentire a chi gestisce la macchina di agire verso il sistema in ogni momento[14]. Un intervento umano senza che siano noti il set di dati e i fattori che possono comportare le loro inesattezze sarebbe inutile[15].

Su di un piano giuridico, data la varietà dei processi decisionali automatizzati, è difficile specificare in termini generali e al tempo stesso precisi quali informazioni debbano essere obbligatoriamente comunicate all'interessato[16].

Inoltre, il problema degli effetti che possono derivare da trattamenti automatizzati incorretti non riguarda solo i dati o la logica utilizzata ma va più a fondo. Proveremo ad esprimerlo con una domanda: siamo così sicuri che anche con la trasparenza e la spiegazione potremmo evitare – o scovare – gli errori o i pregiudizi algoritmici? Sfogliando la letteratura e analizzando gli esempi più noti sul tema sembra emergere una risposta negativa sul punto. Gli algoritmi di apprendimento automatico impiegati per calcolare il tasso di recidiva usati negli Stati Uniti, ad esempio, non funzionano secondo un nesso di causazione ma operano trovando correlazioni. In questi casi è molto facile sbagliare il peso dei «falsi positivi» o dei «falsi negativi», cioè coloro ai quali i giudici, riferendosi al risultato delle macchine, hanno applicato un tasso di recidiva alto ma non sono tornati in carcere o viceversa[17]. Anche volendo depurare tali sistemi

[14] M. Palmirani, *Big Data e conoscenza*, in «Rivista di filosofia del diritto», 2020, n. 1; B. Goodman e S. Flaxman, *European Union regulations on algorithmic decision-making and a «right to explanation»*, in «arXiv», 2016.

[15] Come è stato fatto notare, per consentire tali garanzie è necessario poter «notificare adeguate informazioni all'interessato in merito ai dati trattati, ai modelli utilizzati, alla logica o la famiglia di algoritmi adottati, i dati inferiti e i dati derivati che hanno poi condotto alla decisione finale»; cfr. Palmirani, *Big Data e conoscenza*, cit. Con riguardo alla protezione dei dati personali si veda G. Finocchiaro, *Intelligenza artificiale e protezione dei dati personali*, in «Giurisprudenza italiana», 2019, n. 7.

[16] Simoncini, *Art. 22*, cit., p. 389.

[17] L'algoritmo ci dice solo che per i primi, ad esempio, è molto probabile ripetere l'errore, ma non sanno che alcuni possono essere stati

da tutti i possibili errori, non si arriverà mai a definire un criterio che possa evitare tali distorsioni statistiche.

È molto difficile e forse impossibile affidare a una macchina decisioni complesse nelle quali occorre trasparenza dei motivi che hanno portato a un determinato risultato. Al netto delle questioni legate alla logica inferenziale *data-driven*, l'idea di fondare le decisioni future su regolarità riscontrate in dati che descrivono il passato si scontra con problemi inerenti alla natura stessa dell'atto di decidere, che non è solo frutto di una ponderazione tra variabili ma richiede una serie di caratteristiche che solo l'essere umano può avere[18].

Il diffondersi dei calcoli predittivi senza adeguate tutele tende a contribuire, specie con i fenomeni di *self-prophecy*, in maniera occulta e potenzialmente inconsapevole al materializzarsi degli stessi errori che la tecnologia vuole evitare[19]. Perciò, l'elemento che il diritto costituzionale dovrebbe indicare come primo principio-limite del potere algoritmico è la stessa non autosufficienza della decisione automatica e il connesso principio della responsabilità umana per le decisioni automatizzate, tanto in ambito privato che pubblico.

3. *La proposta di «Artificial Intelligence Act»: verso un nuovo raccordo uomo-macchina*

Le osservazioni e gli esempi riportati consigliano che, quando usiamo l'intelligenza artificiale in modo predittivo

giudicati male. A differenza di un essere umano, i sistemi intelligenti non sanno cosa determina la recidiva e non basano il giudizio su una interpretazione, ma su una tecnica che si limita a leggere serie storiche di dati attraverso meccanismi che riproducono in buona sostanza il «metodo degli analoghi»; cfr. S. Amato, *Emozioni sintetiche e sortilegi al silicio*, in «Ars interpretandi», 2021, n. 1.

[18] C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in «Diritto pubblico comparato ed europeo», numero speciale, 2019; A. Garapon e J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021.

[19] N. Lettieri, *Contro la previsione. Tre argomenti per una critica del calcolo predittivo e del suo uso in ambito giuridico*, in «Ars interpretandi», 2021, n. 1.

per decisioni sociali, dobbiamo essere estremamente cauti e approcciare il problema in modo diverso da come succederebbe se un sistema intelligente stesse prevedendo una mossa nel gioco degli scacchi o la traduzione di un brano da un'altra lingua. Nel caso delle decisioni automatizzate occorre trasparenza, conoscibilità, comprensibilità e necessariamente l'intervento umano da parte di chi conosce bene il meccanismo, i dati di allenamento e i limiti di tali strumenti. Questi principi vanno accompagnati però da un altro criterio, quello della responsabilità decisionale umana, che vuol dire anzitutto consapevolezza delle decisioni. Nel campo delle politiche criminali, ad esempio, non può essere ammesso che la recidiva venga decisa solo grazie a calcoli effettuati da macchine programmate da chi magari non conosce cosa è la recidiva o non è consapevole di come si giudica il rischio relativo[20].

A dire il vero la prospettiva di un nuovo incontro uomo-macchina non è del tutto estranea al diritto europeo. Dopo aver tracciato a partire dal 2018 una strategia per la creazione di regole armonizzate per l'intelligenza artificiale[21], il 21/4/2021 la Commissione europea ha presentato una Proposta di Regolamento generale sull'intelligenza artificiale chiamata *Artificial Intelligence Act*[22]. È il primo atto giuridico che mira a stabilire norme armonizzate (art. 114 TFUE) per lo sviluppo, l'inserimento nel mercato e l'uso di sistemi di IA.

[20] A. Simoncini e S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in «Rivista di filosofia del diritto», 2019, n. 1.

[21] La strategia è stata lanciata nel 2018 con la *Communication on Artificial Intelligence for Europe*, SWD(2018) 137 final ed è poi seguita con il *White paper on «Artificial Intelligence. A European approach to excellence and trust»*, COM(2020) 65 final.

[22] Nella primavera del 2020 la Commissione aveva presentato una proposta legislativa per la consultazione pubblica comprendente quattro opzioni che andavano dalla semplice *soft law* a norme obbligatorie che prevedevano veri e propri divieti, passando per la regolazione dei rischi legati allo sviluppo e all'uso di determinate applicazioni di IA. Un anno dopo la Commissione ha presentato la Proposta di *AI Act* che si concentra principalmente sulla regolamentazione dei sistemi «ad alto rischio» attraverso requisiti obbligatori e misure di divieto.

Si tratta ancora una volta di un Regolamento, segno che nell'ambito della protezione dei dati, delle piattaforme e del mercato digitale, l'UE intende costruire una posizione di leadership attraverso norme valevoli per tutti e lasciando agli Stati membri la definizione di una porzione minima della disciplina, pur nel rispetto dei principi di proporzionalità e sussidiarietà[23].

L'approccio europeo mira ad introdurre una regolazione che distingue tra «pratiche di intelligenza artificiale vietate» e sistemi che prevedono un rischio «alto», «limitato» e «minimo»[24].

Il cuore della proposta contiene le misure per la gestione dei sistemi ad *alto rischio*, secondo l'idea, già presente nel GDPR, che tali sistemi sono parte di un nuovo sistema economico che merita tanto un incentivo quanto regole molto stringenti[25]. Sul piano della *governance*, ogni Stato membro provvederà ad istituire una autorità nazionale responsabile per la supervisione dell'IA. Tuttavia, viene creato anche un nuovo European Artificial Intelligence Board composto da membri designati dagli Stati.

Tale approccio, già presente nella proposta dell'High-Level Expert Group on Artificial Intelligence[26], prevede la predisposizione di un «sistema di gestione dei rischi», la formulazione di «pratiche di *governance* e gestione dei

[23] La dottrina parla a tale proposito di *acti-fication*; cfr. V. Papakonstantinou e P. De Hert, *EU lawmaking in the Artificial Intelligent Age: Act-ification, GDPR mimesis, and regulatory brutality*, in «European Law Blog», 8/7/2021, https://europeanlawblog.eu/2021/07/08/eu-lawmaking-in-the-artificial-intelligent-age-act-ification-gdpr-mimesis-and-regulatory-brutality/.

[24] I sistemi non vietati sono raggruppati in rischio «alto», «limitato» e «minimo». Sul punto si veda J. Liboreiro, *«The higher the risk, the stricter the rule»: Brussels' new draft rules on artificial intelligence*, in «Euronews», https://www.euronews.com/2021/04/21/the-higher-the-risk-the-stricter-the-rule-brussels-new-draft-rules-on-artificial-intellige.

[25] M. Veale e F. Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach*, in «Computer Law Review International», 2021, n. 4.

[26] Cfr. High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI*, Bruxelles, 8/4/2019, p. 8.

dati» che soddisfino criteri di «qualità» e la creazione di una «documentazione tecnica» redatta prima della immissione nel mercato.

Da un punto di vista etico-giuridico, la proposta eredita lo stesso approccio di bilanciamento tra la tutela dei diritti e la garanzia dell'innovazione tecnologica con il quale è stato forgiato il GDPR, ma abbraccia anche il sistema di gestione della sicurezza dei prodotti stabilita nella Decisione del Parlamento e del Consiglio 768/2008/CE.

In linea con tale impostazione, i sistemi ad alto rischio devono garantire un funzionamento sufficientemente trasparente, tale da consentire agli utenti di «interpretare l'*output* del sistema e utilizzarlo adeguatamente». A ciò si aggiunge la previsione di adeguati strumenti di interfaccia uomo-macchina che possano essere «supervisionati da persone fisiche» durante il periodo di uso dell'IA. Si prevedono «misure di sorveglianza umana» anteriori e successive alla immissione nel mercato o della messa in servizio che consentano alle persone fisiche di «comprendere appieno le capacità e i limiti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento». Sono previste poi una serie di misure che possono essere considerate come corollari del principio di non esclusività, tra cui il rendere consapevole l'utilizzatore della possibilità di essere «catturati» dalla automazione facendo involontariamente affidamento o eccessivo affidamento sull'output prodotto da un sistema di IA (*automation bias*), il potere di decidere di non usare il sistema di IA o di ignorare o non considerare l'output o di interpretarlo correttamente ed essere in grado di arrestare il sistema in caso di pericolo. Non mancano infine previsioni circa la qualità dei dati con i quali i sistemi vengono alimentati (si veda il Considerando n. 44).

Per garantire il rispetto delle norme che disciplinano i sistemi ad alto rischio la proposta prevede l'applicazione di un sistema di co-regolazione. Come indica il testo della proposta ed in particolare il Considerando n. 61 sono previsti «standard armonizzati» stabiliti da enti di certificazione

privati, come il CEN e il CENELEC[27]. Su tale aspetto la proposta si espone ad alcune critiche, soprattutto alla luce della passata esperienza nei settori dove tali meccanismi sono usati[28].

La scelta di affidarsi alla gestione interna del rischio e a meccanismi di co-regolazione appare allo stesso tempo il punto di maggiore forza e di maggiore debolezza di questo testo. Da un lato, la proposta limita il controllo e le norme stringenti alle tecnologie più rischiose e lo gradua rispetto all'evenienza del rischio. Dall'altro, la scelta di questo sistema espone a minacce concrete, in quanto si affida tutto il controllo a una valutazione *ex ante* della conformità, che non è effettuato da soggetti terzi pubblici ma dalle stesse aziende unitamente alla presunzione di conformità, che scatta nel caso in cui il fornitore segua standard armonizzati, i quali poi sono da sviluppare da parte degli enti privati di standardizzazione in corrispondenza con la Proposta di Regolamento. Come abbiamo visto, la regolazione dei sistemi di IA usati nelle decisioni automatizzate non è una questione puramente tecnica, ma richiederebbe un dibattito politico che coinvolga la società nel suo insieme visto che si tratta di regolare strumenti – o addirittura agenti – che prendono decisioni in ambiti pubblici, come la salute, l'istruzione, la sicurezza, la giustizia, i diritti fondamentali e la protezione dei consumatori.

La proposta, quindi, offre solo in parte una soluzione per i possibili problemi derivanti dall'impiego dell'intelligenza artificiale e delle decisioni automatizzate, lasciando molte delle questioni o dei problemi a standard da elaborare secondo procedure che si prestano ad essere prive di controllo democratico e giudiziale[29]. Tra questi il tema della

[27] Veale e Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, cit.

[28] M. Ebers, *Standardizing AI-The Case of the European Commission's Proposal for an Artificial Intelligence Act*, in L.A. DiMatteo, M. Cannarsa e C. Poncibò (a cura di), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, Cambridge, 2022.

[29] Veale e Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, cit.

spiegazione appare come uno dei nervi scoperti dell'attuale Proposta di Regolamento europeo. Non esiste nel testo un «diritto alla spiegazione» ma solo una serie di previsioni che hanno a che fare con la trasparenza, l'interpretazione degli output dei sistemi di IA e con l'intervento umano, ma che però non toccano il possibile uso di tecnologie di *explainable AI*.

4. *Conclusioni (provvisorie)*

Abbiamo evidenziato che il problema principale generato dalle ADM riguarda la trasparenza degli algoritmi che le fanno funzionare. Sebbene abbiano una notevole forza pratica, siano facili da sviluppare e si presentino estremamente convenienti, il funzionamento degli algoritmi e i motivi di ciascuna decisione non possono essere compiutamente spiegati a partire dal loro codice sorgente. Quest'ultimo, infatti, rivela solo il funzionamento dell'algoritmo di apprendimento e non la configurazione finale del modello che volta per volta il sistema crea.

Ciò vuol dire che in questi sistemi sono (astrattamente) osservabili e interpretabili solo input e output, ma non lo sono altrettanto le ragioni per le quali il sistema ha prodotto una certa decisione. Il processo decisionale automatizzato e algoritmico è solitamente difficile da comprendere per un essere umano e la sua logica sarà difficile da spiegare a posteriori. È paradossale che strumenti usati per predire non consentano ai beneficiari di tali decisioni di conoscere il perché di certe valutazioni o previsioni. Per come è implementata attraverso gli strumenti dell'apprendimento automatico la predizione rischia di mettere in crisi la stessa prevedibilità, generando più problemi di quelli che si vorrebbero risolvere.

Come ha lucidamente osservato alcuni anni fa Stefano Rodotà[30], ci si trova di fronte ad una nuova forma di *arcana imperi*. Le tecnologie dell'informazione, infatti, sono usate

[30] S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2012.

per rendere la società più trasparente perché permettono controlli diffusi di qualsiasi potere. Tuttavia, come questi algoritmi vengono miscelati per ottenere i risultati migliori è il segreto di ogni azienda[31]. Un'alchimia che ha fatto parlare di «società delle scatole nere» in cui i flussi informativi viaggiano all'insaputa delle persone che producono tali dati e di coloro che utilizzano tali sistemi per prendere decisioni[32]; mentre chi gestisce tali masse di dati e li elabora con i propri codici può acquisire un potere economico immenso grazie alla protezione offerta dalle leggi sulla proprietà intellettuale e la complessità stessa di tali sistemi informatici[33].

Nel caso dell'IA, dunque, non si verifica semplicemente un problema di opacità ma di «massima» asimmetria cognitiva, poiché spesso di tali sistemi sono conoscibili solo le conseguenze delle loro elaborazioni e, operando in ambienti complessi, diviene assai difficile comprendere se tali output siano il frutto di anomalie, oppure derivino dai dati immessi, dal codice o da tutti e tre gli elementi.

Anche in virtù di ciò è necessaria una regolamentazione che sappia contemperare l'iniziativa economica privata, necessaria all'innovazione, con gli interessi dei singoli e della collettività, in uno sforzo di elaborazione che sappia mettere insieme i giuristi alle altre professionalità che operano in questi settori. La sfida è tanto tecnica quanto culturale perché riguarda la costruzione di un nuovo nesso uomo-macchina. In una prospettiva di questo tipo chi progetta, costruisce e vende la macchina non può mirare a sostituire l'uomo ma a valorizzare le sinergie tra intelligenza umana e artificiale in tutte le fasi decisionali.

[31] A. Vespignani e R. Rijtano, *L'algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*, Milano, 2019.

[32] Pasquale, *The black box society: The secret algorithms that control money and information*, cit.

[33] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, 2019.