medialaws

Anticipazioni

La proposta di regolamento europeo in materia di Intelligenza Artificiale: verso una "discutibile" tutela individuale di tipo *consumer-centric* nella società dominata dal "pensiero artificiale"

Daniela Messina

La proposta di regolamento europeo in materia di Intelligenza Artificiale: verso una "discutibile" tutela individuale di tipo *consumer-centric* nella società dominata dal "pensiero artificiale"

Abstract

Consapevole delle imponenti sfide derivanti dalla diffusione sempre più ampia del "pensiero artificiale" che è penetrato nel tessuto delle società tecnologicamente avanzate innovandone processi produttivi, tecniche comunicative, meccanismi relazionali e modalità di erogazione dei servizi pubblici e privati, l'Unione europea ha avviato, negli ultimi anni, una articolata strategia in materia di Intelligenza Artificiale il cui primo e rilevante "approdo" normativo è costituito dalla proposta di regolamento sull'IA COM(2021) 206. L'obiettivo perseguito è quello di contemperare l'esigenza di massimizzare le opportunità offerte da tale tecnologia con l'imperativo di incanalare tale complesso fenomeno all'interno di un quadro regolamentare ed etico adeguato, basato su principi e valori condivisi a livello europeo e coerente con la Carta dei diritti fondamentali. Tuttavia, nonostante le promettenti premesse orientate a garantire uno sviluppo dell'IA di tipo *human-centric*, tale atto risulta essere debole proprio nella costruzione dell'impianto di garanzie destinato agli utenti finali che subiscono la decisione "artificiale". La proposta, infatti, sembra delineare un sistema *consumer-centric* nel quale la tutela degli utenti viene a strutturarsi quasi esclusivamente attorno agli operatori che agiscono all'interno dell'ecosistema dell'intelligenza artificiale. Tale specifica impostazione impone una seria riflessione sulle molteplici criticità che possono emergere in un panorama digitale che assume sfumature sempre più deterministiche e rischia di comprimere la sfera di libertà del singolo con inevitabili ripercussioni sulla sua dignità, incidendo sul percorso evolutivo della sua persona nella duplice dimensione di individuo in grado di autodeterminarsi e come partecipante attivo e consapevole di una società democratica.

Aware of the challenges arising from the ever-widening diffusion of "artificial thought", the European Union has recently launched an articulated strategy on Artificial Intelligence. In this perspective, the proposal for an EU regulatory framework on artificial intelligence COM (2021) 206 constitutes the first relevant regulatory act aiming to balance the need to make the most of the opportunities offered by AI with the imperative of channeling this complex phenomenon within an adequate ethical and legal framework, based on shared values and consistent with the Charter of Fundamental Rights of the European Union. However, despite the promising premises aimed at guaranteeing a human-centric European development of AI, this act appears weak precisely in building the protection system intended for the addressees of the "artificial decision". The proposal seems to outline a consumer-centric approach in which the protection is articulated almost exclusively around the operators who act within the AI ecosystem. This specific approach requires a deep reflection in a panorama that takes on increasingly deterministic shades and risks increasingly compressing the individual's sphere of freedom with inevitable repercussions on his dignity both as an autonomous individual and an active and conscious participant in a democratic society.

Sommario

Keywords

Intelligenza artificiale – dignità – diritto all'autodeterminazione – identità personale - tutela dei dati personali

1. L'Intelligenza Artificiale come "trasformatore intelligente" di nuova conoscenza

Ci sono delle innovazioni che più di altre sono in grado di cambiare il corso della storia. La loro portata rivoluzionaria è talmente ampia e significativa da travolgere in maniera improvvisa ed innovare in modo indelebile dinamiche e processi anche consolidati, delineando nuove traiettorie evolutive, prima inimmaginabili.

Con il suo incedere impetuoso e capillare, l'Intelligenza Artificiale (IA) può sicuramente considerarsi una di queste. È indiscutibile, infatti, che il "pensiero artificiale"[1]

[1] L'espressione "pensiero artificiale", in questo lavoro, è impiegata nella consapevolezza che ad oggi la più grande aspirazione degli studiosi dell'IA, vale a dire la replicazione delle capacità cognitive umane all'interno di sistemi che siano in grado di sviluppare un ragionamento in maniera autonoma e consapevole, costituisca un obiettivo non ancora conseguito e che, presumibilmente, mai lo sarà (sull'evoluzione dell'IA e, in particolare, sul cambio di prospettiva determinata dal passaggio da una cd. IA "forte" a una "debole" cfr. nt. 3 e 4). Tuttavia, il termine "pensiero" si presta particolarmente bene all'analisi che si vuole effettuare perché è in grado di dar conto e rappresentare figurativamente la complessità associata a strumenti capaci di mettere in atto processi decisionali che, seppur non paragonabili a quelli umani, sono in grado di sviluppare logiche inferenziali talmente avanzate da sfuggire, in alcuni casi, alla comprensibilità degli stessi programmatori. Si fa riferimento, infatti, ad un capacità di elaborazione estremamente sofisticata che, come si desume anche dalle definizioni riportate nel corso del lavoro delineate a livello sovranazionale dalla Commissione europea e dall'*High-Level Expert Group on Artificial Intelligence* e ribadite dalla giurisprudenza nazionale, si innesta su una interazione attiva della macchina con l'ambiente esterno e su una intensa attività di elaborazione e di valutazione dei dati che contempla anche un'attività di "ragionamento" sulle conoscenze già acquisite e sui risultati precedentemente conseguiti, nonchè di determinazione delle migliori azioni da intraprendere per il perseguimento di un obiettivo (non sempre compiutamente prefissato). Le attività svolte mediante i meccanismi di *machine learning*, pertanto, non si sostanziano in un processo passivamente rivolto al conseguimento di un fine predeterminato sulla base di uno schema di azioni limitate e stabilite dall'esterno, come avviene solitamente nel caso di impiego di algoritmi di base, bensì di un processo decisionale che si caratterizza per uno sviluppo autonomo della logica implementata all'interno di sistemi che imparano a identificare processi e schemi via via sempre più complessi e a individuare nuovi percorsi conoscitivi, "facendo tesoro" dell'esperienza nel frattempo acquisita. Tutto ciò anche in

sia ormai penetrato nel tessuto delle moderne società, innestando un ciclo di profonde trasformazioni che coinvolgono ad ampio spettro processi produttivi, tecniche comunicative, meccanismi relazionali e modalità di erogazione dei servizi pubblici e privati. Da obiettivo settoriale e quasi utopistico, in poco tempo, è diventato un fenomeno concreto e pervasivo, imponendosi come nuova direttrice di sviluppo delle odierne comunità.

Secondo opinione comune, tale salto "evolutivo" è stato determinato prevalentemente dall'incessante sviluppo delle tecnologie digitali e il connesso processo di datificazione della società umana[2]. Dopo anni di studi caratterizzati da risultati dalle alterne fortu-

assenza dell'uomo come nel caso dei processi di apprendimento di tipo non supervisionato. Pertanto, pur tenendo conto dell'attuale inesistenza di un "pensiero artificiale" che rispecchi pedissequamente i canoni propri dei processi cognitivi umani e la conseguente assenza di coscienza e consapevolezza di sé di tali macchine, la definizione permette di cogliere l'essenza di un fenomeno che si fonda su sistemi di IA sempre più avanzati che sono in grado di innestare, spesso autonomamente, specifici percorsi logico-deduttivi, di formulare valutazioni sui dati che attingono direttamente dall'ambiente esterno, di imparare dall'esperienza e di pervenire, infine, a un risultato che si rileva (almeno per la macchina stessa) il migliore tra i risultati possibili. Dato il carattere trasversale delle applicazioni dell'IA, la letteratura in materia è ovviamente vasta e multisettoriale. Per una visione generale del fenomeno del "pensiero artificiale" e sulle implicazioni sulle società tecnologicamente avanzate si v. A.M. Turing, *Computing Machinery and Intelligence*, in *Mind*, 49, 1950, 433 ss.; A. M. Turing, (trad.) N. Dazzi, *Intelligenza meccanica,* Torino, 1994; S. J. Russell – P. Norvig, (trad.) S. Gaburri, *Intelligenza artificiale. Un approccio moderno,* Torino, 2005; P. Ongsulee, *Artificial intelligence, machine learning and deep learning*, 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 2017; M. Boden, (trad.) F. Calzavarini, *L'intelligenza artificiale*, Bologna, 2019; L. Floridi - F. Cabitza, *L' intelligenza artificiale. L'uso delle nuove macchine*, Milano, 2021.

[2] Gli studi in materia di IA hanno avuto inizio, come è noto, nel secondo dopoguerra allorquando un gruppo di ricercatori indipendenti iniziò a lavorare sulla programmazione di macchine che fossero in grado di emulare il pensiero umano, analizzando le capacità di percezione, di ragionamento, di apprendimento e di interazione di artefatti in ambienti complessi. In tale filone, in particolare, si inserisce il lavoro sviluppato da *Alan Mathison Turing* che nel 1950 con il suo famoso *Imitation game* sollecitò la curiosità degli esperti in materia (e non solo) analizzando, nell'ambito di un gioco, l'effettiva capacità di una macchina di sostituirsi a un essere umano e di non essere "scoperta" dagli altri giocatori. La prospettiva di realizzare macchine "pensanti", nonostante le difficoltà e i limiti evidenziati dallo stesso matematico inglese, diede inizio a una vivace stagione di ricerca che portò ben presto alla nascita di una nuova disciplina scientifica, formalmente inaugurata in occasione del "*Darthmouth Summer Research Project on Artificial Intelligence*" nel 1956 e dedicata allo studio della capacità dei sistemi automatizzati di compiere ragionamenti e funzioni tipicamente umane. Il filone della "strong IA" destinata alla creazione di sistemi artificiali capaci non solo di replicare esattamente le attività cognitive, ma anche di imparare dall'esperienza, di risolvere problemi specifici e di pianificare decisioni future, tuttavia, si trovò a scontrarsi ben presto con la limitatezza di fatto delle tecnologie a disposizione e con le aspirazioni dei ricercatori ritenute sempre più diffusamente eccessive. Le scoperte compiute, infatti, presentavano uno straordinario valore scientifico, ma incorrevano in notevoli criticità operative una volta spostate dai laboratori e calate all'interno di ambienti reali e complessi o poste dinanzi a problemi di più difficile risoluzione. Dinanzi ad un diffuso sentimento di sfiducia si aprì quello che è stato definito "l'inverno dell'IA", un periodo di profonda crisi degli studi condotti in tale settore che ha avuto termine solo negli anni '80 in seguito a un decisivo ridimensionamento degli obiettivi della ricerca connesso a una intensa e non prevedibile evoluzione tecnologica e a una riscoperta degli studi in materia di reti neurali artificiali. Lo straordinario ampliamento del potere computazionale delle macchine e dei processi di digitalizzazione hanno favorito, quindi, una nuova stagione dell'IA che ha condotto all'attuale "entusiasmo scientifico" attorno a tale peculiare tecnologia consentendone una diffusione capillare anche in ambiti tradizionalmente lontani e distinti. Sulla storia della nascita e della evoluzione dell'IA vi è un'ampia e articolata letteratura in materia, soprattutto di carattere settoriale. Per una ricostruzione di carattere settoriale si faccia riferimento, tra gli altri, a A.M. Turing, *Computing Machinery and Intelligence*, *cit.*.; J. McCarthy - M. L. Minsky- N. Rochester - C.E. Shannon, *A proposal for the Dartmouth summer research project on artificial intelligence*, in *AI Magazine,* 27(4), 1955, 12 ss.; A. L. Samuel, *Some Studies in*

ne[3], la grande aspirazione di replicare artificialmente il funzionamento dei processi cognitivi umani, seppur ridimensionata[4], è divenuta realtà grazie al potenziamento della capacità di calcolo degli elaboratori, alla creazione della Rete Internet con le sue innumerevoli ramificazioni applicative e alla diffusione sempre più penetrante dell'*Internet of Things*, (IOT), che ha consentito di connettere e, quindi, di "far dialogare" oggetti di uso comune. L'incremento costante delle capacità di memorizzazione degli strumenti digitali supportati da ambienti sempre più capillarmente connessi ha favorito, come è noto, un ampliamento esponenziale delle attività di elaborazione e trattamento delle informazioni grazie alla creazione di masse sempre più estese di dati, i c.d. *big data*, la cui ampiezza, velocità e varietà costituiscono oggi la linfa vitale dell'ecosistema digitale e si pongono come elemento imprescindibile per lo sviluppo di tecniche di apprendimento automatizzato sempre più sofisticate e avanzate[5]. Un processo inar-

Machine Learning Using the Game of Checkers in *IBM Journal of Research and Development,* 3(3), 1959, 210 ss.; M. Minsky - S. Papert, *Perceptrons: An Introduction to Computational Geometry* in *The MIT Press*, Cambridge MA, 1969; J. Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation,* San Francisco, 1976; L. Bazzocchi, *Intelligenza Artificiale e sistemi esperti* in *Rivista trimestrale di analisi e critica*, 1-2, 1988; J. McCarthy, *What is Artificial Intelligence,* Computer Science Department, Stanford University, 2004; S.J. Russell - P. Norvig, *Intelligenza artificiale. Un approccio moderno*, Torino, 2005; Y. LeCun - Y. Bengio - G.Hinton, *Deep Learning* in *Nature*, 2015, 436 ss.; P. Ongsulee, *Artificial intelligence, machine learning and deep learning*, *cit.*; Y. Xin et al., *Machine Learning and Deep Learning Methods for Cybersecurity* in *IEEE Access*, vol. 6, 2018, 35365 ss.; Y. Bengio - Y. LeCun - G.Hinton, *Deep Learning for AI* in *Communications of the ACM*, Vol. 64, 7, 2021, 58 ss.; G. Alpa (a cura di), *Diritto e intelligenza artificiale. Profili generali, soggetti, contratti, responsabilità civile, diritto bancario e finanziario, processo civile,* Pisa, 2020; S. Dorigo (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pisa, 2020.

[3] Se durante i primi anni di studi in materia di IA vennero conseguiti rilevanti risultati che portarono, ad esempio, nel 1959 alla creazione del primo programma di apprendimento automatico applicato al gioco della dama, alcuni importanti esperimenti come quello destinato alla creazione di un traduttore automatico con il quale gli Stati Uniti avrebbero dovuto tradurre velocemente i documenti russi durante la guerra fredda fallirono clamorosamente. Ad aggravare la situazione intervenne nel 1969 la pubblicazione di un articolo scientifico dal titolo *Perceptrons* in cui due illustri esperti del settore dimostrarono l'inadeguatezza applicativa delle prime reti neurali artificiali da loro stessi realizzati. L'aspra critica rivolta alle effettive potenzialità dell'IA ebbe grosso eco all'interno degli ambienti scientifici decretando il crollo definitivo dei cospicui finanziamenti fino ad allora ottenuti per le attività di ricerca e una sensibile diminuzione dell'interesse generale per questo campo. Sul punto cfr. A. Newell - H. Simon, *Human Problem Solving*, Englewood Cliffs, NJ, 1972; M. Minsky - S. Papert, *Perceptrons: An Introduction to Computational Geometry*, cit.; S.J. Russell - P. Norvig, *Intelligenza artificiale. Un approccio moderno,* cit.

[4] Durante il cd. "inverno dell'IA", gli studiosi, ormai disillusi, furono costretti a ridimensionare le proprie aspettative orientando i propri lavori verso la realizzazione di sistemi operanti in ambiti esperienziali umani più ristretti e dotati, quindi, di una IA "debole". In altri termini, venne abbandonata l'idea di poter creare macchine senzienti in grado di replicare perfettamente e interamente l'esperienza cognitiva umana e risolvere autonomamente problemi con un livello di intelligenza pari o addirittura superiore all'uomo, per passare all'obiettivo di realizzare sistemi capaci di emulare i meccanismi del ragionamento solo in campi specifici in modo da individuare le soluzioni migliori in presenza di problemi determinati. Sul punto, tra gli altri, L Bazzocchi, *Intelligenza Artificiale e sistemi esperti,* cit.; P. Mariani - D. Tiscornia (a cura di), *Sistemi esperti giuridici L'Intelligenza Artificiale applicata al diritto*, Milano, 1989.

[5] L'*Organisation for Economic Cooperation and Development* ha definito l'utilizzo dei *big data* come «*the use of large-scale computing power and technologically advanced software in order to collect, process and analyze data characterized by a large volume, velocity, variety, and value*». Organisation for Economic Co-operation and Development, *Big Data: Bringing Competition Policy to the Digital Era, Executive Summary of the 126th meeting of the Competition Committee held on 29 November 2016*. Sul punto cf., tra gli altri, P. Savona, *Administrative decision-making after big data revolution*, in *Federalismi.it*, 19, 2018; V. Mayer-Shönberger - K. Cukier, *Big Data*, London, 2013; I.S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?* in *International Data*

restabile, ancora in *fieri*, in cui la datificazione di ampie sfere dell'esperienza umana ha favorito - e continua a favorire - la creazione di nuovi e sempre più affascinanti sentieri di crescita conoscitiva ed economica grazie all'elaborazione avanzata di tali preziose e diffuse informazioni.

All'interno di tale panorama, gli individui sono diventati dei *walking data generators*[6]: produttori incessanti di dati grazie all'uso intensivo di piattaforme *social*, di siti e di strumenti connessi che consentono di trasformare esperienze personali e di carattere sociale, eventi, decisioni di acquisto e di consumo in una fiumana incessante di *input* preziosi per la nuova architettura tecnologica[7]. Tali tracce digitali, inoltre, presentano una specifica peculiarità consistente nella loro capacità di essere granulari, intesa come la possibilità di esprimere una molteplicità di informazioni differenti dalla diversa sfumatura conoscitiva e specifico livello di dettaglio. Più il dato è granulare, quindi, maggiore è la sua rilevanza per il sistema digitale perché più ampie sono le relative potenzialità di sfruttamento attraverso l'elaborazione e l'incrocio dello stesso con altri frammenti informativi. Tale versatilità, quindi, non solo consente di ottimizzare servizi, prodotti e processi, ma permette di individuare nuove relazioni significative tra *set* di dati anche estremamente differenti, sfruttando altresì quelle informazioni ritenute solitamente residuali e dal significato limitato, se non addirittura nullo. Ed è proprio su questo innovativo connubio che insiste tra granularità del dato e potenza di elaborazione delle nuove tecnologie che si innesta il valore aggiunto apportato da tali sistemi. Al di là di una indiscutibile capacità di rendere più efficienti processi produttivi e decisionali già esistenti o comunque noti, tali sistemi operano come "trasformatori intelligenti" di informazioni grazie a una interazione attiva con l'ambiente circostante. Tale impatto "trasformativo" emerge chiaramente nel documento stilato dall'*High-level expert group on artificial intelligence*, il gruppo di esperti nominato dalla Commissione europea a supporto della strategia sovranazionale in materia, nel quale l'IA è defi-

Privacy Law, 3(2), 2013, 74 ss.; V. Zeno Zencovich - G. Codiglione, *Ten legal perspectives on the "Big Data Revolution"*, in F. Di Porto (ed.), *Big Data e concorrenza*, special issue in *Concorrenza e Mercato*, 23, 2016, 29 ss.; B. Van Der Sloot - S.Van Schendel, *Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study* in *Jipitec - Journal of Intellectual Property, Information Technology and E-Commerce Law*, 7, 2016.

[6] A. McAfee - E. Brynjolfsson, *Big Data: The management revolution* in *Harvard Business Review*, 10, 2012. L'espressione, che evidenzia la facilità con cui l'individuo produce continuamente dati personali mediante lo svolgimento di semplici attività ormai di carattere quotidiano come l'impiego dello smartphone, la realizzazione di acquisti online, l'utilizzo dei social networks o di strumenti connessi di vario tipo, è presente anche in S. Calzolaio, *Protezione dei dati personali* in R. Bifulco - A. Celotto - M. Olivetti (a cura di), *Digesto delle Discipline Pubblicistiche*, 2017, spec. 598.

[7] La conversione in dati di gran parte della vita materiale degli utenti unita al potenziamento e all'ampliamento delle capacità di memorizzazione delle macchine consente, infatti, di estrarre dalle abitudini sociali, culturali e lavorative degli individui non solo un numero incredibilmente più elevato di informazioni rispetto al passato, ma anche di estrapolare nuove fonti di conoscenza, spesso non strettamente connesse all'originaria finalità di trattamento, ma che possono essere reinvestite in maniera diversa nel più ampio scenario digitale E' quello che, ad esempio, è accaduto alla società americana *Google* che, inizialmente intenta a ottimizzare le attività in Rete dei propri utenti attraverso il proprio motore di ricerca, nel tempo ha compreso che quelli che considerava erroneamente "dati di scarto" della propria attività, come la tipologia delle informazioni ricercate, i tempi di permanenza e la formulazione delle domande, rappresentavano, invece, un *surplus* di straordinario valore da impiegare per la realizzazione di altri obiettivi tra cui la creazione di spazi pubblicitari personalizzati. Sul punto S. Zuboff, *Il capitalismo della sorveglianza*, (trad.) P. Bassotti, Roma, 2019, spec. 84 ss.

nita[8] come «sistemi software (ed eventualmente hardware) progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulle conoscenze, o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato. I sistemi di IA possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti»[9]. Si fa riferimento, quindi, a un comportamento intelligente[10] che presenta caratteristiche ben più avanzate rispetto a quanto precedentemente esistente, in particolare nell'ambito delle tecniche algoritmiche, dal momento che il "pensiero artificiale" non si sostanzia unicamente in «una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato», bensì si caratterizza per un apporto attivo ed osmotico nei confronti dell'ambiente esterno dal momento che «elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico»[11].
Ne consegue che, il contributo più rilevante apportato dall'IA, nella sua veste più evoluta, consiste proprio nella capacità di individuare, in tempi estremamente ridotti, relazioni sottese e ignote tra gli innumerevoli *dataset* che sono oggi a disposizione, portando alla luce e "trasformandole" in informazione significative, connessioni che

[8] Negli ultimi anni, a livello sovranazionale, si sono susseguite diverse definizioni di IA. Quella riportata nel testo, ad esempio, rappresenta, come confermato dalla stessa *High-Level Expert Group on Artificial Intelligence,* un perfezionamento della definizione tracciata solo l'anno precedente dalla Commissione europea in occasione della Comunicazione intitolata "L'intelligenza artificiale per l'Europa" del 25 aprile 2018, COM(2018) 237 final, nella quale tali tecnologie vengono indicate come «sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi» e, inoltre, viene specificato che "i sistemi basati sull'IA possono consistere solo in software che agiscono nel mondo virtuale (ad esempio assistenti vocali, software per l'analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure incorporare l'IA in dispositivi hardware (per esempio in robot avanzati, auto a guida autonoma, droni o applicazioni dell'Internet delle cose)». La necessità di perfezionare la definizione proposta dalla Commissione secondo l'*High-Level Expert Group on Artificial Intelligence* è emersa al fine di «chiarire alcuni aspetti dell'IA intesa come disciplina scientifica e come tecnologia, con il triplice obiettivo di evitare fraintendimenti, favorire una conoscenza comune e condivisa dell'IA che sia fruibile anche dai non esperti e fornire dettagli utili in vista delle discussioni sugli orientamenti etici dell'IA e sulle raccomandazioni strategiche in materia».
Rileva sottolineare, inoltre, che ancora diversa, è la definizione presente nella proposta di regolamento dell'IA COM(2021) 206 final analizzata nel corso del lavoro. All'art. 3, par. 1, infatti, l'IA è indicata come un «un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono». I molteplici tentativi di definizione a livello sovranazionale sono la più palese espressione di una evoluzione tecnologica che risulta difficilmente delineabile perfino nei suoi tratti essenziali e giustifica, pertanto, la difficoltà di regolamentare un fenomeno che assume così importanza per le società tecnologicamente avanzate, ma che è foriera di altrettante criticità per il loro sviluppo democratico.

[9] High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*, 18 dicembre 2018, spec.8.

[10] Comunicazione della Commissione, *L'intelligenza artificiale per l'Europa*, COM(2018) 237 final, spec. 1.

[11] Cons. Stato, sez. III, 4-25 novembre 2021, n. 7891. Cfr., e*x multis,* Cons. Stato, sez. VI, 13 dicembre 2019, n. 8472; Cons. Stato, sez. VI, 4 febbraio 2020, n. 882.

l'uomo, anche quello più esperto, non sarebbe mai in grado di cogliere da solo a causa della naturale limitatezza delle sue capacità di calcolo e di comprensione. Un processo questo che è evocativamente riassunto nel termine *data mining*: in un mondo basato sulla incessante produzione di dati, il motore dell'economia è, infatti, rappresentato da una vera e propria attività di estrazione e di elaborazione "intelligente" di frammenti conoscitivi estratti da immense "miniere informative" con il fine di ricavare nuovo sapere e individuare relazioni significative, non ancora note.

È evidente che tale straordinaria abilità inferenziale che consente di scoprire nuovi e inaspettati percorsi di conoscenza è destinata ad assumere un ruolo sempre più centrale ed estremamente prezioso per lo sviluppo economico, sociale e culturale delle odierne società, anche alla luce delle ulteriori e ancora pressoché ignote potenzialità che potrà esprimere nell'immediato futuro grazie ai crescenti studi in materia.

Tuttavia, come ogni evento rilevante, il "pensiero artificiale" richiede di essere analizzato con cura, non solo esaltandone gli indiscutibili vantaggi, ma anche approfondendo in ottica evolutiva le conseguenze che possono derivare da un utilizzo non democraticamente orientato di tali strumenti[12]. In sistemi incardinati sul principio personalista che pongono al centro del loro divenire l'individuo e la relativa possibilità di evolversi all'interno di un quadro di libertà e diritti costituzionalmente riconosciuti e tutelati[13], assume, infatti, particolare importanza la possibile incidenza di tali tecnologie sulla capacità dell'individuo di autodeterminarsi liberamente nello scenario digitale e, quindi, di realizzare pienamente e consapevolmente la propria persona sia come singolo, sia come soggetto che partecipa attivamente e, soprattutto, consapevolmente all'interno della comunità di appartenenza[14].

---

[12] Molteplici sono le riflessioni che negli ultimi anni si sono sviluppate in ambito dottrinale intorno alla possibile incidenza dell'impiego dell'IA sull'esercizio di diritti e libertà fondamentali. Sul tema cfr., tra gli altri, A. D'Aloia (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2021; A. Simoncini - S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale* in *Rivista di filosofia del diritto*, 8, 2019; A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà* in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019; P. Zuddas, *Brevi note sulla trasparenza algoritmica* in *Amministrazione in cammino*, 2020; G. Alpa (a cura di), *Diritto e intelligenza artificiale.* cit.; S. Dorigo (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, cit.; U. Pagallo, *Intelligenza Artificiale e diritto. Linee guida per un oculato intervento normativo* in *Sistemi intelligenti*, 3, 2017. Più specificamente sulla possibile compatibilità dei sistemi di IA con i meccanismi della democrazia rappresentativa, cfr. A. Cardone, *«Decisione algoritmica» vs decisione politica? A.I., legge, democrazia*, Napoli, 2021;

[13] Sulla rilevanza del principio personalista e della sua "eccedenza assiologica" che ne estende il relativo valore in modo diffuso in seno alla trama costituzionale cfr. A. Ruggeri, *Il principio personalista e le sue proiezioni*, in *Federalismi.it*, 17, 2013. Sulla centralità della persona all'interno delle architetture democratiche si v., tra gli altri, N. Lupo, *I diritti costituzionali di libertà nell'ordinamento giuridico attuale* ne *Il Foro Italiano*, 1958; C. Pinelli, "*Diritto di essere se stessi" e "pieno sviluppo della persona umana"* in *Rivista AIC*, 4, 2021; A. Spadaro, *I due volti del costituzionalismo di fronte al principio di auto-determinazione*, in *Politica del diritto*, 3, 2014; M. Ruotolo, *Appunti sulla dignità umana***,** in Studi in onore di Franco Modugno, IV, Napoli, 2011; G. Silvestri, *Considerazioni sul valore costituzionale della dignità della persona*, Intervento al Convegno trilaterale delle Corti costituzionali italiana, portoghese e spagnola, Roma, 2007.

[14] È, infatti, la possibilità di determinare autonomamente e consapevolmente l'orizzonte evolutivo delle proprie vicende personali e sociali all'interno di una comunità giuridicamente organizzata al riparo da indebite interferenze da parte di soggetti terzi, pubblici e privati, a delineare quel peculiare percorso di sviluppo della personalità dell'individuo che è oggetto di riconoscimento e di tutela nelle società democraticamente avanzate. L'autodeterminazione intesa quale pretesa del singolo di determinarsi liberamente nella scelta dei propri percorsi esistenziali costituisce, pertanto, il pilastro fondamentale su cui si innesta l'impianto personalista di tali assetti ed è espressione di quella dimensione "dinamica"

La possibilità offerta dai sistemi di IA di elaborare una elevata mole di dati al fine di estrapolare nuova e significativa conoscenza da reinvestire nell'ecosistema digitale, in assenza di regole che siano in grado di orientare la logica algoritmica verso principi e valori condivisi, preservandola da potenziali illegittime alterazioni determinate dalla qualità dei dati immessi[15] o da "pregiudizi" già esistenti all'interno del sistema (*bias*)[16], rischia di condurre a esiti decisionali che possono determinare una compressione degli spazi di esercizio di libertà fondamentali, a una intensificazione di situazioni discriminatorie già esistenti o alla creazione di nuove forme di discriminazione. Inoltre, nel caso di valutazioni di carattere predittivo possono condurre alla determinazione di situazioni di "penalizzazione delle propensioni"[17] e di condizionamento delle "aspi-

della dignità umana che costituisce un valore "supercostituzionale" e per questo non bilanciabile con gli altri diritti fondamentali. Sul diritto all'autodeterminazione e, in particolare, sul legame che insiste con la dignità intesa quale elemento immanente delle architetture democratiche cfr., *ex multis,* S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2013; A. Spadaro, *I due volti del costituzionalismo di fronte al principio di auto-determinazione*, cit.; E. Denninger, *Il diritto all'autodeterminazione individuale nell'ordinamento costituzionale tedesco* in *dirittifondamentali.it*, n. 2, 2018; L. Antonini, *L'autodeterminazione nel sistema dei diritti costituzionali* in Aa. Vv., *Autodeterminazione. Un diritto di spessore costituzionale?,* Torino, 2012; G.P. Dolso (a cura di), *Dignità, Eguaglianza e Costituzione*, Trieste, 2019; P. Becchi, *Il principio della dignità umana*, Brescia, 2009; G. Monaco, *La tutela della dignità umana: sviluppi giurisprudenziali* in *Politica del diritto*, 1, 2011; A. Ruggeri, *Dignità dell'uomo, diritto alla riservatezza, strumenti di tutela (prime notazioni)* in *ConsultaOnline*, 2016; R. Bin, *Dignità umana e biodiritto* in *BioLaw Journal – Rivista di BioDiritto*, 2, 2017; A. Ruggeri, *La dignità dell'uomo e il diritto di avere diritti (profili problematici e ricostruttivi)* in *ConsultaOnline*, 2018; C. Pinelli, "*Diritto di essere se stessi" e "Pieno sviluppo della persona umana"*, cit.; M. Ruotolo, *Appunti sulla dignità umana,* cit.; A. Papa, *La problematica tutela del diritto all'autodeterminazione informativa nella big data society* in *ConsultaOnline*, 2020; A. Ruggeri, Determinazione (il principio di) in Digesto delle discipline pubblicistiche - VIII agg., 2021; C. Casonato, *Il Principio di autodeterminazione. Una modellistica per inizio e fine vita* in *Rivista AIC,* 1, 2022.

[15] L'IA raccoglie ed elabora dati del passato per produrre decisioni che valgono per il presente o per il futuro. È evidente che tale percorso, in mancanza di un significativo apporto umano che controlli l'evolversi della logica decisionale o intervenga su esiti palesemente iniqui, distorti o discriminatori, dipende fortemente dal livello qualitativo dei dati impiegati. Informazioni inesatte, datate o alterate conducono inevitabilmente ad altrettanti esiti inesatti, datati o alterati che nel caso di decisioni che impattano direttamente sull'individuo possono avere conseguenze estremamente rilevanti sul suo percorso evolutivo. In tale contesto, si fa riferimento al fenomeno "*garbage in garbage out*". Sul punto cfr. M.F. Kilkenny - K.M.,Robinson, *Data quality: "Garbage in – garbage out"* in *Health Information Management Journal*, 47, 2018. In dottrina, sul tema cfr. A. Simoncini - S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale,* cit.

[16] Con il termine *bias* nell'ambito dell'IA si fa riferimento a possibili distorsioni del processo decisionale o valutativo determinate da "pregiudizi" insiti all'interno di tali strumenti tecnologici. Infatti, dal momento che le macchine sono il frutto della creazione di essere umani caratterizzati a loro volta da specifici sistemi di valori, principi e credenze avviene spesso che tale bagaglio esperienziale venga trasferito – più o meno consapevolmente - all'interno di tali sistemi con la conseguenza di renderli *a priori* non neutrali perché "intrisi", sin dalla loro progettazione, della "visione del mondo" dei loro programmatori. Una visione, inoltre, che è umana e, quindi, fallibile. Relativamente all'impatto dei *bias* sui meccanismi decisionali di tipo automatizzano risulta particolarmente significativa un'analisi compiuta nel 2019 dall'*AI Now Institute* dal titolo "*Disability, Bias, and AI*" nella quale è riportato, tra gli altri, il caso di un sistema di IA che non è stato in grado di riconoscere come umani alcuni soggetti diversamente abili in quanto il loro tipo di interazione non rispecchiava fedelmente i parametri con i quali la tecnologia era stata addestrata per distinguere una macchina da un umano. AI Now Institute, *Disability, Bias, and AI,* New York, 2019. Per un approfondimento del test condotto nell'esempio citato, cfr. K. Nakamura, *My Algorithms Have Determined You're Not Human: AI-ML, Reverse Turing-Tests, and the Disability Experience* in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility Association for Computing Machinery*, New York, 2019.

[17] V. Mayer-Shönberger - K. Cukier, *Big Data*, *A Revolution That Will Transform How We Live, Work, and Think*, trad. R. Merlini; Milano, 2013, spec.213.

razioni" del singolo[18]. In altri termini, l'impiego di sistemi di IA rischia di insinuarsi tra le pieghe del percorso evolutivo dell'individuo e di alterarne l'effettiva capacità di determinare la propria esistenza autonomamente e contribuire attivamente all'evoluzione della società all'interno di assetti democratici intesi, invece, «non soltanto come ideale o valore ma come sistema di realizzazioni concrete […], "luogo" di maturazione della personalità dei componenti la comunità statale»[19]. Dinanzi a tale frizione emerge la necessità che la portata innovativa degli strumenti tecnologici basati sull'IA, pur nell'importanza del loro sviluppo all'interno delle moderne società, venga ricondotta all'interno dei principi e dei valori che costituiscono il perno delle architetture democratiche, attraverso l'individuazione di un idoneo punto di equilibrio tra il significato economico espresso dai dati e la tutela dei diritti fondamentali, nel tentativo di evitare che gli straordinari benefici apportati da tali innovazioni si rilevino, in realtà, un ostacolo all'esercizio di libertà che rappresentano l'essenza stessa di tali assetti.

2. Le ombre dell'IA sulla capacità di autodeterminazione dell'individuo tra *epistemic bubbles*, *eco-chambers* e *nudging algoritmico*

Si è detto che il valore aggiunto dell'IA consiste nella capacità di creare nuova conoscenza mediante l'individuazione di collegamenti non noti tra le informazioni raccolte da molteplici *dataset*. Eppure, proprio la capacità di sviluppare nuovi sentieri conoscitivi rappresenta oggi l'aspetto più critico di tali tecnologie, in quanto il loro pervasivo impiego dimostra che, in assenza di adeguate regole e misure di tutela, tali strumenti sono in grado di incidere in maniera sempre più significativa sulla effettiva capacità degli individui di autodeterminarsi all'interno di società datocentriche arrivando ad insidiare pericolosamente istituti e presidi di garanzia che costituiscono l'infrastruttura determinante dell'attuale Stato di diritto.

In un panorama sempre più digitalizzato, infatti, l'individuo vive un costante processo di destrutturazione della propria identità personale, la quale risulta continuamente e inevitabilmente sezionata e suddivisa in migliaia di frammenti informativi che, a loro volta, si riverberano all'esterno mediante la circolazione di dati in grado di proiettare aspetti più o meno intimi della propria persona[20]. Si tratta di un fenomeno che pre-

[18] Tale punto è stato sottolineato recentemente dalla Dichiarazione europea sui diritti e i principi digitali per il decennio digitale COM (2022) 28 approvata nel gennaio 2022. Tale manifesto esprime l'intento di promuovere nel prossimo decennio un modello europeo per la transizione digitale che metta al centro le persone, conferendo loro maggiore autonomia e responsabilità all'interno di un panorama che garantisca non solo il rispetto dei diritti, ma cosa che assume particolare rilievo ai fini dell'analisi condotta, appunto anche le *aspirazioni* delle persone (considerando 4) e l'importanza della *libertà di scelta* (considerando 5).

[19] A. Ruggeri, *Il principio personalista e le sue proiezioni* in *Federalismi.it*, 2013, 24.

[20] Sul punto G. Finocchiaro, *Identità personale (diritto alla)*, cit., spec. 731 ss.; F. Pizzetti, *Privacy e il diritto europeo alla protezione dei dati personali – dalla Direttiva 95/46 al nuovo Regolamento europeo*, Milano, 2016; O. Pollicino – T. E. Frosini – E. Apa – M. Bassini (a cura di), *Diritti e libertà in Internet*, Milano, 2017; S. Rodotà, *Intervista su privacy e libertà*, Roma-Bari, 2005; S. Niger, *Le nuove dimensioni della privacy: dal diritto alla riservatezza alla protezione dei dati personali*, Padova, 2006; S. Rodotà, *La vita e le regole. Tra diritto e non*

senta sicuramente radici "analogiche"[21], ma che con la digitalizzazione e l'evoluzione dell'intelligenza artificiale ha subito un'accelerazione straordinaria, arrivando a spingere la frontiera delle attività di profilazione verso confini la cui compatibilità con assetti democratici risulta essere sempre più discutibile.

La capacità garantita dai nuovi metodi di apprendimento automatico di agire come "trasformatori intelligenti" scovando nuove relazioni nel sottobosco dei dati digitali, infatti, estende enormemente, l'ambito di applicazione di tali attività di trattamento arrivando a toccare le corde più profonde delle società democraticamente avanzate con conseguenze estremamente rilevanti per l'evoluzione dell'individuo, sia come singolo, sia come consociato. In particolare, quello che più colpisce è che l'incalzante ricerca di risultati perfettamente rispondenti alle esigenze e alle preferenze degli utenti finisce col travalicare la sfera delle attività economiche tradizionalmente intese, trascinando nel suo incessante vortice non solo prodotti e servizi, ma anche, e per la prima volta in maniera così significativa, idee, opinioni e valori. La dematerializzazione dei risultati della profilazione può avere come conseguenza una compressione degli orizzonti conoscitivi della persona, favorendo la creazione di barriere che impediscono di volgere lo sguardo al di là di quanto sia già noto e, pertanto, considerato sicuro.

Ne sono esempio i fenomeni sempre più diffusi di *epistemic bubbles*[22] che spingono il singolo a rifugiarsi in una vera e propria *comfort zone* destinata in breve tempo a trasformarsi in una gabbia "dorata" perché priva di quel dinamico confronto di idee, valori e opinioni differenti che da sempre costituisce il motore vitale degli assetti democratici. Tali processi di chiusura verso l'esterno, come è noto, possono anche portare - nelle forme più estreme, ma sempre meno rare - alla creazione di *ecochambers* a seguito di intensi processi di radicalizzazione delle opinioni che conducono i partecipanti a rifiutare perfino ciò che è realmente accaduto e/o è stato scientificamente e ufficialmente dimostrato[23].

A favorire ulteriormente tali situazioni interviene, inoltre, un fenomeno meno manifesto, ma ulteriormente rilevante, che può essere definito come "*nudging algoritmico*"[24]. È

diritto, Milano, 2006.

[21] È noto, infatti, che la profilazione non è emersa all'interno del nuovo panorama artificiale, ma costituisce un'attività ormai risalente nel tempo, sviluppata per decenni nell'ambito delle scienze economico-aziendalistiche, ancor prima della rivoluzione apportata dalla digitalizzazione.

[22] Con tale termine si fa riferimento a delle vere e proprie "bolle epistemiche" in cui gli utenti, soprattutto nel variegato panorama dei *social network*, si rifugiano spesso inconsapevolmente mediante un processo di selezione "per omissione" delle informazioni, conoscenze, relazioni e valori che contrastano la propria visione del mondo. Cfr. C. Nguyen, *Echo chambers and epistemic bubbles* in *Episteme*, 17(2), 2020, 141 ss.

[23] Le *eco-chambers* sono strutture costruite anch'esse a partire da attività di allontanamento dell'individuo da idee e prospettive contrastanti con il proprio pensiero, ma, a differenza delle *epistemic bubbles*, prevedono meccanismi di selezione più strutturati e consapevoli. In tali "camere", infatti, il processo di esclusione avviene in maniera attiva attraverso meccanismi di rinforzo del disaccordo (*disagreement-reinforcement mechanism*) che spingono non solo a rifiutare, ma addirittura a screditare il pensiero contrario. Ivi, 144.

[24] Tale termine fa riferimento alla *nudge theory*, un filone di studi particolarmente diffuso nei settori della psicologia e dell'economia comportamentale, che sostiene la possibilità di influenzare i meccanismi decisionali di un individuo in modo non coercitivo, mediante la previsione di una serie di piccoli "rinforzi positivi" che consistono in suggerimenti velati e non aggressivi al fine di indurre comportamenti

noto, infatti, che il mondo della Rete e, soprattutto dei *social networks*, sia ormai intriso di "spinte gentili", meccanismi automatizzati che mirano ad influenzare inconsciamente i processi decisionali dei naviganti, trasmettendo loro il falso convincimento di essere liberi di decidere. Si fa riferimento, in particolare, a tutti quei casi in cui le attività *online* risultano caratterizzate da continue interruzioni durante la navigazione determinate da suggerimenti di pagine e/o di informazioni che, pur non essendo richieste o ricercate direttamente dagli utenti, vengono loro mostrate improvvisamente, anche per pochi secondi alla luce di una raffinata attività di profilazione dei loro interessi. Queste sollecitazioni, soprattutto se ripetute nel tempo, incidono inevitabilmente sulla effettiva possibilità di questi ultimi di autodeterminarsi liberamente in quanto operano sull'inconscio incoraggiando decisioni e comportamenti che non necessariamente risultano essere i migliori possibili, ma sicuramente i più affini alle propensioni e alle preferenze di chi li subisce.

A rendere ancora più complesso un panorama già di per sé pieno di ombre vi è un'ulteriore abilità del "pensiero artificiale" che consiste nella capacità di tali sistemi di elaborare informazioni non solo al fine di realizzare *output* decisionali sempre più accurati ed efficienti, come avviene nell'ambito della profilazione, ma anche per svolgere delle vere e proprie attività previsionali. La capacità di creare nuova conoscenza nell'universo dell'IA si concretizza, infatti, anche nella possibilità di anticipare in modo sempre più puntuale *trend*, azioni e comportamenti futuri attraverso l'individuazione di peculiari relazioni e tendenze a partire da dati storici ed attuali.

Sebbene l'ampiezza delle modalità di applicazione e le connesse potenzialità di tali modelli rende tale specifico campo dell'IA di estremo interesse e ne giustifica la continua ascesa nell'era digitale, tuttavia, il suo esplicarsi con sempre maggiore precisione assume sfumature dense di significato quando l'attività previsionale viene a ricadere non su prodotti, servizi o eventi naturali, bensì su comportamenti o decisioni umane. Nel momento in cui si attua questo passaggio, la decisione predittiva perde gran parte del suo "fascino" e acquista un velo di opacità perché viene ad insinuarsi nella sfera di autodeterminazione del singolo trasformandosi inevitabilmente da "questione di efficienza" di processi a "questione di libertà e di esercizio di diritti fondamentali". È indubbio, infatti, che l'adozione di provvedimenti in vista di un comportamento che potrebbe potenzialmente verificarsi in futuro, anche con un'elevata probabilità, ma di cui non si ha una assoluta certezza, svuota di significato il concetto stesso di libertà e determina, come è stato sottolineato, il collasso di un sistema che si fonda, invece, sul baluardo della responsabilità individuale e sul legame assiologico fondamentale esistente tra causa ed effetto[25]. È inevitabile a tal proposito far riferimento al noto caso "Compas" e alle

raccomandabili. Nell'universo dell'IA, come evidenziato, questa tecnica grazie appunto ad analisi inferenziali di tipo avanzato sembra assumere sempre più risalto. Sono molteplici, infatti, i casi in cui l'utente è sottoposto inconsapevolmente a queste "spinte gentili" generate automaticamente che sono in grado di spronarlo verso l'adozione di determinate decisioni e/o comportamenti nella convinzione di non subire influenza di alcun genere. Sulla *nudge theory* si veda R. H. Thaler - C.R. Sunstein, *Nudge - Improving decisions about health, wealth, and happiness*, New Haven & London, 2008; L. Savadori, *Nudge: opportunità o moda passeggera?* in *Giornale italiano di psicologia*, 2, 2020, 355 ss. Sul tema del *nudging* di tipo algoritmico cfr., tra gli altri, A. Simoncini - S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale,* cit., 87 ss.

[25] Ivi, 93.

molteplici riflessioni compiute sull'affidabilità di un sistema algoritmico impiegato per prevedere il tasso di recidività di un individuo sottoposto a giudizio[26]. Ne consegue che l'utilizzo di meccanismi decisionali automatizzati con l'intento di anticipare i potenziali comportamenti futuri di un soggetto sulla base di serie storiche di dati ed esperienze simili rischia di creare, in assenza di dovute e opportune limitazioni, un sistema in cui gli individui sono chiamati sempre più a rispondere non delle proprie azioni, ma delle proprie inclinazioni[27]. Oppure di innescare meccanismi penalizzanti determinati esclusivamente sulla base di specifiche situazioni personali, fisiche o culturali, o peggio alla luce dell'appartenenza ad un determinato contesto economico – sociale, eliminando di fatto quella spinta all'evoluzione della persona che è costituzionalmente tutelata nel rispetto di un'uguaglianza di fatto che costituisce fermamente uno dei capisaldi delle società democratiche.

Infine, come ultimo tassello di un mosaico già di per sé difficile da comporre, interviene il fenomeno delle *black box*. Con tale termine, mutuato dalla teoria dei sistemi[28], si suole identificare tutti quei casi in cui l'utilizzo dell'IA conduce a risultati che sono frutto di dinamiche decisionali non comprensibili dall'esterno e, quindi, difficili da capire nella loro interezza, nonostante si abbia piena conoscenza dei dati impiegati in partenza. In tali situazioni, infatti, l'algoritmo opera proprio come una scatola nera: l'uomo ha la possibilità di osservare il momento in cui le informazioni entrano nel sistema e anche l'esito della loro elaborazione, ma non è in grado di comprendere la logica con cui sono

[26] Il caso ha riguardato l'utilizzo da parte di un tribunale statunitense di un peculiare *software* denominato COMPAS in occasione del giudizio *State v. Loomis* relativo ad un soggetto accusato di furto d'auto e mancato rispetto dell'alt intimato dalla polizia. Per la valutazione del caso, infatti, il giudice Tribunale circondariale di *La Crosse* aveva deciso di avvalersi del supporto del *Correctional Offender Management Profiling for Alternative Sanctions,* un algoritmo che mira a valutare il rischio di recidiva e la pericolosità sociale di un individuo a partire dall'analisi dei suoi precedenti giudiziari, dei risultati dei colloqui con i soggetti interessati e di serie storiche di dati relative a gruppi di individui aventi caratteristiche simili. Nel caso di specie, l'algoritmo aveva riscontrato nel soggetto la presenza di un elevato livello di rischio in tutte e tre le categorie di recidiva sottoposte ad analisi (processuale, generale e violenta) e, pertanto, dichiarato l'imputato pericoloso per la comunità. Alla luce di tale indicazione e tenuto conto della gravità del fatto commesso, il giudice del Tribunale circondariale di *La Crosse* aveva, quindi, deciso di non concedere al soggetto la libertà vigilata e dinanzi alla richiesta della difesa di aver accesso al codice sorgente del software aveva risposto negativamente adducendo alla natura proprietaria del software. Successivamente, la richiesta di revisione del giudizio avanzata dalla parte soccombente era stata rigettata dalla Corte suprema dello Stato del Wisconsin alla luce della considerazione che le eventuali criticità che si riscontrano nell'utilizzo di un processo decisionale completamente automatizzato debbono ritenersi superate nel caso in cui, come effettivamente accaduto nel caso di specie, l'organo giudicante confermi che avrebbe adottato la medesima pronuncia anche in assenza del supporto tecnologico. Sul punto cfr. K. Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, in *N.C. J.L. & Tech.*, 75, 2016; H. Liu – C.F. Lin - Y. Chen, *Beyond State v Loomis: artificial intelligence, government algorithmization and accountability* in *International Journal of Law and Information Technology*, 27 (2), 2019, 122 ss.; A. L. Washington, *How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate* In *The Colorado Technology Law Journal*, 17 (1), 2019; S. Carrer, *Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wiscon*sin, in *Giurisprudenza Penale Web,* 4, 2019; A. Simoncini - S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale,* cit., spec. 93 ss.

[27] V. Mayer-Shönberger - K. Cukier, *Big Data*, *A Revolution That Will Transform How We Live, Work, and Think*, cit., spec. 213.

[28] F. Pasquale, *The Black Box Society -The Secret Algorithms That Control Money and Information* in Harvard University Press, 2016; J.A. Kroll, *The Fallacy of Inscrutability* in *Phil. Trans. R. Soc.,* 376(2133), 2018; H. Shah, *Algorithmic Accountability* in *Phil. Trans. R. Soc.*, 376(2128), 2018; D. Pedreschi - F. Giannotti et al., *Open the Black Box. Data-driven Explanation of Black Box Decision Systems,* in *ArXiv*, 2018.

stati elaborati i dati. Tale opacità, che può assumere sfumature differenti[29], introduce ulteriori criticità nel panorama digitale dal momento che rischia inevitabilmente di privare il destinatario finale della decisione algoritmica della motivazione che ne è alla base e, quindi, di quel fondamentale elemento che, come è noto, è prezioso baluardo di tutela per l'individuo dinanzi a percorsi decisionali caratterizzati da esiti discriminatori o ingiustamente lesivi e «strumento di legittimazione politica della decisione cui essa accede» in presenza di atti di pubblici poteri»[30].

E' evidente che in assetti costituzionali, come quelli che caratterizzano - in gran parte - il panorama europeo, imperniati sulla centralità dell'individuo e sul dovere dello Stato di rimuovere gli ostacoli che impediscono il pieno sviluppo della personalità del singolo[31], la pervasiva capacità dei sistemi di IA di erodere le possibilità decisionali dell'individuo attraverso un costante e spesso oscuro processo di ridimensionamento dell'opportunità di conoscenza e delle alternative di scelta ovvero di subire indebite limitazioni nell'esercizio di diritti riconosciuti e garantiti costituisce oggi un *vulnus* nel percorso evolutivo in senso democratico di tali società. È indubbio, infatti, che da uno scenario del genere caratterizzato da percorsi decisionali, pubblici e privati, pericolosamente opachi e da rischiosi processi di polarizzazione delle opinioni, l'individuo esce svilito e mortificato perché privato degli strumenti necessari per realizzare una piena e matura evoluzione della propria persona, nonché per contribuire attivamente e consapevolmente alla crescita della collettività di cui fa parte. Viene messo in discussione, in particolare, quello specifico «diritto di determinarsi liberamente nella scelta dei propri percorsi esistenziali […] che consente alla persona di avere specifica percezione del sé quale soggetto responsabile e non mero oggetto passivo della propria esperienza

[29] Nonostante l'indubbia rilevanza della conoscibilità del percorso inferenziale che conduce alla decisione algoritmica, sono diverse le motivazioni che spesso impediscono la piena comprensibilità di un algoritmo. A tal proposito P. Zuddas individua tre diverse tipologie di opacità. La prima definita "tecnica" attiene alla diffusa difficoltà per un individuo che non sia esperto in materia di comprendere pienamente il funzionamento del software a causa della particolare complessità della tecnologia utilizzata. L'opacità "intrinseca", invece, opera soprattutto in ambienti di *machine learning* in cui la macchina è in grado di imparare da sola a partire dai dati a disposizione e, quindi, di evolversi autonomamente. In questi casi la piena comprensibilità della logica decisionale sottesa all'esito algoritmico può sfuggire anche agli stessi programmatori. Infine, l'opacità può assumere carattere "giuridico" allorquando, pur essendo potenzialmente conoscibile e comprensibile, il funzionamento dell'algoritmo non può essere rilevato a causa della natura dei dati coinvolti o della presenza di specifici interessi giuridici sullo stesso. P. Zuddas, *Brevi note sulla trasparenza algoritmica,* cit.

[30] A. Romano Tassone, *Sulla c.d. funzione democratica della motivazione degli atti dei pubblici poteri* in A. Ruggeri, *La motivazione delle decisioni della Corte costituzionale. Atti del seminario di Messina 7-8 maggio 1993*, Torino, 1994, 33 ss. Sulle implicazioni della trasparenza algoritmica in materia di motivazione degli atti adottati da soggetti pubblici e, in particolare, delle decisioni giudiziarie e amministrative si v. P. Zuddas, *Brevi note sulla trasparenza algoritmica,* cit.

[31] È palese il riferimento all'art. 2 della Costituzione italiana che consacra e tutela il principio personalista inteso come pilastro dell'architettura ordinamentale delineata dal legislatore costituzionale nel '48. Tuttavia, la centralità dell'individuo emerge come elemento fondamentale e comune anche di molte altre carte costituzionali europee, *in primis* quella tedesca, che pone significativamente in apertura del *Grundgesetz,* all'art. 1, proprio il riconoscimento dell'intangibilità della dignità umana e il dovere di ogni potere statale di rispettarla e proteggerla. Come è noto, la decisione di riconoscere e tutelare in maniera diretta e, soprattutto, con fonte apicale l'uomo nella sua essenza di essere umano, d'altra parte, fu una decisa e condivisa reazione degli Stati agli orrori vissuti durante i due conflitti mondiali. Sul punto v. G.P. Dolso (a cura di) *Dignità, eguaglianza e Costituzione*, cit.

esistenziale»[32].

Pur inaugurando un florido periodo di innovazioni e cambiamenti, quindi, la nuova primavera dell'IA impone necessariamente anche l'avvio di una stagione di azioni che richiedono un agire concertato da parte di tutti i soggetti variamente coinvolti nell'utilizzo del "pensiero algoritmico", alla luce della considerazione che il dibattito sui futuri sviluppi dell'intelligenza artificiale attiene anche e, soprattutto, alla dignità e alla libertà umana, nonché agli spazi in cui quest'ultima concretamente si realizza nella molteplicità delle sue declinazioni possibili. Si tratta di un dibattito indubbiamente complesso che vede ancora una volta confrontarsi, da un lato, l'entusiasmo tipico della ricerca scientifica e, dall'altro, il doveroso rigore della scienza giuridica, ma che risulta sicuramente necessario e, vista la velocità dei cambiamenti apportati, improcrastinabile, dal momento che da esso dipende inevitabilmente il futuro delle società democraticamente avanzate.

3. Il percorso intrapreso dall'Unione europea per lo sviluppo di una IA di tipo antropocentrico

Consapevole delle imponenti sfide derivanti dalla diffusione sempre più ampia del "pensiero artificiale", l'Unione europea, negli ultimi anni, ha avviato una articolata strategia in materia al fine di garantire un approccio coordinato e condiviso tra i vari Stati membri. Partendo dalla convinzione che l'IA non «è fantascienza [ma] è già parte delle nostre vite»[33], l'obiettivo perseguito è quello di sfruttare al massimo le opportunità offerte dall'IA incanalando, tuttavia, tale complesso fenomeno all'interno di un quadro etico e giuridico adeguato, basato su valori condivisi e coerente con la Carta dei diritti fondamentali dell'UE in modo da creare un ecosistema affidabile e sicuro per tutti i cittadini europei.

Punto di partenza della riflessione europea in materia di IA può essere considerata la proposta di risoluzione del Parlamento recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)). Il "Rapporto Delvaux", dal nome del parlamentare proponente, rappresenta, infatti, il primo documento ufficiale in cui il fenomeno del "pensiero artificiale" è affrontato in maniera strutturata e prospettica, evidenziando la necessità di norme in grado di disciplinare, in particolare, «la trasparenza e l'assunzione di responsabilità e che riflettano i valori intrinsecamente europei, universali e umanistici che caratterizzano il contributo dell'Europa alla società»[34]. Ricco di suggestioni futuristiche provenienti dal mondo della letteratura fantascientifica, con riferimenti che vanno dal *Frankenstein* di *Mary Shelley* al robot di *Karel Čapek*, il documento rappresenta, infatti, una chiara presa di coscienza della forza innovativa del fenomeno dell'IA ed esprime l'urgenza di adottare norme dedicate

[32] Cass. civ., sez. III, ord. 23 marzo 2018, n. 7260.

[33] COM (2018) 237, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *L'intelligenza artificiale per l'Europa*, 1.

[34] Considerando V della proposta di Risoluzione del Parlamento europeo divenuto considerando U nella versione finale.

che tengano conto delle implicazioni e delle conseguenze etiche e legali derivanti da tali tecnologie senza ostacolarne, però, lo sviluppo ormai ritenuto indispensabile. In tale ottica, l'atto, quasi interamente confluito nel testo finale della Risoluzione adottata il 16 febbraio 2017[35], introduce una serie di linee guida declinate in ambiti specifici, tra i quali la ricerca e l'innovazione, il diritto di proprietà intellettuale, i veicoli autonomi e della medicina robotica. *Trait d'union* di tutte le indicazioni inoltrate alla Commissione è la convinzione che l'operato del legislatore europeo debba partire dal complesso panorama delle questioni relative alla responsabilità civile «dato lo stadio della robotica e dell'intelligenza artificiale»[36] sulla base di un diffuso convincimento circa la necessità di valutare la possibilità che nel lungo periodo l'intelligenza artificiale possa superare la capacità intellettuale umana. Da tale presa di posizione si muove anche il riferimento specifico alle leggi della robotica di Asimov[37] che la Risoluzione presenta attualmente tra i principi generali di riferimento della nuova materia, rivestendole, così, di una specifica veste giuridica, nella consapevolezza che in uno scenario sempre più caratterizzato da decisioni di tipo automatizzato uno degli obiettivi principali del legislatore consiste nel garantire che l'apporto di tali strumenti si concretizzerà sempre e solo in un supporto alle attività dell'uomo, senza mai operare a detrimento della sua esistenza e della sua evoluzione.

La necessità di affrontare celermente un fenomeno così trasversale e pervasivo come quello dell'IA ha portato poco tempo dopo alla sottoscrizione da parte di 24 Stati membri[38] della Dichiarazione di cooperazione sull'intelligenza artificiale. L'atto, firmato il 10 aprile del 2018, evidenzia la volontà di affrontare le crescenti sfide multisettoriali provenienti dall'uso sempre più massivo di tali strumenti mediante l'adozione di

---

[35] Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

[36] *Ibid.*, considerando Y. In tale prospettiva, la Risoluzione ha sostanzialmente invitato la Commissione non solo a esaminare e valutare l'istituzione di un regime assicurativo obbligatorio per i danni potenzialmente causati dai robot, ma addirittura a considerare l'ipotesi di uno *status* giuridico specifico «di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi».

[37] Come è ampiamente noto, le tre leggi furono ideate negli anni '40 dallo scrittore russo naturalizzato americano Isaac Asimov in seguito ad alcune riflessioni compiute sulle macchine robotiche oggetto dei suoi racconti di fantascienza. Formulate, quindi, in un contesto di fantasia, hanno assunto però un nuovo e più pregnante significato all'indomani delle prime affermazioni degli studi in materia di IA in ambito scientifico. Nel dettaglio essere recitano: «(1) Un robot non può recar danno a un essere umano né può permettere che, a causa del proprio mancato intervento, un essere umano riceva danno. (2) Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Prima Legge. (3) Un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la Prima o con la Seconda Legge». In seguito, l'autore aggiunse una nuova legge che si antepone alle tre già formulate in quanto da queste non può essere contrastata: «(0) Un robot non può recare danno all'umanità, né può permettere che, a causa del proprio mancato intervento, l'umanità riceva danno». I. Asimov, *Runaround*, US, 1942; I. Asimov, *Robots and Empire*, New York, 1985.

[38] Oltre ai 24 Stati membri indicati, l'Atto è stato sottoscritto in quella sede anche dalla Norvegia. A questo elenco si sono aggiunti, nel corso dello stesso anno, Romania, Grecia, Cipro (maggio 2018) e Croazia (giugno 2018) Sul punto è possibile consultare la documentazione presente all'interno della sezione dedicata sul sito della Commissione europea EU Member States sign up to cooperate on Artificial Intelligence | Shaping Europe's digital future (archive-it.org)

un approccio comune a livello sovranazionale, promuovendo la creazione di centri di ricerca europei dedicati allo sviluppo del "pensiero artificiale", l'adozione di programmi mirati di finanziamento e lo scambio di *best practice* tra i paesi partecipanti.

Tali obiettivi sono divenuti a loro volta parte integrante della Comunicazione della Commissione europea adottata il 25 aprile 2018 e intitolata "Intelligenza Artificiale per l'Europa"[39] con la quale l'Unione ha proclamato ufficialmente l'avvio di una strategia condivisa le cui azioni sono state definite successivamente con il "Piano coordinato sull'intelligenza artificiale" formalizzato per il tramite della Comunicazione n. 795 del 7 dicembre 2018.

Il ritmo serrato e l'ampiezza degli ambiti settoriali che hanno caratterizzato l'adozione di tali atti sono evidentemente espressione della complessità di un fenomeno che fa della velocità evolutiva e della varietà delle declinazioni applicative il suo carattere distintivo. Mediante tali documenti l'Unione ha definito le direttrici lungo le quali il "pensiero artificiale" dovrà svilupparsi nel prossimo futuro all'interno del territorio europeo, ponendosi come fine ultimo la creazione di «un'IA "*made in Europe*" etica, sicura e all'avanguardia» che sappia sfruttare «i punti di forza scientifici e industriali dell'Europa»[40].

L'analisi dei primi lavori compiuti a livello sovranazionale evidenzia la volontà del legislatore di sfruttare appieno lo straordinario potenziale derivante dal settore dell'IA per fare in modo che il continente europeo risulti competitivo sullo scacchiere internazionale, ma al contempo affiora anche la necessità che tale forza innovativa operi e si sviluppi nel pieno rispetto dei principi condivisi a livello sovranazionale all'interno di un quadro etico e giuridico adeguato che sia sempre basato sui valori dell'Unione e coerente con la Carta dei diritti fondamentali dell'UE[41].

Il risultato di tale tensione porta a far emergere uno dei tratti fondamentali che caratterizza la strategia europea in tale settore: la ricerca di un legame robusto e una interdipendenza fertile tra ambito scientifico e giuridico. Secondo tale visione, infatti, la risposta alle sfide dell'IA richiede un approccio del tutto nuovo rispetto al passato, imponendo un confronto fattivo tra la logica che muove la ricerca tecnica e quella che anima il legislatore, al fine di raggiungere un indispensabile contemperamento tra interessi spesso confliggenti all'interno di una società che mira ad evolversi tecnologicamente, ma pur sempre nel rispetto dei valori fondanti i diversi Stati membri e che sono cristallizzati nell'art. 2 del TUE.

In tale ottica, pertanto, si giustifica la creazione, sempre nel corso del 2018, dell'*High Level Expert Group on AI* (AIHLEG) costituito da 52 esperti di diversa estrazione professionale[42] con il compito specifico di contribuire concretamente alla strategia europea grazie all'apporto di contributi tecnici e spunti di riflessione settoriali al fine di

[39] COM (2018) 237, cit.

[40] COM (2018) 795, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Piano coordinato sull'intelligenza artificiale,* 1.

[41] COM (2018) 237, cit., 4

[42] Il gruppo comprende soggetti appartenenti al mondo accademico e quello dell'industria, nonché esponenti della società civile e delle associazioni non governative.

garantire risposte quanto più complete e puntuali possibili alle sfide sollevate dall'IA[43]. A tale organo, in particolare, è affidato il compito di adottare raccomandazioni destinate ad indirizzare lo sviluppo delle politiche in materia nel medio-lungo termine e di sostenere l'attività del legislatore europeo.

Il Gruppo, nel suo primo anno di vigenza, ha adottato quattro documenti ufficiali in materia di IA[44] tra cui assumono particolare rilievo le *Ethics guidelines for trustworthy AI* che, ad un'analisi più attenta, rappresentano l'espressione più compiuta della strategia che l'Unione europea intende implementare in tale settore. Le linee guida, infatti, evidenziano la necessità che lo sviluppo dell'IA avvenga all'interno di un ecosistema *trustworthy,* termine che, ben lungi dall'invocare una generica garanzia di affidabilità delle nuove tecnologie digitali, si sostanzia in un principio multiforme che a sua volta si compone di tre anime distinte: la *lawfulness* (liceità), l'*ethics* (eticità) e la *robustness* (robustezza). In tale ottica, affinché l'innovazione tecnologica possa contribuire positivamente al progresso delle moderne società e non operare a detrimento di diritti e libertà fondamentali è necessario che l'anima più spiccatamente tecnica possa ampiamente confrontarsi con quella giuridica avendo come orizzonte di riferimento sempre e comunque uno sviluppo eticamente sostenibile e, quindi, *human-centric* dell'IA.

Simbolo della ricerca di una collaborazione fattiva tra le due sfere è la creazione dell'*European AI Alliance* un forum dedicato allo scambio di opinioni tra il Gruppo di esperti e membri del settore industriale, organizzazioni dei consumatori, sindacati e altri rappresentanti ì della società civile con lo scopo di contribuire attivamente ai lavori dei tecnici e per il loro tramite al processo decisionale in materia realizzato in seno alle istituzioni europee.

Un coinvolgimento di così ampio respiro che si estende concretamente a tutti i soggetti direttamente coinvolti nella catena del valore dell'IA è contemplato anche nelle Linee guida in materia di intelligenza artificiale e protezione dei dati redatte dal Comitato consultivo della Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale (Convenzione 108/1981)[45]. Sulla medesima linea ideale, infatti, il testo, approvato a Strasburgo il 25 gennaio 2019, non solo ribadisce, all'art. 1, che la tutela della dignità umana e delle libertà fondamentali, in particolare il diritto alla protezione dei dati personali, assumono un ruolo imprescindibile nello sviluppo e nell'adozione di applicazioni IA, ma mira ad instaurare un dialogo diretto sia con i legislatori e i diversi decisori politici, sia con gli sviluppatori, i produttori e i fornitori di servizi, al fine di favorire «un approccio volto a tutelare i diritti umani fin dalla progettazione di tali servizi ("*human rights by design*") ed evitare qualsiasi potenziale pregiudizio (*bias*), anche involontario o occulto, il rischio di discriminazione o altri ef-

[43] Da sottolineare anche la creazione dell'Alleanza per l'IA, una piattaforma inaugurata nel giugno del 2018 con lo scopo di consentire ai diversi stakeholders di fornire indicazioni, suggerimenti e proposte all'*High Level Expert Group on AI* in materia di politiche in IA.

[44] Si fa riferimento alle *Ethics guidelines for trustworthy AI* adottate nell'aprile del 2018; le *Policy and investment recommendations for trustworthy Artificial Intelligence* del 2019; *The final Assessment List for Trustworthy AI (ALTAI)* del 2020; le *Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI*. I documenti sono reperibili al link Expert group on AI | Shaping Europe's digital future (europa.eu)

[45] Consiglio d'Europa, *Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale Strasburgo*, (Convenzione 108), 28 gennaio 1981.

fetti negativi sui diritti umani e le libertà fondamentali degli interessati»[46] .
L'anelito europeo verso un ecosistema *trustworthy* nasce, quindi, dalla convinzione che le nuove tecnologie, soprattutto nel caso in cui siano caratterizzate da un elevato livello di automaticità e le cui decisioni possano impattare sull'esercizio di diritti e libertà fondamentali, richiedano sempre una chiara e solida base di legittimazione e debbano innestarsi in uno scenario che sia affidabile, sicuro ed eticamente sostenibile. Nel caso in cui ciò non avvenga ovvero manchi anche solo uno degli elementi precedentemente indicati, si concretizza il rischio di esporre l'utente a conseguenze che possono inficiare il pieno e consapevole sviluppo della propria persona, sia come singolo sia come membro di una comunità, oppure al contrario creare un clima di sfiducia diffusa tale da impedire a strumenti che presentano profili di interesse collettivo di apportare i propri contributi positivi all'evoluzione della società.
Gli obiettivi e gli strumenti predisposti dall'Unione europea in materia di IA sono stati, infine, ulteriormente rafforzati nel *White Paper on AI: a European approach to excellence and trust* nel febbraio 2020, con il quale, in uno scenario non ancora rivoluzionato dall'avvento della pandemia da Covid-19, la Commissione ha ribadito la centralità di tali tecnologie per il futuro dell'Europa e la necessità di creare una «IA antropocentrica che sia al servizio delle persone e cha ha come fine ultimo quello di migliorare il benessere degli esseri umani»[47]. Con tale prospettiva, il Libro Bianco ha individuato due direttrici fondamentali di sviluppo: l'eccellenza della ricerca in materia e la creazione di un ecosistema tecnologico affidabile mediante la previsione di un quadro normativo chiaro e prospettico che sia in grado di infondere quella fiducia necessaria ai cittadini per adottare i nuovi strumenti tecnologici e alle imprese e alle organizzazioni pubbliche per investire in tali innovazioni. In attesa, quindi, del fondamentale intervento del legislatore europeo in materia realizzatosi solo l'anno successivo, il testo ha definito un quadro di azioni complesse e multisettoriali a sostegno della ricerca e dell'innovazione, sottolineando ancora una volta la necessità, nonché l'urgenza, data la velocità di sviluppo di tale settore, di un lavoro congiunto tra sfera giuridica e sfera tecnica in modo che i valori che animano la società europea «siano pienamente integrati nelle modalità di sviluppo dell'IA»[48].

4. La proposta di regolamento europeo COM (2021) 206 in materia di IA

L'intensa attività svolta a livello europeo per la realizzazione di un ecosistema affidabile, sicuro ed etico ha trovato nel corso del 2021 il suo "approdo" normativo con l'adozione della prima proposta di regolamento del Parlamento europeo e del Consiglio

[46] Comitato consultivo (CD. T-PD) della Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale (Convenzione 108), Linee guida in materia di intelligenza artificiale e protezione dei dati, 2.

[47] COM (2019) 168, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni *Creare fiducia nell'intelligenza artificiale antropocentrica*, 2.

[48] *Ibid.*

che stabilisce regole armonizzate sull'intelligenza artificiale[49]. L'atto mira a definire un quadro normativo comune che sia in grado di superare le criticità connesse al possibile affermarsi di «un mosaico di regole nazionali potenzialmente divergenti»[50] ritenuto non adeguato a garantire una idonea tutela dei valori e dei diritti fondamentali dell'Unione dinanzi all'avanzare di un fenomeno sempre più trasversale che si mostra, per di più, incurante dei confini territoriali. In tale ottica, la sfida che si propone l'Unione europea è quella di creare un sistema che sia quanto più completo possibile attraverso una certosina opera di coordinamento e di armonizzazione delle nuove norme con le regolamentazioni già esistenti[51] e, al contempo, la previsione di regole che siano *future-proof*, "a prova di futuro", in grado di disciplinare in maniera prospettica un fenomeno multiforme e cangiante. In tale prospettiva, il (recente) passato regolativo assume un nuovo taglio normativo alla luce delle innovazioni apportate dall'intelligenza artificiale, viene a fondersi con le nuove regole e si proietta verso il futuro nel tentativo di creare una robusta cornice regolatoria che sia in grado di tutelare gli individui dalle inevitabili insidie derivanti dall'utilizzo di tali tecnologie. Un obiettivo, quest'ultimo, particolarmente complesso dal momento che una delle problematiche che maggiormente caratterizzano la sfera di regolazione di tale fenomeno è rappresentato proprio dal rischio di una rapida "obsolescenza normativa" causata non solo dalla straordinaria velocità di sviluppo delle tecniche, ma anche dall'uso massivo di tali strumenti che portano alla luce continuamente nuovi e spesso imprevedibili risvolti critici. Inoltre, sotteso a tale testo vi è anche l'intenzione di creare un quadro normativo che sostenga pienamente la ricerca scientifica allontanando i rischi di una iper-regolazione che potrebbe scoraggiare il progresso tecnologico determinando un nuovo "inverno" nella storia dello sviluppo degli strumenti di IA. In tale prospettiva, la proposta si contraddistingue per la definizione di un sistema di regole di carattere orizzontale destinato a disciplinare gli utilizzi dell'IA che presentano un elevato livello di pericolosità integrato dalla previsione di codici di condotta che i fornitori dei sistemi appartenenti alle altre categorie di rischio individuate dall'atto possono creare e adottare su base volontaria[52]. Lo scopo perseguito è quello di incoraggiare questi ultimi ad applicare spontaneamente i requisiti obbligatori previsti per le applicazioni tecnologiche più rischiose e, al contempo, favorire un quadro regolamentare più snello che sia maggiormente in grado di disciplinare, in maniera proporzionale ed efficiente, un fenomeno cangiante e in con-

[49] COM (2021) 206, proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione.

[50] Ivi, 7.

[51] La proposta di regolamento, difatti, si conclude con l'indicazione dettagliata nel Titolo XII, artt.75-82, delle modifiche che dovranno essere apportate alle disposizioni vigenti che saranno interessate dall'entrata in vigore del nuovo atto in modo di garantire la piena coerenza del nuovo quadro normativo in materia. Inoltre, come indicato nella Relazione introduttiva, la proposta fa parte di un pacchetto più ampio di misure che delineeranno, a livello sovranazionale, il sistema di regole operanti in materia di IA. Si fa riferimento, in particolare, all'Atto sulla *governance* dei dati (COM/2020/767), alla Direttiva sull'apertura dei dati (direttiva (UE) 2019/1024) e alla più ampia "Strategia europea dei dati" annunciata con Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni (COM/2020/66) final.

[52] Proposta di regolamento COM(2021) 206 final, art. 69.

tinua evoluzione anche attraverso lo sviluppo di una visione comune delle modalità di fronteggiamento delle criticità che possono caratterizzare tali strumenti[53].

Orizzontalità delle regole e promozione dell'adozione di codici di condotta portano alla definizione di un approccio peculiare che in parte prosegue e rafforza quello adottato nel 2016 dal legislatore europeo nell'ambito della tutela dei dati personali, sebbene declinandolo in maniera specifica nella realtà dell'IA. Si tratta, infatti, dell'ormai famoso *risk based approach*, basato sulla consapevolezza che l'utilizzo di tali tecnologie, al pari del più generale utilizzo di dati personali, costituisca sempre un'attività potenzialmente pericolosa perché, in assenza di norme e strumenti di tutela adeguati, è in grado di ledere in maniera significativa l'esercizio di diritti e di libertà fondamentali. Da qui la necessità che tutti i soggetti variamente coinvolti nella realizzazione e implementazione di prodotti o servizi che utilizzano l'intelligenza artificiale siano consapevoli di tale situazione e adottino una serie di misure tecniche e organizzative che tengano conto, sin dall'inizio, sul modello della *privacy by design e by default ex* art. 25 del regolamento (UE) 679/2016, dell'esistenza di rischi di rilievo per la normale evoluzione dell'individuo, sia come singolo, sia come membro di una collettività organizzata.

Nella proposta di regolamento, tuttavia, tale approccio subisce una modifica dal momento che le regole che lo caratterizzano non vengono applicate in maniera generalizzata, bensì vengono modulate in maniera proporzionata al rischio esistente o che si presume possa emergere in occasione dell'utilizzo di tali tecnologie. Il quadro normativo predisposto a livello europeo, infatti, si struttura su diversi livelli di obblighi e limiti previsti a carico dei soggetti coinvolti nell'universo dell'IA, il cui grado di intensità varia al variare del livello di pericolosità associata ai diversi strumenti utilizzati.

In tale contesto, quindi, il *risk based approach* diventa *proportionate* ed esprime il tentativo europeo di individuare un delicato punto di equilibrio tra l'imperativo di garantire uno sviluppo *human-centric* del "pensiero artificiale" e la necessità di non spegnere l'afflato evolutivo in un settore che è sempre più centrale nella vita e nelle economie delle società moderne. Partendo dalla convinzione che l'IA abbia anime e finalità diverse, la previsione di regole differenziate a seconda del tipo di utilizzo di tali tecnologie consente, infatti, nelle intenzioni del legislatore europeo, di ridurre, da un lato, la "pressione normativa" sugli utilizzi meno rischiosi, favorendo la creazione di un terreno fertile

[53] Il punto è sottolineato nella Relazione introduttiva alla proposta di regolamento in materia di IA nella quale si evidenzia che tale scelta «limita i rischi di violazione dei diritti fondamentali e della sicurezza delle persone e promuove attività efficaci di controllo e applicazione, concentrando i requisiti soltanto sui sistemi che presentano un rischio alto di occorrenza di tali violazioni». Inoltre, la previsione di un sistema composta da regole orizzontali a carico dei sistemi di IA più pericolosi e la promozione di codici di condotta negli altri casi consente, nell'ottica del legislatore europeo, di mantenere «i costi di conformità al minimo, evitando così un inutile rallentamento dell'adozione dovuto a prezzi e costi di conformità più elevati», la creazione di spazi di sperimentazione normativa e, inoltre, «un incremento della fiducia delle persone nei confronti dell'IA», con un incremento dei vantaggi per le imprese in termini «di certezza del diritto» evitando, infine, azioni unilaterali da parte degli Stati membri che potrebbero frammentare il mercato unico. Proseguendo sulla scia dell'adozione volontaria di codici di condotta per i sistemi di IA meno rischiosi il considerando 81), inoltre, promuove anche all'adozione da parte dei fornitori di «requisiti supplementari relativi, ad esempio, alla sostenibilità ambientale, all'accessibilità per le persone con disabilità, alla partecipazione dei portatori di interessi alla progettazione e allo sviluppo di sistemi di IA e alla diversità dei gruppi che si occupano dello sviluppo».

per la ricerca in materia. Dall'altro, di stabilire regole rigide sino a vietarne l'impiego nel caso di strumenti che rischiano di essere lesivi dei diritti e delle libertà fondamentali dei cittadini europei.

Per conseguire tale obiettivo complesso, la proposta di regolamento individua quattro specifiche categorie di possibile uso dell'IA a cui corrispondono diversi livelli di pericolosità e, quindi, diversi obblighi per i soggetti interessati.

La prima tipologia indicata nel Titolo II è denominata in maniera inequivocabile "pratiche di intelligenza artificiale vietate" e comprende tutti quei prodotti o servizi a cui viene associato un livello di rischio considerato talmente elevato da risultare incompatibile con il sistema di valori e principi condivisi a livello europeo. Per questo tipo di intelligenza artificiale, pertanto, l'immissione sul mercato, la messa in servizio e l'uso all'interno dei confini dell'UE sono assolutamente proibiti.

Più nel dettaglio, ai sensi dell'art. 5, par. 1, lett. c) rientrano in tale categoria innanzitutto i sistemi utilizzati dalle autorità pubbliche per la valutazione o la classificazione dell'affidabilità delle persone in base al loro comportamento sociale o alle caratteristiche della personalità, note o previste, mediante un punteggio sociale. Si tratta di una decisione di particolare rilievo, in quanto, proibendo tale specifico tipo di utilizzo, l'Unione europea ha deciso di assumere una posizione netta di rifiuto nei confronti di pratiche di *social rating* implementati in alcuni contesti nazionali – l'esempio più eclatante è il *social credit system* adottato dal governo cinese[54] – che consistono in meccanismi di profilazione estremamente invasivi perché mirano ad attribuire un punteggio sociale (*social score*) ai cittadini con lo scopo ultimo di individuare gli individui meno "virtuosi" ed escluderli dall'accesso a determinati servizi o prevedere per essi trattamenti differenziati o addirittura sfavorevoli.

In secondo luogo, la proposta all'art. 1, par. 2, lett. b) e c) vieta l'implementazione di sistemi di intelligenza artificiale che utilizzano tecniche subliminali o sfruttano le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica e mentale, al fine di sollecitare dei comportamenti in grado di provocare un danno fisico o psicologico per la propria persona o per altri.

Con tale divieto l'UE decide di proseguire anche all'interno dell'ecosistema dell'IA quella specifica attività regolatoria di contrasto, inaugurata nel settore radiotelevisivo con la nota direttiva 89/552/CEE "Televisione senza frontiere"[55] e confermata

[54] Il *Social credit system* è un programma governativo cinese di vaste proporzioni che mira a realizzare un sistema di credito sociale omnicomprensivo con l'obiettivo di creare una società in cui gli individui, le imprese e lo stesso governo agiscono tutti con integrità per la realizzazione di un'economica florida e di un regime sociale stabile. Il progetto ruota intorno ad uno specifico meccanismo di ricompense e sanzioni incentrato sull'ideale confuciano dello "Xin" (信,) una delle cinque virtù costanti del gentiluomo cinese, basata sull'importanza della costruzione della reputazione personale e dello sviluppo della sincerità. La creazione di una "società della reputazione" (*xinyong shehui*), pertanto, viene garantita attraverso un'analisi avanzata dei *big data* provenienti da specifiche piattaforme in modo che il rispetto della legge, l'affidabilità creditizia e la buona condotta vengono premiate e le infrazioni, invece, punite, eliminando di fatto i "costi" derivanti da una società caratterizzata da una "bassa fiducia". Sul punto si veda S. Fei, *Social Credit System in China* in *Panorama. Insights into Asian and European Affairs*, 2, 2018; G. Kostka, *China's social credit systems and public opinion: Explaining high levels of approval* in *New Media & Society*, 21 (7), 2019; 1565 ss.; G. Kostka - C. Zhang, *Tightening the grip: environmental governance under Xi Jinping* in *Environmental Politics*, 5 (27), 2018, 769 ss.

[55] Direttiva 89/552/CEE del Consiglio, del 3 ottobre 1989, relativa al coordinamento di determinate

nell'attuale panorama degli audiovisivi[56], che, come è noto, è diretta ad evitare attività persuasive che mirano a indurre specifici comportamenti d'acquisto o di consumo mediante la diffusione di frammenti visivi e sonori di matrice pubblicitaria che non sono percepiti in maniera cosciente dagli utenti finali. Una pratica che con l'avvento dell'IA assume indubbiamente proporzioni e un grado di incisività molto più elevati anche alla luce dei più ampi ambiti di applicazione e della varietà delle tecniche realizzabili e che richiede, quindi, interventi mirati che tutelino gli individui da meccanismi manipolativi che inficino la relativa capacità decisionale.

L'elenco elle tecnologie vietate all'interno dell'UE comprende, infine, i sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico per fini di attività di contrasto. Si tratta di meccanismi di riconoscimento a distanza basati sull'individuazione di alcune caratteristiche fisiche dei soggetti come il volto, le impronte digitali, nonché di comportamenti caratteristici come l'inflessione della voce e l'andatura. Rileva sottolineare che ricadono in tale categoria solo quegli strumenti impiegati in luoghi pubblici che, attraverso il rilevamento dei dati biometrici, consentono il riconoscimento istantaneo o comunque con breve ritardo di soggetti presenti all'interno di un *database* creato a fini di prevenzione, indagine, accertamento o perseguimento di reati o esecuzione di sanzioni penali, comprese anche le minacce alla sicurezza pubblica e la prevenzione delle stesse. È evidente che tale divieto è volto a evitare l'attuazione di metodi di trattamento di dati personali, spesso di carattere estremamente sensibile tenuto conto della tipologia di attività potenzialmente oggetto di monitoraggio (manifestazioni politiche, culturali, religiose, sociali), che potrebbero integrare meccanismi automatizzati di sorveglianza di massa, considerati a rischio elevato ai sensi dell'art. 35 del regolamento (UE) 679/2016[57].

Tale veto, tuttavia, incontra delle limitazioni nel caso in cui gli strumenti siano utilizzati per la ricerca mirata di potenziali vittime di crimini, inclusi i bambini scomparsi; la prevenzione di specifiche e imminenti minacce alla vita di persone o di attacchi terroristici e l'accertamento, la localizzazione, l'identificazione o l'azione penale nei confronti di un autore del reato o sospettato di un reato punibile con una pena o una misura massi-

disposizioni legislative, regolamentari e amministrative degli Stati membri concernenti l'esercizio delle attività televisive.

[56] Direttiva 2010/13/UE del Parlamento europeo e del Consiglio, del 10 marzo 2010, relativa al coordinamento di determinate disposizioni legislative, regolamentari e amministrative degli Stati membri concernenti la fornitura di servizi di media audiovisivi (direttiva sui servizi di media audiovisivi).

[57] Come è noto, l'art. 35, par. 3, del regolamento (UE) 679/2016 fornisce alcuni esempi specifici nei quali un trattamento possa presentare rischi elevati tra i quali rientra, al punto c), anche la sorveglianza sistematica su larga scala di una zona accessibile al pubblico. Il riconoscimento di tale situazione come altamente rischiosa è giustificato all'interno delle *Guidelines* in materia di valutazione di impatto redatte dall'allora Gruppo 29 operante in materia nelle quali si specifica che tale valutazione deriva dalla constatazione che «i dati personali possono essere raccolti in circostanze nelle quali gli interessati possono non essere a conoscenza di chi sta raccogliendo i loro dati e di come li utilizzerà. Inoltre, può essere impossibile per le persone evitare di essere soggette a tale trattamento nel contesto di spazi pubblici (o accessibili al pubblico)». Cfr. Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, spec. 7 ss.

ma di almeno tre anni. Al fine, comunque, di evitare utilizzi distorsivi delle tecnologie di identificazione biometrica, tali eccezioni possono essere validamente attuate solo in presenza di una previa autorizzazione da parte dell'autorità giudiziaria o dell'autorità competente in materia in cui vi sia indicazione specifica del luogo in cui si intende posizionare lo strumento, del tempo di utilizzo e delle motivazioni che giustificano l'implementazione di tali dispositivi.

La seconda categoria di regole previste dalla proposta di regolamento riguarda le pratiche alle quali l'Unione europea associa un rischio elevato per la salute e la sicurezza o per i diritti fondamentali delle persone fisiche. Si tratta di un bacino di strumenti particolarmente ampio e articolato in quanto la relativa individuazione, basata sulle normative attualmente vigenti in materia di sicurezza dei prodotti, tiene conto non solo della funzione svolta dai sistemi di IA, ma anche delle finalità e delle modalità specifiche di utilizzo degli stessi. Per gli strumenti presenti all'interno di tale categoria, il legislatore consente la relativa immissione sul mercato europeo subordinatamente al rispetto di un quadro dettagliato di requisiti obbligatori volti a limitare i possibili danni derivanti da un loro impiego diretto o indiretto e a una valutazione della conformità *ex ante* alle norme in materia.

Per facilitare il riconoscimento di tali strumenti, il regolamento individua all'art. 6 due distinte categorie. Nella prima confluiscono tutti quei sistemi destinati ad essere utilizzati come componenti di sicurezza di prodotti soggetti a valutazione della conformità *ex ante* da parte di terzi. La seconda, invece, racchiude quelle applicazioni del "pensiero artificiale" che presentano, secondo il legislatore europeo, delle implicazioni principalmente in relazione ai diritti fondamentali. Di queste viene fornito un elenco dettagliato nell'allegato III alla proposta con la premessa, però, sancita all'art.7 del testo principale che, data la velocità di evoluzione di tali pratiche, lo stesso potrà essere in futuro aggiornato e implementato dalla Commissione con atti delegati adottati nel rispetto della procedura di cui all'art.73.

Nello specifico, gli strumenti di IA a rischio elevato riguardano in primo luogo i sistemi di identificazione e categorizzazione biometrica in tempo reale o a posteriori delle persone fisiche utilizzate, in questo caso, non per attività di contrasto e il cui impiego, in assenza di adeguati misure di garanzia e di prevenzione, possono comunque sfociare in decisioni discriminatorie o lesive dei diritti fondamentali. Seguono i sistemi utilizzati come componenti di sicurezza nelle infrastrutture critiche individuate nella gestione del traffico stradale e nella fornitura di acqua, gas, riscaldamento ed elettricità. In tale caso è evidente la necessità di prevedere meccanismi appropriati di tutela dal momento in cui eventi anomali e tentativi di manomissione, alterazione o di sottrazione di dati utilizzati in tali infrastrutture possono arrecare un danno alla salute o avere conseguenze serie sulla vita degli utenti.

Il terzo ambito riguarda il settore dell'istruzione e della formazione professionale con riferimento ai sistemi di IA impiegati al fine di determinare l'accesso o l'assegnazione di persone fisiche agli istituti di istruzione e formazione professionale ovvero per valutare le attività degli studenti o le relative prove d'ammissione. In questo caso l'utilizzo di strumenti di IA in assenza di specifici meccanismi di verifica dell'*iter* decisionale algoritmico e di controllo umano possono inevitabilmente integrare situazioni discri-

minatorie perpetuando *bias* e limitando l'esercizio di diritti fondamentali come quello appunto all'istruzione.

Ragionamento analogo si applica per i sistemi utilizzati in materia di occupazione, gestione dei lavoratori e accesso al lavoro autonomo indicati al punto 4). L'obiettivo anche in questo caso, come evidenziato recentemente in un'ordinanza del Tribunale di Bologna nel dicembre 2020 in occasione di un famoso caso che ha coinvolto una società di *food delivery*[58], è quello di evitare che l'impiego dell'IA si trasformi da strumento di supporto alle attività lavorative complessivamente intese a mezzi di perpetrazione di situazioni lesive che vanno dal mancato rispetto delle parità di accesso alle proposte lavorative all'esercizio di diritti tutelati costituzionalmente come quello di scioperare, fino alla determinazione di condizioni non rispettose della dignità dei lavoratori.

Costituiscono, ancora, forme di IA ad alto rischio quelle tecnologie impiegate per determinare l'accesso a prestazioni e servizi pubblici e a servizi privati essenziali. Rientrano in tale categoria, in particolare, i sistemi di calcolo dell'affidabilità creditizia e quelli utilizzati dalle autorità pubbliche o per loro conto al fine di valutare l'ammissibilità delle persone fisiche alle prestazioni e ai servizi di assistenza pubblica e, infine, i sistemi di IA destinati a essere utilizzati per inviare servizi di emergenza di primo soccorso o per stabilire priorità in merito all'invio di tali servizi, compresi vigili del fuoco e assistenza medica. Al punto 6) è, invece, indicata una lista di tecnologie di IA utilizzati per le attività di contrasto, tra i quali emergono i sistemi destinati a valutare il rischio di reato o di recidiva ovvero il rischio per le vittime; quelli impiegati dalle autorità per rilevare lo stato emotivo di una persona fisica e quelli utilizzati per individuare immagini, contenuti audio o video manipolati che possono risultare falsamente autentici o veritieri (cd. *deep fake*). Chiudono l'elenco delle tecnologie ad alto rischio quelle destinate alla gestione della migrazione, dell'asilo e del controllo delle frontiere mediante un'attività di sostegno nella valutazione dell'autenticità dei documenti e dell'ammissibilità delle domande di asilo, di visto e di permesso di soggiorno e, infine, i sistemi di IA destinati ad assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge.

La terza categoria, disciplinata all'interno del Titolo IV, racchiude, invece, tre specifici utilizzi dell'IA che il legislatore europeo ha deciso di sottoporre a stringenti obblighi di trasparenza al fine di limitare i rischi specifici di manipolazione che essi sono in grado di perpetrare. Si fa riferimento, in primo luogo, a tutte quelle tecnologie che sfruttano il "pensiero artificiale" per interagire direttamente con gli utenti, si pensi ad esempio alle *chat-box*, simulando la presenza di una persona fisica. Per tali applicazioni, l'art. 52, par.

[58] Trib. Bologna, sez. lavoro, ord. 31 dicembre 2020. Come è noto, il Tribunale ha ritenuto discriminatorio l'algoritmo utilizzato da una nota azienda di *food delivery* per organizzare le prenotazioni e definire le sessioni di lavoro dei propri *riders* mediante il supporto di una piattaforma digitale. In particolare, il Giudice ha evidenziato che «il sistema di profilazione dei rider adottato dalla piattaforma *Deliveroo*, basato sui due parametri della affidabilità e della partecipazione, nel trattare nello stesso modo chi non partecipa alla sessione prenotata per futili motivi e chi non partecipa perché sta scioperando (o perché è malato, è portatore di un handicap, o assiste un soggetto portatore di handicap o un minore malato, ecc.) in concreto discrimina quest'ultimo, eventualmente emarginandolo dal gruppo prioritario e dunque riducendo significativamente le sue future occasioni di accesso al lavoro». Sul punto G. Fava, *L'ordinanza del Tribunale di Bologna in merito alla possibile discriminatorietà dell'algoritmo utilizzato da Deliveroo* in *Lavoro Diritti Europa, Rivista nuova del diritto del lavoro*, 1, 2021.

1, stabilisce l'obbligo per i fornitori di implementare meccanismi conoscitivi di *default* che consentano agli utenti di comprendere immediatamente di star interagendo con una macchina e non con un individuo reale. Seguono poi i commi 2 e 3 che impongono la massima trasparenza nei confronti dei destinatari in caso di utilizzo di sistemi di riconoscimento delle emozioni, di categorizzazione biometrica e di attività di *deep fake*. Ne consegue che l'implementazione di tali specifiche applicazioni richiede sempre che l'utente sia sufficientemente edotto circa le modalità di funzionamento dei sistemi di IA e nel caso specifico del deep fake sia consapevole che il contenuto che sta visionando è stato generato o manipolato artificialmente al fine di evitare di infondere la falsa convinzione che le immagini siano vere.

Rileva sottolineare che le disposizioni di cui al Titolo IV sollevano qualche perplessità in quanto, ad una prima lettura sotto il profilo sistematico, le tecnologie analizzate sembrerebbero caratterizzate da un livello di rischiosità minore rispetto a quanto indicato per i sistemi precedentemente analizzati vista la loro assenza nell'elenco di cui all'allegato III della proposta di regolamento. Tuttavia, l'art. 52 al par. 4 stabilisce che «i paragrafi 1, 2, 3 lasciano impregiudicati i requisiti e gli obblighi di cui al Titolo III del presente regolamento» a testimonianza del fatto che le disposizioni in materia di trasparenza dovrebbero – il condizionale è d'obbligo – aggiungersi a quelle previste per i sistemi ad alto rischio. Tale incertezza normativa assume particolare rilievo in quanto le tecnologie prese in considerazione appaiono tutt'altro che limitatamente incisive, considerata la possibilità di realizzare, mediante l'implementazione di sistemi di riconoscimento delle emozioni, profilazioni sofisticate ed estremamente penetranti la sfera di intimità dei singoli ovvero di determinare conseguenze negative in termini di pluralismo informativo e di veridicità delle informazioni nel caso di situazioni non correttamente regolamentate (e sanzionate) di *deep fake*.

L'ultima categoria, infine, assume carattere residuale e accoglie tutte quelle applicazioni dell'IA che presentano un rischio minimo o nullo, come nel caso di videogiochi o dei filtri spam nei messaggi di posta elettronica, e per il qual motivo non sono sottoponibili al quadro normativo delineato dal regolamento.

5. Il futuro quadro di tutela dell'individuo nella società del "pensiero artificiale"

In presenza di un atto normativo che si dichiara "*human-centric*" assume un ruolo inevitabilmente cruciale il quadro di garanzie approntato dal legislatore al fine di proteggere l'individuo da conseguenze negative derivanti dalla decisione algoritmica. In tale prospettiva, la proposta di regolamento innesta tale tipo di tutela, in primo luogo, all'interno di uno specifico percorso di responsabilizzazione dei soggetti che producono o che si servono dei sistemi di IA riprendendo quel fortunato filone dell'"*accountability*" che, come è noto, rappresenta oggi la spina dorsale del regolamento (UE) 679/2016. In tale ottica, i sistemi algoritmici sin dalla loro progettazione vengono inseriti all'interno di un processo di tipo iterativo che mira a minimizzare l'impatto sull'esercizio di diritti e libertà fondamentali lungo l'intero ciclo di vita di tali strumenti attraverso un

percorso di garanzia di conformità alle regole in materia da attuarsi in maniera costante e sistematica (art. 9). Grazie a tale impostazione la creazione, l'immissione sul mercato e la messa a disposizione di sistemi di intelligenza artificiale a cui è associato un rischio elevato vengono subordinate al soddisfacimento di un insieme di misure tecniche ed organizzative predisposte a partire dai soggetti fornitori e comprovate a cascata dagli operatori variamente coinvolti lungo la catena del valore dell'IA che puntano a blindare l'impiego di tali strumenti all'interno del quadro dei principi e dei valori condivisi a livello sovranazionale.

Ed è proprio lungo tale complessa architettura, sostenuta da un istituendo sistema di *governance* a livello sovranazionale caratterizzato dalla creazione di un apposito Comitato europeo in materia e di specifiche autorità di controllo nazionali (Titolo VI della proposta), che si struttura in maniera decisiva il quadro di tutela prevista dal legislatore europeo a favore del destinatario finale del processo algoritmico.

Nell'articolato sistema di nuove regole destinate a disciplinare tale mutevole scenario emerge in particolar modo l'obbligo a carico dei fornitori *ex* art.17 di implementare un sistema di gestione della qualità che garantisca la piena e, soprattutto, duratura conformità delle tecnologie alle disposizioni previste nella proposta attraverso la definizione di politiche, procedure e istruzioni che definiscano in dettaglio la strategia di controlli e di garanzie che tali soggetti intendono attuare nel corso dell'intera vita dello strumento. Si tratta di una richiesta particolarmente impegnativa dal momento che il conseguimento degli standard qualitativi richiesti dal legislatore europeo impone una visione omnicomprensiva e dettagliata del funzionamento dei sistemi di IA che consenta al fornitore di definire *ex ante* una serie di specifiche tecniche e organizzative finalizzate non solo a prevenire i rischi connessi a tali strumenti anche attraverso un'attenta e corretta gestione dei dati il cui utilizzo, come è noto, rappresenta un elemento vitale, ma altresì di intervenire immediatamente ed efficientemente in caso di eventi dannosi. In tale prospettiva si inserisce, mediante l'art. 9, il *Risk management system*, un percorso volto ad assicurare il livello qualitativo dei sistemi algoritmici articolato in quattro fasi fondamentali costituite dall'identificazione e analisi dei rischi noti e prevedibili associati all'utilizzo dell'IA; dalla stima di quelli che possono emergere in caso di usi impropri ragionevolmente prevedibili; dalla valutazione delle criticità ulteriori che appaiono in occasione dell'analisi dei dati raccolti e del monitoraggio effettuato una volta immesso il sistema sul mercato e, infine, dalla predisposizione di adeguate misure di gestione dei rischi individuati.

È palese l'affinità che lega tale meccanismo di controllo al *Data Protection impact assessment* (DPIA) previsto dall'art. 35 del regolamento (UE) 679/2016 con cui condivide sia la ragion d'esistenza determinata dalla rischiosità intrinseca dell'utilizzo di tali tecnologie, sia la necessità di un controllo che sia *by design* e *by default*. Comune è, infatti, l'obiettivo di garantire, ancor prima dell'immissione sul mercato di tali strumenti, la predisposizione di misure tecniche ed organizzative che consentano di intervenire immediatamente in occasione di un evento dannoso e di limitare il più possibile conseguenze lesive per i destinatari. Più ampi, invece, risultano essere i confini applicativi, in quanto, a differenza di quanto non avvenga all'interno dell'ecosistema del trattamento dei dati personali, il *risk management system* non si limita alla fase iniziale, ma opera per

l'intero ciclo di vita di un sistema di IA richiedendo un aggiornamento costante e sistematico, giustificato dal continuo progredire delle modalità applicative e delle caratteristiche di tali tecnologie.

L'articolata rete di protezione delineata a favore del destinatario finale si sviluppa, inoltre, anche attraverso il coinvolgimento diretto degli altri punti nodali della catena del valore dell'IA. Gli importatori, i distributori e gli utenti, infatti, non solo diventano, nel rispetto delle loro specificità attoriali, "garanti di secondo grado" della conformità degli strumenti di IA attraverso la verifica dell'avvenuto soddisfacimento degli adempimenti imposti al fornitore dal regolamento[59], ma operano anche come "sentinelle" perché sono tenuti a impedirne la messa a disposizione sul mercato in assenza dei requisiti richiesti o, nel caso specifico degli utenti, a sospenderne l'impiego in seguito a malfunzionamento o incidente grave.

A completare, infine, il quadro di regole che la proposta stabilisce ai fini di un utilizzo etico e affidabile dei sistemi di IA che presentano un rischio elevato intervengono: la redazione della documentazione tecnica necessaria da sottoporre alle autorità nazionali competenti e agli organismi notificati al fine di consentire la valutazione di adeguatezza di tali strumenti (art.11); l'obbligo di utilizzo di *set* di dati di addestramento, convalida e prova che siano pertinenti, rappresentativi, esenti da errori e completi (art. 10); la registrazione automatica degli eventi ("log") in modo da monitorare il funzionamento del sistema di IA una volta immesso nel mercato e intercettare tempestivamente situazioni di rischio a livello nazionale (art.12) e la previsione di un adeguato livello di accuratezza, robustezza e cibersicurezza da garantire sin dalla fase della loro progettazione e sviluppo (art.15).

L'insieme delle summenzionate norme previste a carico dei soggetti che sviluppano o si affidano al "pensiero artificiale" per la realizzazione delle proprie attività, sommato al divieto di utilizzo previsto dal Titolo II di alcuni specifici strumenti ritenuti proibiti in quanto incompatibili con il sistema dei valori e dei principi condivisi a livello europeo e alla previsione di ulteriori obblighi di trasparenza per determinati strumenti al fine di tenere conto dei rischi specifici di manipolazione che essi comportano (Titolo IV), definisce pertanto l'intelaiatura fondamentale del quadro di tutela previsto dalla proposta di regolamento a favore dell'individuo che subisce la decisione algoritmica.

Secondo le intenzioni del legislatore europeo sarà, quindi, il pedissequo rispetto di tali norme a favorire la nascita di un ecosistema affidabile, sicuro ed etico in cui le criticità dell'IA saranno affrontate e celermente risolte in un'ottica di sviluppo *human-centric* che pone l'individuo, come detto, al centro della rivoluzione del "pensiero artificiale".

[59] L'art. 26 stabilisce, infatti, che «prima di immettere sul mercato un sistema di IA ad alto rischio, gli importatori di tale sistema garantiscono che: a) il fornitore di tale sistema di IA abbia eseguito l'appropriata procedura di valutazione della conformità». Per i distributori, invece, l'art. 27 stabilisce che: «prima di mettere a disposizione sul mercato un sistema di IA ad alto rischio, i distributori verificano che il sistema di IA ad alto rischio rechi la necessaria marcatura CE di conformità, che sia accompagnato dalla documentazione e dalle istruzioni per l'uso necessarie e che il fornitore e l'importatore del sistema, a seconda dei casi, abbiano rispettato gli obblighi di cui al presente regolamento». L'utente, invece, è tenuto ai sensi dell'art. 29 ad usare tali strumenti conformemente alle istruzioni per l'uso che accompagnano i sistemi e le misure di sorveglianza umana indicate dal fornitore.

6. Aspetti critici della proposta di regolamento: un quadro normativo *human-centric* che non contempla il destinatario della decisione algoritmica

Nonostante le fiduciose premesse orientate a garantire uno sviluppo europeo di tipo umano-centrico dell'Intelligenza artificiale, obiettivo più volte rimarcato anche nei suoi numerosi considerando, la proposta di regolamento, tuttavia, risulta essere debole proprio nella costruzione dell'impianto di tutela destinato agli utenti finali che subiscono la decisione "artificiale". Sebbene sia da plaudire la scelta di definire un quadro normativo indirizzato al complesso e cangiante universo dell'IA mediante un laborioso percorso di responsabilizzazione di tutti gli operatori coinvolti nella realizzazione, distribuzione e utilizzo di tali strumenti, non si ritiene ragionevole che la concreta attuazione del sistema di garanzie destinate agli utenti finali venga lasciata completamente alle capacità tecniche ed organizzative di tali soggetti, privando quello che dovrebbe essere il vero "protagonista" della proposta, l'individuo, di specifici e adeguati strumenti di azione nel caso in cui subisca una decisione algoritmica discriminatoria, erronea o in grado di ledere ingiustamente l'esercizio di un suo diritto fondamentale.
Nel testo, infatti, non si riscontrano regole specifiche volte a prevedere strumenti rimediali effettivi che il destinatario possa attivare autonomamente a seguito di un evento dannoso, né tantomeno si intravede un quadro di diritti che il soggetto possa esercitare in maniera non mediata alla stregua di quelli sanciti nel Capo III del regolamento (UE) 679/2016 che rappresentano, invece, un elemento centrale nello scenario della tutela dei dati personali. In altri termini, nonostante la proposta faccia più volte riferimento alla necessità di porre l'uomo al centro dello sviluppo dell'IA, alla fine lo relega in un angolo non riconoscendogli quegli strumenti indispensabili che gli consentirebbero di difendersi da un'evoluzione invasiva di tali tecnologie.
Una mancanza questa estremamente rilevante che è stata sottolineata anche in occasione della *Joint Opinion* 5/2021 presentata dall'*European Data Protection Board* e dall' *European Data Protection Supervisor* il 18 giugno 2021[60]. In tale documento, infatti l' EDPB e l'EDPS non solo sottolineano l'assenza nel futuro quadro normativo di specifici diritti e rimedi da attivare a disposizione degli individui sottoposti ai sistemi di IA, ma evidenziano un'ulteriore significativa lacuna consistente nella mancata definizione delle modalità con cui dovrà concretamente realizzarsi il coordinamento tra il nuovo quadro normativo e il preesistente complesso di norme sancite dal GDPR, l'EUDPR, l'ePrivacy Directive e la LED[61] nel caso in cui il processo decisionale algoritmico dovesse prevedere il trattamento di dati personali. In particolare, si pone il problema del raccordo, estremamente rilevante in questo scenario, tra quanto stabilito dall'art. 22 del GDPR che, come è ampiamente noto, riconosce il diritto per un soggetto di non essere sottoposto ad una decisione completamente automatizzata e il sistema di rego-

[60] EDPB-EDPS Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

[61] Si fa riferimento al regolamento (UE) 2016/679, "*General Data Protection Regulation*" (GDPR), al regolamento (UE) 2018/1725, "*Data Protection Regulation for the European Union institutions, offices, bodies and agencies*" (EUDPR) e alla direttiva (UE) 2016/680, c.d. "*Law Enforcement Directive*" (LED).

le stabilite dalla proposta che incardinano, come evidenziato, la tutela dell'individuo nel quadro degli obblighi tecnici ed organizzativi dei soggetti operanti nello scenario dell'IA. In tale prospettiva si inserisce anche la riflessione dal Garante italiano per la protezione dei dati personali che, in occasione della memoria presentata alla Camera dei deputati nel marzo 2019[62], ha evidenziato come la già citata "rete di protezione" che prevede il coinvolgimento attivo dei diversi soggetti che operano lungo la catena del valore nell'attività di supervisione delle norme in materia di IA, riducendo di fatto l'ambito di intervento delle autorità amministrative indipendenti, potrebbe determinare un «affievolimento delle garanzie dei diritti fondamentali e, in particolare, del diritto alla protezione dei dati personali, la cui tutela effettiva si fonda appunto (ai sensi degli artt. 8 CDFUE e 16 TFUE) sulla necessaria supervisione di Autorità indipendenti».

Ne consegue che, la mancata indicazione di come i due quadri normativi si coordineranno nel prossimo futuro, unita alla riscontrata assenza di strumenti rimediali per i destinatari di tali strumenti, rischia infatti di indebolire ulteriormente la posizione dei soggetti che subiscono passivamente la decisione algoritmica ponendoli all'interno di un limbo normativo conteso tra la tutela dei dati personali e le nuove norme in materia di IA.

A tal fine, non sembra essere risolutiva la previsione *ex* art. 14 che impone l'intervento umano nell'utilizzo di strumenti ad alto rischio secondo la ben nota visione dello "*human in the loop*". Nella consapevolezza che i sistemi appartenenti a tale categoria risultano potenzialmente in grado di arrecare rischi alla salute, alla sicurezza e ai diritti fondamentali anche quando sono utilizzati conformemente alla loro finalità, infatti, la proposta impone la predisposizione di adeguati meccanismi di sorveglianza umana che consentano di intervenire celermente ai primi segnali di anomalie, disfunzioni o prestazioni inattese ovvero nel caso in cui sia necessario ignorare, annullare o ribaltare l'esito prodotto dalla macchina. Tale misura, che viene ulteriormente rafforzata in occasione dell'impiego di sistemi per l'identificazione biometrica remota "in tempo reale" e "a posteriori" delle persone fisiche[63], consente, quindi, di introdurre in maniera stabile il prezioso contributo esperienziale e professionale umano all'interno del circuito artificiale e, al contempo, permette di combattere il fenomeno della "distorsione dell'automazione", inteso come la possibile tendenza a fare automaticamente o eccessivamente affidamento al risultato prodotto dall'IA. Tuttavia, l'obbligo del coinvolgimento umano, per quanto necessario, non solo non contribuisce a tutelare pienamente la capacità di autodeterminazione del destinatario del risultato algoritmico in quanto la decisione di intervenire o meno rimane comunque interamente nella discrezionalità del soggetto tenuto a controllare il processo decisionale, ma potrebbe risultare di dubbia efficacia nel caso in cui la macchina sia in grado di sviluppare percorsi cognitivi non comprensibili esternamente e, in particolare, in presenza di decisioni discriminatorie cd. indirette

[62] Memoria del Garante per la protezione dei dati personali - COM 2021(206) Proposta di regolamento (UE) sull'intelligenza artificiale, Camera dei Deputati - Commissioni IX e X riunite 9 marzo 2022.

[63] L'art. 14, punto 5, della proposta stabilisce, infatti, che nel caso dei sistemi di IA ad alto rischio elencati nell'allegato III, punto 1, lettera a), le misure generalmente previste in materia di sorveglianza umana debbono essere integrate in modo tale da garantire che l'utente non compia azioni o adotti decisioni sulla base dell'identificazione risultante dal sistema, a meno che essa non sia stata verificata e confermata da almeno due persone fisiche.

che risultano non immediatamente riconoscibili all'esterno perché esprimono il loro valore distorsivo solo nel corso del tempo[64].
Simili criticità si evidenziano anche nell'obbligo destinato ai produttori all'art.13 della proposta di progettare e sviluppare i sistemi di IA in modo da garantire che il loro funzionamento sia sufficientemente trasparente all'esterno e consentire, quindi, agli utenti finali di interpretare correttamente il risultato del processo decisionale e utilizzarlo adeguatamente.
Si tratta, evidentemente, di una previsione estremamente importante, ma, anche in questo caso, di dubbia efficacia se calata in alcuni specifici contesti applicativi perché non tiene conto del diffuso fenomeno dell'opacità dei percorsi inferenziali seguiti dalle macchine algoritmiche (il già citato fenomeno della *black box*) che rendono difficile, se non impossibile in alcuni casi, conoscere la motivazione che è alla base della decisione.
Di tale "preoccupazione" se ne fa parzialmente carico l'art. 15 che impone *standard* adeguati di accuratezza, robustezza e cibersicurezza durante tutto il ciclo di vita dell'IA ad altro rischio. In particolare, il punto 3 stabilisce esplicitamente che i sistemi che proseguono il loro apprendimento anche dopo la loro immissione sul mercato siano sottoposti ad un monitoraggio mirato delle prestazioni mediante l'utilizzo di *feedback loops* che prevedono l'utilizzo di tecniche di attenuazione degli output distorti in modo da limitare i casi in cui i risultati viziati siano utilizzati come *input* per operazioni successive.
Ma è evidente che tale previsione, in un panorama evolutivo che vede l'IA sempre più orientata verso percorsi di sviluppo autonomo che rischiano di sfuggire al controllo e alla stessa comprensibilità umana, risulta essere estremamente debole e poco risolutiva rispetto alle criticità evidenziate.

7. Riflessioni conclusive: verso una "discutibile" tutela dell'individuo di tipo *consumer-centric*?

La previsione di un percorso di responsabilizzazione dei soggetti variamente coinvolti nella catena del valore dell'IA connessa all'assenza di un quadro di diritti e un sistema di strumenti rimediali attivabili dal destinatario finale della decisione algoritmica svela quello che è l'elemento di maggiore criticità della proposta di regolamento adottata nell'aprile 2021: la decisione di calare la tutela dell'individuo nel panorama dell'IA all'interno della dialettica produttore-consumatore
Si tratta di una scelta che in realtà traspare sin dalla relazione di accompagnamento all'atto laddove il legislatore dichiara esplicitamente che la base giuridica «è costituita innanzitutto dall'articolo 114 del Trattato sul funzionamento dell'Unione europea (TFUE), che prevede l'adozione di misure destinate ad assicurare l'instaurazione ed il funzionamento del mercato interno».
L'utilizzo del termine "innanzitutto" è chiaramente espressione della volontà di chiarire che l'impianto regolatorio in materia dovrà soddisfare in primo luogo le esigenze di un mercato europeo sempre più desideroso di affermarsi come protagonista nello

[64] Si fa riferimento in particolare al fenomeno della *proxy discrimination* che verrà analizzato *infra*.

scenario internazionale in materia di IA, sebbene nel rispetto del quadro dei principi e dei valori condivisi dagli Stati membri.

Tale visione risulta ancora più evidente e, al tempo stesso, portatrice di criticità se si procede all'analisi del Titolo VII della proposta dedicato alla disciplina della fase di monitoraggio degli strumenti che impiegano il "pensiero artificiale" in seguito alla loro immissione sul mercato con il fine di garantire che sugli stessi operi un controllo continuativo lungo tutto l'arco della loro vita operativa.

In tale contesto assume rilievo, in particolare, la previsione *ex* art. 63 che inserisce i sistemi algoritmici nel novero dei prodotti a norma ai sensi del regolamento (UE) 2019/1020[65]. Tale rimando normativo a un atto che agisce in materia di vigilanza del mercato e di conformità dei prodotti, dietro ad un apparente linearità argomentativa, nasconde in realtà delle ricadute rilevanti sulla futura disciplina del "pensiero artificiale" perché inserisce, non senza forzature, il multiforme e complesso universo applicativo dell'IA all'interno di una categoria regolatoria peculiare e preesistente, finendo col parificare il rischio associato all'utilizzo di tali strumenti a quello di un qualsiasi altro prodotto immesso sul mercato europeo Se la proposta verrà confermata, infatti, un sistema di IA, in caso di pericolo a livello nazionale, sarà considerato alla stregua di «un prodotto che presenta un rischio definito all'articolo 3, punto 19, del regolamento (UE) 2019/1020», vale a dire al pari di un bene che potenzialmente è in grado di «pregiudicare la salute e la sicurezza delle persone in generale, la salute e la sicurezza sul posto di lavoro, la protezione dei consumatori, l'ambiente e la sicurezza pubblica, nonché altri interessi pubblici tutelati dalla normativa di armonizzazione dell'Unione». Sebbene si faccia riferimento ad un quadro di implicazioni negative che effettivamente possono derivare da una applicazione diffusa del "pensiero algoritmico", tale insieme di effetti da solo non è in grado di cogliere appieno la complessità del variegato mondo dell'IA, in quanto riflette solo una parte del quadro delle conseguenze - ad oggi prospettabili - connesse ad un utilizzo non correttamente regolamentato di tali sistemi in un'ottica umano-centrica.

Come evidenziato nel corso dell'analisi, infatti, lo spettro dei rischi che si annida dietro l'utilizzo di tali tecnologie non solo risulta estremamente complesso da identificare in quanto non lascia "tracce visibili" a differenza di un prodotto malfunzionante o difettoso, ma soprattutto viene a toccare uno degli ambiti maggiormente connessi alla vitalità di un sistema democratico: il libero determinarsi di un individuo nella sua duplice sfera di singolo e di componente di una società. La creazione di *epistemic bubbles* ed *eco-chambers,* i meccanismi di *nudging* algoritmico, i risultati decisionali di tipo predittivo e il moltiplicarsi di situazioni di *black box* sono, infatti, tutti espressione di un fenomeno che, favorito dall'indiscutibile contributo positivo che apporta al progresso tecnologico ed economico, non si manifesta nel suo lato oscuro solo in termini di pericolo per la

[65] L'art. 63 della proposta stabilisce, infatti, che «il regolamento (UE) 2019/1020 si applica ai sistemi di IA disciplinati dal presente regolamento. Tuttavia, ai fini dell'efficace applicazione del presente regolamento: a) ogni riferimento a un operatore economico a norma del regolamento (UE) 2019/1020 si intende fatto anche a tutti gli operatori di cui al titolo III, capo 3, del presente regolamento; b) ogni riferimento a un prodotto a norma del regolamento (UE) 2019/1020 si intende fatto anche a tutti i sistemi di IA che rientrano nell'ambito di applicazione del presente regolamento».

sicurezza o per la salute dei destinatari finali, bensì anche e soprattutto in termini di rischio per l'effettiva tutela di quel principio personalistico che irradia le costituzioni democratiche e intorno al quale si dispiegano diritti e libertà fondamentali.
Di tale aspetto, che è parte integrante della più ampia rivoluzione apportata dall'IA e di cui rappresenta una delle conseguenze sicuramente più insidiose per l'evoluzione in senso democratico delle società tecnologicamente avanzate, non vi è nessun riflesso concreto nel quadro di regole previste dal legislatore, ad eccezione della suddivisione in categorie dei sistemi che integrano il "pensiero artificiale" secondo il già citato *risk-based approach* di tipo proporzionale. Non vi è e, a questo punto, non potrebbe esserci un riferimento normativo esaustivo in grado di contemplare anche quelle applicazioni che possono e potrebbero incidere sulla autodeterminazione dei singoli, sul libero dispiegarsi della loro personalità e sull'esercizio dei propri diritti.
In tale prospettiva, quindi, la proposta di regolamento piuttosto che definire un quadro regolatorio di tipo *human-centric* sembra delineare un sistema *consumer-centric* nel quale la tutela viene a strutturarsi quasi esclusivamente attorno alle figure operative che agiscono all'interno dell'ecosistema dell'intelligenza artificiale con l'obiettivo di garantire agli utenti-consumatori la possibilità di utilizzare prodotti, sì dotati di IA, ma che al pari degli altri presenti sul mercato europeo, siano soprattutto sicuri e affidabili.
Ne è un esempio specifico la previsione di cui all'art. 62 che affida ai fornitori il compito di segnalare alle autorità di vigilanza «qualsiasi incidente grave o malfunzionamento di tali sistemi che costituisca una violazione degli obblighi previsti dal diritto dell'Unione». Tale previsione è corredata dalla richiesta che la notifica venga effettuata immediatamente, non appena sia stato individuato un nesso causale ovvero una ragionevole probabilità di collegamento tra quanto accaduto e l'utilizzo del sistema di IA.
Sebbene tale disposizione ricalchi quanto già stabilito analogamente dal regolamento 679/2016 in caso di eventi di *data breach*[66], è intuitivo comprendere la rilevante differenza che intercorre tra le due situazioni disciplinate dal legislatore europeo. Nel panorama della tutela dei dati personali, infatti, l'evento dannoso consiste, come riportato nell'art. 4, n.12 del GDPR, in «una violazione di sicurezza che comporta accidentalmente o in modo illecito la distruzione, la perdita, la modifica, la divulgazione non autorizzata o l'accesso ai dati personali trasmessi, conservati o comunque trattati». Ne consegue che la tipologia di rischio è ragionatamente prevedibile (violazione di sicurezza) e le possibili conseguenze rimangono all'interno di un *range* piuttosto definito (distruzione, perdita, modifica, divulgazione o accesso non autorizzato di dati).
Nel panorama dell'IA, invece, non solo lo spettro dei rischi potenziali varia al variare del tipo di sistema utilizzato e delle finalità perseguite, ma, come già evidenziato, risulta estremamente più ampia e dalle molteplici sfumature la gamma dei possibili effetti sulla persona e sulla società. Ciò comporta, in primo luogo, che non è sempre agevole, se non addirittura possibile, individuare una relazione certa tra impiego del "pensiero

[66] L'art. 33 del regolamento (UE) 679/2016 prevede, infatti, che «in caso di violazione dei dati personali, il titolare del trattamento notifica la violazione all'autorità di controllo competente a norma dell'articolo 55 senza ingiustificato ritardo e, ove possibile, entro 72 ore dal momento in cui ne è venuto a conoscenza, a meno che sia improbabile che la violazione dei dati personali presenti un rischio per i diritti e le libertà delle persone fisiche. Qualora la notifica all'autorità di controllo non sia effettuata entro 72 ore, è corredata dei motivi del ritardo».

artificiale" e impatto sull'esercizio di diritti e libertà fondamentali. Si pensi ad esempio ai fenomeni di *proxy discrimination* in cui la decisione algoritmica, nonostante utilizzi agli occhi degli osservatori parametri tendenzialmente neutri, assume indirettamente, appunto "per procura", una valenza discriminatoria celata dietro un risultato apparentemente corretto, ma che di fatto è in grado di ledere determinate categorie o gruppi di soggetti e che si manifesta solo con il tempo[67].

Ma soprattutto, in un contesto così insidioso, in cui risulta arduo perfino individuare gli effetti dannosi delle singole applicazioni algoritmica è indiscutibile la criticità insista nell'affidare ai fornitori e agli altri operatori coinvolti nella catena del valore dell'IA, perlopiù rappresentati da società private, gran parte del sistema di gestione dei rischi quando i pericoli connessi a tali strumenti contemplano anche e soprattutto la sfera dei diritti e delle libertà fondamentali e possono mettere in seria difficoltà la tenuta degli assetti democratici.

Per quanto la proposta preveda, come indicato, uno specifico sistema di *governance* europea in materia di IA[68], il contributo del futuro Comitato e delle singole autorità nazionali si inserirà solo a completamento di un percorso di garanzia e di controllo della conformità degli strumenti algoritmici che partirà e si strutturerà soprattutto attraverso l'azione di tali operatori.

Spetterà, infatti, a quest'ultimi, nell'ambito del sistema di conformità dei prodotti procedere all'identificazione e all'analisi dei pericoli non solo noti e prevedibili, ma anche potenziali associati a ciascun sistema di IA ad alto rischio, nonché attuare misure tecniche tali da garantire che i pericoli residuali connessi al loro utilizzo risultino accettabili. Ci si chiede a questo punto quali potranno essere i parametri di riferimento a cui gli operatori privati affideranno tali valutazioni e quale sarà il livello di rischio considerato "accettabile" per soggetti che di fatto perseguono finalità di tipo lucrativo. Una problematica che risulta particolarmente rilevante soprattutto con riferimento a quei sistemi algoritmici per i quali la proposta di regolamento riconosce la possibilità di effettuare internamente, senza l'intervento *ex ante* di un organismo notificato, il controllo di conformità al quadro normativo sulla base della constatazione «che l'intervento normativo è in fase iniziale e che il settore dell'IA è molto innovativo e soltanto ora si stanno

[67] Un esempio semplice, ma che è in grado di esplicitare il potenziale impatto di tale pratica è rappresentato dal caso in cui il *software* di una banca, addestrato utilizzando la serie storica dei dati relativi alla solvibilità della propria clientela, individui una correlazione tra casi di mancata restituzione delle somme prese in prestito e codice di avviamento postale e sulla base di tale legame rigetti sistematicamente le richieste che provengono da tale area territoriale. Se in tale contesto ambientale si concentra un numero elevato di persone appartenenti ad una determinata etnia, il processo decisionale automatizzato, pur partendo da informazioni sostanzialmente neutre (solvibilità- area abitativa) e, quindi, non ricadenti in alcuna delle categorie normalmente protette, si risolve inevitabilmente, ma non palesemente, in una discriminazione razziale comportando una lesione della sfera personale e professionale di tali soggetti che si vedranno privati della possibilità di accedere a importanti opportunità di concessione del credito. F.Z. Borgesius, *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, Anti-discrimination Department of the Council of Europe, Strasbourg, 2018. Per una dettagliata analisi del fenomeno della *proxy discrimination* e sulle sue molteplici sfumature applicative cfr. A. E.R. Prince – D. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *105 Iowa L. Rev.*, 2020, 1257 ss.

[68] Sul punto cfr. C. Casonato - B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale in BioLaw Journal –Rivista di BioDiritto*, 3, 2021.

maturando le competenze di audit»[69]. Questa peculiare "concessione", che concentra il controllo iniziale esclusivamente nelle mani dei fornitori, risulta estremamente critica perché coinvolge paradossalmente gran parte dei sistemi di IA a rischio elevato elencati nell'allegato III e, pertanto, strumenti che possono assumere risvolti di carattere pubblicistico come quelli rientranti nella gestione e nel funzionamento delle infrastrutture critiche (punto 2) ovvero quelli destinati all'accesso a prestazioni e servizi pubblici e a servizi privati essenziali e fruizione degli stessi (punto 5). È indiscutibile che proprio tali applicazioni che risultano più strettamente connesse alla realizzazione in senso democratico delle società tecnologicamente avanzate e, di conseguenza, all'effettiva possibilità di un esercizio pieno e consapevole dei diritti e delle libertà fondamentali, avrebbero dovuto essere oggetto di una disciplina più stringente soprattutto nella fase di valutazione iniziale della conformità al quadro normativo. Questo al fine di evitare sin dall'origine l'introduzione e l'applicazione, proprio negli spazi di intervento di carattere pubblico, di strumenti sicuramente innovativi, ma di fatto non ammissibili perché in contrato con il sistema dei valori e dei principi condivisi a livello europeo.

Ne consegue che in un panorama così complesso il futuro assetto normativo in materia di IA, se non modificato, rischia di imprigionare l'individuo all'interno di una categoria, quella del consumatore, che non è in grado di per sé di garantire in futuro un solido sistema di tutele tale da consentire uno sviluppo consapevole e libero della propria personalità all'interno di assetti in cui, parimenti, in assenza di una specifica regolamentazione, sarà messa sempre più a dura prova la tenuta di istituti cardine dei sistemi democratici.

Dinanzi a tale rischio risulta necessario intervenire sin da subito riportando immediatamente al centro della rivoluzione determinata dal "pensiero artificiale" l'uomo, nella sua veste non di semplice utente finale di prodotti o servizi "intelligenti", bensì di soggetto titolare di diritti e libertà fondamentali anche alla luce dello specifico quadro delineato dalla Dichiarazione europea sui diritti e i principi digitali per il decennio digitale approvata nel gennaio 2022 con la quale l'Unione europea si prefigge di realizzare un "percorso per il decennio digitale" significativamente caratterizzato da «una trasformazione digitale che mette al centro le persone; che si basa sulla solidarietà e sull'inclusione; che ribadisce l'importanza della libertà di scelta; che promuove la partecipazione allo spazio pubblico digitale; che garantisce la sicurezza, la protezione e il conferimento di maggiore autonomia e responsabilità, e la sostenibilità»[70]. In tale prospettiva, risulta fondamentale procedere alla previsione di meccanismi di azione diretta che consentano all'individuo di tutelarsi da logiche decisionali che sono in grado di determinare una

[69] La relazione accompagnatoria alla proposta di regolamento al paragrafo 5.2.3., specifica, infatti, «per quanto riguarda i sistemi di IA ad alto rischio indipendenti di cui all'allegato III, sarà istituito un nuovo sistema di conformità e applicazione. Tale scelta segue il modello della legislazione del nuovo quadro normativo attuata mediante verifiche di controllo interno da parte di fornitori, fatta eccezione per i sistemi di identificazione biometrica remota che sarebbero soggetti a una valutazione della conformità da parte di terzi. Una valutazione completa della conformità ex ante attraverso controlli interni, combinata con una forte applicazione ex post, potrebbe costituire una soluzione efficace e ragionevole per tali sistemi, considerato che l'intervento normativo è in fase iniziale e che il settore dell'IA è molto innovativo e soltanto ora si stanno maturando le competenze di audit».

[70] Commissione europea, *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, COM (2022) 28, considerando 5.

limitazione della propria capacità di autodeterminarsi in maniera piena e consapevole all'interno dello scenario dell'IA.

Occorre, altresì, intervenire immediatamente nella definizione delle modalità di coordinamento con il quadro normativo dettato dal regolamento 679/2016 nella consapevolezza che l'utilizzo di tali tecnologie presenta inevitabili momenti di contatto - se non di vera e propria sovrapposizione - con le attività di trattamento dei dati personali e, in assenza di idonee regole, rischia di lasciare l'individuo sprovvisto di fatto di idonee garanzie proprio in occasione di applicazioni del "pensiero artificiale" particolarmente pericolose che uniscono la potenza inferenziale dei sistemi algoritmici all'utilizzo di frammenti informativi della propria identità personale. Inoltre, il citato sistema di conformità - che lascia di fatto privi di un controllo *ex ante* di carattere esterno proprio ambiti di intervento pubblico particolarmente vulnerabili, anche alla luce dei fenomeni di opacità giuridica e delle conseguenze spesso imprevedibili del "pensiero artificiale" - spinge ad un ripensamento della classificazione dei sistemi di IA basata non solo sulla rischiosità effettiva e potenziale di tali strumenti, ma anche e soprattutto dell'ambito di applicazione degli stessi, tenendo in debito conto i casi in cui il relativo utilizzo contempli decisioni di carattere pubblico.

In tale contesto, emerge, infatti, la necessità di una regolamentazione specifica e più penetrante che definisca sin dall'inizio i parametri di riferimento dell'impiego del "pensiero artificiale" e maggiori garanzie nel caso di un eventuale uso distorsivo al fine di evitare che l'individuo risulti destinatario di una decisione completamente automatizzata e determinata alla luce di un percorso inferenziale che rischia di non essere noto e gestibile perfino dal decisore pubblico.

In tale prospettiva, è fondamentale riflettere anche su quale sarà il ruolo degli Stati che, pur dinanzi ad un atto così penetrante come il regolamento europeo, saranno tenuti a convogliare la portata innovativa di tali strumenti all'interno dei confini dettati dai principi e dai valori che definiscono il proprio orizzonte ordinamentale, definendo soluzioni normative nazionali che preservino i processi decisionali di carattere pubblicistico dalle criticità evidenziate in un'ottica ovviamente non di distonia, ma di necessaria complementarietà con il quadro regolamentare che sta emergendo a livello sovranazionale[71].

L'obiettivo ultimo è inevitabilmente quello di creare un panorama che si arricchisca dei contributi apportati dall'IA, ma che al contempo presenti concretamente e non solo

[71] Sul punto cfr. O. Pollicino, *Audizione sulla legge sull'Intelligenza artificiale tenutasi alla Camera dei deputati presso le Commissioni riunite Trasporti e Attività produttive*, 8 febbraio 2022. In tale occasione, il Prof. Pollicino ha sottolineato la delicatezza degli interventi statali in materia di IA alla luce soprattutto del considerando 68) nel quale si stabilisce che «per motivi eccezionali di pubblica sicurezza o di tutela della vita e della salute delle persone fisiche nonché della proprietà industriale e commerciale, gli Stati membri possano autorizzare l'immissione sul mercato o la messa in servizio di sistemi di IA che non sono stati sottoposti a una valutazione della conformità». In tale prospettiva, il rischio paventato consiste non solo nella difficoltà di garantire una piena *compliance* con il futuro regolamento, ma soprattutto una possibile frammentazione applicativa a livello nazionale che metterebbe a rischio la tenuta dell'intero quadro normativo in materia di IA a causa proprio del moltiplicarsi nel panorama europeo di eccezioni nazionali giustificate dal considerando. L'audizione è visionabile sul sito della Camera dei Deputati al link evento | WebTV (camera.it).

sulla carta una impostazione *human-centric* allontanando il pericolo che regole *consumer oriented* declassino la tutela dell'individuo da soggetto libero di decidere e autodeterminarsi a comune destinatario di un prodotto o di un servizio.
Solo in questo modo sarà possibile porre un primo argine in senso democratico ad una trasformazione straordinariamente importante in un'ottica evolutiva, ma che corre impetuosa e che, in assenza di adeguati paletti normativi, rischia di trasformare le attuali società tecnologicamente avanzate in contesti dalle sfumature sempre più deterministiche, dove il mito della certezza e dell'efficienza è destinato a prevalere inesorabilmente sull'autonomia e sul libero determinarsi degli individui.