BioDiritto

Online First
In corso di pubblicazione in *BioLaw Journal 3/2021*

PRIME OSSERVAZIONI SULLA PROPOSTA DI REGOLAMENTO DELL'UNIONE EUROPEA IN MATERIA DI INTELLIGENZA ARTIFICIALE

*Carlo Casonato, Barbara Marchetti**

THE PROPOSAL FOR A REGULATION OF THE EUROPEAN UNION ON ARTIFICIAL INTELLIGENCE: PRELIMINARY NOTES

ABSTRACT: *Artificial Intelligence has great potential in all areas of our lives, but it also presents risks for fundamental rights and the rule of law. The European Union is trying to create a balanced regulatory framework between the pros and cons of AI. On 21 April 2021 EU published a comprehensive proposal for an AI regulation, which should protect and promote European rights and values, without impeding the technological, industrial, and commercial development of AI.*
This article aims to give a first analysis of the proposal, focusing on its main positive and negative aspects.

KEY WORKDS: *Artificial Intelligence, European Union Law, Fundamental Rights, Proposal for a Regulation on AI*

1. Il contesto globale

La letteratura scientifica mondiale, sia sul fronte della *computer science*, sia su quello della riflessione filosofica, giuridica ed etica, concorda sull'esigenza di prevedere un quadro regolatorio in grado di sostenere e consolidare le potenzialità e controllare e minimizzare i rischi

* L'articolo è stato concepito e sviluppato in modo condiviso dagli Autori. Mentre i paragrafi 1 e 9 sono riferibili ad entrambi, i paragrafi 2, 3, e da 6.1 a 6.5 sono attribuibili a Carlo Casonato; i paragrafi 4, 5, 6.6, 7 e 8 sono attribuibili a Barbara Marchetti. carlo.casonato@unitn.it barbara.marchetti@unitn.it Il contributo è stato accettato per la pubblicazione sul n. 3/2021 di BioLaw Journal – Rivista di BioDiritto.

collegati all'impiego dell'IA[1]; anche molte compagnie private paiono d'accordo sull'opportunità di un intervento normativo[2]. I molteplici vantaggi e le evidenti opportunità che la tecnologia promette ed è già in grado di assicurare richiedono, infatti, una parallela consapevolezza dei pericoli che essa dischiude in ragione della sua opacità, della parziale e crescente sottrazione al controllo umano, della possibilità di errori ed esiti discriminatori, potenzialmente lesivi dei diritti fondamentali.

La proposta di regolamento che la Commissione dell'Unione europea ha reso pubblica il 21 aprile 2021 (*Artificial Intelligence Act*, d'ora in poi AIA) rappresenta il primo tentativo compiuto di regolare in termini generali l'IA: essa costituisce l'esito di un processo preparatorio che ha visto, a livello europeo, l'emanazione di numerosi atti di impulso e di soft law in materia di intelligenza artificiale[3]. Tra i molti, si ricordano le risoluzioni del Parlamento europeo sui principi etici dell'IA, della robotica e della tecnologia correlata e sul regime di responsabilità civile per l'IA (entrambi del 20 ottobre 2020) e, più recentemente, sull'uso dell'IA (20 gennaio 2021)[4]. Anche il Libro Bianco sull'Intelligenza artificiale della Commissione (19 febbraio 2020) aveva già indicato un approccio rivolto a combinare eccellenza e fiducia verso la IA e le sue linee generali sono state discusse attraverso un'intensa fase di consultazioni, conclusasi nel maggio del 2020. La proposta di regolamento per la previsione di norme armonizzate per l'IA, inoltre, ha beneficiato del lavoro dell'*High-Level Expert Group on AI* e delle linee guida che tale gruppo ha elaborato in materia di *Trustworthy AI* (8 aprile 2019). Prima della sua approvazione, ancora, la proposta è stata oggetto di una valutazione di impatto da parte del *Regulatory Scrutiny Board* della Commissione (p. 3.3 del *explanatory memorandum*).

Prima di trattare dell'impianto complessivo della disciplina ed esaminarne i contenuti

[1] Fra i più recenti, cfr. B. BRAUNSCHWEIG, M. GHALLAB (Eds.), *Reflections on Artificial Intelligence for Humanity*, Springer, 2021. Molte anche le riviste specialistiche che si dedicano ai rapporti fra diritto e IA: fra le altre, *Artificial Intelligence and Law*, *AI and Society*, oltre che, in Italia, la sezione dedicata a *Artificial Intelligence and Law* ospitata da questa *Rivista*.

[2] Lo stesso Elon Musk, che con Neuralink, fra l'altro, sta progettando elettrodi che collegano il cervello umano con microchip collegati al computer, ha riconosciuto l'urgenza di un intervento regolatorio. Secondo il CEO di Tesla, Spacex e Neuralink i sistemi di intelligenza artificiale richiedono «some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish»: M.U. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competences and Strategy*, in *Harvard Journal of Law & Tec*hnology, 29, 2016, 353.

[3] F. RODI, *Gli interventi dell'Unione europea in materia di intelligenza artificiale e robotica: problemi e prospettive*, in G. ALPA (a cura di), *Diritto e intelligenza artificiale*, 2020, 187-210.

[4] L. PARONA, *Prospettive europee e internazionali di regolazione dell'intelligenza artificiale tra principi etici, soft law e self regulation*, in *Rivista della Regolazione dei Mercati*, 1, 2020, 70.

fondamentali, è opportuno ricordare il contesto in cui essa si pone, caratterizzato dalla sostanziale assenza di discipline normative di portata generale. Sia la Cina che gli Stati Uniti (i due principali competitor globali nel campo dell'IA) stanno fortemente promuovendo la ricerca e lo sviluppo di sistemi di IA, anche grazie ad una regolazione che, laddove esistente, è comunque frammentata o estremamente leggera. Benché negli USA si sia sviluppato un ampio dibattito scientifico e politico in materia di IA, e alcuni utilizzi, come il riconoscimento facciale, siano già stati proibiti a livello cittadino o statale in ragione della loro portata potenzialmente discriminatoria[5], la strategia statunitense appare condizionata dalla preoccupazione che un approccio normativo eccessivamente precauzionale possa inibire lo sviluppo dell'IA, facendo perdere terreno agli Stati Uniti, con implicazioni economiche, di sicurezza nazionale e complessivamente geopolitiche. Sebbene l'amministrazione Biden sembri ora orientata verso una politica meno preoccupata dal confronto con la Cina e più attenta a porre le basi, anche normative, per rendere la IA maggiormente rispettosa dei diritti fondamentali e della rule of law[6], resta il fatto che la Cina è diretta a promuovere un forte sviluppo economico dell'IA (anche grazie alle tecnologie militari) in un quadro autoritario caratterizzato da violazioni della

[5] Sette Stati e oltre una ventina di città USA, ad oggi, hanno disciplinato e parzialmente vietato, a diverse condizioni, il riconoscimento facciale. Si vedano, a titolo esemplificativo, il divieto di riconoscimento facciale da parte (solo) pubblica approvato a San Francisco, su cui A. CHEN, *Why San Francisco's ban on face recognition is only the start of a long fight*, in *MIT Review*, May 16, 2019, L. BARRETT, *Ban facial recognition technologies for children – and for everyone else*, in *Boston University Journal of Science & Technology Law*, 26, 2 2020, 223-285. A livello statale, si veda il divieto del Vermont di utilizzo destinato a *law enforcement* e le prospettive del Massachusetts: N. STATT, *Massachusetts on the verge of becoming first state to ban police use of facial recognition*, in *The Verge*, Dec 2, 2020. In generale sul riconoscimento, ora, G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, 2021. La *Federal Trade Commission* (FTC) ha adottato nel 2020 alcune linee guida per le società che utilizzano IA e algoritmi: in particolare, vengono stabiliti doveri di *transparency, explainability, fairness, robusteness e accuracy* (dei dati) e *accountability*. In particolare, considerata la portata potenzialmente discriminatoria degli algoritmi, l'impiego di un algoritmo deve essere preceduto da un'analisi volta a verificare se e quanto rappresentativo sia il data set, quanto esso consideri la possibilità di *bias*, quanto accurate siano le predizioni in base ai dati e quanto i big data possano essere corretti dal punto di vista etico. Le *guidelines* sono reperibili al sito https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.
Per un confronto tra la strategia USA e quella europea cfr. E. CHITI, B. MARCHETTI, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Rivista della Regolazione dei Mercati*, 1, 2020, 29 e ss.; C. CATH, S. WATCHER, B. MITTELSTADT, M. TADDEO, L. FLORIDI, *Artificial intelligence and the "Good Society": the US, EU and UK approach, in Science and Engineering Ethics*, 2, 2018, 505.

[6] Si veda, ad esempio, il piano *Innovate in America* (https://joebiden.com/made-in-america/). Con diverse prospettive, fra gli altri, A. BURT, *New AI Regulations Are Coming. Is Your Organization Ready?*, in *Harvard Business Review*, April 30, 2021 (https://hbr.org/2021/04/new-ai-regulations-are-coming-is-your-organization-ready); S.A. AARONSON, *America's uneven approach to AI and its consequences*, in *Institute for International Economic Policy Working Paper Series*, George Washington University, April 2020 (https://www2.gwu.edu/~iiep/assets/docs/papers/2020WP/AaronsonIIEP2020-7.pdf).

privacy e dalla mancanza di trasparenza[7].

Un primo dato da considerare nel descrivere e valutare la proposta AIA, quindi, si riferisce al contesto globale ed alle relative dinamiche, spinte e frizioni in cui è destinata a porsi.

Sulla base di queste considerazioni, una disciplina che si proponga di regolare con efficacia e realismo il fenomeno dell'IA deve essere in grado di bilanciare accuratamente diversi interessi e concezioni: essa non deve inibire la ricerca e lo sviluppo dell'IA, per i quali l'Europa si attende investimenti economici pari a 20 miliardi di euro[8], dovendo al contempo affermare e consolidare i principi della *rule of law*; deve essere flessibile e adattabile ai cambiamenti tecnologici e al rapido sviluppo che caratterizzano la tecnologia, assicurando anche quel grado di certezza e prevedibilità necessario per un campo tanto strategico e delicato; non deve farsi inibire dai possibili abusi nell'utilizzo della IA (*abusus non tollit usum*) ma saperne esplorare con coraggio nuovi e benefici domini, promuovendo e rafforzando i diritti fondamentali delle persone e la salute del nostro stesso pianeta[9]. Si tratta, evidentemente, di un equilibrio non facile da trovare a livello nazionale né, tantomeno, a livello europeo (o globale): le regole sono necessarie per assicurare il rispetto dei diritti e dei valori su cui si basa l'Unione europea, ma non devono arrecare un ostacolo sproporzionato rispetto ai margini di sviluppo tecnologico, economico e sociale che la IA può rappresentare. La sfida che la IA pone al diritto, in una formula, riguarda la previsione di una disciplina complessivamente sostenibile[10].

2. La fonte: fra uniformità, certezza e flessibilità

Dal punto di vista dello strumento prescelto, l'Ue ha optato per l'adozione del regolamento in

[7] Si tratta del *New Generation Artificial Intelligence Development Plan' (AIDP)*, su cui, fra gli altri, H. ROBERTS et al., *The Chinese Approach to AI: An Analysis of Policy, Ethics, and Regulation*, in *AI and Society*, 36, 2021, 59-77 (https://doi.org/10.1007/s00146-020-00992-2).

[8] Cfr. M. BROADBENT, *AI Regulation: Europe's Latest Proposal is a Wake-Up Call for the United States*, Center for Strategic and International Studies, May 18, 2021 (https://www.csis.org/analysis/ai-regulation-europes-latest-proposal-wake-call-united-states) in cui peraltro si ricorda che nell'Unione europea hanno sede solo 6 delle prime 100 startup di IA del mondo.

[9] Si vedano, in generale, i diversi contributi raccolti in B. BRAUNSCHWEIG, M. GHALLAB, (Eds.), *Reflections on Artificial Intelligence for Humanity*, 2021; B. AROGYASWAMY, *Big tech and societal sustainability: an ethical framework*, in *AI and Society*, 35, 2020, 829-840.

[10] A. ADINOLFI, *L'Unione europea dinanzi allo sviluppo dell'intelligenza artificiale: la costruzione di uno schema di regolamentazione europeo tra mercato unico digitale e tutela dei diritti fondamentali*, in S. DORIGO (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, 2020, 13-35; M. ZANICHELLI, *Ecosistemi, opacità, autonomia: le sfide dell'intelligenza artificiale in alcune proposte recenti della Commissione europea*, in A. D'ALOIA (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2020, 67-87; A. AMIDEI, *La governance dell'intelligenza artificiale: profili e prospettive di diritto dell'Unione europea*, in U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, cit., 571-588.

luogo della direttiva, in termini analoghi a quanto fatto, fra l'altro, con il GDPR per la disciplina in materia di protezione dei dati: la sua base legale nell'art. 114 del TFUE (che prevede l'adozione di misure volte ad assicurare la realizzazione e il funzionamento del mercato interno) è suscettibile di determinare così vincoli uniformi e direttamente applicabili su tutto il territorio dell'Unione, con l'obiettivo di fissare un quadro normativo omogeneo e tendenzialmente rigido per gli Stati membri, salvo taluni margini di manovra e di apprezzamento per la disciplina delle *sandboxes* e dei codici di condotta, per l'organizzazione interna degli Stati e per il regime sanzionatorio[11].

Il tentativo di dare all'Ue un quadro di regole uniforme e certo si accompagna, tuttavia, all'esigenza di meccanismi di aggiornamento della disciplina: l'IA costituisce, come è noto, un oggetto difficile da regolare, sia perché, ancor più di altre tecnologie innovative, è caratterizzata da incessanti sviluppi che rendono rapidamente obsoleta qualsiasi disciplina volta a regolarla, sia perché, nei suoi sistemi più avanzati (*machine learning*, *deep learning*, *neural networks*) si contraddistingue per una forte dose di autonomia e imprevedibilità di funzionamento, la quale, accompagnata all'inspiegabilità dei processi interni (fenomeno della *black box*) può rappresentare una potenziale fonte di rischi, non calcolabili ex ante[12].

La Commissione ha cercato di tenere in considerazione queste complessità. La proposta, infatti, consapevole dell'estrema mutevolezza delle applicazioni di IA e dei corrispondenti fattori di rischio non tutti prevedibili a priori, introduce due meccanismi di flessibilità del quadro normativo, volti ad assicurare una certa capacità di adattamento e aggiornamento delle regole (si parla, nei documenti di accompagnamento, di disciplina *future-proof*).

In primo luogo, la proposta AIA si completa di alcuni annessi che sono fondamentali, per

[11] Come si vedrà *infra*, inoltre, i poteri di attuazione della disciplina sono attribuiti al livello amministrativo statale.

[12] In questi termini, già M.U. SCHERER, *Regulating artificial intelligence systems: risks, challenges, competencies and strategies*, in *Harvard Journal of Law & Technology*, 29, 2016, 365, secondo cui «one important characteristic of AI that poses a challenge to the legal system relates to the concept foreseeability»; infatti, poichè «AI systems are not inherently limited by the preconceived notions, rules of thumb, and conventional wisdom upon which most human decision-maker rely, AI systems have the capacity to come up with solutions that humans may not have considered, or that they considered and rejected in favor of more intuitively appealing options». Sulle difficoltà di una regolazione dell'IA cfr. J. DANAHER, *Is Effective Regulation of AI Possible? Eight Potential Regulatory Problems*, July 7, 2015 (https://archive.ieet.org/articles/danaher20150707.html). Sui collegamenti con il GDPR, F. PIZZETTI, (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018; A. PAJNO et al., *AI: profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2019, 206 ss.; E. SPILLER, *Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientate*, in questa *Rivista*, 2, 2021, 419-432; G. D'ACQUISTO, *On conflicts between ethical and logical principles in artificial intelligence*, in *AI and Society*, 35, 2020, 895-900; G. SIMEONE, *Machine Learning e tutela della privacy alla luce del GDPR*, in G. ALPA (a cura di), *Diritto e intelligenza artificiale*, cit., 275-295.

esempio, per individuare la categoria dei dispositivi ad alto rischio (*high risk AI systems*) per cui si prevede una specifica procedura di verifica della conformità e a cui dedica – vedremo subito – gran parte delle sue disposizioni. Per la modifica di tali annessi si prevede che sia competente la Commissione ai sensi dell'art. 290 TFUE (atti delegati), la quale è affiancata dai Comitati della Comitologia (art. 74): così facendo, a mano a mano che l'applicazione della normativa evidenzierà eventuali problemi applicativi legati alla individuazione delle categorie dei sistemi ad alto rischio o alle procedure di controllo della conformità, sarà possibile apporre modifiche alla disciplina anche al di fuori del procedimento legislativo ordinario altrimenti richiesto per la revisione formale del regolamento.

In secondo luogo, la proposta prevede un obbligo generale di revisione del regolamento a cinque anni dalla sua entrata in vigore e, successivamente, con cadenza quinquennale[13]. Tale scelta risulta apprezzabile proprio in vista della già menzionata mobilità della materia oggetto di disciplina, permettendo di perseguire un diritto della IA che sia «*stable but not still*»[14]. D'altro canto, simili misure sono ormai consuete all'interno del panorama di diritto comparato che si occupa delle materie fortemente interessate dall'innovazione tecnologica[15]. In tema di IA, ad esempio, si veda la *Directive on Automated Decision-Making* canadese che, entrata in vigore nell'aprile del 2019, è soggetta ad un processo di revisione ogni sei mesi[16].

Perché il processo di revisione del regolamento europeo sia il più possibile avvertito e sostanziale, inoltre, la proposta AIA prevede, al titolo V, l'attivazione del meccanismo delle *sandboxes*. Si tratta di contesti operativi istituiti dagli Stati membri, all'interno dei quali, per un periodo di tempo limitato e sotto il controllo delle autorità competenti nazionali, è possibile sviluppare, testare e validare sistemi di IA innovativi ai fini di una successiva immissione nel mercato. In buona sostanza, si tratta di una sperimentazione tesa a valutare concretamente, ed in caso modificare, il comportamento di nuovi sistemi valutandone i risultati, i benefici e i rischi, e

[13] Cfr. il punto 5.1. dell'*Explanatory Memorandum* e il punto 2.1 delle *Mangement Measures* del *Legislative Financial Statement*.

[14] Il riferimento è a Roscoe Pound: «Law must be stable, and yet it cannot stand still»: R. POUND, *Interpretations of Legal History*, Cambridge, 1923, I.

[15] Cfr. C. Casonato, *21st Century BioLaw: a proposal*, in questa *Rivista*, 1, 2017, 81-95 (http://rivista.biodiritto.org/ojs/index.php?journal=biolaw&page=article&op=download&path%5B%5D=213&path%5B%5D=177).

[16] Per la disciplina canadese, si veda il sito https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592. In termini generali, può menzionarsi la *loi de bioéthique* francese, la quale è rivista ogni 7 anni, con una procedura che prevede anche l'attivazione degli Stati generali della bioetica: https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000038811571/#:~:text=Le%20projet%252.

la compatibilità con la disciplina prevista[17].

Infine, la proposta di regolamento prende in considerazione il carattere plurale e diversificato dell'IA[18]. Pur ricondotta in termini unitari a qualsiasi *software* capace, per un dato insieme di obiettivi definiti dall'uomo, di generare output (contenuti, previsioni, raccomandazioni, decisioni) che influenzano l'ambiente con cui interagiscono (art. 3)[19], la IA comprende tecniche e applicazioni anche molto distanti, il cui funzionamento si caratterizza per gradi variabili di autonomia, imprevedibilità e trasparenza, e il cui utilizzo porta a risultati, potenzialità e rischi anch'essi assai vari[20]. In questo senso, ad esempio, un conto è parlare genericamente di sistemi esperti (tipicamente funzionanti secondo una logica rigida *if-then*), un altro conto di reti neurali, caratterizzate da *trade off* molto diversi in termini di autonomia, trasparenza e spiegabilità. Per ciascun sistema di IA, inoltre, le possibilità concrete di controllo umano sono assai differenti. A un estremo si pongono i sistemi che potrebbero svolgere le proprie funzioni in completa autonomia (*Human out of the loop*), all'altro quelli che sono governati interamente dall'umano (*Human in command*), passando attraverso una serie di posizioni intermedie in cui la dimensione umana gioca un ruolo crescente (*Human post the loop*, *Human on the loop* e *Human in the loop*)[21].

---

[17] Recita il 72 considerando che al fine di assicurare un'attuazione uniforme di tali spazi di sperimentazione, occorre stabilire regole comuni ed una stretta cooperazione tra le autorità nazionali chiamate a monitorare le sandboxes.

[18] Al riguardo, già il parere *Sviluppi della robotica e della roboetica*, del Comitato Nazionale per la Bioetica e del Comitato Nazionale per la Biosicurezza, le Biotecnologie e le Scienze della Vita, del 17 luglio 201, su cui L. d'Avack, *La rivoluzione tecnologica e la nuova era digitale: problemi etici*, in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffrè, 2020, 3-27.

[19] L'art. 3 della proposta va in questo senso letto alla luce dell'annesso I, che richiama i sistemi di *machine learning*, gli approcci *logic- and knowledge-based*, oltre che *statistical approaches, Bayesian estimation, search and optimization methods*.

[20] Sui problemi legati ai profili definitori della IA, fra gli altri, cfr. S. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2020, 17; B.C. SMITH, *The Promise of Artificial Intelligence: Reckoning and Judgment*, MIT press, 2019; B. Marr, *The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance*, in *Forbes*, Feb 14, 2018 (https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#285881804f5d). Una definizione più articolata è fornita dall'*High-Level Expert Group on Artificial Intelligence* nominato dalla Commissione europea nel documento su *A definition of AI: Main Capabilities and Disciplines*, Brussels, aprile 2019 (https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines). Per l'Italia, fra gli altri, si vedano L. PALAZZANI, *Tecnologie dell'informazione e intelligenza artificiale: Sfide etiche al diritto*, Studium, 2020; A. D'ALOIA, *Il diritto verso "il mondo nuovo". Le sfide dell'intelligenza artificiale*, in A. D'ALOIA (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, cit., 14 ss.; A. SANTOSUOSSO, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Mondadori, 2020, 1-20; G. SARTOR, F. LAGIOIA, *Le decisioni algoritmiche tra etica e diritto*, in U. RUFFOLO, (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, cit., 63-91; M.C. CARROZZA et al., *AI: profili tecnologici. Automazione e Autonomia: dalla definizione alle possibili applicazioni dell'Intelligenza Artificiale*, in questa *Rivista*, 3, 2019, 243.

[21] Fra gli altri, S. AMATO, *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, 2020, 88.

Sulla consapevolezza delle complessità menzionate e di tali ultime specificità si basa l'approccio proporzionato al controllo del rischio introdotto dalla proposta AIA, che si traduce in una regolazione differenziata dell'IA. In particolare, si distingue tra sistemi a rischio inaccettabile, per i quali è previsto un regime di divieto salvo deroghe espresse, sistemi ad alto rischio, cui è dedicata la gran parte della disciplina, sistemi a basso e minimo rischio, che, sostanzialmente liberi, sono soggetti a soli oneri di informazione.

3. I destinatari della regolamentazione

Vista la necessità di una protezione diffusa, effettiva e indiscriminata su tutto il territorio dell'Unione, l'art. 2 della proposta AIA specifica la propria applicazione soggettiva, disegnando in termini particolarmente rigorosi (e generosi) l'ampiezza della propria copertura. In questa logica, essa non mira a raggiungere i soli programmatori, disegnatori, produttori e fornitori localizzati in Paesi europei, ma guarda alla collocazione del prodotto (e all'utilizzo dei relativi output) nel mercato europeo, vincolando così qualunque produttore o utilizzatore di sistemi la cui commercializzazione o il cui impiego avvengano all'interno del mercato unico (art. 2)[22].

Sempre con riguardo all'ambito di applicazione, va sottolineato che tra i destinatari delle regole della proposta AIA, oltre ai privati, figurano anche le pubbliche amministrazioni, sia che queste provvedano a dotarsi *in house* di sistemi di IA (operino cioè quali *providers*), sia quando reperiscono tali applicazioni all'esterno e li utilizzano per lo svolgimento dei propri compiti (*contracting out*). Per tale aspetto, dunque, la proposta parifica soggetti pubblici e privati, unificando il tipo di garanzie che l'uso di sistemi di IA impone per le due categorie di soggetti, senza che rilevi il diverso regime giuridico che, invece, contraddistingue generalmente l'azione dei pubblici poteri rispetto a quella dei soggetti privati.

I sistemi impiegati esclusivamente a fini militari sono invece esclusi dalla portata della proposta[23].

4. Le categorie della IA: il risk approach

La proposta, come detto, individua quattro categorie di IA diverse e le sottopone ad altrettanti

[22] Cfr. anche il considerando 10. Sull'applicazione extraterritoriale della regolazione prodotta dall'Unione europea v. A. BRADFORD, *Brussells Effect. How the European Union rules the world*, 2020.
[23] Cfr. il considerando 12 e l'art. 2, terzo comma.

regimi regolatori. Anzitutto considera alcune applicazioni dell'IA a rischio inaccettabile: tra queste, alcune vengono però vietate in assoluto, mentre altre sono vietate in linea di principio, ma sono ammesse eccezioni.

5. I sistemi vietati e quelli a rischio minimo

Anzitutto, sono vietati in termini assoluti i sistemi che mirano a manipolare in base a tecniche subliminali la condotta delle persone oppure fanno leva sulle vulnerabilità di alcuni soggetti (a causa della loro età o disabilità, ad esempio) al fine di condizionarne la condotta (art. 5, rispettivamente lett. a e b). In entrambi i casi, peraltro, il divieto scatta quando il dispositivo possa determinare danni fisici o psicologici all'utente o ad altra persona.

Sono invece vietati solo in linea di principio i sistemi di IA utilizzati da autorità pubbliche per stabilire l'affidabilità delle persone (in termini di trustworthiness e di *social scoring*) in base alla loro condotta sociale o alle caratteristiche personali. In particolare, tali applicazioni (art. 5, lett. c) sono proibite solo se determinano un trattamento pregiudizievole (*detrimental or unfavourable*) in un contesto scollegato a quello in cui i dati sono stati generati, oppure ad un trattamento parimenti pregiudizievole che sia ingiustificato o sproporzionato rispetto alla condotta sociale e alla sua gravità.

La stessa norma vieta altresì (lett. d) l'uso di sistemi di identificazione biometrica *real time* in spazi aperti al pubblico per finalità di polizia (*law enforcement*), a meno che non siano strettamente necessari per la ricerca mirata di potenziali vittime di azioni criminose, come bambini scomparsi, per la prevenzione di un pericolo specifico, sostanziale e imminente alla vita o alla sicurezza di una persona o di un attacco terroristico o, infine, per la individuazione, localizzazione o incriminazione di un soggetto sospetto di reati previsti dall'art. 2(2) della decisione quadro del Consilio 2002/584 per i quali lo Stato membro interessato preveda una pena detentiva pari o superiore a tre anni[24].

[24] Si tratta dei reati per i quali è previsto il mandato di arresto europeo elencati al comma 2 dell'art. 2 della decisione quadro del Consiglio. Danno luogo a consegna in base al mandato d'arresto europeo, alle condizioni stabilite dalla presente decisione quadro e indipendentemente dalla doppia incriminazione per il reato, i reati seguenti, quali definiti dalla legge dello Stato membro emittente, se in detto Stato membro il massimo della pena o della misura di sicurezza privative della libertà per tali reati è pari o superiore a tre anni: partecipazione a un'organizzazione criminale, terrorismo, tratta di esseri umani, sfruttamento sessuale dei bambini e pornografia infantile, traffico illecito di stupefacenti e sostanze psicotrope, traffico illecito di armi, munizioni ed esplosivi, corruzione, frode, compresa la frode che lede gli interessi finanziari delle Comunità europee ai sensi della convenzione del 26 luglio 1995 relativa alla tutela degli interessi finanziari delle Comunità europee, riciclaggio di proventi di reato, falsificazione di monete,

Peraltro, anche qualora il riconoscimento facciale sia rivolto a raggiungere tali obiettivi, la proposta di regolamento stabilisce (art. 5 comma 2) che l'autorità debba tenere in considerazione la situazione concreta, e in particolare la gravità, la probabilità e l'entità del pregiudizio che sarebbe causato dal mancato uso del sistema di identificazione, e le conseguenze che l'impiego di quest'ultimo potrebbe avere per i diritti e le libertà delle persone coinvolte. L'uso individuale del riconoscimento facciale richiede in ogni caso una previa autorizzazione rilasciata dall'autorità giudiziaria oppure da un'autorità amministrativa indipendente a seguito di istanza motivata e in conformità al diritto nazionale.

Dall'esame della disposizione appare come le condizioni che la proposta AIA fissa per delimitare il regime di divieto siano caratterizzate da un certo grado di indeterminatezza, suscettibile di accordare una porzione significativa di discrezionalità allo Stato e alle sue autorità pubbliche, chiamati a decidere, ad esempio, se vi sia un effettivo collegamento tra il contesto in cui sono utilizzati gli output e quello in cui i dati sono stati raccolti, oppure a considerare se le conseguenze pregiudizievoli derivanti dall'uso del riconoscimento facciale siano di entità tale da giustificarne l'impiego, o se la valutazione di affidabilità sia proporzionata rispetto alle condotte osservate.

La presenza di concetti indeterminati ed interpretabili implica flessibilità applicativa e, di conseguenza, margini di manovra a favore degli Stati membri. D'altro canto, in mancanza di pratiche applicative comuni, il rischio è che tale indeterminatezza generi anche incertezza, aprendo la possibilità di applicazioni difformi della disciplina all'interno del territorio dell'Unione e un conseguente aumento del contenzioso giudiziario.

Il titolo terzo della proposta di regolamento è dedicato ai sistemi ad alto rischio e costituisce la parte preponderante e centrale della disciplina. In termini generali, l'art. 6 contiene una classificazione dei sistemi ad alto rischio, che è completata dagli annessi II e III, senza i quali l'identificazione di ciò che il regolamento intende come sistema ad alto rischio sarebbe

---

compresa la contraffazione dell'euro, criminalità informatica, criminalità ambientale, compreso il traffico illecito di specie animali protette e il traffico illecito di specie e di essenze vegetali protette, favoreggiamento dell'ingresso e del soggiorno illegali, omicidio volontario, lesioni personali gravi, traffico illecito di organi e tessuti umani, rapimento, sequestro e presa di ostaggi, razzismo e xenofobia, furti organizzati o con l'uso di armi, traffico illecito di beni culturali, compresi gli oggetti d'antiquariato e le opere d'arte, truffa, racket e estorsioni, contraffazione e pirateria in materia di prodotti, falsificazione di atti amministrativi e traffico di documenti falsi, falsificazione di mezzi di pagamento, traffico illecito di sostanze ormonali ed altri fattori di crescita, traffico illecito di materie nucleari e radioattive, traffico di veicoli rubati, stupro, incendio volontario, reati che rientrano nella competenza giurisdizionale della Corte penale internazionale, dirottamento di aereo o nave, sabotaggio.

incompleta. Per tali sistemi è previsto il necessario rispetto dei requisiti stabiliti nel capitolo II quale condizione per l'immissione in commercio: tale rispondenza richiede un procedimento di verifica della conformità, il quale tuttavia è per lo più svolto dal provider stesso e, solo in casi eccezionali, da soggetti terzi. Una volta stabilita la conformità del sistema di IA ad alto rischio, può essere rilasciata la dichiarazione di conformità (art. 48) ed è apposto il marchio europeo (art. 49) che consente la circolazione nel mercato. I sistemi ad alto rischio conformi al diritto dell'Unione sono poi registrati in una banca dati, accessibile al pubblico, posta sotto il controllo della Commissione (art. 60). Una volta immesso nel mercato, il sistema di IA ad alto rischio viene sottoposto ad una vigilanza post-market, volta a verificare che per tutto il suo ciclo di vita esso resti rispettoso delle condizioni stabilite dal regolamento.

Accanto ai sistemi vietati ed ai sistemi ad alto rischio, cui sarà dedicato il paragrafo successivo, la proposta di regolamento stabilisce una disciplina anche per i sistemi a basso o minimo rischio, regolati dall'art. 52 della proposta di regolamento. Tale categoria di IA è residuale e ricavabile per sottrazione: sono sistemi a basso o minimo rischio tutti quelli che non sono vietati e non sono ad alto rischio. Essi costituiscono la parte più consistente di applicazioni di intelligenza artificiale all'interno del mercato comune europeo. La loro circolazione nell'Unione europea è libera, salva l'ipotesi in cui, per le loro caratteristiche, non debbano essere stabiliti in capo al provider obblighi di trasparenza. In particolare, ciò accade sia per quei sistemi che interagiscono con le persone (*chatbots*) per i quali, per le circostanze del caso e il contesto in cui operano, non sia evidente la loro natura artificiale; sia per quei sistemi che, manipolando o sovrapponendo immagini, video o audio, fanno sembrare autentiche persone, oggetti o luoghi che invece non lo sono (c.d. *deep fake*). In questo caso, la persona deve essere informata del fatto che è intervenuta una manipolazione artificiale dei contenuti originali, così che non abbia dubbi sulla natura reale o artificiale degli stessi.

Limitate deroghe a tali obblighi di trasparenza possono essere previste solo quando il sistema sia autorizzato dalla legge per finalità di prevenzione e persecuzione dei reati, oppure quando sia impiegato per assicurare le libertà di espressione, artistiche o scientifiche protette dalla Carta dei diritti fondamentali dell'Ue.

6. I sistemi ad alto rischio

Come anticipato, gran parte della proposta di regolamento si dedica al regime applicabile ai

sistemi di IA definiti ad alto rischio (*High-Risk AI Systems*).

La logica complessiva che la Commissione ha inteso seguire al riguardo si articola secondo due linee che si riferiscono alla compatibilità ed armonizzazione con gli altri atti normativi europei e, ancora una volta, alla ricerca di un equilibrio fra certezza del diritto e flessibilità della disciplina.

In primo luogo, la lista e il contenuto concreto dei requisiti richiesti vuole porsi, in buona sostanza, sulla stessa linea dei principi già segnalati in diverse occasioni e sedi. Al riguardo, ad esempio, l'*Explanatory memorandum* richiama il lavoro svolto dal menzionato *High-Level Expert Group on AI*[25] e dichiara una sostanziale compatibilità con le raccomandazioni e con i principi internazionali che già vincolano i partner commerciali dell'Unione (pag. 13). Il punto 12 dei considerando, inoltre, richiama l'allineamento con la Carta dei diritti fondamentali e in più punti si richiama la coerenza con una lunga serie di atti normativi della UE (tipicamente, quelli compresi nel *New Legislative Framework*)[26]. In altri casi, invece, si propone che il principio di armonizzazione venga rispettato attraverso la modifica di altri regolamenti o direttive di settore[27].

L'opportunità di mantenere un determinato grado di flessibilità della disciplina a fronte della necessità di certezza giuridica, in secondo luogo, è perseguita, in termini generali, con il già menzionato riferimento agli annessi (più facilmente emendabili)[28], e in termini specifici, con la possibilità per i *providers* di scegliere discrezionalmente quali strumenti concreti adottare per rispettare i requisiti previsti dalla proposta AIA, scegliendoli sulla base dei più recenti innovazioni scientifiche e tecnologiche.

Su queste basi, la proposta AIA dedica il Titolo III, il più corposo dell'intero documento, alla disciplina dei sistemi ad alto rischio.

6.1. L'individuazione della categoria

Il primo profilo affrontato (art. 6) riguarda l'individuazione di tali sistemi; individuazione che

[25] Si veda il sito https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai, costantemente aggiornato.

[26] Su cui la sezione A dell'annesso II della proposta AIA. L'allineamento delle diverse normative potrebbe non essere sempre del tutto lineare. Cfr., ad esempio, J. ANDRAŠKO, M. MESARČÍK, O. HAMUĽÁK, *The regulatory intersections between artificial intelligence, data protection and cyber security: challenges and opportunities for the EU legal framework*, in *AI and Society*, 36, 2021, 623-636.

[27] Cfr. il punto 29 dei considerando.

[28] In riferimento all'annesso III, si veda l'art. 7.

non si basa solo sulla funzione attribuita al dispositivo, ma sul suo scopo complessivo, comprensivo degli obiettivi specifici previsti, e sulle modalità utilizzate. Inoltre, si precisa come possano essere considerati ad alto rischio sia i singoli dispositivi di IA che possano incidere sui diritti fondamentali (*stand-alone AI systems*), sia i sistemi che costituiscono componenti di sicurezza di prodotti che siano già soggetti a valutazione di conformità ai sensi della normativa di settore[29]. Oltre a questa indicazione, l'art. 6 stabilisce che l'individuazione dei sistemi ad alto rischio è svolta sulla base della lista di prodotti coperti dagli atti normativi (direttive e regolamenti) ricompresi nel *New Legislative Framework* (annesso II) e dell'elenco di aree di utilizzo specificate all'annesso III. In buona sostanza, si tratta dei sistemi ammessi di identificazione biometrica delle persone fisiche, di quelli utilizzati per la gestione delle infrastrutture critiche (come traffico, fornitura di acqua, gas, elettricità), per scopi di istruzione e formazione professionale, di occupazione e gestione dei lavoratori. Ancora, oltre ai sistemi utilizzati nell'ambito di servizi pubblici essenziali (come la sanità) o come strumenti di partecipazione alla vita sociale e di miglioramento delle condizioni di vita (come l'accesso al credito), fanno parte dei sistemi ad alto rischio quelli impiegati per motivi di *law enforcement* da parte delle forze dell'ordine, per l'amministrazione della giustizia, nell'ambito dei processi democratici e nella gestione dei flussi migratori, del diritto d'asilo e dei controlli alle frontiere.
Da ricordare che la proposta AIA, al fine di assicurare una protezione non discriminatoria ed effettiva su tutto il territorio dell'Unione, si applica non solo a tutti i *providers* di sistemi di IA destinati ad essere utilizzati nella UE, a prescindere dall'ubicazione della sede legale (interna o esterna all'Unione) e a tutti gli utilizzatori all'interno della Unione, ma anche ai providers e agli utilizzatori ubicati in Paesi terzi quando gli output generati dal sistema sono comunque utilizzati sul territorio dell'Unione.

6.2. Il sistema di risk management

Al fine di ridurre al massimo i pericoli per i diritti degli utenti, si prevede, che ogni *provider* adotti un sistema per la gestione del rischio, il quale deve essere mantenuto e aggiornato per tutto il ciclo di vita del sistema. Si tratta di: un processo di identificazione e valutazione dei rischi sia conosciuti e prevedibili prima della messa in commercio, sia emersi in fase di

29 Si tratta degli «AI systems intended to be used as safety component of products that are subject to third party ex-ante conformity assessment» (*Explanatory memorandum*, p. 16).

monitoraggio post-market; una gestione degli stessi in termini di eliminazione o riduzione, oltre che di adeguata informazione; appropriate procedure di *testing* che ne assicurino un utilizzo conforme agli scopi previsti e ai requisiti disposti dalla proposta di regolamento. In questo senso, in particolare, si impongono un set di dati di alta qualità, la creazione e il mantenimento della documentazione tecnica, un adeguato livello di trasparenza, una supervisione umana e la garanzia di robustezza, accuratezza e sicurezza del sistema[30].

6.3. I requisiti relativi ai dati impiegati ed alla documentazione richiesta

La proposta di regolamento specifica i requisiti minimi perché i sistemi ad alto rischio siano ammessi all'interno del mercato unico.

Visti i rischi di errore e di discriminazione, legati all'impiego di dati non sufficientemente accurati nel *training*, nella *validation* o nel *testing* del sistema o comunque alla logica statistico-probabilistica utilizzata[31], il primo requisito si riferisce alla qualità dei dati e ad un sistema di

[30] Cfr. art. 9 e considerando 42.

[31] Esiste al riguardo un'ampia letteratura. Cfr., fra i molti altri, in generale, A. ROSENFELD, R. ZEMEL, J.K. TSOTSOS, *The Elephant in the Room*, 9 agosto 2018 reperibile al sito della Cornell University https://arxiv.org/abs/1808.03305); 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, 87-105; D. VARONA, Y. LIZAMA-MUE, J.L. SUAREZ, *Machine learning's limitations in avoiding automation of bias*, in *AI and Society*, 36, 2021, 197-203. In tema di amministrazione della giustizia, J. ANGWIN et al., *Machine Bias, There's software used across the country to predict future criminals. And it's biased against blacks*, nel sito di ProPublica, 23 maggio 2016 (www.propublica.org/article/machine-bias-risk-assessments-in-criminalsentencing); A. VAN DEN BRANDEN, *Les robots à l'assaut de la justice. L'intelligence artificielle au service des justiciables*, Bruylant, 2019; M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018, 872-893; S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, 2020; L. AULINO, *Intelligenza artificiale e giustizia: tra nuove soggettività giuridiche e problematiche etiche e deontologiche*, in A. D'ALOIA (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, cit., 283-295; C. SLOBOGIN, *Just Algorithms. Using Science to Reduce Incarceration and Inform a Jurisprudence of Risk*, 2021; P. HAYES, I. VAN DE POEL, M. STEEN, *Algorithms and values in justice and security*, in *AI and Society*, 35, 2020, 533-555; R. BICHI, *Intelligenza digitale, giurmetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione*, in U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, cit. 423-447; in generale European Commission for the Efficiency of Justice (CEPEJ), *Guidelines on how to drive change towards cyberjustice. Stock-taking of tools deployed and summary good practices*, 2017, in https://bit.ly/3kddNCi. In tema di medicina e tutela della salute, oltre al classico E. TOPOL, *Deep Medicine. How artificial Intelligence can make Healthcare human again*, 2019, R. BENJAMIN, *Assessing risk, automating racism*, in *Science*, 366, 6464, 2019, 421-422; D.A. VYAS, L.G. EISENSTEIN, D.S. JONES, *Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms*, in *The New England Journal of Medicine*, 2020, 383, 874-882; G. TAMBURRINI, *Etica delle machine. Dilemmi morali per robotica e intelligenza artificiale*, 2020, 133; V. DE BERNARDIS, *L'impiego delle nuove tecnologie in medicina*, in G. ALPA (a cura di), *Diritto e intelligenza artificiale*, cit., 489-501; K. ASTROMSKĖ, E. PEIČIUS, P. ASTROMSKIS, *Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations* e H. SMITH, *Clinical AI: opacity, accountability, responsibility and liability*, entrambi in *AI and Society*, 36, 2021, rispettivamente 509-520 e 535-545; M. FASAN, *La tecnologia ci salverà? intelligenza artificiale, salute individuale e salute collettiva ai tempi del coronavirus*, nel numero speciale di questa *Rivista* 1, 2020, 677-683; E.A. FERIOLI, *Digitalizzazione, intelligenza artificiale e robot nella tutela della salute*, e L. RUFO,

data governance. In particolare, si prevede l'utilizzo di dati rilevanti, completi ed esenti da errori[32]. Inoltre, si precisa il carattere rappresentativo degli stessi dati, in termini di proprietà statistiche adeguate anche a possibili specificità dei contesti di riferimento (art. 10). Sulla stessa linea, la proposta prevede l'accuratezza e robustezza complessiva dei sistemi ad alto rischio, anche in termini di cybersicurezza e di resilienza, richiamando la capacità di gestire i rischi connessi a possibili debolezze del sistema che possano generare output non affidabili (art. 15).
Vista l'importanza che le informazioni relative allo sviluppo ed al concreto funzionamento dei sistemi ad alto rischio rivestono ai fini di assicurare il rispetto della disciplina proposta, in secondo luogo, si richiede la predisposizione di una completa e costantemente aggiornata documentazione tecnica, che permetta di valutarne anche l'impatto sui diritti fondamentali (art. 11)[33]. Inoltre, è disposto l'obbligo di predisporre la funzionalità per la registrazione automatica dei *logs* (una sorta di diario di bordo della navigazione del sistema) in modo da mantenere il tracciamento del funzionamento del sistema e delle operazioni svolte e verificarne l'appropriatezza per tutto il ciclo di vita (art. 12).

6.4. La trasparenza

Uno dei principali e più noti problemi che i sistemi avanzati di IA presentano riguarda la cd. *black box*. Si tratta del fatto che dispositivi capaci di *machine* o *deep learning* o che utilizzano reti neurali, sono caratterizzati da una estrema complessità che rende praticamente impossibile tracciare la catena dei passaggi seguiti per generare il risultato finale. Se l'input e l'output del processo sono noti non lo è, a motivo della sostanziale opacità delle dinamiche interne allo stesso, l'*iter* che ha generato la decisione né le modifiche che il dispositivo ha autonomamente

---

L'intelligenza artificiale in sanità: tra prospettive e nuovi diritti, entrambi in A. D'ALOIA (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, cit., rispettivamente 423-450 e 451-459; oltre che il numero speciale 2/2021 di questa *Rivista* a cura di L. PALAZZANI su *The impact of Covid-19 on informed consent*. Sull'impatto della IA in campo politico, fra gli altri, M. FASAN, *Intelligenza artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico*, in A. D'ALOIA (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, cit., 345-366; A. CARDONE, *"Decisione algoritmica" vs. decisione politica? AI, legge, democrazia*, 2021.

[32] Le pratiche di gestione dei dati devono concernere: a) le scelte rilevanti operate nel design del sistema; b) la raccolta dei dati; c) le operazioni preparatorie quali l'annotazione, l'etichettatura, la pulizia, l'arricchimento e l'aggregazione dei dati; d) la formulazione degli assunti ricavabili dai dati; e) un accertamento previo della disponibilità, quantità e appropriatezza dei dati necessari; f) un esame in vista di possibili *biases*; g) la individuazione di possibili *gaps*.

[33] Le componenti della documentazione sono precisate all'annesso IV.

assunto per raggiungere con maggior efficacia il compito assegnato[34].

Tale fenomeno presenta una particolare criticità: impedendo la conoscibilità delle singole fasi interne al procedimento e pregiudicando la possibilità di un controllo sulla congruità delle motivazioni alla base della decisione assunta, la *black box* si pone come un ostacolo importante ad una piena legittimazione e ad un complessivo riconoscimento delle menzionate tecniche di IA (Trustworthy AI)[35].

A fronte di tale pericolo, sulla linea di quanto adottato in Canada[36], la proposta di regolamento prevede che ogni sistema ad alto rischio sia disegnato e sviluppato in modo da assicurare un appropriato livello di trasparenza (*sufficiently transparent*). Se tale qualificazione può risultare vaga, va detto che imporre una totale trasparenza avrebbe imposto un obbligo attualmente insostenibile dal punto di vista tecnologico, vista la impossibilità di rendere del tutto decifrabile e intelligibile il processamento interno. Pur a fronte di tale necessitata approssimazione, comunque, la proposta indica le componenti minime che devono informare tale requisito. Ogni sistema ad alto rischio, così, deve essere accompagnato da istruzioni per l'uso che contengano, del dispositivo, informazioni sintetiche, complete, corrette e comprensibili agli utenti. In particolare, oltre ai contatti del *provider*, dovranno essere specificati elementi come gli scopi previsti, i livelli di accuratezza, robustezza e cybersicurezza su cui il dispositivo è stato testato, le caratteristiche degli *input data* e dei *training*, *validation* e *testing data*, la durata attesa e ogni circostanza che possa condurre ad un rischio per la salute, la sicurezza o altri diritti fondamentali (art. 13).

In ogni caso, la previsione di un appropriato livello di trasparenza fra i requisiti necessari perché un dispositivo ad alto rischio possa essere immesso sul mercato europeo va salutata con favore. Al riguardo, larga parte della dottrina ha già da tempo indicato la necessità che il problema della *black box* venga affrontato con misure proporzionate che, senza imporre un obbligo

[34] Cfr., per tutti, F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, 2015.

[35] Molti sono i documenti al riguardo; per tutti, cfr. il report *Ethics guidelines for trustworthy AI*, presentato dall'*High-Level Expert Group on AI* della Commissione europea nell'aprile 2019: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Fra i molti in dottrina, A. BIBAL et al., *Legal requirements on explainability in machine learning*, in *AI and Law*, 29, 2021, 149-169.

[36] La *Directive on Automated Decision Making* (in vigore dall'aprile del 2019 e significativamente soggetta a revisione ogni 6 mesi) prevede «a meaningful explanation to affected individuals of how and why the decision was made»: «In addition to any applicable legal requirement, [any Automated decision Making process must ensure] that a meaningful explanation is provided for common decision results»: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

irrealizzabile, a motivo delle caratteristiche tecnologiche menzionate, permetta una conoscibilità di base della logica complessiva adottata dal dispositivo.

Al riguardo, si potrebbe proporre un diritto in capo ai destinatari della decisione anche parzialmente automatizzata a conoscere perlomeno la logica di fondo alla base della stessa[37]. Solo in questo modo, infatti, verrebbero resi possibili il controllo e la verifica dei motivi alla base dell'output generato; controllo e verifica senza i quali risulta difficile considerare qualsiasi dispositivo affidabile. In alcuni settori come l'amministrazione della giustizia, peraltro, ogni decisione che si poggiasse sulla IA in mancanza di una qualche forma di motivazione sarebbe automaticamente illegittima[38].

6.5. La supervisione umana

Un secondo tema che ha fortemente impegnato la riflessione etica e giuridica internazionale è costituito dalla possibilità che i più avanzati sistemi di IA possano svolgere le funzioni assegnate con crescenti gradi di autonomia rispetto al controllo umano. Tale circostanza è giudicata in maniera assai differente a seconda dei contesti di impiego e dei presupposti di partenza. Tipicamente, un dispositivo che funzioni in totale autonomia, a meno che non si tratti di funzioni operative semplicissime e che non vengano interessati i diritti fondamentali, suscita oggi una diffusa perplessità. Secondo alcuni, però, è bene limitare l'autonomia dei dispositivi solo a motivo dei difetti che ancora segnano la tecnologia; ma nel momento in cui si sarà risolto il problema della *black box* e si potranno scongiurare i rischi di errori e di *bias*, sarà bene assegnare le attività oggi assolte da "imperfetti" esseri umani a macchine che avranno raggiunto

[37] A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in questa *Rivista*, 1, 2019, 76; C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in questa *Rivista*, special issue, 2, 2019, 723. In generale, cfr. il Focus dedicato a IA e amministrazione della giustizia, ospitato in questa *Rivista*, 2, 2021, 359-417.

[38] Il riferimento è all'art. 111, sesto comma della Costituzione, secondo cui «Tutti i provvedimenti giurisdizionali devono essere motivati». Al riguardo, fra gli altri, C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE online*, 3, 2020, 3369-3389: http://www.dpceonline.it/index.php/dpceonline/article/view/1082/1038; F. DONATI, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020, 415-436: https://www.rivistaaic.it/images/rivista/pdf/1_2020_Donati.pdf; S. ARDUINI, *La "scatola nera" della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in questa *Rivista*, 2, 2021, 453-479. Interessante, al riguardo, il divieto francese, assistito dalla reclusione fino a cinque anni, di qualsiasi utilizzo di dati che permettano di identificare i magistrati, e le sedi giudiziarie, al fine di predire il possibile esito della relativa giurisprudenza (art. 33 della *Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*, che stabilisce una modifica ai commi 1 e 2 dell'articolo L. 10 del *code de justice administrative*). In tema, M. FASAN, *L'intelligenza artificiale nella dimensione giudiziaria. Primi profili giuridici e spunti dall'esperienza francese per una disciplina dell'ai nel settore della giustizia*, in corso di stampa nella *Rivista Gruppo di Pisa*.

la capacità di funzionare in modo perfettamente razionale, coerente e imparziale. Secondo altri, viceversa, una lunga serie di funzioni delicate, dalla medicina alla giustizia, richiedono una componente ed una responsabilità tipicamente umane e non delegabili[39].

Da questo ultimo punto di vista, già il regolamento generale sulla protezione dei dati 2016/679 (*General Data Protection Regulation*: GDPR) ha attribuito all'interessato 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»

(art. 22)[40]. La portata di tale diritto, peraltro, è nel GDPR fortemente depotenziata dalle eccezioni relative ai casi in cui la decisione «sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento» oppure «si basi sul consenso esplicito dell'interessato». Tale ultima clausola, in particolare, tende a svuotare di tutela sostanziale gran parte del diritto a una decisione non completamente automatizzata, visto la logica disinformata con cui gestiamo regolarmente i nostri consensi nei confronti dei fornitori dei servizi internet (*blind consent*)[41].

È quindi apprezzabile che l'AIA abbia ripreso e precisato tale elemento, disponendo come ogni sistema ad alto rischio debba essere programmato e sviluppato in modo da assicurare un'efficace supervisione umana (*human oversight*: art. 14). In particolare, si prevede che la presenza umana sia mirata a prevenire e minimizzare qualsiasi rischio per la salute, la sicurezza o altri diritti fondamentali che possano concretizzarsi anche quando il sistema sia utilizzato in conformità ai requisiti della proposta o in modo comunque ragionevolmente prevedibile.

Anche in questo caso, nell'articolare la posizione in oggetto, il testo cerca di coniugare rigore e flessibilità, indicando come la persona incaricata della supervisione debba essere in grado di

[39] Il dibattito non è in questa sede riassumibile nemmeno per sommi capi. Si permetta il rinvio a C. CASONATO, *AI and Constitutionalism: The Challenges Ahead*, in B. BRAUNSCHWEIG, M. GHALLAB, (Eds.), *Reflections on Artificial Intelligence for Humanity*, cit., 138 ss.

[40] Il considerando 71 del GDPR dispone che l'interessato «dovrebbe avere il diritto di non essere sottoposto a una decisione, che possa includere una misura che valuti aspetti personali che lo riguardano, che sia basata unicamente su un trattamento automatizzato e che produca effetti giuridici che lo riguardano o incida in modo analogo significativamente sulla sua persona, quali il rifiuto automatico di una domanda di credito online o pratiche di assunzione elettronica senza interventi umani». L'art. 22 del GDPR trova un diretto precedente nell'art. 15 della (ora abrogata) direttiva europea n. 95/46/CE relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati.

[41] Si permetta il rinvio a C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, cit., 719.

svolgere, in maniera appropriata secondo le circostanze, determinate funzioni (art. 14, comma 4)[42]. In primo luogo, dovrà poter comprendere appieno le capacità e i limiti del sistema, in modo da poterne monitorare le operazioni e rilevare e affrontare tempestivamente possibili anomalie e conseguenze inaspettate (lett. a). Allo stesso modo, il supervisore dovrà essere messo nella condizione di poter interpretare correttamente gli output del sistema (lett. c). Di particolare interesse, in secondo luogo, il fatto che una effettiva sorveglianza umana presupponga la consapevolezza della tendenza ad affidarsi in modo automatico e acritico alla IA, facendo un affidamento eccessivo sui relativi risultati (lett. b)[43]. Rispetto a questa constatazione, la proposta di regolamento indica come le misure adottate per la *human oversight* debbano mettere concretamente il supervisore (umano) nella condizione di non utilizzare il sistema, oppure in quella di non considerarne, superarne o rovesciarne i risultati (lett. d), fino alla possibilità di interromperne il funzionamento (*stop button*: lett. e). Per i sistemi destinati a essere utilizzati per l'identificazione biometrica a distanza in tempo reale e a posteriori di persone fisiche (punto 1(a) dell'annesso III) si prevede, infine, che ogni decisione debba essere verificata e confermata da almeno due supervisori (comma 5).

Tale parte della proposta di regolamento, e in particolare quanto previsto alle lettere b) e d), dimostra un apprezzabile realismo e, al contempo, lancia una sfida culturale impegnativa. In particolare, si riconosce come la mera attribuzione di un diritto ad una decisione che sia frutto di una verifica umana (un diritto all'umanità nella decisione, potremmo sintetizzare) sia destinata a rimanere lettera morta (come, temiamo, l'art. 22 del GDPR) se non assistita da una serie di misure che ne rendano concretamente possibile l'applicabilità. Anche in presenza di tale diritto, infatti, ci si è chiesti quale supervisore voglia assumersi il rischio e la responsabilità personale di disattendere una decisione che viene comunemente percepita come esatta e imparziale[44]. Su questa linea, si è efficacemente messo in guardia rispetto ad un «effet moutonnier» (che potremmo tradure come effetto pecorone) in base al quale la decisione sarebbe "catturata" dalla IA, la verifica umana sarebbe svolta secondo modelli di bieco conformismo nei confronti della macchina e il diritto alla supervisione sarebbe formalmente garantito, ma di fatto svuotato di

[42] Logica analoga è utilizzata dalla citata direttiva canadese, in cui si dispone che «an Automated Decision System allows for human intervention, when appropriate».
[43] Al riguardo, già S. RODOTÀ, *Il diritto di avere diritti*, Roma-Bari, 2012, 401 e ss.
[44] Sulla percezione della IA, T. ARAUJO et al., *In AI we trust? Perceptions about automated decision-making by artificial intelligence*, in *AI and Society*, 35, 2020, 611-623.

qualsiasi contenuto garantista[45]. Le articolazioni del diritto previste al comma 4, quindi, contengono una concretezza operativa che può contribuire a combattere efficacemente tale pericolo.

Allo stesso tempo, peraltro, va detto che non basta un pur articolato e dettagliato articolo di legge ad assicurare che chi è investito di una funzione delicata e impegnativa la svolga secondo modalità che potrebbero esporlo a rischi di responsabilità personale anche pesante. Perché il *law in the book* (o il *right in the book*) diventi *law* (*right*) *in action*, è quindi necessario intraprendere un'azione decisa tesa a promuovere una formazione diffusa sulle potenzialità, ma anche sui rischi della IA; una formazione che, coniugata ad un approfondimento sull'etica del lavoro, ponga le basi per una consapevolezza collettiva capace di permettere a quanti sono coinvolti nella *human oversight* di svolgere responsabilmente e serenamente una funzione di concreto ed effettivo controllo ed eventuale scostamento dai risultati del sistema di IA.

6.6. La procedura di verifica di conformità e il ruolo delle autorità pubbliche

Si è osservato nel paragrafo precedente come l'immissione nel mercato di sistemi di intelligenza artificiale ad alto rischio sia condizionata al rispetto dei requisiti fissati nel capitolo II della proposta di regolamento.

Il modello prescelto dalla Commissione per tale verifica è rappresentato dalla procedura di marchio di conformità europea (marchiatura CE), già sperimentata per regolare la circolazione di numerosi prodotti nel mercato europeo. È il caso, per esempio, della direttiva 2009/48 sulla sicurezza dei giocattoli, ma anche della direttiva 2001/95 sulla sicurezza generale dei prodotti[46].

Tramite essa un prodotto è ammesso a circolare nel mercato europeo (o nello spazio economico europeo – SEE – in cui sono ricompresi anche Paesi extra Ue quali Norvegia, Islanda e Lichtenstein) solo dopo l'apposizione del marchio di conformità, possibile dopo che è stata

[45] La suggestione è di A. GARAPON, J. LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, PUF, 2018, 239. Riferendosi all'impiego della IA nell'amministrazione della giustizia, scrivono: «C'est l'effet moutonnier de la justice prédictive: elle pousse au conformisme et réclame plus d'indépendance d'esprit aux juges qui estiment qu'ils doivent aller à contre-courant, c'est-à-dire qui veulent simplement faire leur métier». Cfr. anche A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, cit., 69.

[46] R. TRICKER, *CE Conformity Marking and New Approach Directives*, Butterworth-Heinemann, 2000; D. HANSON, *CE Marking, Product Standards and World Trade*, Edward Elgar, 2005; A. NOWAK-FAR, *National versus European Component in the European Union Product Conformity Regulation*, in R. GRZESZCZAK (editor), *Renationalisation of the Integration Process in the Internal Market of the European Union*, 2018, 37; O. BORRAZ, *Governing Standards: The Rise of Standardization Processes in France and in the EU*, in *Governance*, 20, 2007, 1.

compiuta una procedura di verifica della conformità agli standard e alle regole stabilite dall'Unione effettuata dal produttore stesso o da un organismo certificatore terzo. Diversamente da quanto stabilito per l'immissione nel mercato Ue dei medicinali (dir. 2001/83 per la procedura decentrata e reg. 726/2004 per la procedura accentrata[47]) o dei prodotti OGM (reg. 1829/2003), per i quali pure la componente di rischio è significativa, dunque, non si prevede il rilascio di un'autorizzazione pubblica, nazionale o europea, ma il rispetto dei requisiti posti a protezione della sicurezza, della salute e dei diritti fondamentali dei cittadini è garantita da un'auto-valutazione da parte del soggetto che ha interesse a commercializzare il prodotto.

Sono evidenti i vantaggi immediati di tale scelta regolatoria: se, infatti, l'autorizzazione pubblica dà maggiori certezze e garanzie in ordine alle verifiche sulla sicurezza del prodotto, presentando però costi non irrilevanti in termini amministrativi e di efficienza, la procedura di marchio di conformità europea, ponendo i controlli a carico del produttore, rende più agevole l'immissione del prodotto nel mercato e sposta sugli operatori economici la responsabilità di assicurare il rispetto dei requisiti di sicurezza stabiliti dalla normativa.

La scelta di questa procedura più "leggera", tuttavia, presenta anch'essa un temperamento, che è conseguenza nuovamente di un approccio proporzionato ai rischi. Infatti, per alcuni sistemi di IA, in particolare per quelli biometrici di cui all'annesso III n. 1, la procedura di conformità deve essere condotta da organismi terzi, secondo quanto prevede l'art. 43, comma 1, della proposta AIA[48]. Inoltre, vale comunque, anche per tale aspetto procedurale quanto già osservato sulla modificabilità della disciplina AIA, dato che, nell'ipotesi di una inadeguata tenuta del sistema, gli stessi annessi VI e VII, dedicati alla procedura interna o esterna di verifica di conformità, potranno essere agevolmente sottoposti a revisione[49].

---

[47] Cfr. J. ABRAHAM, *Regulating Medicines in Europe: Competition, Expertise and Public Health*, 2000.

[48] Stabilisce, peraltro, l'art. 43 che, anche nel caso di sistemi biometrici, quando il provider applica gli standard armonizzati di cui all'art. 40 della proposta di regolamento o, se applicabili, le *common specifications* di cui all'art. 41, per dimostrare la conformità del sistema *high r*isk ai requisiti di cui al capitolo II possa scegliere tra la procedura di verifica di conformità interna oppure quella affidata ad un organismo terzo.

[49] Prevede espressamente l'art. 43, comma 5, che «The Commission is empowered to adopt delegated acts in accordance with article 73 for the purpose of updating Annexes VI and VII in order to introduce elements of the conformity assessment procedures that become necessary in the light of technical progress». Ad ulteriore precisazione, il comma 6 stabilisce che «The Commission is empowered to adopt delegated acts to amend paragraphs 1 and 2 in order to subject high-risk AI systems referred to in points 2 to 8 of Annex III to the conformity assessment procedure referred to in Annex VII or parts thereof». Nell'adozione di tali atti delegati la Commissione terrà conto, in particolare, della effettività della procedura di verifica interna dell'annesso VI «in preventing or minimizing the risks to health and safety and protection of fundamental rights posed by such systems as well as the availability of adequate capacities and resources among notified bodies».

La procedura iniziale di controllo della conformità non costituisce, in ogni caso, l'unico momento di controllo dei rischi. La proposta della Commissione, infatti, affianca alle verifiche precedenti all'immissione nel mercato, un sistema di monitoraggio *post market* che ha l'obiettivo di garantire la rispondenza del prodotto ai requisiti stabiliti dalla disciplina nel corso dell'intero ciclo di vita dei sistemi di IA (art. 61). A questo fine, da un lato, viene previsto che ciascuno Stato membro si doti di un'autorità per la gestione del sistema di sorveglianza *post market*; dall'altro, si riconoscono precisi obblighi sia in capo all'utilizzatore, sia in capo al *provider*. Quest'ultimo, come anticipato, deve conservare, per un periodo di tempo adeguato, da individuarsi in base alle caratteristiche del sistema, le registrazioni di funzionamento dell'IA (*automatically generated logs*), che potranno essere oggetto di accesso da parte delle autorità nazionali competenti per finalità di controllo: egli deve inoltre adottare immediatamente le misure correttive opportune in caso di difformità e deve altresì comunicare all'autorità nazionale competente (ed eventualmente all'organismo certificatore) eventuali violazioni riscontrate nel funzionamento del sistema (artt. 20-23).

La disciplina relativa agli organismi certificatori e ai requisiti che questi devono possedere per poter dichiarare la conformità del sistema, così come la normativa sulle relazioni tra questi ultimi e le autorità nazionali competenti, ricalca quella prevista in altre discipline sul marchio di conformità comunitario[50]. Ciascuno Stato membro, in particolare, ha il compito di designare un'autorità di notifica responsabile per l'istituzione e l'esecuzione delle procedure necessarie per la valutazione e la notifica degli organismi certificatori (art. 30), la quale deve operare secondo modalità che garantiscano l'assenza di qualsiasi conflitto di interesse con gli organismi certificatori.

L'art. 33, inoltre, stabilisce dettagliatamente i requisiti strutturali che l'organismo certificatore è chiamato a rispettare per poter esercitare i propri compiti: oltre ad avere le risorse organizzative e finanziarie adeguate alla verifica di conformità, esso deve essere indipendente rispetto al *provider* del sistema ad alto rischio soggetto alla sua valutazione (art. 33, comma 4). Tale indipendenza deve altresì sussistere rispetto ad ogni altro operatore avente un interesse economico rispetto all'applicazione esaminata, oltre che nei confronti di eventuali concorrenti del *provider*. L'autorità di notifica nazionale comunica alla Commissione e agli altri Stati

[50] Si consideri, ad esempio, la disciplina stabilita sugli organismi notificati e sulle autorità di notifica prevista nella citata direttiva 2009/48 sulla sicurezza dei giocattoli (agli artt. 22 e ss.).

membri la lista degli organismi di valutazione della conformità utilizzando lo strumento elettronico di notifica elaborato e gestito dalla Commissione. In caso di dubbi sulla rispondenza degli organismi di verifica ai requisiti previsti nel regolamento, la Commissione potrà effettuare indagini e verifiche suscettibili di produrre misure correttive, comprensive della revoca della notificazione qualora ciò risulti necessario (art. 37).

Agli organismi certificatori spetta assicurare la propria collaborazione all'autorità di notifica per tutta la durata della loro attività, fornendo la documentazione del *provider* affinché l'autorità stessa possa compiere le eventuali attività di verifica, di monitoraggio e di sorveglianza idonee ad assicurare la correttezza dei procedimenti seguiti. Essi, inoltre, hanno significativi obblighi informativi, sia nei confronti dell'autorità di notifica sia nei confronti degli altri organismi certificatori. Rispetto alle loro decisioni, gli Stati membri sono chiamati a garantire la possibilità per chiunque abbia un interesse legittimo (*legitimate interest*) di esperire una procedura di revisione (*appeal procedure*) (art. 45). È evidente dall'impiego del termine *appeal*, che non si tratta di un mero sindacato di legittimità, ma di un ricorso che possa portare alla revisione di merito della decisione dell'organismo certificatore, il quale potrebbe anche chiamare in causa una specifica divisione amministrativa dell'autorità di notifica.

7. La governance complessiva e la banca dati sull'IA

La proposta della Commissione dedica il titolo VI alla *Governance*, il titolo VII alla banca dati europea e il titolo VIII alla vigilanza *post market*, con il relativo regime sanzionatorio. Benché nella logica di questo primo commento non sia possibile approfondire ciascuna di queste parti, un sintetico sguardo d'insieme anche a questi tratti della disciplina sembra necessario per dare conto del disegno regolatorio complessivo.

Con riguardo alla *governance*, va rilevato anzitutto come la Commissione abbia scelto di attribuire al livello amministrativo statale i poteri di attuazione della disciplina. Benché, infatti, la proposta AIA contempli l'istituzione dello *European Artificial Intelligence Board* (art. 56), sono le autorità nazionali ad assicurare «the application and implementation of this Regulation» (art. 59 comma I), secondo il noto modello dell'amministrazione comunitaria indiretta[51]. In tale

[51] Sui modelli di amministrazione europea cfr. P. CRAIG, *Amministrazione comunitaria. Storia, tipologia, e "accountability"*, in M. D'ALBERTI (a cura di), *Le nuove mete del diritto amministrativo*, 2010, 11; S. CASSESE, *Poteri divisi: amministrazione europea e amministrazioni nazionali*, in *Studi in ricordo di Enzo Capaccioli*, 1988, 23;

logica, la Commissione, pur lasciando agli Stati la decisione in ordine all'assetto istituzionale ottimale, incide in vario modo sulle scelte organizzative interne, sia suggerendo l'individuazione, tra le *national competent authorities*, di un'unica autorità di sorveglianza e di notifica (considerato l'onere di precisare specificamente le ragioni, amministrative e organizzative, per la designazione di più autorità, stabilito nei commi 2 e 3 dell'art. 59)[52], sia indicando, nel comma 4, le competenze del personale da impiegare presso queste autorità, comprensive di «in-depth understanding of artificial intelligence technologies, data and data computing, fundamental rights, health and safety risks and knowledge of existing standards and legal requirements».

L'indicazione è interessante sotto due profili: da un lato perché mostra la crescente tendenza del diritto dell'Unione a conformare i sistemi amministrativi nazionali anche in relazione ai profili, organizzativi e procedurali, che storicamente erano considerati terreno riservato all'autonomia procedurale degli Stati[53]; e dall'altro, perché conferma la necessità, sempre più avvertita, di competenze e abilità trasversali e interdisciplinari all'interno delle pubbliche amministrazioni in ragione della complessità tecnica delle funzioni da esercitare.

Sul fronte dell'amministrazione europea, lo *European Artificial Intelligence Board* esercita, invece, funzioni di consulenza e assistenza finalizzate a favorire la cooperazione tra le diverse autorità degli Stati e la Commissione, a promuovere l'analisi e la soluzione di questioni emergenti nell'attuazione del regolamento e a garantire un'applicazione conforme del Regolamento. Esso dunque non ha compiti di amministrazione della disciplina europea né si

---

S. CASSESE, *Diritto amministrativo europeo e diritto amministrativo nazionale: signoria o integrazione?*, in S. CASSESE, *Il diritto amministrativo: storia e prospettive*, 2010, 399; G. FALCON, *Dal diritto amministrativo nazionale al diritto amministrativo comunitario*, in *Rivista italiana di diritto pubblico comunitario*, 1991, 351; E. CHITI, C. FRANCHINI, *L'integrazione amministrativa europea*, 2003; G. DELLA CANANEA, *L'Unione europea. Un ordinamento composito*, Roma-Bari, 2003; L. DE LUCIA, B. MARCHETTI, *L'amministrazione europea e le sue regole*, 2015.

[52] Fa eccezione quanto stabilito dall'art. 63 comma 4 con riguardo al mercato finanziario, per il quale si prevede che «for AI systems placed on the market, put into service or used by financial institutions regulated by Union legislation on financial services, the market surveillance authority for the purposed of this regulation shall be the relevant authority responsible for the financial supervision of those institution under that legislation». Il comma 5 stabilisce, altresì, vincoli in ordine alla individuazione dell'autorità di sorveglianza con riguardo ai sistemi biometrici utilizzati per pubblica sicurezza (*law enforcement*).

[53] S. CASSESE, *L'influenza del diritto amministrativo comunitario sui diritti amministrativi nazionali*, in *Rivista italiana di diritto pubblico comunitario*, 1993, 329; M.P. CHITI, *Le forme di azione dell'amministrazione europea*, in F. BIGNAMI, S. CASSESE (a cura di), *Il procedimento amministrativo*, 2004, Quaderno della Rivista trimestrale di diritto pubblico, 2004; D.U. GALETTA, *L'autonomia procedurale degli Stati membri dell'Unione europea: "Paradise lost"?*, 2009; L. SALTARI, *Amministrazioni nazionali in funzione comunitaria*, 2007; P. CHIRULLI, *Amministrazioni nazionali ed esecuzione del diritto europeo*, in L. DE LUCIA, B. MARCHETTI, *L'amministrazione europea e le sue regole*, cit., 145 e ss.; nello stesso volume v. anche S. TORRICELLI, *L'europeizzazione del diritto amministrativo italiano*, 247 e ss.

pone in una posizione di preminenza rispetto alle autorità nazionali. La sua composizione è il riflesso della logica di cooperazione istituzionale tra le diverse amministrazioni nazionali e l'autorità europea di riferimento[54]. Siedono, infatti, nel *Board* i vertici delle amministrazioni nazionali di vigilanza ed il Garante europeo per la protezione dei dati, mentre la presidenza è riservata alla Commissione.

I poteri del *Board* sono indicati espressamente all'art. 58 della proposta AIA e mirano a garantire un coordinamento tra le diverse autorità nazionali nella formazione e condivisione di buone pratiche, l'elaborazione comune degli standard e delle *common specifications* in materia di IA e la predisposizione di linee guida in materia di sanzioni. Pur non dando vita ad atti giuridicamente vincolanti, le raccomandazioni ed i pareri del *Board* appaiono fondamentali nel supportare la Commissione nell'esercizio di alcune sue funzioni chiave per la tenuta unitaria del sistema. Quest'ultima, infatti, oltre ad esercitare i poteri normativi attribuiti dall'art. 73 della proposta, da esercitarsi, come detto, in conformità all'art. 5 del Regolamento 182/2011 e dunque seguendo la procedura di esame della Comitologia[55], ha sia importanti funzioni di predisposizione del piano per il monitoraggio *post market* (art. 61[56]), sia compiti di tenuta e controllo della banca dati europea sui sistemi di intelligenza artificiale ad alto rischio (art. 60). Tale banca dati, che sarà creata e gestita dalla Commissione in collaborazione con gli Stati membri, deve contenere tutte le informazioni riguardanti i sistemi di IA ad alto rischio destinati al mercato europeo che il provider è tenuto a comunicare[57].

La funzione di tale archivio è evidente: consentire la pubblicità e la controllabilità dei sistemi circolanti nell'Unione non solo a vantaggio delle Istituzioni competenti, ma anche del pubblico,

[54] L. DE LUCIA, *Strumenti di cooperazione per l'esecuzione del diritto europeo*, in L. DE LUCIA, B. MARCHETTI, *L'amministrazione europea e le sue regole*, cit., 171; E. CHITI, *Le Agenzie europee. Unità e decentramento nelle amministrazioni comunitarie*, Padova, 2002.

[55] La procedura d'esame, come noto, comporta che la Commissione non possa adottare l'atto normativo se il Comitato non esprime parere favorevole sulla proposta. A fronte del parere negativo del Comitato interessato, dunque, essa ha due alternative: o ritira la proposta e la ripresenta in una versione modificata entro due mesi dall'adozione del parere negativo, oppure presenta il progetto di atto entro un mese al comitato di appello per una nuova espressione di parere (art. 5 comma III).

[56] Stabilisce, in particolare, l'art. 61 comma 3 che «the post-market monitoring system shall be based on a post-market monitoring plan. The post-market monitoring plan shall be part of the technical documentation referred to in Annex IV. The Commission shall adopt an implementing act laying down detailed provisions establishing a template for the post-market monitoring plan and the lists of elements to be included in the plan». Sulla base di questo piano per la sorveglianza dei sistemi ad alto rischio il provider dovrà predisporre un sistema di vigilanza post -market in base al quale sono raccolti e analizzati i dati di funzionamento di un sistema di IA per tutto il ciclo di vita allo scopo di verificare la continua rispondenza del sistema ai requisiti previsti dal capitolo II del Regolamento.

[57] La registrazione nella banca dati costituisce una delle obbligazioni del *provider*, che, ai sensi dell'art. 51, deve farvi fronte come condizione per la commercializzazione o la messa in servizio del sistema di IA.

considerato che l'art. 60, comma 3, espressamente prevede l'accessibilità al pubblico delle informazioni in esso contenute. Tuttavia, esso può porre anche problemi di tutela della riservatezza: non a caso, la proposta cerca di conciliare accessibilità e privacy nel comma 4 della norma, in cui è precisato che «the EU database shall contain personal data only insofar as necessary for collecting and processing information in accordance with this regulation»[58]. Ovviamente il grado di effettiva pubblicità e controllabilità della Banca Dati dipenderà da moti fattori, tra cui la quantità di informazioni destinate ad esservi inserite. Non è chiaro, infatti, se la registrazione comporti la conoscibilità di tutta la documentazione del *provider* (art. 50) oppure la sola dichiarazione di conformità con il numero identificativo dell'organismo certificatore (quando previsto).

8. Gli strumenti di enforcement

La proposta di regolamento stabilisce, poi, poteri e misure per assicurare il rispetto dei requisiti e degli obblighi previsti da parte dei *providers*. In particolare, l'autorità di sorveglianza che abbia conoscenza di una difformità del sistema di IA con i requisiti e le obbligazioni fissati nel regolamento potrà imporre al soggetto interessato l'adozione delle misure appropriate per far cessare la violazione, ritirare il sistema di IA dal mercato o richiamarlo per un tempo ragionevole commisurato alla natura del rischio (art. 65, comma 2). Di tali azioni deve essere data comunicazione alla Commissione e agli altri Stati membri, per l'eventuale attivazione di misure corrispondenti di salvaguardia[59].

La disciplina sulle sanzioni previste in caso di violazione della disciplina è prevista, infine, negli artt. 71 e 72 della proposta AIA. La disciplina specifica per la determinazione di tali sanzioni è

[58] La creazione della banca dati è strategica per la attuazione di tutta la proposta AIA e dovrà essere predisposta in termini rapidissimi, senza che si ripeta l'inescusabile inerzia nella predisposizione del *Clinical Trials Information System* (CTIS) che ancora impedisce l'applicazione del regolamento 536/2014 in tema di sperimentazione clinica.

[59] Tale comunicazione ha il senso di aprire il confronto con gli Stati membri, secondo la procedura di salvaguardia ben nota nel processo di integrazione europeo (la medesima è contemplata, ad esempio, anche nella direttiva sulla sicurezza dei giocattoli più volte ricordata, 2009/48 CE). Se gli Stati e la Commissione non sollevano alcuna obiezione rispetto alle misure adottate dall'autorità di sorveglianza statale entro 3 mesi dal ricevimento, la misura è considerata giustificata. In tal caso, tutte le autorità di sorveglianza degli Stati membri assicureranno le misure restrittive appropriate (compreso l'eventuale ritiro del sistema di IA) dal mercato senza ritardo. Viceversa, se vengono sollevate obiezioni, la Commissione valuterà con gli Stati interessati e gli operatori economici coinvolti la misura adottata a livello nazionale e deciderà se detta misura è giustificata oppure no entro 9 mesi dal momento in cui le è stata comunicata, informando lo Stato interessato. Nel caso in cui la Commissione consideri giustificata la misura nazionale, tutti gli Stati membri sono tenuti ad assicurare che il sistema in questione sia ritirato dal mercato, informandone la Commissione. Nel caso, invece, in cui la misura sia considerata ingiustificata dalla Commissione, lo Stato interessato è tenuto a ritirarla (art.66).

demandata agli Stati membri, che devono determinarla in una misura adeguata ad assicurarne la natura effettiva, proporzionata e dissuasiva. Tuttavia, la proposta della Commissione individua, secondo un approccio proporzionato, diversi regimi sanzionatori a seconda della gravità delle violazioni.

In particolare, viene stabilita una sanzione fino a 30 milioni di euro o (in caso di società) fino al 6% del fatturato totale annuo nel caso di immissione nel mercato o di messa in servizio di sistemi di IA vietati (art. 5) o di violazione della disciplina stabilita nell'art. 10 relativamente a *data* e *data governance*; ed una sanzione fino a 20 milioni di euro o una percentuale del 4% del fatturato societario totale per il mancato rispetto degli altri requisiti fissati per i sistemi ad alto rischio (cap. II). Un regime sanzionatorio ulteriore, e più leggero, viene stabilito per la violazione da parte dei *providers* degli obblighi informativi (art. 71, comma 5), con una sanzione amministrativa in tal caso ulteriormente dimezzata (10 milioni di euro e 2 % del fatturato).

Secondo la proposta AIA, la determinazione della sanzione da comminare non dovrebbe essere mai automatica. Tre gli elementi che dovrebbero essere considerati, si contano: la natura, gravità e durata della violazione, oltre alle sue conseguenze; la circostanza che la sanzione sia già stata comminata da un'altra autorità di sorveglianza per la stessa violazione e allo stesso operatore; la dimensione del mercato in cui la violazione è stata commessa. Nel caso in cui le violazioni siano state compiute da un'autorità pubblica, lo Stato deve valutare se e in che misura applicare le medesime sanzioni (art. 71, comma 6).

Una disciplina a parte è prevista per il caso in cui a violare il regolamento sia un'istituzione, un'agenzia o un organismo amministrativo dell'Unione. In tal caso, infatti, tra i criteri da considerare nella determinazione della sanzione applicabile si prevede, oltre agli elementi sopra menzionati dell'art. 71 comma 6, la volontà mostrata dall'autorità di collaborare con il Garante per i dati personali sia nell'opera di mitigazione degli interessi avversi sia nell'assicurare la *compliance* alle misure adottate dal medesimo Garante (art. 72 comma I).

La norma in questione interessa anche il tema dell'applicazione (e della estensione) delle sanzioni alle amministrazioni pubbliche, stabilendo che, nel caso di violazioni del regolamento da parte delle amministrazioni dell'Unione, il massimale sia fissato fino a 500.000 euro per le violazioni più gravi, riferite alla messa in servizio di IA vietata o in contrasto con le previsioni sui dati e la *data governance*, e fino a 250.000 euro per le altre difformità rispetto ai requisiti

previsti per i sistemi ad alto rischio.

Ultimo aspetto di rilievo consiste nel richiamo delle fondamentali garanzie del contraddittorio nei procedimenti sanzionatori. La proposta AIA parla della «opportunity to be heard on the matter regarding the possible infringement», e della necessità che l'autorità garante dei dati basi «his or her decisions only on elements and circumstances on which the parties concerned have been able to comment» (art. 72, comma 4).

9. Conclusioni: una proposta sostenibile?

Già oggi, la IA facilita e allo stesso tempo condiziona le nostre vite più di quanto possiamo percepire. In un futuro assai vicino, grazie ai continui progressi scientifici e al moltiplicarsi delle applicazioni tecnologiche, essa permeerà la maggior parte delle nostre attività, da quelle banali a quelle più delicate e complesse[60]. La limitata visibilità e la velocità di tale rivoluzione tecnologica non favoriscono l'elaborazione di un quadro regolatorio specifico; quadro che, d'altro canto, si impone a motivo della pervasività del fenomeno e del suo impatto sul nostro vivere[61]. Di fronte a tale situazione paradossale, la scelta dell'Unione europea di dotarsi di un regolamento generale è coraggiosa e apprezzabile.

Nel contenuto, la disciplina proposta pare disegnata in base ai caratteri specifici dell'oggetto da regolare. In particolare, la complessità e la pluralità della tecnologia impiegata hanno suggerito un approccio basato sui rischi; i suoi costanti mutamenti hanno consigliato l'utilizzo di strumenti tipicamente caratterizzati da flessibilità e adattabilità; l'esigenza di una tutela effettiva ha imposto una copertura soggettiva particolarmente ampia; il potenziale positivo a fronte dell'entità delle incognite ha ispirato un bilanciamento fra strumenti di sostegno e misure di garanzia che pare complessivamente ragionevole e dotato di una buona dose di resilienza.

Il percorso della proposta AIA è all'inizio e certamente il dibattito che ne scaturirà potrà apportare elementi di miglioramento del testo. In molte parti, infatti, la proposta AIA si affida a clausole generali che dovranno essere interpretate in modo almeno tendenzialmente uniforme per non pregiudicare le finalità di armonizzazione e di certezza del diritto su cui si basa l'affidabilità dell'impiego della tecnologia e senza le quali si innescherebbe un insostenibile

[60] Si tratta di un fenomeno collegato alla cd. *onlife*, su cui L. FLORIDI, (a cura di), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, 2015.

[61] Fra gli altri, S. GREENSTEIN, *Preserving the rule of law in the era of artificial* intelligence, in *AI and Law*, 17 July 2021.

contenzioso giudiziario.

Inoltre, i codici di condotta e le *sandboxes* avranno un ruolo cruciale nell'adattare le disposizioni previste ad una realtà complessa, sempre in movimento e in forte crescita come quella della IA. Al riguardo, sarà necessario alimentare un confronto fra diverse forme di sapere e un dibattito dai forti tratti interdisciplinari, a partire dal superamento dell'incomunicabilità delle diverse terminologie specialistiche. E anche i modelli di formazione e di carriera universitaria italiani dovranno essere profondamente ripensati, al fine di aprire i recinti dei propri settori scientifico-disciplinari ad una realtà che può essere capita, interpretata e disciplinata solo se approcciata in termini plurali e comunicanti[62].

Allo stesso tempo, pare cruciale la necessità di attivare percorsi di sensibilizzazione e formazione di base e avanzata sui vantaggi e sui rischi della IA[63]. Per l'estensione e la profondità con cui tale tecnologia sempre più influirà sul nostro vivere, infatti, è in gioco non solo la mera certificazione di nuovi dispositivi tecnologici, ma l'individuazione delle coordinate su cui costruire un nuovo modello di società; una società in cui componenti umane e non umane saranno chiamate a convivere in modo sostenibile, facendo al meglio, secondo un rinnovato principio di sussidiarietà, quanto più si addice alle rispettive caratteristiche con l'obiettivo di un mondo complessivamente più equo[64].

[62] Su cui, fra gli altri, A. SANTOSUOSSO, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, cit., 149-175; J. ROCHEL, F. EVÉQUOZ, *Getting into the engine room: a blueprint to investigate the shadowy steps of AI ethics*, in *AI and Society*, 36, 2021, 609-622; G. PASCUZZI, *Quale formazione per la ricerca interdisciplinare?*, in questa *Rivista*, 1, 2021, 337-343.

[63] In generale, S.B. KULIKOV, A.V. SHIROKOVA, *Artificial intelligence, culture and education*, e D. SCHIFF, *Out of the laboratory and into the classroom: the future of artificial intelligence in education*, entrambi in *AI and Society*, 36, 2021, rispettivamente 305-318 e 331-348.

[64] Fra i molti, oltre a quanto già citato, si vedano J. MACLURE, S. RUSSEL, *AI for Humanity: The Global Challenges*, e A. DENGEL, L. DEVILLERS, L.M. SCHAAL, *Augmented Human and Human-Machine Co-evolution: Efficiency and Ethics*, entrambi in B. BRAUNSCHWEIG, M. GHALLAB (Eds.), *Reflections on Artificial Intelligence for Humanity*, cit., rispettivamente 116-126 e 203-227; R. CINGOLANI, D. ANDRESCIANI, *Robots, macchine intelligenti e sistemi autonomi: analisi della situazione e delle prospettive*, in G. ALPA (a cura di), *Diritto e intelligenza artificiale*, cit., 23-56; la *Presentazione* di L. VIOLANTE e i contributi del *Focus sull'intelligenza artificiale* raccolti in questa *Rivista*, 3, 2019, 179-254, oltre al *Forum* dedicata a *AI and Law*, in questa *Rivista*, 1, 2020, 463-515.