PROTEZIONE DEI DATI PERSONALI E USO DEGLI ALGORITMI. INDAGINE A PARTIRE DALLA PRASSI: LA DIMENSIONE EUROPEA*

COSTANZA MASCIOTTA**

Sommario

Abstract

The essay hereto aims at outlining useful guidelines for future regulation of algorithms and artificial intelligence, starting from the analysis of some reference acts from independent authorities guaranteeing the protection of personal data in the European sphere, furthermore looking at the contribution they have provided to the regulatory framework and assessing whether the indications provided by them on the subject have been incorporated in the regulations proposed by the European political institutions.

Some acts of the European Data Protection Supervisor (EDPS) and the Commission Nationale de l'Informatique et des Libertés (CNIL) will be analyzed, such as EDPS opinion n. 4/2020 on the white paper of the European Commission concerning AI, the EDPS opinion on the proposal for the Digital Services Act regulation, the CNIL report on ethical issues raised by AI and some decisions of the French National Commission with reference to algorithmic decisions and the use of facial recognition devices: these are acts which have played a key role in the creation of a new "European algorithmic Constitutionalism".

Suggerimento di citazione

C. MASCIOTTA, *Protezione dei dati personali e uso degli algoritmi. Indagine a partire dalla prassi: la dimensione europea*, in *Osservatorio sulle fonti*, n. 2/2021. Disponibile in: http://www.osservatoriosullefonti.it

* Il contributo costituisce la rielaborazione della relazione tenuta al *webinar* "Autorità amministrative indipendenti e regolazione delle decisioni algoritmiche" svoltosi il 7 maggio 2021 e organizzato dal Dipartimento di Scienze Giuridiche dell'Università di Firenze, nell'ambito del Progetto PRIN 2017 *Self- and Co-regulation for Emerging Technologies: Towards a Technological Rule of Law* (SE.CO.R.E TECH).

** Assegnista di ricerca in Diritto costituzionale presso l'Università degli Studi di Firenze.
Contatto: costanza.masciotta@unifi.it

Registrazione presso il Tribunale di Firenze n. 5626 del 24 dicembre 2007
Rivista inclusa nella classe A delle Riviste scientifiche di Area 12 – Direttore Prof. Paolo Caretti
ISSN 2038-5633

1. Introduzione

La società in cui viviamo è attualmente dipendente dai dati e dalle tecnologie digitali, come ha dimostrato anche la contingente situazione pandemica: non possiamo rinunciare all'intermediazione delle *ICT*, quindi, il tema della regolazione della società dei dati e delle tecnologie su di essi fondate si impone in tutta la sua cogenza.

In questa dimensione il piano della tutela dei diritti fondamentali e quello delle istanze economiche di mercato si intersecano reciprocamente[1]: nel contesto eurounitario emerge con forza l'esigenza di un bilanciamento tra tali esigenze in modo da realizzare una correzione del mercato in senso conforme ai diritti fondamentali dell'individuo. Si è, così, assistito all'affermazione della protezione dei dati personali come diritto della persona *ex* art. 8 Carta di Nizza, art. 16 TFUE e successivamente nel regolamento GDPR. In tale contesto i dati rappresentano il motore dei sistemi algoritmici e dell'intelligenza artificiale i quali attraverso la loro elaborazione riescono a raggiungere determinate previsioni o addirittura decisioni. Dati, algoritmi e IA costituiscono i lineamenti di una "rivoluzione digitale" dinanzi alla quale occorre un approccio tale da garantire sia la massima protezione della persona che la circolazione libera e sicura dei dati[2]. L'erompere di questa rivoluzione digitale rovescia paradigmi consolidati e richiede talora all'ordinamento di usare strumenti "vecchi" per finalità nuove[3]: le decisioni algoritmiche, quindi, costituiscono un fenomeno che da un lato richiede di rivedere categorie giuridiche tradizionali, come lo stesso concetto di fonte del diritto, ma dall'altro non deve indurre a ritenere che i fondamenti ordinamentali consolidati siano del tutto superati[4]. Si tratta, all'evidenza, di una realtà ancora fluida che talvolta il diritto stenta a regolare per la velocità delle trasformazioni conseguenti all'innovazione tecnologica e al progresso scientifico. La domanda di fondo è fino a che punto l'impiego dell'IA può rappresentare un supporto, una implementazione per la sfera delle libertà della persona e quando, invece, può divenire un rischio, una limitazione di tale sfera[5]. Solo il costituzionalismo, chiamato a porre limiti al potere a garanzia della persona umana, può tentare di dare una risposta a tale quesito di fondo e l'erompere di questa rivoluzione digitale sta aprendo la via ad una nuova stagione del costituzionalismo[6], fondata su presupposti nuovi che mira a rafforzare gli strumenti giuridici tradizionali, prendendo atto delle trasformazioni che la società algoritmica determina sulla sfera fisica, psichica e

[1] Come ben evidenzia C. COLAPIETRO, in questo fascicolo.
[2] Cfr. C. COLAPIETRO, *ult. cit*.
[3] Sul punto cfr. anche G. TARLI BARBIERI ed E. CREMONA, in questo fascicolo.
[4] Cfr. G. TARLI BARBIERI, *ult. cit*.
[5] Profili evidenziati anche da E. CHELI, in questo fascicolo.
[6] Cfr. E. CHELI, *ult. cit*.

relazionale della persona. Compito del diritto è guidare le trasformazioni digitali per arricchire le libertà tradizionali anch'esse in trasformazione, riequilibrando il sistema dei poteri e delle libertà nella cornice di questa rivoluzione digitale.

Il quadro giuridico di riferimento in ordine all'impiego degli algoritmi e dell'intelligenza artificiale è piuttosto complesso ed è costituito da una serie di atti vigenti (primo fra tutti il GDPR) e *in fieri* quali le proposte di regolamento adottate dalla Commissione europea in riferimento all'intelligenza artificiale, al *Governance Act*, al *Digital Services Act* e al *Digital Market Act*. A dare attuazione ai principi normativi contribuiscono sia le autorità indipendenti nazionali ed europee sia le autorità giudiziarie che insieme concorrono a chiarire il quadro normativo europeo di tutela dinanzi all'uso di tali tecnologie.

In questo contesto accanto alla funzione fondamentale affidata alle autorità politiche emerge l'importante ruolo svolto, appunto, dalle Autorità indipendenti (AAI), organi dotati di ampie funzioni di vigilanza in settori "tecnologicamente sensibili" come la *privacy*, caratterizzati dall'esercizio di numerosi diritti fondamentali[7]. Le AAI non godono di una legittimazione democratica diretta e non rientrano nel circuito politico tradizionale e l'analisi del loro ruolo di regolazione è particolarmente utile in quanto assumono decisioni che presentano una elevata complessità tecnica, esercitano un potere qualitativamente diverso dai poteri tradizionali dello Stato e hanno a disposizione uno strumentario rapido, flessibile e tecnico per la tutela dei diritti fondamentali della persona[8]. Inoltre, tra i vantaggi di affidare la regolazione di settori tecnologicamente sensibili alle AAI vi è quello di poter favorire meccanismi più efficienti di auto- e co-regolazione[9] e per questo tali autorità offrono un punto di vista qualificato per studiare i processi di regolazione inerenti alle nuove tecnologie.

Tali autorità da anni svolgono un'importante funzione di difesa delle libertà della persona sia a livello europeo che nazionale. L'attribuzione di potere a tali autorità è finalizzata alle istanze che emergono dal settore da regolare, pertanto,

[7] Sul ruolo delle Autorità amministrative indipendenti specialmente in tema di contrasto alla disinformazione in rete cfr. G. PITRUZZELLA, *La libertà di informazione nell'era di internet*, in *MediaLaws*, n. 1/2018, 19 ss.

[8] Come rilevato anche da E. CHELI, *ult. cit*.

[9] Su quale sia il "dosaggio" di autonormazione, *soft law*, *hard law* e *co-regulation* migliore per intervenire si possono vedere le considerazioni di D. DE GRAZIA, *Il governo di Internet*, FrancoAngeli, Milano, 2010, 272 ss.; M. BETZU, *Regolare Internet. Le libertà di informazione e di comunicazione nell'era digitale*, Giappichelli, Torino, 2012, 19 ss.; T.E. FROSINI, *Internet come ordinamento giuridico*, in M. NISTICÒ, P. PASSAGLIA (a cura di), *Internet e Costituzione*, Giappichelli, Torino, 2014, 59 ss.; P. COSTANZO, *Osservazioni sparse su nodi, legami e regole su Internet*, in P. PASSAGLIA, D. POLETTI (a cura di), *Nodi virtuali, legami informali: Internet alla ricerca di regole*, Pisa University Press, Pisa, 2017, 17 ss. e, più di recente, G. DE MINICO, *Libertà in Rete. Libertà dalla Rete*, Giappichelli, Torino, 2020, 263 ss. e G. MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *BioLaw Journal – Rivista di BioDiritto*, n. 2/2020, 401 ss.

le regole sono ricavate dall'oggetto della regolazione e la regolazione dello specifico segmento di mercato è spesso partecipata dagli operatori economici, con conseguente osmosi tra settore pubblico e privato. Si assiste, inoltre, ad una regolazione integrata che chiama in causa molteplici livelli di governo, quello unionale e al suo interno quello delle autorità indipendenti europee e quello statale, in un intreccio di competenze[10].

Le autorità indipendenti garanti della protezione dei dati personali nella dimensione europea hanno un duplice ruolo fondamentale: forniscono un contributo significativo per la corretta interpretazione e attuazione degli atti normativi esistenti e danno indicazioni utili delle quali potrà tenere conto il legislatore europeo nel disciplinare gli algoritmi e l'intelligenza artificiale.

La presente attività di indagine mira a delineare degli orientamenti e degli indirizzi utili ai fini della futura regolazione di tali tecnologie, a partire dall'analisi degli atti di riferimento di alcune autorità indipendenti nella dimensione europea, guardando, altresì, al contributo da esse dato al quadro normativo di riferimento e valutando se le indicazioni da esse fornite in materia siano state recepite nelle proposte di regolamento adottate dagli organi politici europei.

L'analisi è condotta sulla base degli atti reperibili nei siti internet delle autorità indipendenti preposte alla garanzia dei dati personali e di alcune piattaforme quali la *Digital Clearing House*, una piattaforma *online* che facilita la cooperazione, il dialogo e lo scambio di migliori pratiche tra autorità di regolamentazione, responsabili politici, ricercatori e altri *stakeholders*.

Dalla indagine sono emersi alcuni atti di riferimento per la regolazione delle decisioni algoritmiche[11] e dell'intelligenza artificiale[12]: tali atti si distinguono

[10] Profili evidenziati anche da G. TARLI BARBIERI, *ult. cit*.

[11] Per un approfondimento generale sull'uso di algoritmi e il loro impatto cfr. S.C. OLHEDE, P. J. WOLFE, *The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations* (2018) 376 *Philos. TR Soc.* A 1; S. VALENTINE, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control* (2019) 46 *Fordham Urb. Law J.* 364, 370-93; K. CRAWFORD, J. SCHULTZ, *AI Systems as State Actors* (2019) 119 *Columbia Law Rev.* 7, 1941-72; G. SARTOR, *Human rights and information technologies*, in R. BROWNSWORD, E. SCOTFORD, K. YEUNG (a cura di), *The Oxford Handbook on the Law and Regulation of Technology*, Oxford University Press, Oxford, 2016; M. HILDEBRANDT, *Algorithmic regulation and the rule of law* (2018) 376 *Philosophical transactions of the Royal Society A,* 2128 ss.; K. YEUNG, *Algorithmic regulation: a critical interrogation* (2018) 12 *Regulation & Governance,* 505-23.

[12] L'IA può essere definita come «La scienza del fare fare alle macchine cose che richiederebbero intelligenza se fatte da uomini» (Marvin Minsky). Per una definizione di IA in dottrina cfr. B. MARR, *The Key Definitions of Artificial Intelligence (AI) that Explain Its Importance*, Forbes (February 14, 2018), www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain- its-importance/#15081bd34f5d; F. ZUIDERVEEN BORGESIUS, *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*, Study for the Council of Europe (2018), https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision- making/1680925d73; S. WACHTER, B. MITTELSTADT, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI* (2019) 2 *Colum. Bus. Law Rev.* 494; A. GARAPON, J. LASSEGUE, *Justice digitale*, Presses universitaires de France, Paris, 2018. Cfr. sul tema anche A.

per la rilevanza dei principi in essi espressi, oltrechè per l'autorevolezza della autorità di provenienza e per i profili problematici da essi desumibili.

In particolare risultano di fondamentale rilievo alcuni atti dell'*European Data Protection Supervisor* (EDPS) e della *Commission nationale de l'Informatique et des Libertés* (CNIL): l'EDPS, come noto, è l'autorità indipendente per la protezione dei dati dell'Unione europea, chiamata a monitorare e garantire la protezione dei dati personali e della *privacy* quando le istituzioni e gli organi dell'UE trattano le informazioni personali e fornisce pareri alle istituzioni e agli organi dell'UE su tutte le questioni relative al trattamento dei dati personali, rappresenta, pertanto, un importante "anello di trasmissione" del diritto unionale nella dimensione nazionale[13].

La CNIL, *Commission Nationale de l'Informatique et des Libertés*, è, invece, l'autorità garante francese in materia di protezione dei dati personali.

Proprio in riferimento ad IA e processo decisionale automatizzato sono particolarmente interessanti ai fini della indagine le linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione, adottate dal gruppo di lavoro Articolo 29 per la protezione dei dati[14], l'*opinion* n. 4/2020 dell'EDPS sul libro bianco della Commissione europea concernente l'IA, l'*opinion* dell'EDPS sulla proposta di regolamento *Digital Services Act*, il report della CNIL su questioni etiche sollevate dall'IA e alcune decisioni della Commissione nazionale francese su casi problematici che si sono verificati in riferimento a decisioni algoritmiche e all'utilizzo di dispositivi di riconoscimento facciale.

2. Le linee guida sul processo decisionale automatizzato: un'importante specificazione dei diritti degli interessati

Il tema del trattamento automatizzato dei dati ha origine risalente: nell'ambito del Consiglio d'Europa, come noto, è stata adottata la Convenzione di

SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, n. 1/2019, 63 ss.; ID., *Diritto costituzionale e decisioni algoritmiche*, in S. DORIGO (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pacini, Pisa, 2020, 61 ss.; A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista internazionale di filosofia del diritto*, n. 1/2019; G. DE MINICO, *Antiche libertà e nuova frontiera digitale,* Giappichelli, Torino, 2016; con particolare riguardo alle disfunzioni indotte dall'IA, compresi i rischi della profilazione e delle decisioni automatiche cfr. F. LAGIOIA, G. SARTOR, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *Federalismi.it*, n. 11/2020. Sul ripensamento delle categorie umane cui l'IA obbliga il diritto cfr. C. CASONATO, *Potenzialità e sfide dell'intelligenza artificiale*, in *BioLaw Journal*, n. 1/2019. Sull'impatto rispetto al concetto di sovranità cfr. A. SIMONCINI, *Sovranità e potere nell'era digitale*, in T.E FROSINI, O. POLLICINO, E. APA, M. BASSINI (a cura di), *Diritti e libertà in internet*, Le Monnier, Firenze, 2017, 19 ss.

[13] Cfr. sul punto anche G. TARLI BARBIERI, *ult. cit*.

[14] Article 29 Data Protect Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*.

Strasburgo, considerata una delle prime fonti in materia di protezione dei dati. I principi in essa sanciti fanno parte del patrimonio europeo in materia di dati personali, essendo stati prima incorporati nella cd. "direttiva madre" e poi confermati nel GDPR[15]. Già l'art. 15 della direttiva 95/46 prevedeva il diritto di non essere sottoposto ad una decisione che produca effetti giuridici o abbia effetti significativi fondata esclusivamente su un trattamento automatizzato di dati[16]. La tutela è stata poi rafforzata grazie all'art. 22 GDPR e il quadro è divenuto ancor più nitido grazie alle Linee guida del gruppo di lavoro articolo 29 sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679. Il gruppo di lavoro è stato istituito in virtù dell'articolo 29 della direttiva 95/46/CE ed è un organo consultivo dell'UE per la protezione dei dati personali. Tali linee guida sono particolarmente utili in quanto mirano a chiarire le disposizioni del reg. 2016/679 introdotte per far fronte ai rischi derivanti dalla profilazione[17] e dal processo decisionale automatizzato[18]. Le linee guide costituiscono un punto di riferimento fondamentale in materia, perché forniscono innanzitutto una previa definizione di profilazione e processo decisionale automatizzato, cercando di rispondere a un profilo problematico che emerge da molti atti in materia, ovvero la previsione di definizioni chiare e univoche per la regolazione e, in secondo luogo, contengono delle raccomandazioni concrete che possono essere utili ai fini della regolazione delle decisioni algoritmiche.

Dal punto di vista definitorio la profilazione è intesa come qualsiasi forma di trattamento automatizzato di dati personali, consistente nell'utilizzo di tali dati per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi,

[15] Come evidenziato da C. COLAPIETRO, *ult. cit*.

[16] Sul punto cfr. anche C. COLAPIETRO, *ult. cit*

[17] Sulla profilazione cfr. F. BOSCO, N. CREEMERS, V. FERRARIS, D. GUAGNIN, B.J. KOOPS, *Profiling technologies and fundamental rights and values: regulatory challenges and perspectives from European Data Protection Authorities*, in S. GUTWIRTH, R. LEENES, P. DE HERT (a cura di), *Reforming European data protection law*, Springer Science, Dordrecht, 2015. Si veda anche M. HILDEBRANDT, *Profiling and AML*, in K. RANNENBERG, D. ROYER, A. DEUKER (a cura di), *The Future of Identity in the Information Society. Challenges and Opportunities*, Springer, Dordrecht, 2009.

[18] Sul tema, in generale, cfr. O. SESSO SARTI, *Profilazione e trattamento dei dati personali*, in L. CALIFANO, C. COLAPIETRO (a cura di), *Innovazione tecnologica e valore della persona: il diritto alla protezione dei dati personali nel Regolamento UE 2016/670*, Editoriale scientifica, Napoli, 2017, 773 ss.; G.M. RICCIO, G. SCORZA, E. BELISARIO, *GDPR e normativa privacy: commentario al regolamento (UE) 2016/679 del 27 aprile 2016, decreto di armonizzazione- codice privacy – D. Lgs. n. 196/2003*, Wolters Kluwer, 2018, 219 ss. Nel dibattito internazionale cfr. E. PEHRSSON, *The meaning of the GDPR Article 22*, European Union Law Working Papers n. 31, Stanford-Vienna Transatlantic technology law forum, 2018; S. WACHTER, *Normative challenges of identification in the internet of things: privacy, profiling, discrimination, and the GDPR* (2018) 34 *Computer Law & Security Review* 3, 436-449.

l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona. Le linee guida spiegano che la profilazione consiste, quindi, nella raccolta di informazioni su una persona (o un gruppo di persone) e nella valutazione delle loro caratteristiche o dei loro modelli di comportamento al fine di includerli in una determinata categoria o gruppo, in particolare per analizzare e/o fare previsioni. Il processo decisionale automatizzato ha, invece, una portata diversa da quella della profilazione, a cui può sovrapporsi parzialmente o da cui può derivare. Sempre dal punto di vista definitorio il processo decisionale esclusivamente automatizzato consiste nella capacità di prendere decisioni impiegando mezzi tecnologici senza coinvolgimento umano. Le decisioni automatizzate possono essere prese ricorrendo o meno alla profilazione, la quale a sua volta può essere svolta senza che vengano prese decisioni automatizzate.

Si possono, quindi, distinguere: i) profilazione generale; ii) processo decisionale (umano) basato sulla profilazione; iii) decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produce effetti giuridici o incide in modo analogo significativamente sull'interessato (articolo 22, paragrafo 1 GDPR).

Come spiegano le linee guida alcune disposizioni del GDPR si applicano a tutte le profilazioni e a tutti i processi decisionali automatizzati: ad esempio l'articolo 5, paragrafo 1, lettera a) GDPR - liceità, correttezza e trasparenza del trattamento dei dati; l'articolo 5, paragrafo 1, lettera b) - limitazione della finalità; l'articolo 5, paragrafo 1, lettera c) - la minimizzazione dei dati; l'articolo 5, paragrafo 1, lettera d) - esattezza dei dati in tutte le fasi del processo di profilazione; l'articolo 5, paragrafo 1, lettera e) - limitazione della conservazione.

Anche il Garante *Privacy* italiano ha affermato la necessità di garantire i principi europei di esattezza, integrità, aggiornamento del dato, minimizzazione, proporzionalità, liceità, proprio nello spirito più autentico del GDPR.

Per quanto attiene ai diritti degli interessati le linee guida precisano che essi possono essere esercitati nei confronti del titolare del trattamento che crea il profilo e del titolare del trattamento che prende una decisione automatizzata su un interessato con o senza intervento umano, qualora tali soggetti non coincidano.

Tra tali diritti assume particolare rilievo il diritto di accesso: l'articolo 15 GDPR conferisce all'interessato il diritto di ottenere informazioni dettagliate sui dati personali utilizzati per la profilazione, ivi comprese le categorie di dati impiegati per creare un profilo. In particolare, il titolare del trattamento deve rendere disponibili i dati utilizzati come *input* per creare il profilo e consentire l'accesso alle informazioni sul profilo.

Occorre considerare che i dati di *input* potrebbero essere inesatti o irrilevanti oppure avulsi dal contesto, o l'algoritmo utilizzato per individuare le correlazioni potrebbe presentare lacune e in tali casi opera il diritto di rettifica di

cui all'articolo 16 GDPR. I diritti di rettifica e di cancellazione (art. 17 GDPR) si applicano tanto ai "dati personali di input" (i dati personali utilizzati per creare il profilo) quanto ai "dati di output" (il profilo stesso o il "punteggio" assegnato alla persona fisica).

In tale prospettiva assumono particolare rilievo ai fini della indagine le raccomandazioni finali contenute nelle linee guida a proposito del diritto di rettifica, secondo le quali il titolare del trattamento dovrebbe consentire agli interessati di aggiornare o modificare eventuali inesattezze presenti nei dati o nel profilo. Le raccomandazioni suggeriscono una modalità operativa concreta per dare attuazione al diritto di rettifica: il titolare del trattamento potrebbe prendere in considerazione la possibilità di introdurre strumenti di gestione delle preferenze *online*, ad esempio un *dashboard* per la protezione dei dati, permettendo così agli interessati di gestire l'uso delle loro informazioni, di modificare, aggiornare i loro dettagli personali e rivedere o modificare il loro profilo per correggere eventuali inesattezze.

Le linee guida sono, inoltre, particolarmente interessanti là dove si soffermano sulla spiegazione delle disposizioni relative a decisioni basate unicamente sul trattamento automatizzato di cui all'articolo 22 GDPR: secondo tale disposizione l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona. Secondo le linee guida il termine "diritto" contenuto nella disposizione non significa che l'articolo 22, paragrafo 1, si applica soltanto se invocato attivamente dall'interessato: l'articolo 22 stabilisce un divieto generale nei confronti del processo decisionale basato unicamente sul trattamento automatizzato e tale divieto si applica indipendentemente dal fatto che l'interessato intraprenda un'azione in merito al trattamento dei propri dati personali. In sintesi, l'articolo 22 GDPR stabilisce che: i) di norma, esiste un divieto generale all'adozione di decisioni completamente automatizzate relative alle persone fisiche, compresa la profilazione, che hanno un effetto giuridico o che incidono in modo analogo significativamente; ii) esistono eccezioni alla regola; iii) laddove si applichi una di tali eccezioni, devono essere adottate misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato.

L'articolo 22, paragrafo 1 GDPR si riferisce a decisioni "basate unicamente" sul trattamento automatizzato: ciò significa che non vi è alcun coinvolgimento umano nel processo decisionale. Se un essere umano riesamina il risultato del processo automatizzato e tiene conto di altri fattori nel prendere la decisione finale, tale decisione non sarà "basata unicamente" sul trattamento automatizzato.

Il regolamento non definisce, invece, i concetti di "effetto giuridico" o "effetti in modo analogo significativi", tuttavia secondo le linee guida un "effetto giuridico" richiede che la decisione basata unicamente su un trattamento automatico incida sui diritti giuridici di una persona, quali la libertà di associarsi, di votare nel contesto di un'elezione o di intraprendere azioni legali. Un effetto giuridico può, altresì, essere qualcosa che influisce sullo *status* giuridico di una persona o sui suoi diritti ai sensi di un contratto (ad es. le decisioni automatizzate su una persona fisica che portano alla cancellazione di un contratto o alla concessione o alla negazione del diritto a una particolare prestazione sociale).

L'art. 22, parag. 1 GDPR prevede, quindi, un divieto generale a meno che non si applichi una delle eccezioni di cui all'articolo 22, paragrafo 2, nel contesto delle quali la decisione: a) è necessaria per la conclusione o l'esecuzione di un contratto; b) è autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento; o c) si basa sul consenso esplicito dell'interessato. La portata applicativa di tali eccezioni è potenzialmente molto ampia e in una futura prospettiva regolatoria potrebbe essere necessaria una loro definizione in termini restrittivi per garantire maggior effettività al divieto generale.

La parte più significativa delle linee guida concerne i diritti dell'interessato rispetto al trattamento di cui all'art. 22 GDPR, in particolare il diritto di essere informato: l'articolo 13, paragrafo 2, lettera f), e l'articolo 14, paragrafo 2, lettera g), impongono al titolare del trattamento di fornire informazioni specifiche e facilmente accessibili sul processo decisionale automatizzato basato esclusivamente sul trattamento automatizzato, compresa la profilazione. Il titolare del trattamento deve comunicare all'interessato l'esistenza di processi decisionali automatizzati, fornire informazioni significative sulla logica utilizzata, spiegare l'importanza e le conseguenze previste di tale trattamento.

Un profilo di particolare interesse ai fini della indagine consta proprio nella precisa presa di posizione del gruppo di lavoro articolo 29 in ordine alla logica utilizzata: il titolare del trattamento deve fornire informazioni significative sulla logica utilizzata, ma non necessariamente una spiegazione matematica complessa degli algoritmi utilizzati o la divulgazione dell'algoritmo completo. Come vedremo, anche la CNIL assume una posizione analoga in ordine alle informazioni sulla logica utilizzata, ritenendo non necessario l'accesso diretto al codice sorgente.

Molto significative sono anche le raccomandazioni finali contenute nelle linee guida in ordine al diritto ad essere informato: il titolare del trattamento dovrebbe fornire informazioni all'interessato, su, ad esempio, categorie di dati che sono state o saranno utilizzate nella profilazione o nel processo decisionale, motivi per i quali tali categorie sono considerate pertinenti, modalità di creazione del profilo utilizzato nel processo decisionale automatizzato, motivi per

i quali tale profilo è pertinente per il processo decisionale automatizzato e modalità di utilizzo del profilo ai fini di una decisione riguardante l'interessato.

Inoltre, nei casi in cui una decisione è necessaria per la conclusione o l'esecuzione di un contratto o basata sul consenso esplicito, il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione.

Come spiegano le linee guida, l'intervento umano è un aspetto fondamentale: qualsiasi riesame dovrebbe essere effettuato da una persona che dispone dell'autorità e della competenza adeguate per modificare la decisione e il responsabile di tale riesame dovrebbe effettuare una valutazione approfondita di tutti i dati pertinenti, comprese eventuali informazioni aggiuntive fornite dall'interessato. Il considerando 71 GDPR precisa che in ogni caso le garanzie adeguate dovrebbero comprendere anche la specifica informazione all'interessato e il diritto di ottenere l'intervento umano, di esprimere la propria opinione, il diritto di *ottenere una spiegazione della decisione* conseguita dopo tale valutazione[19] e di contestare la decisione. Da ciò emerge la necessità di trasparenza del trattamento: l'interessato sarà in grado di contestare una decisione o esprimere il proprio parere soltanto se comprende pienamente come è stata presa la decisione e su quali basi.

Anche il Garante *privacy* italiano ha da anni avuto ben chiaro il rischio rappresentato dalla esposizione delle persone fisiche a processi di decisione automatizzata basati sulla lettura algoritmica anche di metadati[20]. In particolare, il Garante italiano ha riconosciuto che il trattamento automatizzato deve essere sindacabile, non può rappresentare la modalità esclusiva di trattamento dei dati personali dell'interessato, specialmente quando le decisioni della macchina sono rilevanti per la vita del soggetto e incidono sulla sua sfera privata. Per questo il titolare che preveda di adottare un sistema di trattamento basato sull'IA dovrà assicurare la possibilità di ricorrere contro la decisione, interagendo con un operatore umano, garantendo l'effettivo esercizio dei diritti *ex* artt. 15-23 GDPR.

Occorre osservare che un profilo problematico delle linee guida consta nel fatto che il diritto di ottenere una spiegazione della decisione rimane indefinito

[19] Sul punto cfr. M. E. KAMINSKI, *The right to explanation, Explained* (2019) 34 *Berkeley Tech. L. J.* 189; B. CASEY, A. FARHANGI, R. VOGL, *Rethinking explainable machines: the GDPR's' "right to explanation". Debate and the rise of algorithmic audits in enterprise* (2019) 34 *Berkeley Tech. L. J.*, 143.

[20] Ci sono state decisioni interessanti del Garante *Privacy* italiano ad esempio sul redditometro, in particolare concernenti algoritmi di IA che consistevano in trattamenti automatizzati di dati personali presenti nell'anagrafe tributaria al fine di selezionare i contribuenti da sottoporre ad accertamento e rideterminare il reddito.

e la sua specificazione e cogenza resta una delle grandi sfide sollevate dalla società algoritmica.

Con particolare riguardo alla comprensibilità delle decisioni algoritmiche e alla spiegabilità dei processi decisionali che avvengono nella "*black box*" è in corso un ampio dibattito riguardo all'esistenza o meno di un vero e proprio "*right to an explanation*". Tale concetto può assumere diversi connotati a seconda dell'oggetto cui si riferisce e del momento temporale in cui viene fornita la spiegazione. In riferimento all'oggetto si può distinguere la *explanation* riferibile al funzionamento del sistema, quindi, al meccanismo decisionale nel suo complesso, e la spiegazione relativa a decisioni specifiche, quindi, alle ragioni che hanno condotto alla singola decisione. In base al criterio temporale, invece, la distinzione riguarda l'*explanation ex ante* ed *ex post*: mentre la prima può essere riferita solo al funzionamento complessivo del sistema, la seconda può riguardare sia l'intera struttura decisionale sia le singole e specifiche decisioni[21].

In un'ottica *de iure condendo* è proprio quest'ultima nozione che dovrebbe integrare un *right to an explanation,* poiché la comprensibilità e spiegabilità della decisione sono elementi necessari per la tutela del diritto alla protezione dei dati personali. L'effettiva garanzia della dignità della persona nella odierna società digitale non può inverarsi senza una corretta informazione all'interessato sul trattamento dei suoi dati e vengono, così, in rilievo gli obblighi informativi derivanti dagli artt. 13 e 14 GDPR che, insieme all'art. 15 GDPR relativo al diritto di accedere ai propri dati, sono stati considerati la "Magna Charta" del diritto dell'interessato di essere informato e di controllare la legittimità del trattamento[22].

Nell'ottica della costruzione di un nuovo costituzionalismo della società algoritmica il diritto alla comprensione delle ragioni di una determinata decisione, adottata con il supporto di/o da un sistema automatizzato, rappresenta uno dei "nuovi" e fondamentali diritti da tutelare, soprattutto quando sono in gioco decisioni adottate da pubblici poteri.

Un ulteriore profilo problematico delle linee guida riguarda i controlli sui sistemi algoritmici, poiché ivi si afferma che i sistemi che verificano gli algoritmi e i riesami periodici dell'esattezza e della pertinenza del processo decisionale automatizzato, compresa la profilazione, sono ulteriori misure utili. In realtà, là dove è previsto l'uso di algoritmi, i sistemi di verifica e controllo periodico del loro funzionamento sono misure non solo utili, ma necessarie e indefettibili anche per prevenire errori, inesattezze o discriminazioni, non soltanto in fase di progettazione, ma continuativamente durante l'applicazione e l'esito di tali

[21] Sul punto cfr. C. COLAPIETRO, in questo fascicolo, § 2.
[22] Cfr. C. COLAPIETRO, *ult. cit*.

verifiche dovrebbe andare ad alimentare nuovamente la progettazione del sistema.

Un profilo problematico del GDPR che si ripercuote inevitabilmente anche sulle linee guida riguarda, inoltre, la tutela dei minori: l'articolo 22 GDPR di per sé non opera alcuna distinzione in merito al fatto che il trattamento riguardi adulti o minori. Tuttavia, il considerando 71 afferma che le decisioni basate unicamente sul trattamento automatizzato, compresa la profilazione, che producono effetti giuridici o in modo analogo significativi non dovrebbero riguardare minori. Dato che tale formulazione non si riflette, però, nell'articolato del regolamento, il gruppo di lavoro non ritiene che ciò rappresenti un divieto assoluto di questo tipo di trattamento in relazione ai minori.

In una prospettiva *de iure condendo* sarebbe necessario prevedere la portata delle eccezioni al principio generale in termini restrittivi, onde garantire l'effettività del principio medesimo, con particolare riguardo a gruppi vulnerabili come i minori.

Infine, in un'ottica di *accountability* le linee guida evidenziano l'importanza della valutazione di impatto sulla protezione dei dati, in quanto strumento essenziale per la responsabilizzazione: essa consente al titolare del trattamento di valutare i rischi connessi al processo decisionale automatizzato, compresa la profilazione e permette di evidenziare che sono state messe in atto misure adeguate per affrontare tali rischi nel rispetto del regolamento. Come vedremo l'importanza della valutazione d'impatto è sottolineata anche dall'EDPS nella *opinion* sul Libro bianco della Commissione europea sull'IA ed è in linea anche con l'orientamento del Garante *Privacy* italiano, rappresentando un presupposto indefettibile per ogni futura regolazione delle decisioni algoritmiche. Come evidenzia l'articolo 35, paragrafo 3, lettera a) GDPR, il titolare del trattamento deve effettuare una valutazione d'impatto sulla protezione dei dati in caso di una valutazione sistematica e globale di aspetti personali relativi a persone fisiche, basata su un trattamento automatizzato, compresa la profilazione, e sulla quale si fondano decisioni che hanno effetti giuridici o incidono in modo analogo significativamente su dette persone fisiche. Pertanto, tale disposizione si applica sia a un processo decisionale che abbia effetti giuridici o in modo analogo significativi che non è un processo decisionale interamente automatizzato, sia a una decisione basata unicamente sul trattamento automatizzato di cui all'articolo 22, paragrafo 1 GDPR.

Alla luce delle linee guida si possono, quindi, individuare dei primi orientamenti utili ai fini della regolazione delle decisioni algoritmiche: dall'importanza di dare definizioni chiare e univoche di termini fondamentali in questa materia (i.e. profilazione, processo decisionale automatizzato), alla modalità per declinare alcuni diritti dell'interessato rispetto al trattamento di cui all'art. 22 GDPR, come il diritto di rettifica o a essere informato anche in

ordine alla logica utilizzata, all'importanza della valutazione di impatto sulla protezione dei dati. Le linee guida fanno, altresì, emergere alcuni profili problematici come la definizione del diritto ad una spiegazione della decisione, la necessità di prevedere sistemi di verifica periodici del funzionamento degli algoritmi utilizzati durante tutto il loro ciclo di vita e la definizione di una tutela specifica dei minori in quanto gruppo vulnerabile.

3. L'*opinion* n. 4/2020 dell'EDPS e le criticità espresse rispetto al Libro bianco della Commissione europea sull'intelligenza artificiale

Nel quadro europeo di regolazione delle tecnologie di intelligenza artificiale assume particolare rilievo l'*opinion* n. 4/2020 dell'EDPS del 29 giugno 2020 sul Libro bianco della Commissione europea sull'intelligenza artificiale "Un approccio europeo all'eccellenza e alla fiducia", pubblicato il 19 febbraio 2020. L'*opinion* presenta il punto di vista dell'EDPS sul Libro bianco nel suo insieme, nonché su alcuni aspetti specifici come l'approccio ivi proposto basato sul rischio e l'implementazione della regolamentazione sull'intelligenza artificiale. L'obiettivo delle raccomandazioni contenute nell'*opinion* è chiarire e sviluppare ulteriormente le garanzie e i controlli relativi alla protezione dei dati personali, tenendo conto del contesto specifico dell'IA.

Tale *opinion* deve essere letta anche alla luce della recente proposta della Commissione europea di un regolamento per l'intelligenza artificiale, "*Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*", pubblicata lo scorso 21 aprile[23] che rappresenta il proseguo del libro bianco sull'IA e, quindi, anche della *opinion* dell'EDPS. Tale proposta per la prima volta fissa una cornice forte in materia e si basa sulla suddivisione all'interno dell'IA dei trattamenti per fasce di rischio ed è utile valutare se tale approccio regolatorio si ponga o meno in linea di continuità con le raccomandazioni dell'EDPS sul tema.

Dal punto di vista dell'ambito applicativo di una futura regolazione dei sistemi di intelligenza artificiale l'EDPS nella *opinion* raccomanda chiaramente che qualsiasi nuovo quadro normativo si applichi sia agli Stati membri dell'UE che alle istituzioni, agli uffici, agli organi e alle agenzie dell'UE, profilo questo recepito nella proposta di regolamento sull'intelligenza artificiale.

Alcuni profili di criticità rispetto al Libro bianco sono espressi dall'EDPS per quanto attiene al fondamentale problema definitorio: il termine IA deve

[23] Si precisa che nel corso della pubblicazione del presente contributo è stata adottata l'EDPB-EDPS *Joint Opinion* n. 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)*, consultabile su *www.edps.europa.eu.* In dottrina sulla proposta cfr. A. PAJNO, *Introduzione allo studio della proposta della Commissione europea di Regolamento sull'Intelligenza Artificiale*, in *Astrid Rassegna*, rinvenibile all'indirizzo *www.astrid-online.it*, n.12/2021.

«essere chiaramente definito ai fini di qualsiasi possibile futura iniziativa politica» e, tuttavia, il Libro bianco presenta più di una definizione e non ne abbraccia chiaramente nessuna. In tal senso la proposta di regolamento sull'IA sembra rispondere a tale sollecitazione, fornendo una chiara definizione di intelligenza artificiale.

L'EDPS concorda, invece, con l'approccio presentato nel Libro bianco in base al quale per i sistemi di IA operanti nell'UE «è fondamentale che le norme dell'UE siano rispettate da tutti gli attori ... indipendentemente dal fatto che abbiano sede nell'UE o meno», in quanto ciò è coerente con l'approccio del legislatore europeo scelto per la protezione dei dati personali, in particolare con il GDPR.

Un altro profilo sul quale pone particolare attenzione l'EDPS è l'impatto collettivo dei sistemi di IA: le norme sulla protezione dei dati sono progettate per proteggere principalmente le persone e potrebbero non essere adatte per affrontare i rischi per gruppi di individui. L'EDPS, pertanto, raccomanda che qualsiasi disciplina giuridica relativa all'IA sia concepita per proteggere da qualsiasi impatto negativo non solo gli individui, ma anche la collettività e la società nel suo insieme, perché è evidente che le decisioni algoritmiche coinvolgono sia diritti di singoli sia di gruppi. A questo proposito si potrebbe pensare in una prospettiva *de iure condendo* a modelli di *governance* inclusiva che conferiscano potere alle organizzazioni che rappresentano la società civile (ad esempio ONG e associazioni senza scopo di lucro) in modo che esse possano anche aiutare a valutare l'impatto delle applicazioni di IA su collettività specifiche e sulla società.

Il Libro bianco afferma, inoltre, che quando il sistema di IA "impara" mentre è in funzione «il risultato non potrebbe essere previsto o anticipato in fase di progettazione e i rischi non deriveranno da un difetto nella progettazione originale del sistema ma piuttosto dagli impatti pratici delle correlazioni o dei modelli che il sistema identifica in un set di dati di grandi dimensioni». L'EDPS non è d'accordo con questa valutazione: la progettazione delle applicazioni di intelligenza artificiale dovrebbe tenere conto del potenziale pregiudizio/discriminazione[24] nei dati di addestramento e nei dati operativi. Il *bias* può e deve essere misurato e corretto durante il funzionamento delle applicazioni di IA come può essere misurato e corretto durante il loro sviluppo[25].

[24] Sul problema della discriminazione nell'uso degli algoritmi cfr. R. NUNN, *Discrimination in the Age of Algorithms*, in W. BARFIELD (eds.), *The Cambridge Handbook of the Law of Algorithms*, Cambridge University Press, Cambridge, 2020, 182 ss.; S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact* (2016) 104 *Calif. Law Rev.*, 671; T. Z. ZARSKY, *Understanding Discrimination in the Scored Society* (2014) 89 *Wash. Law Rev.*, 1375; R. WILLIAMS, *Rethinking Deference for Algorithmic Decision-Making*, Oxford Legal Studies Research Paper n. 7/2019, 30 ss.

[25] In dottrina sulla integrazione *by design* dei principi costituzionali nel momento della progettazione delle applicazioni di IA cfr., *ex multis*, G. BUCHOLTZ, *Artificial Intelligence and Legal Tech:*

Un ulteriore profilo di criticità rispetto al Libro bianco riguarda l'approccio per la valutazione del livello di rischio dei sistemi di IA: la proposta del Libro bianco è quella di aggiungere alcuni requisiti legali per le applicazioni di IA "ad alto rischio", in conformità con i due criteri cumulativi di 1) settore ad alto rischio e 2) di impatto dell'applicazione di IA. L'EDPS ritiene che l'approccio per determinare il livello di rischio dell'uso delle applicazioni di intelligenza artificiale dovrebbe essere più solido: i criteri per determinare il livello di rischio dovrebbero riflettere le linee guida del Comitato europeo per la protezione dei dati e dovrebbero, quindi, includere attività di valutazione o punteggio, processo decisionale automatizzato con rilevanti effetti giuridici, monitoraggio sistematico, dati sensibili, dati trattati su larga scala, dati relativi a soggetti vulnerabili, trasferimento di dati oltre confine al di fuori dell'Unione Europea. Inoltre, i rischi dovrebbero essere determinati tenendo conto di criteri oggettivi specifici come la natura dei dati personali (es. sensibili o meno), la categoria di interessati (es. minore), il numero di interessati, lo scopo del trattamento, la gravità e la probabilità che si verifichino impatti sui diritti e sulle libertà dell'interessato.

Criticità sono espresse anche in riferimento alla nozione di "rischio di impatto" contenuta nel Libro bianco: oltre all'"impatto sulle parti interessate", l'EDPS ritiene che la valutazione del livello di rischio di un determinato uso dell'IA dovrebbe essere basata anche su considerazioni sociali più ampie, compreso l'impatto sul processo democratico[26], sul giusto processo e sulla *rule of law*. Sul punto la *Proposal for a Regulation on a European approach for Artificial Intelligence* sembra recepire tale sollecitazione là dove prevede al Considerando 40 che «Certain AI systems intended for [...] democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, rule of law, individual freedoms».

Challenges to the Rule of Law, in T. WISCHMEYER, T. RADEMACHER (eds.), *Regulating artificial intelligence*, Springer, Berlin, 2020, 176 ss., part. 191 ss.

[26] In dottrina si interrogano sugli effetti distorsivi che l'applicazione dell'IA alla comunicazione politica è in grado di produrre sulla costruzione del consenso, *ex multis*, P. CIARLO, *Democrazia, partecipazione popolare e populismo al tempo della Rete*, in *Rivista AIC*, n. 2/2018, 10 ss.; G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di eguaglianza*, in *Politica del diritto*, n. 2/2019, 214; F. LAGIOIA, G. SARTOR, *Profilazione e decisione algoritmica*, cit.,101 ss.; nonché A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, cit., 94 che fanno l'esempio degli *backers* russi che, attraverso la creazione di falsi *accounts* Facebook, «hanno interferito con i risultati del referendum sulla Brexit o delle elezioni presidenziali americane». Per un'ampia disamina degli effetti dell'IA sulla decisione politica cfr. A. CARDONE, "*Decisione algoritmica" vs. decisione politica?*, Editoriale scientifica, Napoli, 2021. Sugli effetti del ricorso agli algoritmi rispetto al processo di produzione del diritto politico e sulla analisi della prassi anche di diritto comparato cfr. il contributo dello stesso Autore in questo fascicolo, dal quale emerge che l'uso di algoritmi incide su almeno tre distinti livelli: il livello del procedimento legislativo, delle strategie dei gruppi e dei partiti in Parlamento e della formazione dell'opinione pubblica.

L'approccio fondato sulle applicazioni "ad alto rischio" è, tuttavia, criticato dall'EDPS, poiché non riflette l'impostazione precauzionale adottata dall'Unione Europea nella legislazione sulla protezione dei dati personali: il Libro bianco sembra adottare un approccio "tutto o niente", proponendo che solo le applicazioni di IA ad alto rischio siano soggette a specifici obblighi aggiuntivi.

L'EDPS suggerisce, invece, che se la Commissione proponesse un nuovo quadro normativo specifico per l'IA, un certo numero di garanzie ragionevoli dovrebbe applicarsi a tutti i sistemi di IA, indipendentemente dal livello di rischio, come l'adozione di misure tecniche e organizzative completamente trasparenti sugli obiettivi, l'uso e la progettazione dei sistemi algoritmici, la garanzia della robustezza del sistema di IA, o meccanismi trasparenti di responsabilità, ricorso e controllo indipendente. L'EDPS ritiene che requisiti normativi come l'assenza di discriminazione o la robustezza e l'accuratezza dell'IA siano così fondamentali che dovrebbero essere applicati a qualsiasi sistema di IA, non solo a quelli "ad alto rischio".

Da questo punto di vista, però, la recente proposta di regolamento sull'IA sembra rimanere ancorata al requisito dell'alto rischio, introducendo un elenco di pratiche di intelligenza artificiale vietate (i.e i sistemi che consentono alle autorità pubbliche di attribuire un "punteggio sociale") e individuando gli specifici settori "ad alto rischio" cui si riferiscono gran parte delle disposizioni del regolamento. Secondo la proposta soltanto i sistemi di intelligenza artificiale ad alto rischio devono essere progettati e sviluppati in modo tale da raggiungere, alla luce dello scopo previsto, un livello appropriato di accuratezza, robustezza e sicurezza informatica e funzionare in modo coerente sotto tali aspetti per tutto il loro ciclo di vita.

Più che un Regolamento sull'Intelligenza Artificiale, quello proposto dalla Commissione sembra un Regolamento "sulle intelligenze artificiali ad alto rischio": il testo non si occupa tanto di regolare in maniera trasversale l'uso dell'IA, ma si concentra, invece, su una serie di specifiche tecnologie considerate ad alto rischio. In questo senso, l'impatto del Regolamento, laddove dovesse essere mantenuto il testo proposto dalla Commissione, potrebbe avere effetti intersettoriali più limitati rispetto, per esempio, al GDPR.

Un altro aspetto fondamentale ai fini della futura regolazione dell'IA concerne la valutazione di impatto: la *Data Protection Impact Assessment* prevista dall'articolo 35 del GDPR è uno strumento fondamentale per la gestione dei rischi per i diritti e le libertà delle persone e l'EDPS si rammarica che il Libro bianco non menzioni esplicitamente le DPIA, nonostante il suo impegno a ridurre al minimo i rischi posti dall'intelligenza artificiale. La DPIA richiede una valutazione della necessità e della proporzionalità del trattamento: occorre, quindi, dimostrare che l'utilizzo dell'IA è effettivamente lo strumento più adatto per raggiungere l'obiettivo di una specifica attività di trattamento dei

dati e la valutazione della proporzionalità dovrebbe tenere conto di una serie di fattori, in particolare l'interesse dei responsabili del trattamento dei dati, i diritti e le libertà delle persone, le ragionevoli aspettative degli individui e lo scopo del trattamento dei dati.

L'EDPS suggerisce, pertanto, che un futuro quadro giuridico preveda il requisito di una valutazione d'impatto per qualsiasi utilizzo di sistemi di IA[27]. Tale profilo è stato solo parzialmente recepito nella proposta di regolamento della Commissione che richiede di svolgere una valutazione d'impatto sulla protezione dei dati, ai sensi dell'art. 35 GDPR, solo per i sistemi di intelligenza artificiale ritenuti ad alto rischio, così come prevede solo limitatamente a tali sistemi un *risk management system*, ossia un sistema che comprenda l'identificazione e analisi dei rischi noti e prevedibili associati a ciascun sistema di IA ad alto rischio, la stima e la valutazione dei pericoli che possono emergere quando il sistema è utilizzato e l'adozione di idonee *risk management measures*. Si tratta, pertanto, di misure di garanzia la cui prescrizione è problematicamente limitata ai sistemi ad alto rischio in parte discostandosi da quanto richiesto dall'EDPS.

Uno dei profili problematici che emergono dall'*opinion* consta, altresì, nell'opacità dei sistemi di intelligenza artificiale, in particolare a causa dell'effetto "*black box*"[28]: l'opacità è correlata all'incapacità umana di spiegare il ragionamento alla base della decisione delle applicazioni di IA. L'EDPS suggerisce, pertanto, che procedure trasparenti di verifica e controllo nel contesto delle valutazioni di conformità preliminari dovrebbero far parte di qualsiasi applicazione di IA che tratta dati personali: rendere pubblicamente disponibili tali procedure garantirebbe che le autorità di vigilanza possano svolgere i propri compiti, aumentando anche la fiducia degli utenti nelle applicazioni di IA.

Occorre, però, considerare che l'applicazione del principio di trasparenza all'algoritmo cela in sé una sorta di paradosso: rimuovere ogni opacità sul modo in cui tali sistemi funzionano riporta l'uomo in una posizione di comando, ma le macchine sono progettate da soggetti che intendono sfruttare proprio tale asimmetria informativa, poiché dalla opacità possono trarre un vantaggio economico[29]. Pertanto, per superare l'*impasse* si dovrebbe percorrere una soluzione mediana che permetta di spiegare il processo seguito dalla macchina, la logica del trattamento, chiarendo in modo semplice e diretto agli interessati il perché delle decisioni che li riguardano, dando loro anche la

[27] Sul punto cfr. H. L. JANSSEN, *An approach for a fundamental rights impact assessment to automated decision-making* (2020) 10 *International Data Privacy Law* 1, 76-106, https://doi.org/10.1093/idpl/ipz028.

[28] Circa l'espressione «*black box*» cfr. F. PASQUALE, *The black box society: the secret algorithms that control money and information*, Harvard University Press, Cambridge, 2015.

[29] Come evidenziato anche da C. COLAPIETRO, in questo fascicolo.

possibilità contestarle, senza però svelare i segreti della macchina. Tra le modalità di spiegazione attraverso le quali accrescere la possibilità di interpretare le singole decisioni assunte sulla base di sistemi di *machine learning*, una metodologia potrebbe consistere nella adozione di prove controfattuali al fine di dare una spiegazione del processo senza svelare i segreti della macchina[30]. Data la complessità di alcuni algoritmi caratterizzati da milioni di variabili non è facile spiegare il perché di una decisione né comprendere la logica interna, pertanto, occorre un approccio funzionale: fornire un minimo di informazioni relative alla capacità di orientare la decisione, potendo, così, mostrare come tale decisione sia basata su certi dati, senza richiedere che gli interessati comprendano nello specifico la logica interna del modello. Tale soluzione potrebbe rendere accessibili e comprensibili anche le decisioni della pubblica amministrazione senza svelare i segreti della macchina.

In ordine al profilo della trasparenza la proposta di regolamento sull'IA prevede ancora una volta precisi e stringenti obblighi limitatamente ai sistemi ad alto rischio: all'art. 13 si dice che tali sistemi «devono essere progettati e sviluppati in modo da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo in modo appropriato». Ne deriva l'obbligo di accompagnare a tali sistemi delle istruzioni per l'uso che includano informazioni concise, complete, corrette, chiare, pertinenti, accessibili e comprensibili per gli utenti (ad es. informazioni sulle caratteristiche, sulle capacità e sui limiti delle prestazioni del sistema di IA ad alto rischio, tra cui il livello di accuratezza, robustezza e cibersicurezza rispetto al quale il sistema è stato testato e convalidato). Per quanto attiene agli "altri" sistemi di IA, non ad alto rischio, si prevede soltanto che quelli destinati a interagire con persone fisiche siano progettati e sviluppati in modo tale che le persone siano informate che stanno interagendo con un sistema di IA, a meno che ciò non risulti ovvio dalle circostanze e dal contesto di utilizzo. Tale obbligo va incontro, però, ad una serie di eccezioni dalla portata potenzialmente molto ampia: esso non si applica comunque ai sistemi autorizzati per legge a rilevare, prevenire, indagare e perseguire reati, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato. Inoltre, in caso di sistemi di riconoscimento delle emozioni o di categorizzazione biometrica occorre informare del funzionamento del sistema le persone fisiche ad esso esposte e per i sistemi di IA che generano o manipolano immagini,

[30] Sul punto cfr. S. WACHTER, B. MITTELSTAND, C. RUSSELL, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, in *Harvard Journal of Law & Technology*, reperibile in *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289*. Cfr. anche M. BRKAN, G. BONNET, *Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas*, in *European Journal of Risk Regulation*, 2020, 35.

audio o video si deve comunicare che il contenuto è stato generato o manipolato artificialmente. Si tratta, quindi, di prescrizioni sulla trasparenza particolarmente stringenti limitatamente ai sistemi ad alto rischio, con conseguenti problemi di garanzia per quanto attiene agli "altri" sistemi di IA.

Alla luce della *opinion* dell'EDPS si possono, quindi, ricavare degli orientamenti utili ai fini della futura regolazione delle decisioni algoritmiche e dei sistemi di IA, anche nell'ottica di "perfezionare" la recente proposta di regolamento della Commissione in materia: qualsiasi nuovo quadro normativo si dovrebbe applicare sia agli Stati membri dell'UE che alle istituzioni, agli uffici, agli organi e alle agenzie dell'UE, dovrebbe essere progettato per proteggere da qualsiasi impatto negativo non solo sugli individui, ma anche sulle comunità e sulla società nel suo complesso, dovrebbe prevedere uno schema di classificazione dei rischi solido e articolato, fondato sulle linee guida del Comitato europeo per la protezione dei dati, assicurando che qualsiasi danno potenziale significativo posto dalle applicazioni di IA sia accompagnato da adeguate misure di mitigazione e, soprattutto, dovrebbe includere una valutazione di impatto e un certo numero di garanzie minime ragionevoli per tutti i sistemi di IA, non soltanto per quelli ad alto rischio.

4. L'*opinion* dell'EDPS sul *Digital Services Act* richiede maggiori garanzie dinanzi all'uso di mezzi automatizzati

A livello unionale si sta formando un quadro normativo composito nel tentativo di disciplinare con una struttura "a mosaico" le piattaforme digitali: sono state introdotte normative settoriali adeguate all'evoluzione tecnologica (i.e. direttiva *copyright*, direttiva sui servizi media audiovisivi, regolamento *Platform to business*) e allo stesso tempo si stanno prevedendo norme trasversali come il *Digital Services Act*, chiamato a colmare le lacune regolatorie delle normative settoriali.

Nel contesto di "*actification*" che sta animando il diritto unionale nella creazione di un nuovo "ecosistema normativo" si colloca, appunto, la proposta di regolamento europeo presentata dalla Commissione il 15 dicembre 2020 in materia di regolamentazione dei servizi digitali e responsabilità delle piattaforme, denominata "*Digital Services Act*", animata da una logica di trasparenza, *accountability* e proporzionalità. Tale proposta prevede una responsabilità graduata per le piattaforme, sulla scia del GDPR, con regole tanto più prescrittive a seconda delle potenzialità lesive dei diritti in gioco, tenendo conto delle possibilità di differenziazione. La proposta mira a introdurre un regime di maggior responsabilità per le piattaforme attraverso una regolazione "a gradoni" per cui al crescere della loro dimensione cresce anche il *set* di meccanismi e *remedies* che possono essere applicati.

Il Regolamento proposto mira a ridefinire la disciplina applicabile alle piattaforme *online*, modificando la Direttiva 31/2000 ed introducendo nuove disposizioni in materia di trasparenza, obblighi informativi e, appunto, *accountability*. La scelta dello strumento regolamentare anzichè della direttiva, peraltro, dimostra l'intenzione della Commissione UE di creare un ecosistema omogeneo e condiviso a livello unionale dinanzi alla pulsione dei vari Stati membri a intervenire in modo autonomo, con l'obiettivo di evitare il frazionamento del mercato.

Ai fini della presente indagine tale proposta assume rilievo là dove contiene specifici obblighi di trasparenza a carico delle piattaforme in caso di utilizzo di sistemi automatizzati e della profilazione, aspetti sui quali è intervenuto l'EDPS con un'*opinion* che richiede misure aggiuntive di garanzia per gli utenti in caso di utilizzo di mezzi automatizzati.

La proposta di regolamento prevede, in particolare, precise regole sulla moderazione dei contenuti e obblighi di trasparenza informativa a carico delle piattaforme, chiamate a spiegare come vengono mostrate le pubblicità e i contenuti raccomandati, come vengono rimossi i contenuti e a informare l'utente circa il funzionamento di meccanismi che consentono la visualizzazione di un certo contenuto pubblicitario raccomandato sulla base della profilazione. In particolare, si prevede che nelle condizioni generali i prestatori di servizi includano informazioni sulle restrizioni che impongono in relazione all'uso dei loro servizi, tra le quali le procedure, le misure, gli strumenti utilizzati ai fini della moderazione dei contenuti, compresi il processo decisionale algoritmico e la verifica umana[31].

Un'altra previsione significativa riguarda i prestatori di servizi di *hosting:* qualora tali soggetti decidano di rimuovere contenuti dovranno informare il destinatario della decisione, fornendo una motivazione chiara e specifica che contenga, ove opportuno, informazioni sugli strumenti automatizzati usati per adottare la decisione. Inoltre, a norma dell'art. 23 della proposta, le piattaforme *online* devono comunicare informazioni su qualsiasi uso di strumenti automatizzati per la moderazione dei contenuti, comprese la descrizione delle finalità, gli indicatori di accuratezza degli strumenti automatizzati nel perseguimento di tali fini e le garanzie applicate.

L'*opinion* dell'EDPS del 10 febbraio 2021 sul *Digital Services Act* ha indubbiamente accolto con favore la proposta della Commissione, compreso il fatto che essa integra e non sostituisce le tutele offerte dal GDPR; l'EDPS ha, inoltre, sostenuto l'obiettivo della Commissione di promuovere un ambiente *online* trasparente e sicuro, definendo in particolare obblighi e responsabilità delle piattaforme *online*.

[31] Cfr. art. 12 della proposta in oggetto.

L'EDPS ricorda, in linea con l'*opinion* n. 4/2020, che alcune attività nel contesto delle piattaforme *online* presentano rischi crescenti non solo per i diritti degli individui, ma per la società nel suo insieme, ciò è tanto più evidente per le piattaforme di grandi dimensioni e ben si riflette nella considerazione che esse dovrebbero sostenere *standard* più elevati di *due diligence*, proporzionati al loro impatto sulla società.

Sebbene la proposta di regolamento includa una serie di misure di mitigazione del rischio l'EDPS ha, però, raccomandato garanzie aggiuntive per tutelare gli interessati, in particolare in caso di pubblicità *online*, moderazione dei contenuti e sistemi di raccomandazione.

Conformemente ai requisiti di minimizzazione dei dati personali e protezione dei dati *by design* e *by default*, secondo l'EDPS la moderazione dei contenuti dovrebbe, per quanto possibile, non comportare alcun trattamento di dati personali. Laddove tale trattamento sia indispensabile dovrebbe riguardare solo i dati necessari per la specifica finalità, applicando tutti gli altri principi del Regolamento (UE) 2016/679.

Nella medesima ottica, nonché al fine di garantire il principio della certezza del diritto, l'*Act* dovrebbe specificare in quali circostanze l'esigenza di contrastare eventuali fenomeni illegali possa legittimare il trattamento dei dati personali e, soprattutto, dovrebbe precisare al ricorrere di quali condizioni potrebbe legittimarsi l'utilizzo di sistemi automatizzati in relazione alle medesime finalità; un utilizzo che, ribadisce l'EDPS, deve essere svolto garantendo il rispetto delle prescrizioni di cui al GDPR in materia di processi decisionali automatizzati. In assenza di ulteriori garanzie, vi è il rischio che la proposta contribuisca indirettamente al trattamento dei dati personali non proporzionato alle finalità perseguite, poiché non qualifica le tipologie di contenuto illegale che possono effettivamente giustificare l'uso di tecniche di rilevamento automatizzato che comportano il trattamento dei dati personali. A seconda delle categorie di dati elaborati e della natura del trattamento, la moderazione automatizzata dei contenuti può avere un impatto significativo sia sul diritto alla libertà di espressione che sul diritto alla protezione dei dati.

L'EDPS sostiene, invece, favorevolmente gli obblighi informativi contenuti nell'articolo 12, paragrafo 1 della proposta, nella misura in cui intendono aumentare ulteriormente la trasparenza delle pratiche di moderazione dei contenuti, nonchè gli obblighi di report sulla trasparenza di cui all'art. 13 e all'art. 23, paragrafo 1, lettera c), secondo i quali il report deve includere anche informazioni su «qualsiasi uso fatto di mezzi automatizzati ai fini della moderazione dei contenuti, compresi una specifica delle finalità precise, indicatori dell'accuratezza dei mezzi automatizzati nell'adempimento di tali finalità ed eventuali garanzie applicate». L'EDPS guarda, altresì, favorevolmente alla previsione

che piattaforme *online* molto grandi siano sottoposte ad *audit* esterni e indipendenti e pubblichino i relativi *reports* di *audit* e valutazione dei rischi.

Con particolare riguardo alla notifica all'*hosting provider* circa la presenza di contenuti illegali e all'utilizzo di mezzi automatizzati per elaborare o prendere decisioni in merito alle notifiche ricevute, l'EDPS accoglie con favore la previsione secondo cui i prestatori di servizi di *hosting* devono fornire, «ove applicabile, informazioni sull'uso fatto di mezzi automatizzati per prendere la decisione» e significativamente questo requisito si applica in situazioni in cui le decisioni non sono prese esclusivamente sulla base di mezzi automatizzati e/o non comportano la profilazione. Tuttavia, per rafforzare la trasparenza, l'EDPS raccomanda di modificare la proposta per affermare in modo inequivocabile che dovrebbero in ogni caso essere fornite le informazioni sui mezzi automatizzati utilizzati per l'individuazione e l'identificazione di contenuti illegali, indipendentemente dal fatto che la decisione successiva abbia comportato l'uso di mezzi automatizzati o meno. In particolare, occorre rafforzare i requisiti di trasparenza, specificando le informazioni da fornire alle persone interessate: le piattaforme *online* che utilizzano mezzi automatizzati per la moderazione dei contenuti o il processo decisionale dovrebbero almeno informare le persone interessate circa la procedura seguita, la tecnologia utilizzata e i criteri e le motivazioni a sostegno della decisione.

Si specifica, inoltre, nella *opinion* che la moderazione dei contenuti non deve comportare il monitoraggio o la profilazione del comportamento delle persone, a meno che il fornitore non possa dimostrare, sulla base di una valutazione del rischio, che tali misure sono strettamente necessarie per mitigare i rischi sistemici identificati nella proposta (ie. diffusione di contenuti illegali; effetti negativi per l'esercizio dei diritti fondamentali; manipolazione intenzionale del servizio).

La proposta di regolamento richiede, altresì, che le piattaforme *online* prevedano un sistema interno di gestione dei reclami contro le decisioni delle piattaforme di rimuovere o disabilitare l'accesso alle informazioni o di sospendere o interrompere la fornitura del servizio e le piattaforme devono garantire che le decisioni in merito a tali reclami non siano prese esclusivamente sulla base di mezzi automatizzati. L'EDPS sostiene tale previsione, ma raccomanda di introdurre una garanzia simile in relazione a tutti i fornitori di servizi di *hosting*, non solo alle piattaforme *online*, ogni volta che il rilevamento e l'identificazione di contenuti illegali comporta il trattamento di dati personali.

Inoltre, l'EDPS raccomanda di specificare che le misure di moderazione dei contenuti devono essere "necessarie" oltre che "proporzionate" alle finalità perseguite, che siano il più mirate possibile e progettate in conformità a principi quali la minimizzazione dei dati: per quanto possibile, nessun dato personale dovrebbe essere elaborato durante la moderazione dei contenuti.

Anche da questa *opinion* emerge, inoltre, l'importanza di una valutazione di impatto: l'EDPS raccomanda di richiedere a tutte le piattaforme *online* che utilizzano strumenti automatizzati di moderazione dei contenuti di pubblicare la DPIA risultante o almeno i rischi identificati e le misure di mitigazione associate.

Per quanto concerne la pubblicità mirata, invece, l'EDPS sostiene fermamente le disposizioni della proposta sugli obblighi informativi che mirano a fornire una maggiore trasparenza e, quindi, responsabilità in modo complementare alla legislazione sulla protezione dei dati.

Considerata, però, la moltitudine di rischi associati alla pubblicità mirata *online*, l'EDPS esorta il legislatore a prendere in considerazione norme aggiuntive che vadano oltre la trasparenza, prevedendo misure ulteriori, quali ad esempio l'introduzione graduale di un divieto di pubblicità mirata basata sul monitoraggio pervasivo, la previsione di restrizioni in ordine alle categorie di dati che possono essere trattati a fini mirati e alle categorie di dati che possono essere comunicate agli inserzionisti.

L'EDPS, inoltre, rileva che la maggior parte, ma non tutta, la pubblicità *online* è gestita automaticamente e raccomanda di aggiungere agli obblighi di trasparenza già previsti nella proposta una nuova voce per informare gli interessati se l'annuncio pubblicitario è stato selezionato utilizzando un sistema automatizzato e, in tal caso, informare sull'identità della persona fisica o giuridica responsabile del sistema.

Da ultimo, l'EDPS ricorda che i sistemi di raccomandazione possono avere un impatto significativo sulla capacità dei destinatari di interagire con le informazioni *online* e possono anche svolgere un ruolo importante nell'amplificazione di determinati messaggi, nella diffusione virale delle informazioni e nella stimolazione del comportamento *online*.

Conformemente ai requisiti di *data protection by design* e *by default* l'EDPS precisa che tali sistemi dovrebbero di *default* non essere basati sulla profilazione e occorre garantire misure ulteriori di trasparenza, assicurando che gli utenti mantengano il controllo sulle proprie scelte.

Per migliorare la trasparenza e il controllo degli utenti, l'EDPS raccomanda di includere nella proposta di regolamento per le piattaforme *online* di dimensioni molto grandi una serie di requisiti aggiuntivi: indicare il fatto che la piattaforma utilizza un sistema di raccomandazione e offrire controlli con le opzioni disponibili in modo facile da usare; informare l'utente della piattaforma se il sistema di raccomandazione è automatizzato e, in tal caso, comunicare l'identità della persona fisica o giuridica responsabile del sistema; consentire agli interessati di visualizzare, in modo intuitivo, qualsiasi profilo utilizzato per curare i contenuti della piattaforma; consentire ai destinatari del servizio di personalizzare i sistemi di raccomandazione sulla base di criteri quali ad

esempio il tempo, i temi di interesse e fornire agli utenti un'opzione facilmente accessibile per eliminare qualsiasi profilo utilizzato per curare il contenuto che vedono.

L'EDPS accoglie, invece, con favore l'inclusione nella proposta di regolamento sia dei rischi sistemici per i diritti individuali che per gli interessi della società, ma esorta il legislatore a chiarire ulteriormente cosa siano i "rischi sistemici" e i "danni per la società", suggerendo di ricomprendere eventuali effetti negativi attuali o prevedibili sulla protezione della salute pubblica, dei minori, del discorso civico, o effetti relativi ai processi elettorali e alla sicurezza pubblica, in particolare in relazione al rischio di manipolazione intenzionale del servizio, anche mediante lo sfruttamento automatizzato del servizio stesso.

L'EDPS ritiene, infine, che l'*Act* dovrebbe garantire una cooperazione istituzionalizzata e strutturata tra le autorità di controllo competenti nel contesto del mercato digitale, comprese le autorità per la protezione dei dati, dei consumatori e le autorità garanti della concorrenza. Occorre assicurare la complementarità nella sorveglianza e nella supervisione delle piattaforme *online* e di altri fornitori di servizi di *hosting,* al fine di mitigare i rischi di indebita interferenza con i diritti fondamentali, ma anche per aumentare la trasparenza e la responsabilità degli attori coinvolti. L'EDPS raccomanda, quindi, che la proposta individui una base giuridica esplicita per la cooperazione tra le autorità competenti, nonché preveda una cooperazione istituzionalizzata e strutturata e faccia esplicito riferimento alle autorità coinvolte, identificando le circostanze in cui tale cooperazione dovrebbe svolgersi. L'EDPS raccomanda, inoltre, misure precise per migliorare la "*governance* digitale": occorre garantire che anche i coordinatori dei servizi digitali e la Commissione abbiano il potere e il dovere di consultare le autorità competenti, comprese le autorità di protezione dei dati, nel contesto delle loro indagini e valutazioni della conformità con la proposta di regolamento, in una sorta di "regolazione integrata" del mercato digitale[32]. Inoltre, la cooperazione dovrebbe riguardare a titolo esemplificativo l'identificazione e valutazione dei rischi sistemici più importanti e ricorrenti, nonché le migliori pratiche per mitigare tali rischi, i codici di condotta per i diversi tipi di contenuti illegali e per la pubblicità *online*.

L'*opinion* è, quindi, molto chiara nell'indicare al legislatore europeo la via per rafforzare le garanzie a favore degli interessati là dove siano impiegati mezzi automatizzati per la pubblicità mirata, la moderazione dei contenuti e i sistemi di raccomandazione. Occorrono ulteriori misure non soltanto in punto di trasparenza, bensì anche nella previsione normativa chiara delle ipotesi che

[32] L'EDPS raccomanda, inoltre, di chiarire che le autorità di controllo competenti ai sensi della proposta dovrebbero essere in grado di fornire alle autorità di controllo competenti ai sensi del regolamento (UE) 2016/679 tutte le informazioni ottenute nel contesto di eventuali *audit* e indagini relative al trattamento dei dati personali, includendo una base giuridica esplicita in tal senso.

consentono di impiegare processi automatizzati e il trattamento dei dati personali per rilevare, identificare e fronteggiare i contenuti illegali, secondo il principio di stretta necessità della profilazione a tal fine. La moderazione dei contenuti dovrebbe, per quanto possibile, non comportare alcun trattamento di dati personali e qualsiasi fornitore di servizi di *hosting* che utilizzi mezzi automatizzati di moderazione dei contenuti dovrebbe garantire che tali strumenti non producano risultati discriminatori o ingiustificati. Le misure di moderazione dovrebbero essere necessarie, oltrechè proporzionate ai fini perseguiti e occorre, altresì, rafforzare i requisiti di trasparenza della proposta, specificando ulteriormente le informazioni da fornire alle persone interessate, in particolare in caso di utilizzo di mezzi automatizzati per la moderazione dei contenuti.

L'EDPS, inoltre, indica misure concrete che vadano oltre la trasparenza in caso di pubblicità mirata, compresa un'eliminazione graduale che porti al divieto di pubblicità mirata sulla base del tracciamento pervasivo e la considerazione di limitazioni in relazione alle categorie di dati che possono essere elaborate a fini mirati e che possono essere comunicate agli inserzionisti, nonché l'informazione agli interessati se l'annuncio pubblicitario è stato selezionato utilizzando un sistema automatizzato.

Del pari i sistemi di raccomandazione dovrebbero di *default* non essere basati sulla profilazione e le informazioni riguardanti il ruolo e il funzionamento di tali sistemi dovrebbero essere presentate in un modo facilmente accessibile, chiaro e conciso.

L'EDPS indica, inoltre, misure specifiche per le piattaforme *online* di dimensioni molto grandi, in particolare obblighi informativi se il meccanismo di raccomandazione è un sistema decisionale automatizzato e richiede di consentire agli interessati di visualizzare, in modo intuitivo, i profili utilizzati per curare i contenuti della piattaforma o di fornire agli utenti un'opzione facilmente accessibile per eliminare i profili utilizzati per curare il contenuto che vedono.

Infine, l'*Act* dovrebbe prevedere una cooperazione istituzionalizzata e strutturata tra le autorità di controllo competenti, incluse le autorità di protezione dei dati, identificando le circostanze in cui la cooperazione dovrebbe aver luogo, nell'ottica di un'"integrazione regolatoria" volta a garantire un maggior dialogo e una stretta interazione tra autorità di controllo competenti, verso la creazione di un "*network* di vigilanza" anche sulle decisioni algoritmiche delle grandi piattaforme *online*.

5. I principi fondamentali in materia di IA e alcune raccomandazioni per il futuro: il report della CNIL sull'intelligenza artificiale come volano per una futura regolazione giuridica

Nel dicembre 2017 la CNIL ha adottato un report "*How can humans keep the upper hand? Ethical matters raised by algorithms and artificial intelligence*" che assume particolare rilievo sia per il contenuto sostanziale che lo contraddistingue sia per le modalità di adozione. Rispetto al secondo profilo il report trae origine da un dibattito pubblico condotto dalla CNIL e tiene conto dei commenti e dei punti di vista espressi in più di quaranta eventi tenuti in tutta la Francia nell'arco di circa un anno. È stato, quindi, adottato un approccio inclusivo, partecipativo e decentralizzato che ha contribuito a migliorare la conoscenza della società francese sulle questioni sollevate dagli algoritmi e dall'IA.

Il report si concentra sull'intelligenza artificiale fondata sull'apprendimento automatico e individua una serie di problematiche con riguardo all'uso di algoritmi: in primo luogo, ci si chiede se le macchine autonome siano una minaccia al libero arbitrio e alla responsabilità. Una delle grandi sfide è come garantire che tali sistemi non "diluiscano" le responsabilità dei numerosi attori coinvolti nel processo algoritmico.

In secondo luogo, gli algoritmi e l'intelligenza artificiale possono creare pregiudizi, discriminazione o addirittura esclusione nei confronti di individui e gruppi di persone. Tutti gli algoritmi sono in un certo senso "distorti", in quanto sono sempre il riflesso - attraverso la loro configurazione e i criteri operativi, o attraverso i loro dati di *input* - di un insieme di scelte e valori sociali[33]. Secondo la CNIL è il *bias* generato dai dati di *input* che pone la sfida maggiore nell'attualità: occorre considerare che un *set* di dati può riprodurre discriminazioni o disuguaglianze preesistenti. La scelta dei dati di *input* da utilizzare per le fasi di formazione dell'algoritmo implica chiaramente l'adozione di decisioni che potrebbero avere conseguenze di vasta portata ed il problema è che spesso le stesse persone che curano i dati di *input* non sono consapevoli del possibile *bias*[34].

[33] Es. nel 2015, Google Foto, un *software* di riconoscimento facciale, ha suscitato scalpore quando due giovani afroamericani si sono resi conto che una delle loro foto era stata etichettata come "Gorilla". Questo problema tecnico può essere spiegato dal tipo di dati con cui l'algoritmo è stato addestrato a riconoscere le persone. In questo caso, è probabile che fosse principalmente - se non esclusivamente - addestrato con fotografie di persone bianche. Del pari possono essere create discriminazioni di genere: uno studio del 2015 ha mostrato che Adsense, il programma pubblicitario di Google, ha generato pregiudizi contro le donne. Utilizzando uno strumento chiamato Adfisher, i ricercatori hanno creato 17.000 profili e simulato la loro navigazione sul *Web* per condurre una serie di esperimenti. Ciò che hanno scoperto è che le donne avevano molte meno probabilità di visualizzare annunci di lavoro per posizioni altamente pagate rispetto agli uomini, per livelli simili di qualifiche ed esperienza.

[34] In dottrina sul tema dell'*algorithmic bias* cfr. D. RESTREPO AMARILES, *Algorithmic Decision Systems. Automation and Machine Learning in the Public Administration*, in W. BARFIELD (eds.), *The*

Il report si occupa anche delle "*filter bubbles*"[35] e delle loro implicazioni sul pluralismo e sulla diversità culturale. Filtrando le informazioni in base alle caratteristiche dei profili utente, gli algoritmi stanno rafforzando le tendenze degli individui ad abbracciare solo quegli oggetti, persone, opinioni e culture che si conformano ai loro interessi. A livello individuale, infatti, il rischio è che ogni persona si veda associata a un *alter ego* digitale che è ricavato dai suoi dati personali, isolandolo all'interno di una bolla di raccomandazioni che è sempre conforme a questo profilo. A livello sociale, i diversi modi in cui gli individui sono "protetti dall'alterità", dalle opinioni diverse dalle proprie - in termini politici in particolare - potrebbero essere problematici per la qualità e la vitalità del dibattito pubblico, per la diversità delle informazioni, in generale per il buon funzionamento delle democrazie e non a caso il dibattito sulle *filter bubbles* e sui loro esiti politici ha avuto particolare importanza durante le elezioni presidenziali statunitensi del 2016 e il referendum sulla *Brexit* di pochi mesi prima.

A ciò la CNIL aggiunge la sfida della selezione dei dati utilizzati dall'IA: la posta in gioco è alta quando si tratta di scegliere i dati che saranno elaborati da un algoritmo ed ogni processo di selezione dovrebbe ricercare quantità, accuratezza e assenza di *bias*. La questione della qualità dei dati elaborati dagli algoritmi e dall'intelligenza artificiale è la più semplice: dati errati o semplicemente superati porteranno a errori o malfunzionamenti di gravità variabile a seconda del settore, dal mero invio di pubblicità mirata che non si abbina al profilo a una diagnosi medica errata.

Errori o distorsioni nei dati producono effetti sul modello decisionale finale e sullo sviluppo di un sistema decisionale algoritmico affidabile[36]. Lo stesso Consiglio di Stato italiano[37] ha osservato che occorre rettificare i dati in "ingresso" per evitare effetti discriminatori nell'*output* decisionale, operazione che richiede la necessaria cooperazione di chi istruisce le macchine.

Cambridge, cit., 280 ss.; K. HAO, *This Is How AI Bias Really Happens – and Why It's So Hard to Fix*, in *MIT Technology Review*, February 4, 2019; S. BAROCAS, A.D. SELBST, *Big Data's Disparate Impact*, cit., 671; J. DUSTIN, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women*, in *Reuters*, October 11, 2018, *www.reuters.com*; T. SPEICHER, M. ALI, G. VENKATADRI, et al., *Potential for Discrimination in Online Targeted Advertising* (2018) 81 *Proc. Mach. Learn. Res.* 1.

[35] Sulle *"filtering bubbles"* cfr. S. FLAXMAN, S. GOEL, J.M. RAO, *Filter Bubbles, Echo Chambers, and Online News Consumption*, in *Public Opinion Quarterly*, March 2016, 298-320, consultabile su: https://doi.org/10.1093/poq/nfw006; E. PARISER, *The Filter Bubble: What the Internet Is Hiding from You*, Penguin Group, New York, 2011; C.R. SUNSTEIN, *#republic: Divided Democracy in the Age of Social Media*, Princeton University Press, Princeton, 2017, 3 ss.

[36] Come evidenziato anche da C. COLAPIETRO, in questo fascicolo, § 3. Cfr. anche EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *Data quality and artificial intelligence – mitigating bias and error to protect fundamental right*, 7 June 2019; INFORMATION COMMISSIONER'S OFFICE, *Guidance on the AI auditing framework – Draft guidance for consultation*, 2020, 26 ss.

[37] Consiglio di Stato, Sez. VI, n. 8472/2019.

È bene ricordare che obiettività e imparzialità sono doti che l'IA ancora non possiede, ma che potrebbe sviluppare se l'uomo si muoverà verso questa consapevolezza, intervenendo proprio sulla qualità dei dati. Il ricorso agli algoritmi, infatti, non garantisce *ex se* imparzialità della decisione né necessariamente la sua efficienza, poichè spesso i malfunzionamenti sono originati da dati di bassa qualità o da processi mal congeniati: occorre, quindi, garantire che i dati utilizzati siano il più possibile di buona qualità[38]. Lart. 5 GDPR prevede un principio fondamentale in materia, quello di esattezza dei dati personali[39]: l'importanza di tale principio nell'ambito di decisioni automatizzate è vitale, dal momento che dati viziati all'origine possono provocare previsioni errate relativamente a aspetti significativi della vita delle persone e decisioni inadeguate o incongrue.

Proprio in riferimento alla qualità dei dati occorre osservare che la proposta di regolamento sull'IA introduce dei profili di innovazione: per la prima volta è prevista una disciplina concreta della fase di *training* dei sistemi di *machine learning*, quindi, del momento in cui vengono usati i dati per costruire o validare algoritmi, prevedendo che i *set* di dati di addestramento devono essere «pertinenti, rappresentativi, esenti da errori e completi» e sarà interessante capire come verrà garantita effettività a tale disciplina.

Un'altra problematica che emerge dal report concerne l'identità umana dinanzi alla sfida dell'IA: l'idea di un'unicità umana irriducibile è messa in discussione dall'autonomia di tali sistemi, da un lato, e dalla crescente ibridazione degli esseri umani con le macchine, dall'altro. Innanzitutto, c'è la questione della "macchina etica": si indaga se sia possibile stabilire un quadro etico per la programmazione in una macchina. Per quanto attiene alla questione dell'ibridazione un confine poco chiaro è emerso nel contesto di alcuni tentativi nel campo della robotica, volti a dare ai robot l'aspetto di esseri umani e, inoltre, un intero campo di ricerca è orientato alla progettazione di robot empatici in grado di percepire le emozioni umane, analizzando ad esempio il volto o la voce, per adattare il loro comportamento all'umore dell'interlocutore. Ciò solleva la questione di dove stia il limite tra, da un lato, i benefici di una forma di IA capace di comprendere e adattarsi agli stati d'animo dei destinatari e, dall'altro, una forma di manipolazione che si basa sull'ingegneria tecnica che è in grado di sfruttare le vulnerabilità emotive.

[38] Cfr. C. COLAPIETRO, *ult. cit*.

[39] Sulla importanza del principio di esattezza dei dati con particolare riguardo all'utilizzo di sistemi di IA, cfr. G. D'ACQUISTO, *Qualità dei dati e Intelligenza Artificiale: intelligenza dai dati e intelligenza dei dati*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, Torino, 2018, 265 ss.; H. ZHU, E. MADNICK, Y.W. LEE, R.Y. WANG, *Data and Information Quality Research,* in H. TUCKER, T. GONZALEZ, A. TOPI (eds.), *Computing handbook*, 2014; J. STEVENS, *Datenqualität bei algorithmischen Entscheidungen*, in *Computer und Recht* 2020, 73; sul punto anche il Garante per la protezione dei dati personali, cfr. provv. n. 515/2013.

Un ulteriore profilo problematico di grande rilievo evidenziato dalla CNIL consta nel fatto che la regolamentazione giuridica in materia attualmente riguarda gli effetti che gli algoritmi hanno sugli interessati - cioè da una prospettiva individuale - ma non fa menzione diretta di eventuali effetti collettivi: si pensi all'impatto che gli algoritmi utilizzati per il *marketing* elettorale possono avere sulla democrazia stessa. In tal caso sorgono problemi di tutela delle posizioni giuridiche soggettive, ma anche questioni delicate di tenuta democratica e di "inquinamento" nei processi informativi dell'opinione pubblica[40].

L'*opinion* n. 4/2020 dell'EDPS si pone, pertanto, in linea di continuità con il report della CNIL nell'evidenziare l'importanza di una valutazione degli impatti collettivi dei sistemi di IA, ad esempio proprio con riguardo al principio democratico.

Nella prospettiva di una futura regolazione degli algoritmi e dell'IA un aspetto di particolare rilievo del report consta nei principi fondamentali ivi delineati per garantire che l'intelligenza artificiale sia al servizio degli esseri umani: il principio di *fairness*, di una *fair* IA e il principio di continua attenzione e vigilanza da applicare a ogni singolo *stakeholder* (*designer*, aziende, utenti finali) coinvolto in "catene algoritmiche".

Secondo la CNIL un algoritmo *fair* non dovrebbe finire per generare, replicare o aggravare alcuna forma di discriminazione, anche se ciò dovesse avvenire senza che i progettisti ne siano consapevoli. Il principio sostanziale di *fairness* degli algoritmi dovrebbe, inoltre, tener conto dell'idea di correttezza nei confronti degli utenti non solo come consumatori, ma anche come cittadini e anche nei confronti delle comunità che potrebbero essere influenzate da algoritmi, indipendentemente dal fatto che questi elaborino o meno dati personali.

L'altro principio fondamentale riguarda un'attenzione e vigilanza continuativa ed è un principio metodologico: la natura modificabile degli algoritmi di apprendimento automatico e il loro potenziale impatto aumentano l'imprevedibilità dei risultati. Promuovere un principio di attenzione e vigilanza continue potrebbe essere un modo per affrontare questa sfida, dovendo i progettisti e gli operatori di intelligenza artificiale tenere conto di questa nuova caratteristica. Un ulteriore scopo di tale principio sarebbe quello di compensare l'eccessiva fiducia e l'indebolimento della responsabilità che possono sorgere di fronte ai *black box algorithms*.

Tale principio dovrebbe riguardare i sistemi algoritmici, quindi, complesse e lunghe "catene algoritmiche" composte da una miriade di *stakeholders* (sviluppatori, utenti finali, aziende che raccolgono dati per scopi di *machine learning*, professionisti che svolgono il "processo di apprendimento", acquirenti di una soluzione di *machine learning* che intendono implementare): si tratta di

[40] Sul punto cfr. anche A. CARDONE, in questo fascicolo.

essere "*vigils*" *of digital society*. Questo principio deve essere inteso come una risposta concreta a tre sfide centrali che la società digitale deve affrontare: la natura mutevole e imprevedibile degli algoritmi nell'era del *machine learning*[41]; in secondo luogo la "*silo mentality*" che influenza l'organizzazione delle catene algoritmiche e porta ad azioni che si svolgono in isolamento, senza guardare agli impatti complessivi del sistema algoritmico ed il rischio che venga riposta un'eccessiva fiducia nelle macchine.

Ulteriori principi fondamentali individuati nel report sono il principio di esplicabilità o intellegibilità[42]: come evidenziato nelle linee guida del gruppo

[41] Il *machine learning* consente di svolgere compiti molto più complessi di un algoritmo convenzionale: Andrew Ng della Stanford University definisce l'apprendimento automatico come: «la scienza per far agire i computer senza essere programmati esplicitamente». L'intelligenza artificiale fondata sull'apprendimento automatico riguarda, quindi, algoritmi che sono stati specificatamente progettati affinché il loro comportamento possa evolversi nel tempo, sulla base dei dati di *input*. La dottrina sulla *governance* algoritmica con particolare riguardo al *machine learning* è sconfinata, cfr. *ex multiis* C. COGLIANESE, D. LEHR, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era* (2017) 105 *Geo. Law J*., 1147; C. COGLIANESE, D. LEHR, *Transparency and Algorithmic Governance* (2019) 71 *Admin. Law Rev*. 1. Per un'eccellente ricostruzione introduttiva del *machine learning* cfr. D. LEHR, P. OHM, *Playing with the Data: What Legal Scholars Should Learn about Machine Learning* (2017) 51 *UC Davis Law Rev*., 653; T. MILLS, *Machine Learning vs Artificial Intelligence: How Are They Different?*, in *Forbes*, June 11, 2018, www .forbes.com/sites/forbestechcouncil/2018/07/11/machine-learning-vs-artificial-intelligence-how-are-they-different /#4b961f153521; C. KUMAR GN, *Artificial Intelligence vs Machine Learning*, in *Medium*, September 1, 2018, https://medium.com/@chethankumargn/artificial-intelligence-vs-machine-learning-3c599637ecdd; M.I. JORDAN, T.M. MITCHELL, *Machine Learning: Trends, Perspectives, and Prospects* (2015) 349 *Science*, 255; H. SURDEN, *Machine Learning and Law* (2014) 89 *Wash. Law Rev*., 87.

[42] Sulla trasparenza degli algoritmi occorre evidenziare che in Francia una autorità indipendente, il CSA (*Conseil supérieur de l'audiovisuel*) si è vista conferire *ex lege* una funzione di garanzia in materia di trasparenza algoritmica. La legge di riferimento è la *loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information*. L'art. 11 della legge prevede che gli operatori di piattaforme *online* attuano misure sulla trasparenza dei propri algoritmi; l'art. 14 specifica meglio i dati che tali piattaforme sono tenute a pubblicare relativamente all'impiego di algoritmi di raccomandazione, classificazione, reindirizzamento di contenuti informativi relativi a un dibattito di interesse generale. Il CSA vigila sulla applicazione della legge e gli operatori delle piattaforme sono tenuti a comunicare al medesimo quali misure abbiano preso per garantire la trasparenza degli algoritmi impiegati (cfr. art. 11, ultimo periodo). Molto interessante anche la *Recommandation n° 2019-03 du 15 mai 2019 du Conseil supérieur de l'audiovisuel aux opérateurs de plateforme en ligne dans le cadre du devoir de coopération en matière de lutte contre la diffusion de fausses informations* del CSA, in particolare il punto 2 della Raccomandazione che è dedicato alla trasparenza degli algoritmi: «*Les utilisateurs doivent pouvoir exercer de manière éclairée leur esprit critique sur les contenus qui leur sont proposés par les plateformes en ligne. Ils doivent pouvoir accéder aux informations leur permettant de connaître et de comprendre les principes de fonctionnement des algorithmes qui régissent l'organisation, la sélection et l'ordonnancement de ces contenus*». Inoltre, il punto 7 della Raccomandazione prevede che gli operatori di piattaforme debbano trasmettere al CSA ogni anno, entro il 31 marzo, una dichiarazione in cui precisano «*les modalités de mise en œuvre de chacune des mesures prises en application de l'article 11 de la loi du 22 décembre 2018", telles qu'éclairées par la présente recommandation*». Sempre in base al punto 7, il CSA assume anche alcuni poteri ispettivi, potendo richiedere informazioni alle piattaforme in caso di episodi di manipolazione dell'informazione *online*. Il CSA, inoltre, pubblica ogni anno un bilancio basato sulle dichiarazioni degli operatori che sono state trasmesse ed è molto

articolo 29 secondo la CNIL piuttosto che avere accesso diretto al codice sorgente, ciò che importa è la capacità di comprendere la logica generale alla base del modo in cui funziona l'algoritmo, questa logica deve, quindi, essere spiegata a parole anziché in codice.

Infine, si afferma il principio dell'intervento umano nelle decisioni algoritmiche[43]: il principio che vieta qualsiasi processo decisionale che produca effetti giuridici su un soggetto interessato, se basato esclusivamente sul trattamento automatizzato dei dati personali è previsto dal GDPR, tuttavia, è privato di gran parte della sua sostanza a causa di eccezioni molto ampie, come evidenziato da tempo in dottrina[44]. Sul punto la proposta di regolamento europeo sull'IA prevede una "*human oversight*" che sembra avere, però, una portata sostanziale minore rispetto al più ampio "intervento umano" e comunque limita problematicamente, come già evidenziato, tale garanzia ai sistemi di IA ad alto rischio.

Nella prospettiva nazionale anche il Consiglio di Stato[45] ha posto in evidenza l'importanza del principio di non esclusività della decisione algoritmica, specificando che deve comunque esistere nel processo decisionale un contributo umano capace di controllare, validare ovvero smentire la decisione automatica secondo il modello *human in the loop* nel quale per produrre il suo risultato è necessario che la macchina interagisca con l'essere umano.

In una prospettiva *de iure condendo* ai fini di una futura regolazione dell'IA assumono particolare rilievo anche le raccomandazioni finali espresse dalla CNIL, rivolte sia alle autorità pubbliche che alla società civile: è necessaria un'azione lungo tutta la catena algoritmica, dal progettista all'utente finale, attraverso i formatori dei sistemi e una gamma diversificata di azioni da parte dei vari *stakeholders*.

In primo luogo, occorre favorire l'educazione di tutti gli attori coinvolti nelle "catene algoritmiche" con riguardo ai profili etici sollevati dall'IA. I cittadini hanno un ruolo chiave, perché possono individuare eventuali abusi: consentire loro di capire queste nuove tecnologie in modo che possano usarle in modo sicuro, attivo e informato è importante. Occorre, quindi, sviluppare una

interessante il primo bilancio, pubblicato il 30 luglio 2020. Sul ruolo del CSA nel sistema francese e sul bilancio pubblicato nel 2020 cfr. il contributo di E. CATERINA, in questo fascicolo.

[43] Sull'idea di una "*collaborative intelligence*" cfr. H.J. WILSON, P.R. DAUGHERTY, *Collaborative intelligence: humans and AI are Joining forces*, in *Harvard Business Review*, July-August, 2018.

[44] Cfr. S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation* (2017) 7 *International Data Privacy Law* 2, 76 ss.

[45] Consiglio di Stato, Sez. VI, n. 8472/2019, punto 15.2. In tale decisione il Consiglio di Stato si è soffermato anche sul principio della trasparenza «da intendersi sia per la stessa p.a. titolare del potere per il cui esercizio viene previsto il ricorso allo strumento dell'algoritmo, sia per i soggetti incisi e coinvolti dal potere stesso».

"nuova alfabetizzazione digitale" dalla scuola primaria fino al livello universitario che includa una familiarità di base con gli algoritmi. È, inoltre, fondamentale che i progettisti di algoritmi (sviluppatori, programmatori, codificatori, *data scientists*, ingegneri) abbiano la massima consapevolezza possibile delle conseguenze etiche e sociali del loro lavoro: la formazione è un primo passo essenziale affinché tali soggetti siano in grado di cogliere le implicazioni a volte molto indirette della loro azione sia per gli individui che per la società, rendendoli così consapevoli delle proprie responsabilità e imparando a mostrare attenzione e vigilanza continue.

Un'altra importante raccomandazione consiste nel rendere comprensibili i sistemi algoritmici, rafforzando i diritti esistenti: un modo per affrontare la sfida sarebbe vincolare i responsabili del sistema all'obbligo (piuttosto che semplicemente laddove richiesto dagli interessati) di comunicare le informazioni in modo chiaro e comprensibile, in un modo che consenta di comprendere la logica coinvolta, anche per algoritmi che non elaborano i dati personali dei loro utenti, nella misura in cui possono avere impatti collettivi significativi. Sul punto la proposta di regolamento sull'IA parzialmente risponde alla sollecitazione, prevedendo che i sistemi di IA ad alto rischio siano progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'*output* del sistema e utilizzarlo adeguatamente e siano accompagnati da istruzioni per l'uso che comprendano informazioni concise, complete, corrette, chiare, pertinenti e comprensibili per gli utenti concernenti anche le caratteristiche, le capacità e i limiti delle prestazioni del sistema di IA.

Un'ulteriore importante raccomandazione della CNIL concerne il miglioramento della progettazione dei sistemi algoritmici per prevenire il *black box effect*: occorre lavorare sulla progettazione per contrastare la natura "simile a una scatola nera" che gli algoritmi possono assumere, presentandosi come sistemi imperscrutabili che mostrano risultati senza mettere in prospettiva i propri limiti o spiegare il modo in cui sono costruiti. A titolo esemplificativo si potrebbe pensare alla creazione di sistemi di visualizzazione che restituiscano un maggiore controllo agli utenti, dando loro informazioni migliori. A tal riguardo il concetto di "testabilità" potrebbe rappresentare un principio che governa la progettazione di sistemi algoritmici virtuosi, *user-friendly*: occorre consentire agli utenti di "testare" i sistemi, giocando con le loro impostazioni, ad esempio dando loro l'opportunità di eseguire un test, vedendo quali risultati vengono forniti per le diverse scelte prima di compiere le loro scelte finali.

Profili problematici sono, invece, sollevati dall'ulteriore raccomandazione della creazione di una piattaforma nazionale per controllare gli algoritmi, per verificare la conformità alla legge e la *fairness*.

Sul punto il report resta piuttosto vago per quanto attiene al soggetto legittimato a effettuare tali controlli: essi, secondo la CNIL, potrebbero essere eseguiti da un ente pubblico formato da esperti che monitorerebbe e testerebbe gli algoritmi, oppure un'altra soluzione potrebbe essere che le autorità pubbliche accreditino società private di *audit* sulla base di un quadro giuridico di riferimento.

Da questo punto di vista la prospettiva di affidare a terzi la valutazione della *fairness* degli algoritmi è di particolare interesse e per affrontare il problema del *black box effect* si potrebbe ipotizzare *de iure condendo* di richiedere almeno due pareri distinti ad imprese di *audit* che diano specifiche garanzie di indipendenza e imparzialità, con comprovata esperienza e competenza per valutare la *fairness* dell'algoritmo.

Da ultimo, un aspetto di particolare rilievo del report concerne l'implementazione dell'etica all'interno delle imprese, ad esempio creando comitati etici, diffondendo buone pratiche in ogni settore o rivedendo codici etici: una soluzione potrebbe essere la creazione di comitati etici all'interno delle aziende che utilizzano algoritmi con impatti di vasta portata.

Un profilo problematico del report consta, invece, nel fatto che la CNIL non specifica se per l'attuazione dei principi e delle raccomandazioni delineate si debba dare la precedenza alla legge o alle iniziative volontarie dei vari *stakeholders*. Il suggerimento della fonte più idonea a regolare i singoli aspetti e a dare attuazione alle raccomandazioni espresse avrebbe potuto fornire elementi di maggior chiarezza in una prospettiva *de iure condendo*.

Il report rappresenta, tuttavia, un importante volano per la regolazione giuridica dei sistemi di IA, là dove indica i principi fondamentali che dovrebbero disciplinare la materia e suggerisce una via per rafforzare i diritti degli interessati e arginare il *black box effect*, introducendo il principio di testabilità e prevedendo sistemi di *audit* della *fairness*.

6. Casi problematici in tema di decisioni algoritmiche: l'APB e la mancanza di un intervento umano

Tra i Paesi che hanno introdotto sistemi di decisione algoritmica nel settore pubblico, la Francia merita un'analisi particolarmente approfondita: dall'indagine della prassi in materia sono emersi due casi problematici, particolarmente significativi, concernenti proprio l'ordinamento giuridico francese.

Il primo caso ha riguardato la piattaforma di ammissione alle università francesi "*Admission Post-Bac*", utilizzata dal Ministero dell'istruzione per l'assegnazione di posti universitari. Nel 2009, il Governo francese ha lanciato, per la prima volta, questa piattaforma *online,* APB: un *software* che consente ai diplomati delle scuole superiori di cercare un posto in un programma

universitario. L'algoritmo APB combina le preferenze dei candidati con i requisiti dei programmi delle università[46] ed è stato progettato per applicare le regole e i criteri contenuti nel Codice francese dell'istruzione superiore, delegando al *software* ciò che in precedenza veniva svolto dai funzionari pubblici, per garantire un'applicazione uniforme e imparziale delle regole. Chiaramente, però, APB ha un impatto diretto sul futuro professionale e sulle opportunità degli studenti.

Tale dispositivo assume particolare rilievo ai fini della presente indagine, poiché la CNIL è intervenuta, demolendo punto per punto il funzionamento di APB nel 2017[47].

La Commissione ha ritenuto che l'APB avesse violato il divieto di emettere decisioni amministrative individuali basate unicamente sul trattamento automatizzato: la legge francese, infatti, non autorizza ad assegnare posti nelle università e nei programmi di formazione agli studenti senza alcun intervento umano.

In base a tale dispositivo ogni candidato poteva formulare fino a 24 preferenze sulla piattaforma e, una volta registrate, l'elaborazione automatizzata era realizzata per classificare i candidati. Un algoritmo stabiliva il profilo delle persone a partire da tre criteri di importanza decrescente, vale a dire la loro accademia di collegamento, l'ordine delle preferenze formulate e la loro situazione familiare.

Il trattamento permetteva, così, di indirizzare automaticamente ai candidati, a partire dalla classifica effettuata dall'algoritmo, una proposta di formazione senza che gli istituti di istruzione disponessero di un controllo sull'assegnazione finale proposta.

Nella Decisione 2017-053 del 30 agosto 2017 la CNIL ha constatato che nessuna informazione relativa al trattamento dei dati personali è presente sul sito, le informazioni relative in particolare ai destinatari dei dati non figurano nelle note legali del sito Internet e le risposte fornite ai candidati che contestano l'assegnazione proposta non fanno riferimento all'uso di un algoritmo per procedere alla classificazione. Secondo la CNIL, inoltre, risulta che nessun riesame della decisione finale presa sulla sola base del trattamento viene effettuato alla luce degli elementi forniti dai candidati che intendono contestare le decisioni prese nei loro confronti. Pertanto, le proposte di assegnazione a corsi di formazione nell'istruzione superiore vengono svolte sulla base di un'elaborazione completamente automatizzata e senza alcun intervento umano. Tali fatti

[46] Per un'ampia disamina del funzionamento di APB e della sua collocazione nel sistema educativo francese cfr. P.H. GUILLAUD, *Admission post-bac, cas d'école des algorithmes publics?*, in *internetactu.net*, 28.07.2017.

[47] CNIL, *Decision* No. 2017-053, *August* 30, 2017, www.legifrance.gouv.fr/affichCnil.do?&id=CNILTEXT 000035647959.

costituiscono un inadempimento del quadro legislativo di riferimento, secondo il quale nessuna decisione che produca effetti giuridici nei confronti di una persona può essere presa sulla sola base di un trattamento automatizzato di dati inteso a definire il profilo dell'interessato o a valutare determinati aspetti della sua personalità.

La CNIL, come osservato, constata, inoltre, il mancato rispetto dell'obbligo di informare le persone: i candidati devono fornire un numero significativo di dati personali quali cognome, nome, data di nascita, paese di nascita, situazione familiare, ma non sono fornite informazioni relative in particolare all'identità del responsabile del trattamento, allo scopo del trattamento e ai diritti di cui dispongono le persone ai sensi della legge sulla protezione dei dati. Inoltre, le note legali del sito sono incomplete, in quanto non indicano i destinatari dei dati dei candidati e tali fatti costituiscono una violazione del quadro legislativo di riferimento.

Un altro inadempimento constatato dalla CNIL concerne il diritto di accesso: se i candidati chiedono i motivi di un rifiuto di assegnazione al termine della procedura prevista dal trattamento, nessuna informazione relativa all'uso di un algoritmo e al suo funzionamento per la classificazione e l'assegnazione (in particolare il metodo che ha permesso di sviluppare l'algoritmo, il suo tasso di errore o il punteggio ottenuto dal candidato) è fornita ai candidati. Questi fatti costituiscono una violazione dell'articolo 39-I-5 della legge del 6 gennaio 1978 come modificata, secondo il quale ogni persona fisica ha il diritto di interrogare il responsabile del trattamento al fine di ottenere le informazioni che consentano di conoscere la logica alla base del trattamento automatizzato in caso di decisioni prese in base ad esso che producano effetti giuridici nei confronti dell'interessato.

Da tale caso problematico si possono, quindi, trarre orientamenti particolarmente utili in ordine alla regolazione delle decisioni algoritmiche: decisioni che producono effetti giuridici rilevanti sulle persone non dovrebbero basarsi sul solo trattamento automatizzato di dati e occorre che l'intervento umano tenga conto delle osservazioni delle persone interessate. Occorre, inoltre, garantire l'informazione degli interessati, con particolare riguardo all'identità del responsabile del trattamento, alle finalità del trattamento, ai diritti delle persone, indicando i destinatari ai quali i dati sono trasmessi e occorre, altresì, attuare un'efficace procedura per il trattamento delle richieste di diritto di accesso, garantendo la trasmissione di informazioni che consentano di conoscere e contestare la logica sottostante al trattamento.

7. Parcoursup e il parere favorevole della CNIL sul suo funzionamento

A seguito di questa decisione *tranchant* della CNIL, il Governo francese ha istituito una nuova piattaforma denominata Parcoursup[48] che ha, invece, ricevuto un parere favorevole dalla Commissione nazionale[49]. Parcoursup assegna automaticamente gli studenti agli istituti di istruzione superiore, ma si è previsto un certo grado di intervento umano[50]. Gli studenti registrano le università e/o i programmi di formazione sulla piattaforma Parcoursup e li classificano in ordine di preferenza e priorità. L'università, quindi, classifica le domande secondo i propri criteri interni e, infine, il sistema fornisce le proposte ottimali ai candidati, tenendo conto delle preferenze dell'università e dello studente.

La Deliberazione della CNIL n. 2018-119 del 22 marzo 2018 è relativa ad una richiesta di parere da parte del Ministro dell'istruzione superiore in merito al decreto che autorizza il trattamento di dati personali denominato Parcoursup. Sul funzionamento del dispositivo la CNIL spiega che esso opera in tre fasi: la raccolta delle preferenze, la loro classificazione e l'assegnazione dei candidati nelle formazioni di insegnamento superiore. I candidati si collegano alla piattaforma Parcoursup per esprimere le loro preferenze, in secondo luogo, gli istituti di istruzione superiore accedono ai dati e alle informazioni relative ai candidati che fanno domanda per uno dei corsi di formazione offerti. Alla luce di queste informazioni gli istituti di istruzione superiore forniscono un parere sulla piattaforma su ciascuna delle domande ricevute. In terzo luogo, Parcoursup consente, grazie ad un algoritmo, la gestione delle risposte ai candidati secondo i pareri formulati dagli istituti di istruzione superiore e, a differenza del sistema APB, la graduatoria può essere modificata dal Rettore dell'Accademia.

In merito alle garanzie apprestate da Parcoursup la CNIL rileva diversi elementi relativi alla trasparenza e all'intervento umano: per quanto attiene al primo profilo, secondo la CNIL informazioni sull'elaborazione algoritmica più accessibili e intellegibili del solo codice sorgente sono fornite ai candidati, ciò che costituisce un passo avanti rispetto al precedente sistema di preregistrazione.

Tuttavia, nella decisione, la Commissione non indica quali siano le informazioni sull'elaborazione algoritmica fornite ai candidati e permangono, pertanto, problemi di trasparenza in ordine alla logica alla base della decisione algoritmica. Inoltre, la CNIL valuta positivamente il fatto che a valle del processo di

[48] *Arre^te´, January* 19, 2018 *autorisant la mise en œuvre d'un traitement automatisé de données à caractère personnel dénommé «Parcoursup»*, www.legifrance.gouv.fr/affichTexte.do? cidTexte=JORFTEXT000036520954&categorieLien=id.

[49] CNIL, *Decision* No. 2018-119, *March* 22, 2018.

[50] Sulle differenze tra i due sistemi cfr. R. GRIBONVAL, *D'APB à Parcoursup: quelles méthodes d'affectation post-bac?*, in *interstices.info*, 2.2.2021. Cfr. anche M. PEREL, N. ELKIN-KOREN, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement* (2017) 69 *Fla. Law Rev.*, 181.

assegnazione dei candidati, un comitato etico e scientifico ha il compito di garantire il "rispetto dei principi giuridici ed etici" su cui si basa la piattaforma Parcoursup: il comitato formula eventuali proposte atte a migliorare la trasparenza delle procedure e la loro buona comprensione da parte dei candidati.

La CNIL si concentra, però, più analiticamente sul profilo dell'intervento umano, rilevando che il dispositivo consente tale intervento sotto due aspetti: da un lato, è prevista la possibilità per un candidato di far valere, a determinate condizioni, circostanze eccezionali relative alla sua condizione di salute, disabilità, o oneri familiari e ciò può portare al riesame della sua domanda e, dall'altro, è prevista la possibilità di avviare un dialogo con il Rettore in caso di assenza di una proposta di assegnazione, al fine di vedersi proporre l'iscrizione in una formazione il più vicino possibile ai desideri espressi. Alla luce di questi elementi, la CNIL ritiene che il trattamento attuato dal Ministero dell'istruzione superiore sia conforme al quadro legislativo di riferimento, poichè tale sistema, considerato nel suo insieme, dovrebbe consentire che la decisione di assegnazione non venga presa sulla sola base del trattamento automatizzato.

Per quanto attiene ai dati trattati in Parcoursup vengono registrati i dati relativi alla formazione dei candidati, al loro *curriculum vitae* e al progetto formativo. Osserva la CNIL che gli istituti di istruzione superiore devono disporre di informazioni quali, in particolare, i risultati scolastici per valutare la necessità di proporre un sistema di sostegno educativo o la coerenza di una domanda rispetto alle aspettative richieste per una determinata formazione: alla luce di questi elementi la CNIL ritiene che la raccolta di queste informazioni sia giustificata e legittima[51].

La Commissione si sofferma, inoltre, sull'individuazione dei destinatari di Parcoursup, previsti dall'articolo 3 del progetto di decreto: accedono ad esempio a Parcoursup i soggetti autorizzati della Direzione dell'Istruzione Superiore e dell'Inserimento Professionale (DGESIP) e della Direzione Generale dell'Istruzione Scolastica del Ministero della istruzione nazionale (DGESCO). Nella misura in cui i candidati che devono ricorrere alla procedura di preregistrazione nazionale sono molto numerosi, la CNIL ritiene che non sia eccessivo

[51] Per quanto attiene ai periodi di conservazione dei dati, la bozza di decreto prevede che i dati registrati in Parcoursup siano conservati per un periodo massimo di due anni e poi inseriti in una banca dati di archivio intermedio per ulteriori quattro anni. Per quanto riguarda la durata di due anni, il Ministero ha indicato che essa dovrebbe consentire a un candidato che non sia riuscito a iscriversi a un corso di formazione o che volesse riorientarsi dopo un primo anno in una formazione di istruzione superiore, di riutilizzare i dati già inseriti. Per quanto concerne, invece, la conservazione in archivi intermedi, il Ministero ha indicato che, una volta inserite in questa banca dati, le informazioni relative all'identità e all'indirizzo dei candidati sono cancellate, quindi, fatta salva questa effettiva cancellazione e tenuto conto della natura indirettamente identificativa dei dati memorizzati, la CNIL considera tali dati pseudonimizzati e ritiene che i periodi di conservazione siano giustificati rispetto alle finalità perseguite dal trattamento.

che così tanti dipendenti, dell'amministrazione centrale e dei servizi decentrati, accedano ai dati dei candidati[52].

Per quanto attiene, infine, ai diritti delle persone interessate, l'articolo 5 del progetto di decreto esclude il diritto di opposizione: questa esclusione secondo la CNIL è conforme alla legge e si comprende nella misura in cui la preregistrazione *online* è obbligatoria ai sensi del Codice dell'istruzione.

Per quanto attiene al diritto di accesso, invece, la CNIL si limita a ricordare che nella misura in cui il trattamento utilizza un algoritmo di attribuzione, tale diritto implica di fornire tutti gli elementi che permettono di comprendere la logica alla base dell'algoritmo. Un profilo problematico della deliberazione consta, però, proprio nel fatto che la CNIL non indica quali informazioni siano fornite al fine di garantire l'esplicabilità dell'algoritmo, permanendo, quindi, problemi di trasparenza in ordine alla logica utilizzata[53].

Alla luce dei due casi esaminati emergono, quindi, alcuni profili problematici: entrambi i sistemi, APB e Parcoursup, sollevano preoccupazioni circa l'opacità in cui operano e, quindi, sulle regole e i criteri che vengono di fatto applicati per assegnare gli studenti a specifici programmi di formazione. Tali dispositivi portano, inoltre, in primo piano questioni relative al livello desiderabile di automazione nella pubblica amministrazione[54] e alla necessità di

[52] Ad accedere a Parcoursup è anche il personale degli istituti in cui vengono formati i candidati e quelli degli istituti di istruzione superiore in cui i candidati hanno espresso le proprie preferenze. Nella misura in cui tali soggetti devono rispettivamente fornire informazioni riguardanti il candidato e decidere sulle candidature, dopo averle esaminate, la CNIL ritiene che tali accessi siano giustificati.

[53] Anche sulla sicurezza dei dati e sulla tracciabilità delle azioni il parere è positivo: in primo luogo, la CNIL raccomanda che le autorizzazioni di accesso siano concesse per un periodo fisso e limitato, vengano rimosse non appena un utente non è più autorizzato e che venga eseguita regolarmente una revisione di tutte le autorizzazioni concesse. La CNIL ricorda che l'elaborazione di Parcoursup si basa su un'architettura *client-server* e avvengono numerosi scambi di dati tramite canali di comunicazione cifrati che garantiscono l'autenticazione della fonte e del destinatario. In alcuni casi, è possibile risalire alle informazioni tramite posta elettronica: la CNIL raccomanda che il contenuto e gli allegati dei messaggi vengano cifrati, utilizzando un algoritmo all'avanguardia. L'accesso al sito Parcoursup, inoltre, avviene tramite protocollo HTTPS che garantisce la riservatezza dei dati scambiati nonché l'autenticazione del responsabile del trattamento. La Commissione osserva che sono implementati meccanismi di controllo per garantire l'integrità dei dati trattati, l'accesso remoto è protetto tramite una VPN crittografata e una forte autenticazione dell'utente. Aggiornamenti *software* vengono installati regolarmente e vengono eseguiti *backup* giornalieri. La CNIL ritiene, quindi, che le misure descritte rispettino il requisito di sicurezza previsto dall'articolo 34 della legge del 6 gennaio 1978 e successive modifiche.

[54] Sul tema del *machine learning* nella pubblica amministrazione cfr. D. RESTREPO AMARILES, *Algorithmic Decision Systems*, cit., 273 ss.; C. COGLIANESE, D. LEHR, *Regulating by* robot, cit., 1161 ss. Sull'impatto delle tecnologie digitali rispetto alle funzioni pubbliche cfr. C. HARLOW, R. RAWLINGS, *Proceduralism and authomation: challanges to the values of Administrative Law*, in A. FISHER, J. KING, A. YOUNG (eds.), *The Foundations and future of Public Law (in honour of Paul Craig)*, Oxford University Press, Oxford, 2019; B.W. WIRTZ, J.C. WEYERER, C. GEYER, *Artificial intelligence and the Public sector-Applications and challenges* (2019) 42 *International Journal of Public Administration* 7, 596-615. Su IA e procedimento amministrativo nella prospettiva nazionale cfr. L. VIOLA,

garantire che i sistemi di decisione algoritmica rimangano il più aperti e responsabili possibile. Da questo punto di vista sia il codice sorgente APB che quello di Parcoursup sono stati resi pubblici, tuttavia, ciò ha reso disponibile la parte algoritmica dello strumento, ma non tutte le informazioni necessarie per comprendere l'algoritmo e la logica utilizzata, permanendo, quindi, problemi di intellegibilità ed esplicabillità dell'algoritmo medesimo[55].

8. Il riconoscimento facciale al vaglio della CNIL: elementi di continuità con la posizione dell'EDPS

Il riconoscimento facciale è uno strumento definito dal *Data Protection Working Party* come «il trattamento automatico di immagini digitali contenenti volti di individui, per scopi di identificazione, autenticazione/verifica, o categorizzazione dei suddetti individui»[56]. Questa tecnologia appartiene alla più generale categoria della biometria, progettata al fine di raccogliere, trattare e conservare dati biometrici, soggetti a una disciplina speciale ai sensi del GDPR. I suoi possibili utilizzi sono molteplici: dalla sicurezza al *marketing*, al controllo sociale, all'identificazione. Si tratta di un *software* probabilistico, ciò implica la produzione di un risultato elaborato a livello statistico e non incontrovertibile. Un sostanziale limite è il grado di errore che questo tipo di tecnologia può produrre, come nel caso di *screening* di volti femminili o di colore[57]. Nonostante le problematiche correlate al suo impiego, tanto a livello pubblico quanto a livello privato si sono verificati diversi tentativi di sperimentazione del suo utilizzo.

La CNIL è intervenuta in un caso relativo al progetto di sperimentazione di dispositivi per la scansione facciale degli studenti agli ingressi delle scuole superiori Les Eucalyptus di Nizza e Ampère di Marsiglia. Nell'ottobre 2019, l'ufficio scolastico delle Provincie di Marsiglia e Nizza ha firmato una *partnership*

L'intelligenza artificiale nel procedimento e nel processo amministrativo: lo stato dell'arte, in *Foro Amm.*, n. 9/2018, 1598 ss. Sui problemi costituzionali connessi all'amministrazione algoritmica cfr. A. SIMONCINI, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di Diritto Pubblico*, 2019, 1149 ss.

[55] In tal senso anche D. RESTREPO AMARILES, *Algorithmic Decision Systems*, cit., 286 ss.; J.A. KROLL, J. HUEY, S. BAROCAS, et al., *Accountable Algorithms*, cit., 633.

[56] Tale definizione proviene dall'Article 29 Data Protection Working Party, *Opinion n. 2/2012 on facial recognition in online and mobile services*, Brussels, 22 *March* 2012, 2. Per un'ampia disamina della regolazione del riconoscimento facciale cfr. il recente contributo di G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Editoriale scientifica, Napoli, 2021.

[57] Sul punto cfr. l'analisi di J.G. CAVAZOS, P.J. PHILLIPS, C.D. CASTILLO, A. J. O'TOOLE, *Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?*, in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, n. 3/2021, 101-111. Si veda, inoltre, R. SINGH, A. AGARWAL, M. SINGH, S. NAGPAL, M. VATSA, *On the Robustness of Face Recognition Algorithms Against Attacks and Bias*, in *AAAI Conference on Artificial Intelligence*, 2020, 13583-13589.

con *Cisco International Limited* con l'intento di installare i suddetti dispositivi nelle due scuole superiori delle rispettive città, allo scopo di identificare gli studenti.

La CNIL ha ricevuto dalla regione una richiesta di parere sulla sperimentazione del "portale virtuale" per il controllo degli accessi mediante riconoscimento facciale all'ingresso di tali due scuole superiori della regione.

Questo sistema doveva riguardare solo gli studenti che avessero precedentemente acconsentito ed essere testato per un intero anno scolastico, con lo scopo di assistere gli agenti incaricati del controllo degli accessi alle scuole superiori al fine di prevenire intrusioni, furti di identità e ridurre la durata di questi controlli.

La Commissione ha riconosciuto che il dispositivo è contrario ai fondamentali principi di proporzionalità e minimizzazione dei dati, stabiliti dal Regolamento generale sulla protezione dei dati (GDPR): la CNIL ha evidenziato che il riconoscimento facciale è una tecnologia particolarmente invasiva e non proporzionale allo scopo per cui se ne è progettato l'utilizzo[58], ritenendo che il controllo della presenza degli studenti potesse essere effettuato con delle modalità meno lesive della loro *privacy*. Secondo la CNIL questo dispositivo riguardante gli alunni, per lo più minorenni, con l'unico scopo di facilitare e garantire l'accesso non è né necessario né proporzionato rispetto ai suddetti obiettivi. L'esigenza di garantire e razionalizzare l'ingresso nelle scuole poteva essere raggiunta con mezzi molto meno invasivi, come il dispiegamento di agenti di sorveglianza aggiuntivi o il controllo del *badge*. La Commissione ha ricordato che il trattamento dei dati biometrici è particolarmente delicato, giustificando una maggiore protezione delle persone: in particolare, i dispositivi di riconoscimento facciale sono invasivi, presentano gravi rischi per la vita privata e le libertà individuali degli interessati[59] ed è anche probabile che creino

[58] Cfr. CNIL, *Expérimentation de la reconnaissance faciale dans deux lycées: la CNIL précise sa position*, 29 ottobre 2019, in https://www.cnil.fr/fr/experimentation-de-la-reconnaissance-faciale-dans-deux-lycees-la-cnil-precise-sa-position. Sul punto, si veda L. KAYALI, *French privacy watchdog says facial recognition trial in high schools is illegal*, in *Politico*, 29 ottobre 2019, https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/.

[59] Simile considerazione è espressa nella decisione del 20 agosto 2019 della autorità di controllo svedese per la protezione dei dati che ha condannato una scuola a una multa di 200000 SEK (circa 18.500 euro). La scuola aveva testato il riconoscimento facciale per tre settimane per sostituire il sistema di appello degli studenti. L'istituto non aveva precedentemente consultato l'autorità di controllo, né effettuato uno studio d'impatto preliminare. Nell'ambito di questo esperimento, la scuola aveva ottenuto il consenso di 22 studenti. A tal proposito, l'autorità ha ritenuto che il consenso non fosse la base corretta per la liceità del trattamento a causa dell'evidente squilibrio tra l'interessato e il responsabile del trattamento. L'autorità, come la CNIL, conclude che il riconoscimento facciale è un trattamento particolarmente sensibile e che questo dispositivo potrebbe violare gli articoli 5 e 9 del GDPR se l'istituzione decidesse di continuare l'esperimento. In un altro caso molto recente del febbraio 2021, l'autorità svedese per la protezione della *privacy* ha, inoltre, rilevato che l'autorità di polizia

un senso di sorveglianza rafforzata. Questi rischi aumentano quando i dispositivi di riconoscimento facciale vengono applicati ai minori che sono oggetto di una protezione speciale nella legislazione nazionale ed europea. In questo contesto e in presenza di mezzi alternativi meno invasivi, l'uso di un dispositivo di riconoscimento facciale per controllare l'accesso a una scuola superiore appare sproporzionato. La CNIL conclude, quindi, che un tale dispositivo non può essere implementato e spetta alla regione e alle scuole interessate, responsabili del dispositivo, trarne le relative conseguenze.

L'*avis* della CNIL è, quindi, un monito severo e si incentra sui principi di necessità e proporzionalità della misura rispetto agli scopi perseguiti, oggetto di uno *strict scrutiny* quando sono in gioco i minori. Si tratta di principi fondamentali come riconosciuto anche dal Comitato della Convenzione 108 nelle Linee guida sul riconoscimento facciale: il trattamento dei dati deve essere proporzionato, in termini di impatto sui diritti e sulle libertà delle persone, rispetto all'obiettivo che persegue e riguardare esclusivamente i dati "necessari" per raggiungerlo.

Sebbene la CNIL avesse giudicato negativamente l'iniziativa, la decisione è, comunque, ricaduta nelle mani dell'ufficio scolastico che ha rinviato la richiesta all'autorità regionale la quale ha autorizzato il progetto di sorveglianza, classificandolo come "sperimentale". Successivamente alcune associazioni hanno depositato un ricorso amministrativo per chiedere l'annullamento della delibera regionale che autorizzava l'attuazione della sperimentazione. Nel febbraio 2020 il Tribunale Amministrativo di Marsiglia[60] si è, infine, pronunciato in favore dei ricorrenti, soffermandosi su diversi aspetti, *in primis*, la violazione del GDPR tramite l'installazione di tali dispositivi, in quanto lesivi dell'art. 9 del Regolamento. Gli studenti, difatti, non potevano in alcun modo fornire un libero consenso al trattamento dei loro dati biometrici, vista la relazione di autorità con le amministrazioni scolastiche. Inoltre, la Corte ha richiamato la posizione espressa dalla CNIL, ribadendo come il riconoscimento facciale non

svedese ha elaborato dati personali in violazione del *Criminal Data Act* svedese quando ha utilizzato *Clearview AI* per identificare le persone. L'autorità ha concluso che la polizia non ha adempiuto ai propri obblighi di responsabile del trattamento dei dati per una serie di *account* per quanto riguarda l'uso di *Clearview AI* e non ha attuato misure organizzative sufficienti per garantire e dimostrare che il trattamento dei dati personali in questo caso era stato effettuato in conformità con il *Criminal Data Act*. Quando ha utilizzato *Clearview AI* la polizia ha elaborato illegalmente dati biometrici per il riconoscimento facciale e non è riuscita a condurre una valutazione dell'impatto sulla protezione dei dati. L'autorità ha imposto una sanzione amministrativa per violazione del *Criminal Data Act*, inoltre, ha ordinato alla polizia di rafforzare la formazione dei propri dipendenti al fine di evitare qualsiasi trattamento futuro dei dati personali in violazione delle norme e dei regolamenti sulla protezione dei dati.

[60] *Trib. Adm. de Marseille*, n. 1901249, 3 febbraio 2020. Il testo può essere consultato al link che segue: https://www.laquadrature.net/wp-content/uploads/sites/8/2020/02/1090394890_1901249.pdf. A commento cfr. F. PAOLUCCI, *Riconoscimento facciale a scuola: il caso francese*, in *iusinitinere*, 8.12.2020.

possa essere considerato una modalità proporzionale e necessaria per controllare l'accesso degli studenti all'edificio scolastico, rappresentando, altresì, una grave violazione dei diritti fondamentali degli individui coinvolti: posizione aggravata dal fatto che questi ultimi fossero per la maggior parte minori. Tali diritti sono la tutela della *privacy*, rafforzata dai principi di proporzionalità e necessità del trattamento, ed il diritto all'istruzione. Il Tribunale ha, quindi, dato seguito alla deliberazione della CNIL, ritenendo che l'utilizzo di sistemi connotati da forte invasività, quali quelli di riconoscimento facciale, sarebbe del tutto sproporzionato rispetto alla finalità di controllare gli accessi in un liceo e ha, dunque, invitato l'ufficio scolastico a prendere in considerazione modalità meno invasive per raggiungere il medesimo scopo, come controlli tramite *badge*. Qualsiasi uso di questa tecnologia, sebbene in via sperimentale, come sottolineato anche dalla CNIL, deve essere rispettoso dei principi posti dalla normativa a tutela dei dati personali, *in primis* dei principi di proporzionalità e necessità: il Tribunale ha, dunque, annullato la delibera regionale.

Anche le *Guidelines* n. 3/2019 «*on processing of personal data through video devices*» adottate dall'EDPB evidenziano che l'uso intensivo di dispositivi video ha forti implicazioni per la protezione dei dati e l'enorme quantità di dati generati, unita a nuovi strumenti tecnici per sfruttare le immagini, aumenta il rischio di utilizzo secondario. Come evidenziato anche dalla CNIL, l'EDPB mette in luce i rischi legati al possibile malfunzionamento di questi dispositivi e al *bias* che possono produrre. Alcuni studi hanno evidenziato che il *software* utilizzato per l'identificazione, il riconoscimento e l'analisi del viso funziona in modo diverso in base all'età, al sesso e all'etnia della persona, pertanto, il *bias* è uno dei principali problemi connessi a tali dispositivi[61]. I responsabili del trattamento devono valutare regolarmente la rilevanza di tali metodi di identificazione e vigilare sulle garanzie necessarie. Sulla stessa linea d'onda della CNIL, l'EDPB mette in luce l'importanza dei principi di proporzionalità e necessità, sottolineando che «*video surveillance is not by default a necessity when there are other means to achieve the underlying purpose*».

Poco dopo il caso dei licei francesi, nel novembre 2019, la CNIL ha espresso la propria posizione sul riconoscimento facciale[62] con una dichiarazione che mette in evidenza come tale tecnologia richieda scelte politiche precise su come conciliare la tutela dei diritti e delle libertà fondamentali con gli imperativi della sicurezza o questioni economiche e sulla opportunità di determinare quando il riconoscimento facciale è necessario in una società democratica e

[61] Un settore in cui la discriminazione algoritmica è molto rilevante è proprio quello della intelligenza artificiale applicata al riconoscimento facciale: si vedano i risultati del progetto "*gender shades*" promosso dal MIT disponibili su *http://gendershades.org/*.

[62] CNIL, *Reconnaissance faciale: pour un débat à la hauteur des enjeux*, 15 novembre 2019, consultabile su https://www.cnil.fr/fr/reconnaissance-faciale-pour-un-debat-la-hauteur-des-enjeux.

quando non lo è. La scelta politica non dovrebbe essere dettata puramente e semplicemente dalle possibilità tecniche: il ruolo del politico è quello di determinare, tra i possibili utilizzi di queste tecnologie, quelli realmente desiderabili. Il contributo della CNIL su questo tema persegue diversi obiettivi: *in primis*, fronteggiare il problema definitorio[63], presentando, tecnicamente, cos'è il riconoscimento facciale e a cosa serve. Questa tecnica biometrica di riconoscimento automatizzato, basata sulle caratteristiche del viso, non va confusa con altre tecniche di elaborazione delle immagini (ad esempio, con dispositivi "video intelligenti" che consentono di rilevare eventi o emozioni senza riconoscere, però, gli individui). Evidenzia, inoltre, la CNIL che dietro al riconoscimento facciale c'è un'ampia varietà di possibili utilizzi che vanno dallo sblocco di uno *smartphone* al riconoscimento di una persona ricercata dalla polizia in mezzo alla folla e questi usi sollevano problematiche diverse, soprattutto in termini di controllo delle persone sui propri dati, del loro margine di iniziativa nell'utilizzo della tecnologia e delle conseguenze che ne derivano: è, quindi necessario ragionare caso per caso. Per determinare se il trattamento dei dati personali è legale, è necessario partire dalla sua finalità: è solo per uno scopo specifico che si può valutare se i dati sono pertinenti, proporzionati, se i periodi di conservazione sono appropriati e se la sicurezza è adeguata.

La CNIL cerca, inoltre, di evidenziare i rischi tecnologici, etici e sociali associati a questa tecnologia, legati alla natura biometrica del riconoscimento facciale: i dati biometrici sono dati "sensibili" ai sensi della legislazione sulla protezione dei dati, hanno la particolarità di consentire l'identificazione dell'interessato in qualsiasi momento sulla base di una realtà biologica, permanente nel tempo. A differenza di una *password* o di un identificativo, i dati biometrici non sono, quindi, modificabili in caso di compromissione (smarrimento, intrusione nel sistema): non sono revocabili e qualsiasi appropriazione indebita o abuso di questi dati pone, quindi, rischi sostanziali significativi per la persona da cui provengono (i.e. blocco del suo accesso a servizi o luoghi, furto della sua identità a fini criminali). Suggerisce la CNIL che la memorizzazione dei dati biometrici su un supporto individuale tenuto dall'utente dovrebbe essere sempre privilegiata rispetto alle soluzioni di archiviazione centrale, al fine di

[63] La CNIL definisce il riconoscimento facciale come una tecnologia biometrica per il riconoscimento dei volti, una tecnica informatica e probabilistica che riconosce automaticamente una persona in base al suo volto, al fine di autenticarla o identificarla e colloca il riconoscimento facciale all'interno della categoria più ampia delle tecniche biometriche. Il riconoscimento facciale può svolgere due funzioni distinte: l'autenticazione per verificare che una persona sia chi dice di essere e l'identificazione per trovare un soggetto all'interno di un gruppo di individui, in un luogo, un'immagine o un *database*. In entrambi i casi, le tecniche di riconoscimento facciale si basano su una stima della corrispondenza tra i modelli e sono, da questo punto di vista, probabilistiche: dal confronto si deduce una probabilità, più o meno forte, che la persona sia proprio quella che si cerca di autenticare o identificare; se questa probabilità supera una soglia determinata nel sistema, questo considererà che c'è una corrispondenza.

ridurre al minimo i rischi connessi: è solo in caso di assoluta necessità, in mancanza di alternative, che può essere preso in considerazione lo stoccaggio centralizzato, soggetto a severe misure di sicurezza.

Inoltre, il riconoscimento facciale si basa su stime statistiche della corrispondenza tra gli elementi confrontati, su una probabilità, e non su una certezza assoluta di corrispondenza, è quindi intrinsecamente fallibile e possono scaturire conseguenze molto importanti per le persone, laddove esse non siano ben riconosciute[64], o in caso di *bias* significativo (i.e. i tassi di errore commessi dagli algoritmi di riconoscimento facciale possono variare in base al sesso o al colore della pelle). Un altro problema è che questa tecnologia consente di elaborare i dati da remoto, senza contatto, anche all'insaputa delle persone, può consentire il monitoraggio in tempo reale dei movimenti di tutti, senza interazione con la persona e può, quindi, diventare uno strumento di sorveglianza indifferenziato, onnipresente e intrusivo, riducendo l'anonimato dei cittadini nello spazio pubblico[65]. La valutazione del rischio è, quindi, necessaria per determinare quali pericoli non sono accettabili in una società democratica e quali possono essere affrontati con adeguate garanzie.

La CNIL indica tre requisiti essenziali che devono guidare l'uso di tali dispositivi per garantire il rispetto dei principi di tutela della *privacy* dei cittadini e dei loro dati personali: in primo luogo, il riconoscimento facciale, sperimentale o meno, deve essere conforme al quadro europeo, al GDPR e alla direttiva «*police justice*». La CNIL ricorda il quadro normativo imposto ai dispositivi di riconoscimento facciale: la necessità deve essere stabilita caso per caso, occorre garantire un'elevata affidabilità nella verifica dell'identità delle persone, nonché la proporzionalità dei mezzi impiegati e la protezione speciale dei minori e delle persone vulnerabili. I principi di legittimità degli obiettivi perseguiti, di stretta necessità e proporzionalità sono requisiti insuperabili: il riconoscimento facciale non può essere legalmente utilizzato, anche su base sperimentale, se non si basa sulla necessità di garantire un elevato livello di affidabilità dell'autenticazione o dell'identificazione delle persone interessate e sulla dimostrazione dell'inadeguatezza di altri mezzi meno intrusivi.

[64] Il riconoscimento facciale, come altre tecniche della stessa natura, implica, quindi, necessariamente "falsi positivi" e "falsi negativi". A seconda della qualità e della configurazione del dispositivo, i tassi di falsi positivi e falsi negativi possono variare: la scelta degli operatori nella configurazione di questi sistemi è, quindi, di fondamentale importanza.

[65] Lo spazio pubblico è un luogo in cui vengono esercitate molte libertà individuali e collettive: il diritto alla *privacy* e alla protezione dei dati personali, ma anche la libertà di espressione e di riunione, il diritto di manifestare, la libertà di coscienza: la violazione dell'anonimato, da parte delle autorità pubbliche o di organizzazioni private, rischia, quindi, di mettere in discussione alcuni di questi principi fondamentali e deve essere oggetto di una riflessione approfondita secondo la CNIL.

Si tratta, appunto, dei medesimi principi affermati dalla CNIL con riferimento al caso concernente i dispositivi di controllo degli accessi degli alunni nelle scuole[66].

In secondo luogo, occorre che il rispetto delle persone sia posto al centro di tali sistemi: dato l'impatto degli strumenti di riconoscimento facciale, il rispetto dei diritti delle persone deve essere centrale, ad esempio ottenendo il loro consenso, soprattutto nel contesto sperimentale, o garantendo loro il controllo sui propri dati e la trasparenza, fornendo informazioni chiare, comprensibili e facilmente accessibili. Devono essere garantiti i diritti di recedere dal sistema, di accedere alle informazioni e di ricorrere all'intervento umano in caso di controllo automatico.

Il terzo requisito essenziale consiste nell'adozione di un approccio autenticamente sperimentale: ciò implica in particolare una limitazione nel tempo e nello spazio di tali dispositivi, un'identificazione esatta degli obiettivi perseguiti e delle loro probabilità di successo, la definizione precisa dei loro metodi di valutazione che devono essere rigorosi, basati sul contraddittorio, multidisciplinari ed eseguiti entro un ragionevole lasso di tempo: un vero e proprio approccio sperimentale consentirà di testare e perfezionare soluzioni tecniche che rispettano il quadro normativo di riferimento[67].

Occorre osservare che anche l'*opinion* n. 4/2020 dell'EDPS è in linea con la presa di posizione della CNIL in materia. L'EDPS si sofferma sui rischi per i diritti fondamentali determinati dall'identificazione biometrica remota (RBI) e in particolare su due questioni sollevate da tali sistemi: l'identificazione (distante, scalabile e talvolta nascosta) degli individui e l'elaborazione (distante,

[66] La CNIL ricorda, appunto, che certi usi sono proibiti, come recentemente indicato per l'implementazione di sistemi di identificazione mediante riconoscimento facciale dei bambini allo scopo di controllare l'accesso alle scuole, poiché gli obiettivi di garantire e razionalizzare gli ingressi in tali istituti possono essere raggiunti con mezzi altrettanto efficaci e molto meno invadenti in termini di *privacy* e libertà individuali e tenendo conto della protezione speciale di cui dovrebbero godere i minori.

[67] Infine, la CNIL precisa la propria posizione, il proprio ruolo nella regolazione del riconoscimento facciale: è dotata dal diritto europeo e nazionale di funzioni consultive, in particolare per le autorità pubbliche, e di controllo del rispetto della legge. Per fare ciò dovrà garantire in ogni fase il rispetto delle regole specifiche stabilite dall'attuale quadro normativo per il trattamento dei dati biometrici che, quindi, si applicano a qualsiasi dispositivo di riconoscimento facciale: rigorose eccezioni al principio del divieto del trattamento di tali dati, la natura libera e informata del consenso delle persone, l'esecuzione di un'analisi di impatto prima dell'attuazione di tale trattamento al fine di limitare i rischi. La CNIL intende svolgere pienamente il suo ruolo rispetto a questa tecnologia, in particolare fornendo consulenza indipendente, contribuendo, nell'ambito delle sue competenze, alla valutazione di questi dispositivi ed esercitando, se occorre, poteri di indagine. La CNIL potrà consigliare le autorità pubbliche a monte su qualsiasi quadro sperimentale e dovrebbe essere consultata su qualsiasi progetto legislativo o regolamentare adottato per consentire o facilitare eventuali sperimentazioni; potrebbe anche essere consultata preventivamente e sistematicamente sui casi concreti di sperimentazione previsti, al fine di garantire che i progetti siano conformi al quadro giuridico sperimentale.

scalabile e talvolta nascosta) dei loro dati biometrici. Anche l'EDPS, come la CNIL, evidenzia i rischi posti per i diritti e le libertà degli individui da sistemi come il riconoscimento facciale dal vivo in luoghi pubblici e sottolinea la necessità di identificare adeguatamente tali pericoli e mitigarli. In linea con quanto affermato dalla CNIL alcuni dei rischi derivano dal fatto che i sistemi RBI sono nascosti, spesso vengono presentati come semplici "esperimenti", ma potrebbero facilmente essere trasformati in complessi di sorveglianza onnipresenti e pervasivi. Una volta che l'infrastruttura che supporta l'RBI è in atto, inoltre, può essere facilmente utilizzata per altri scopi ("*function creep*") e tali *function creeps* dovrebbero, quindi, essere adeguatamente affrontati in qualsiasi regolamentazione sull'intelligenza artificiale. Come evidenziato dalla CNIL, inoltre, è della massima importanza valutare se la tecnologia è necessaria e proporzionata nella situazione rilevante in cui verrà impiegata[68].

In linea con tali orientamenti della CNIL e dell'EDPS occorre osservare che nella proposta di regolamento sull'IA i sistemi di identificazione biometrica remota sono soggetti a requisiti rigorosi come, appunto, i principi di stretta necessità e proporzionalità. Con particolare riferimento all'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, tale uso è vietato a meno che e nella misura in cui esso sia strettamente necessario per la persecuzione di una serie di obiettivi di sicurezza pubblica e prevenzione della criminalità e deve rispettare le garanzie e le condizioni necessarie e proporzionate in relazione all'uso, in particolare per quanto riguarda i limiti temporali, personali e geografici, previa autorizzazione di un'autorità giudiziaria o di un'autorità amministrativa indipendente dello Stato membro in cui deve avvenire l'uso.

Alla luce del caso concernente i due licei francesi, della dichiarazione della CNIL sul riconoscimento facciale e della *opinion* dell'EDPS emerge, pertanto, una "via" per la regolazione dei dispositivi di riconoscimento facciale che la proposta di regolamento europeo sull'IA sembra seguire: una via basata sui principi di stretta necessità e proporzionalità del trattamento rispetto agli obiettivi perseguiti, su una limitazione nel tempo, su un approccio precauzionale e sperimentale rispetto all'uso di tale tecnologia, improntato alla massima tutela dei diritti fondamentali. Il riconoscimento facciale può essere utilizzato, anche su base sperimentale, solo se si basa sulla necessità di garantire un elevato livello di affidabilità dell'autenticazione o dell'identificazione delle persone interessate e sulla dimostrazione dell'inadeguatezza di altri mezzi meno intrusivi.

[68] L'EDPS sostiene, inoltre, l'idea di una moratoria sull'applicazione nell'UE del riconoscimento automatizzato negli spazi pubblici delle caratteristiche umane, non solo dei volti ma anche dell'andatura, delle impronte digitali, del DNA, della voce, di altri segnali biometrici o comportamentali, in modo che possa aver luogo un dibattito informato e democratico sul tema.

9. Considerazioni conclusive

Nella costruzione di questo nuovo "costituzionalismo algoritmico europeo" gli atti e i casi problematici analizzati consentono di individuare alcuni indirizzi utili per la definizione di una "legalità algoritmica", quindi, per la regolazione delle decisioni algoritmiche e dell'IA: le linee di fondo di questo nuovo costituzionalismo emergono dai principi volti a garantire la trasparenza degli algoritmi e la loro controllabilità. Tali principi sono affidati non soltanto all'applicazione giurisdizionale, ma anche alla attività delle autorità amministrative indipendenti nella dimensione europea in ragione della loro natura, in quanto soggetti chiamati a fronteggiare l'innovazione tecnologica e legati sia al contesto nazionale sia a quello sovranazionale, veri e propri "organi di trasmissione" del diritto europeo negli ordinamenti interni[69].

In tale contesto le autorità indipendenti in materia di protezione dei dati nella dimensione europea svolgono un importante ruolo nel chiarire la portata della disciplina vigente e nel formulare indirizzi utili per il legislatore europeo e l'EDPS, in particolare, sembra orientare anche le autorità nazionali nella prospettiva di un'attuazione uniforme del quadro normativo europeo.

Dagli atti esaminati emerge, innanzitutto, l'importanza di prevedere definizioni chiare e univoche di IA, algoritmo, *machine learning*, profilazione, processo decisionale automatizzato. Da questo punto di vista le linee guida del gruppo articolo 29 forniscono definizioni chiare di processo decisionale automatizzato e profilazione e sono, altresì, molto utili per declinare i diritti dell'interessato rispetto al trattamento di cui all'art. 22 GDPR, in particolare il diritto ad essere informato e alla rettifica.

Dalla indagine emerge, altresì, la necessità di prevedere sistemi di verifica del funzionamento degli algoritmi utilizzati e una definizione chiara dell'ambito applicativo di un nuovo quadro giuridico per l'IA (cfr. *opinion* n. 4/2020).

Come evidenziato sia dall'EDPS sia dalla CNIL, inoltre, un futuro quadro per l'IA dovrebbe guardare anche ad una valutazione degli impatti collettivi di tali sistemi e non soltanto alle loro implicazioni in una prospettiva individuale. In particolare, i due aspetti degli effetti dell'IA sulle decisioni individuali e sulle decisioni generali e astratte sono strettamente collegati tra loro e ciò incide anche sul ruolo delle autorità garanti nella regolazione delle decisioni algoritmiche[70]. Muovendo dal presupposto che gli effetti degli algoritmi non si esauriscono sul piano dei diritti individuali, ma incidono ad esempio anche sulla tenuta della forma di stato democratica, ne discende il ruolo fondamentale dei sistemi di regolazione di tipo "misto", di co-regolazione, perché in grado di coniugare da un lato la responsabilizzazione dei gestori delle piattaforme

[69] In tal senso cfr. anche E. CHELI, *ult. cit.* e G. TARLI BARBIERI, *ult. cit.*

[70] Come ben evidenzia A. CARDONE, *ult. cit.*

rispetto alla protezione dei diritti individuali e dall'altro il controllo pubblicistico sanzionatorio anche a presidio delle conseguenze che le violazioni di diritti individuali possono avere su interessi generali della collettività come la democraticità dell'ordinamento[71]. La disciplina legislativa può porre degli argini la cui violazione deve poter essere sanzionata da parte delle autorità garanti proprio perché sono in gioco non soltanto diritti individuali, ma anche istanze generali come la tenuta della forma di stato democratica.

Ciò che emerge, inoltre, dalla *opinion* dell'EDPS sul libro bianco della Commissione europea è la necessità di un approccio solido per delineare il livello di rischio dei sistemi di IA, secondo i criteri previsti nelle linee guida del Comitato europeo per la protezione dei dati, con particolare riguardo ad attività di valutazione o punteggio, monitoraggio sistematico, dati sensibili, dati relativi a soggetti vulnerabili.

In una prospettiva *de iure condendo* sembra auspicabile anche la previsione di garanzie minime e di una valutazione d'impatto per tutte le applicazioni di IA a prescindere dal livello di rischio: la valutazione di impatto dovrebbe includere una valutazione dei rischi per i diritti e le libertà degli individui ed è, quindi, un mezzo utile per anticipare eventuali effetti discriminatori. Tra le garanzie minime da apprestare si potrebbero prevedere misure tecniche e organizzative trasparenti sugli obiettivi, l'uso e la progettazione di sistemi algoritmici e la garanzia della loro robustezza. La previsione di garanzie minime per tutte le applicazioni di IA a prescindere dal livello di rischio non sembra, però, recepita nella proposta di regolamento sull'IA che prevede una serie di garanzie dettagliate limitatamente ai sistemi di IA ad alto rischio.

Nella prospettiva di una futura regolazione occorre, inoltre, rafforzare la tutela degli interessati là dove siano impiegati mezzi automatizzati per la pubblicità mirata *online*, la moderazione dei contenuti e i sistemi di raccomandazione. Oltre alle misure volte a garantire la trasparenza, è necessaria anche una previsione normativa chiara delle ipotesi che consentono di impiegare processi automatizzati e il trattamento dei dati personali per rilevare, identificare e fronteggiare contenuti illegali, secondo il principio di stretta necessità della profilazione a tal fine. La moderazione dei contenuti dovrebbe, per quanto possibile, non comportare alcun trattamento di dati personali e le misure di moderazione dovrebbero essere necessarie oltrechè proporzionate ai fini perseguiti.

L'EDPS ha, inoltre, indicato misure concrete che vadano oltre la trasparenza in caso di pubblicità mirata, compreso un graduale divieto di pubblicità mirata sulla base del tracciamento pervasivo e la considerazione di limitazioni in relazione alle categorie di dati che possono essere elaborate a fini mirati e che possono essere comunicate agli inserzionisti, nonché l'informazione agli

[71] In tal senso cfr. le riflessioni di A. CARDONE, *ult. cit*.

interessati se l'annuncio pubblicitario è stato selezionato utilizzando un sistema automatizzato. Del pari i sistemi di raccomandazione dovrebbero di *default* non essere basati sulla profilazione e dovrebbero essere previsti stringenti obblighi informativi nel caso in cui il meccanismo di raccomandazione sia un sistema decisionale automatizzato.

Dal report della CNIL si possono, inoltre, desumere i principi fondamentali che dovrebbero regolare la materia: il principio di una *fair* IA che in una prospettiva *de iure condendo* potrebbe essere garantito mediante la previsione di due distinti pareri di valutazione della *fairness* dell'algoritmo da parte di società di *audit* che diano adeguate garanzie di imparzialità e indipendenza per fronteggiare il "*black box effect*"; il principio dell'attenzione e vigilanza continue e dell'intellegibilità dell'algoritmo in forza del quale la logica deve essere spiegata a parole anziché in codice e il principio dell'intervento umano sul quale si sono incentrate le decisioni della CNIL su alcuni casi problematici analizzati. Infine, alcune raccomandazioni potrebbero essere utili per la futura regolazione dell'IA: favorire l'educazione di tutti i soggetti coinvolti in catene algoritmiche sui profili etici sollevati dall'IA; rendere più comprensibili i sistemi algoritmici, vincolando i responsabili all'obbligo di comunicare informazioni in un modo che consenta di comprendere la logica coinvolta, anche per algoritmi che non elaborano dati personali, nella misura in cui possono avere impatti collettivi significativi; migliorare la progettazione di tali sistemi, introducendo il fondamentale principio di testabilità ed, infine, rafforzare l'etica nelle imprese, ad esempio creando comitati etici all'interno delle aziende che utilizzano algoritmi con impatti di vasta portata. Occorre, altresì, osservare che l'esigenza di garantire la trasparenza dei sistemi di IA, posta in evidenza dalla CNIL, sembra essere stata in parte recepita dal legislatore europeo nella proposta di regolamento sull'intelligenza artificiale là dove è prevista una disciplina dettagliata degli obblighi di trasparenza anche se limitatamente ai sistemi ad alto rischio: gli indirizzi della autorità francese sono stati in parte ascoltati dal legislatore europeo in punto di trasparenza dei sistemi di intelligenza artificiale.

Sul riconoscimento facciale la CNIL è, inoltre, molto chiara nel dire che tali dispositivi devono essere conformi al quadro europeo e, in particolare, al principio di necessità da valutare caso per caso, alla proporzionalità dei mezzi impiegati e alla protezione speciale dei minori e delle persone vulnerabili. In linea con l'*opinion* n. 4/2020 dell'EDPS, i principi di legittimità degli obiettivi perseguiti, di proporzionalità e di necessità del trattamento sono ritenuti requisiti indefettibili: il riconoscimento facciale non può essere utilizzato, anche su base sperimentale, se non si fonda sulla necessità di garantire un elevato livello di affidabilità dell'autenticazione o dell'identificazione delle persone interessate e sulla dimostrazione dell'inadeguatezza di altri mezzi meno intrusivi. Inoltre, occorre che il rispetto delle persone e dei loro diritti sia posto al centro di tali

dispositivi, ad esempio ottenendo il loro consenso, soprattutto nel contesto sperimentale, o garantendo loro il controllo sui propri dati e la trasparenza, ed è, altresì, necessario un approccio sperimentale che implica, in particolare, una limitazione nel tempo e nello spazio di tali dispositivi, un'identificazione esatta degli obiettivi perseguiti e delle loro probabilità di successo e la definizione precisa dei loro metodi di valutazione.

Si tratta all'evidenza di orientamenti utili e molto precisi, desumibili da "atti" spesso formalmente non vincolanti di autorità indipendenti garanti della protezione dei dati personali nella dimensione europea: tali autorità stanno indicando una via per la futura regolazione giuridica dell'IA e delle decisioni algoritmiche e stanno, quindi, assumendo un fondamentale ruolo di "guida" e "pungolo" verso il legislatore nazionale ed europeo per rafforzare i diritti, le libertà individuali e collettive dinanzi all'uso delle nuove tecnologie.