

Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano

di Alberto Oddenino

Abstract: Algorithmic decision making and international perspectives of valorization of human intervention – This contribution investigates the effectiveness as well as the actual and potential role of human intervention in algorithmic decision-making. Moving from an assessment of the ever-increasing power of algorithms, coped with the peculiarities of algorithmic decision making when compared to human reasoning, the analysis focuses on the right not to be subject to a purely automatic algorithmic decision. This right, formally provided in its legal dimension in art. 22 GDPR, is evaluated in terms of effectiveness as well as in its potential evolution beyond its current scope of application as a global standard for algorithmic decision making. A more explicit acknowledgement seems important as a tool to protect human autonomy in the ongoing process of establishment of an international regulatory framework for artificial intelligence.

199

Keywords: Artificial Intelligence; Automated Decisions; Human Intervention; Human Rights; Human Autonomy.

1. Introduzione: l'affermazione crescente del potere algoritmico

La rivoluzione digitale basata sulla elaborazione algoritmica dei dati incide profondamente su ogni profilo della nostra società, tanto in dimensione individuale quanto in dimensione collettiva. Si tratta, come è stato ben argomentato, di una nuova era della storia umana¹ che, se è capace di portare grandi benefici in campo non solo economico², appare al contempo suscettibile di modificare profondamente i processi decisionali, innovandone le logiche e rimodulando i rapporti di potere,

¹ Sul punto il riferimento è al concetto di 'Iperstoria' come nuova era dello sviluppo umano, che è descritta in L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, 2017, spec. 1-26 e 99-111, ove l'accento è posto sulla progressiva perdita di centralità dell'uomo nel corso delle successive rivoluzioni della storia umana: copernicana, darwiniana, freudiana e, da ultimo, ad opera di quella che l'autore definisce la 'quarta rivoluzione' basata sull'intelligenza artificiale, che fa dell'uomo un organismo informazionale. E' su queste premesse che il contributo intende innestarsi per proporre, attraverso una valorizzazione di precetti giuridici, il recupero almeno parziale di una dimensione maggiormente 'umanista'.

² Il termine 'economia globale dei dati' è divenuto di uso comune: non a caso già da tempo il *World Economic Forum* ha anche classificato i dati come fattori di produzione (*World Economic Forum, Report 2011: Personal Data, The Emergence of a New Asset Class*). In effetti, e a ben vedere, i dati, il loro flusso e la loro aggregazione sono essenziali per l'economia digitale e dell'informazione dal momento che rappresentano la principale opportunità post-industriale di crescita economica.

con un importante impatto non solo sulla privacy e sui diritti umani, ma anche, e più in generale, sul modello economico e di sviluppo umano³.

Sullo sfondo di questo tema problematico, vi è la natura stessa degli algoritmi e la questione relativa al loro potere e alla loro pretesa neutralità. Il tema, vasto e variegato, può essere affrontato da molte prospettive anche extra-giuridiche. Se è vero, tuttavia, che attraverso gli algoritmi si evidenziano connessioni e si determinano inferenze assai rilevanti per adottare decisioni suscettibili di incidere significativamente sulle sfere individuali, il tema non può non rientrare pienamente nel perimetro di azione del diritto⁴. Ciò è ancor più vero se si considera che gli algoritmi incorporano scelte non scontatamente neutrali, che segnano la fase della loro concezione e perpetuano, talora amplificandoli, i limiti cognitivi e valoriali dei dati su cui si innesta la loro operatività. Ciò rimanda al problema, delicato ed ampio, di ripensare lo stesso concetto di discriminazione all'epoca degli algoritmi⁵ e di elaborare strumenti per contrastare specificamente la discriminazione algoritmica⁶. Questo tema deve poi essere letto in connessione con il sostrato materiale di operatività degli algoritmi, ossia il fenomeno Big Data, tema a sua volta vasto e complesso su cui non indugeremo in questa sede, facendo rinvio ad ampia letteratura in materia che dà conto della dimensione e pervasività del fenomeno, nonché delle esigenze di una sua efficace regolazione giuridica⁷.

Ciò che interessa sottolineare, nell'economia del presente lavoro, è che il potere algoritmico si appalesa come potere vieppiù basato su una automaticità di decisione che, se è suscettibile di incidere fortemente sulle sfere individuali, presenta tuttavia caratteristiche di difficile contestabilità e quindi di altrettanto difficile reversibilità⁸. La sfida, che si presenta come giuridica e tecnologica

³ Su questi ultimi profili cfr. la penetrante critica di S. Zuboff, *The Age of Surveillance Capitalism. The Fight for a human future at the new frontier of power*, New York, 2019, in cui sono evidenziate molte delle problematiche indotte da quello che è efficacemente definito come 'modello capitalistico di sorveglianza'.

⁴ In tema si veda da ultimo l'illuminante saggio di M. Durante, *Potere computazionale. L'impatto delle ICT su diritto, società, sapere*, Milano, 2019, ove l'analisi muove proprio dal presupposto di una crescente tendenza a delegare decisioni e compiti ad agenti artificiali.

⁵ Sul punto cfr. di recente l'ampia ricostruzione, in termini di rischi e di opportunità, condotta in J. Kleinberg, J. Ludwig, S. Mullainathan, C.R. Sunstein, *Discrimination in the Age of Algorithms*, NBER Working Paper n. 25548, February 2019.

⁶ In tema cfr. di recente P. Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law*, in 55 *CMLR* 1143 (2018).

⁷ In questo senso cfr. già A. Mantelero, *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, in *Il diritto dell'informazione e dell'informatica*, 2012, 135 ss. Per una prospettiva internazionalistica in relazione alla tutela dei diritti cfr. l'ampio studio di G. Della Morte, *Big Data e protezione internazionale dei diritti umani*, Napoli, 2018. In tema si veda altresì in generale C. Focarelli, *La Privacy. Proteggere i dati personali oggi*, Bologna, 2015 e più specificamente il recente contributo di Fernanda Faini in questa rivista (F. Faini, *Big data, algoritmi e diritto*, in *DPCE online*, 2019, 1869 ss.). In tema di Big Data cfr. altresì V. Zeno Zencovich, *Ten Legal Perspectives on the "Big Data Revolution"*, Napoli, 2017 e sia consentito altresì rinviare alla *Special Issue: Data Flow, Big Data and International Economic Law* (a cura di A. Oddenino) in *Diritto del Commercio internazionale*, 2017, 775-879, che compendia quattro contributi dedicati a vari profili di rilevanza del fenomeno *Big Data* nella prospettiva del diritto internazionale economico, specificamente focalizzati sull'impatto sistemico, sul tema del flusso transfrontaliero, sul problema della proprietà e sull'impatto in materia di diritto della concorrenza.

⁸ Sul punto cfr. M. Durante, *Potere Computazionale*, cit., p. 231 ss.

insieme, è pertanto quella di predisporre, in chiave preventiva o rimediale, meccanismi idonei a bilanciare il portato automatico del potere algoritmico con le istanze di tutela dell'umano. Per questo il contributo suggerisce di concentrarsi sul diritto all'intervento umano, e sulla sua attuale normativizzazione, con lo scopo di indagare le prospettive di riequilibrio del potere algoritmico proprio attraverso una valorizzazione dell'intervento umano che vada al di là dello spazio normativo che è ad oggi ad esso riconosciuto .

Muovendo da una preliminare considerazione delle peculiarità dei processi decisionali algoritmici rispetto al ragionamento umano, l'analisi si concentra sul diritto a non essere sottoposti a decisioni algoritmiche totalmente automatizzate. Tale diritto è vagliato nell'effettività della sua portata, attualmente formalizzata all'art 22 GDPR, ma anche nella sua possibile proiezione come standard, non solo europeo ma globale, in materia di decisioni algoritmiche. Per questo viene approfondita, nella parte conclusiva, la possibilità che il diritto all'intervento umano sia declinato in modo più ampio come volano a tutela del principio di autonomia umana, anche nel più ampio quadro degli attuali tentativi di regolazione internazionale dell'intelligenza artificiale.

2. La peculiarità dei processi decisionali algoritmici rispetto al ragionamento umano

La definizione di algoritmo in termini di *'qualunque schema o procedimento matematico di calcolo; più precisamente, un procedimento di calcolo esplicito e descrivibile con un numero finito di regole che conduce al risultato dopo un numero finito di operazioni, cioè di applicazioni delle regole'*⁹ è certamente suscettibile di abbracciare il fenomeno in tutta la sua ampiezza e nella moltitudine delle sue manifestazioni¹⁰, che comprendono svariate, e più o meno sofisticate, forme di intelligenza artificiale¹¹.

Con particolare riferimento ai processi decisionali automatizzati, rilevano i cd. algoritmi predittivi, basati su funzioni di 'apprendimento automatico' (*machine learning*): con tale termine si intende ogni strumento che sia in grado di impiegare dati esistenti per trarne non solo conoscenza attuale ma anche modelli e capacità di previsione¹². L'elemento qualificante dell'algoritmo basato sul *machine learning*

⁹ La definizione tratta dal vocabolario Treccani (www.treccani.it/vocabolario/algoritmo/) sembra sottolineare, e non appare un caso per quanto si analizzerà criticamente di seguito, la definitezza e quindi conoscibilità dei vari passaggi computazionali.

¹⁰ A proposito della varietà e vaghezza definitoria, cfr. R.K. Hill, *What an algorithm is*, in 29 *Philosophy & Technology*, 35, 36 (2015).

Sulla pretesa neutralità e sulla correlata distinzione fra *«managed, policy-directed algorithms»* e *«policy-neutral algorithms»*, con i secondi che offrirebbero risultati non manipolati né orientati a valori predeterminati, cfr. O. Tene, J. Polonetsky, *Taming The Golem: Challenges of Ethical Algorithmic Decision-Making*, in 19 *North Carolina Journal of Law & Technology*, 125, 132 (2018).

¹¹ Anche il termine 'intelligenza artificiale' può avere varie definizioni e declinazioni. In questa sede ci si può per semplicità riferire alla «science of making machines do things that would require intelligence if done by persons» (cfr. S. Chopra, L.F. White, *A Legal Theory of Autonomous Artificial Agents*, University of Michigan Press, 2011, 5).

¹² Cfr. M. Van Otterlo, *A machine learning view on profiling*, in M. Hildebrandt, K. de Vries (eds.) *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology*, Abingdon, 2013, 41 ss.

è pertanto l'attitudine ad elaborare modelli predittivi a partire da dati storici, impiegando al contempo il modello in chiave decisionale rispetto a dati nuovamente acquisiti¹³.

Ai fini della presente indagine occorre considerare i processi decisionali algoritmici come volti, in particolare anche se non esclusivamente, alla c.d. «profilazione»¹⁴: tale attività, nata con finalità commerciali di personalizzazione delle pubblicità o dei risultati di ricerca in relazione ai prodotti disponibili in Rete, ha ampiamente travalicato detti confini, proiettandosi in sfere che coinvolgono diritti individuali ben più fondamentali nonché valori pubblicistici¹⁵. È evidente come la profilazione sia resa sempre più capillare in ragione dell'aumento esponenziale del volume dei dati personali raccolti, conservati e trasferiti attraverso servizi e applicazioni *online*, e come essa si faccia forza di una strutturale asimmetria informativa e della capacità di aggregare dati personali e dati non personali per potenziare la propria capacità predittiva¹⁶.

Su questi presupposti, è chiaro che il successo e la crescente diffusione dei processi decisionali automatizzati basati su algoritmi è certo dovuto alla loro tendenziale maggiore efficienza rispetto alle capacità analitiche umane: gli algoritmi predittivi sono infatti in grado di analizzare quantità elevatissime di dati, anche fra loro eterogenei sotto il profilo qualitativo, individuandone preziose correlazioni che sfuggirebbero alla mente umana. Siamo, come già anticipato, nella dimensione dei cd. *Big Data*, che determinano un salto che è al contempo quantitativo e qualitativo poiché la crescita esponenziale di informazioni conduce a inferenze che nessuna valutazione umana sarebbe in grado di comprendere¹⁷. Ciò evidenzia quella che è definibile come una “ontologica differenza” fra l'attività decisionale degli algoritmi e la decisione umana¹⁸, un profilo che come si vedrà è qualificante ai fini della presente indagine: i rapporti, le correlazioni, le inferenze e i *pattern* comportamentali, utilizzati dall'algoritmo con finalità predittive e

¹³ Cfr. I. Zliobaite, *Measuring discrimination in algorithmic decision making*, in 31 *Data Mining and Knowledge Discovery*, 1060, 1067 (2017).

¹⁴ Occorre peraltro precisare che, come non ogni processo decisionale algoritmico necessariamente comporta profilazione, così non ogni attività di profilazione comporta un processo decisionale algoritmico né è pertanto soggetto alle norme che a tale processo si applicano.

¹⁵ Cfr. in questo senso B. van der Sloot, *Decisional privacy 2.0: the procedural requirements implicit in Article 8 ECHR and its potential impact on profiling*, in 7 *International Data Privacy Law*, 190, 200 (2017).

¹⁶ In questo senso, come è stato correttamente notato, può venire meno anche una precisa distinzione fra dati personali e non personali e possono risultare inefficaci le tutele approntate in relazione ai dati personali, quali in particolare il principio del consenso al trattamento o la limitazione di scopo nel trattamento stesso. In tema cfr. F. Faini, *Big data, algoritmi e diritto*, cit., 1877, e G. Ruotolo, *I dati non personali: l'emersione dei big data nel diritto dell'Unione europea*, in *Studi sull'integrazione europea*, 2018, 97 ss. spec. 108.

¹⁷ Cfr. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi, *The ethics of algorithms: Mapping the debate*, in 3 *Big Data & Society*, 3 (2016), ove si osserva: «Analytics demonstrates how algorithms can challenge human decision-making and comprehension even for tasks previously performed by humans».

¹⁸ In linea con questa ricostruzione, non deve sorprendere l'inquietante assimilazione che si ritrova in F. Pasquale, *The Black Box Society: The Hidden Algorithms That Control Money and Information*, Cambridge, 2015, 187, secondi cui «the perfect search engine would be like the mind of God».

decisionali¹⁹, segnano un incremento verticale dell'elemento quantitativo che implica, in tema di probabilità, un salto qualitativo sostanziale rispetto alla previsione umana²⁰.

Non deve a questo punto sorprendere la crescente diffusione e ubiquità dei processi decisionali algoritmici: lo sviluppo dell'intelligenza artificiale ha seguito in questo senso un percorso, quello che la proietta in una dimensione cibernetica immateriale e capillarmente diffusa, che è almeno in parte diverso da quella in passato ipotizzato²¹. Ciò, se da un lato ne ha resi meno visibili lo sviluppo e la diffusione, ne ha per converso ampliato in modo considerevole l'impatto, nella misura in cui gli algoritmi sono divenuti una mediazione sostanzialmente necessaria di ogni attività umana, non solo informativo-cognitiva, ma anche di interazione con l'ambiente circostante²².

Per questo la diffusione dei processi decisionali algoritmici ha destato crescente attenzione nei tempi recenti²³, e non solo in ambito strettamente giuridico²⁴: la proliferazione delle decisioni algoritmiche e l'inarrestabile allargamento del loro campo di azione hanno infatti accresciuto anche le preoccupazioni circa le possibili ricadute che sembrano potere derivare dallo sviluppo di modelli basati su una sempre più intensa e pervasiva *governance* algoritmica²⁵.

¹⁹ In sintesi, tale processo viene così descritto in D. K. Citron, F. Pasquale, *The scored society: Due process for automated predictions*, in 89 *Washington Law Review*, 1, 3 (2014): «predictive algorithms mine personal information to make guesses about individuals' likely actions and risks. A person's on- and offline activities are turned into scores that rate them above or below others. Private and public entities rely on predictive algorithmic assessments to make important decisions about individuals».

²⁰ In questo senso cfr. K. Brennan-Marquez, "Plausible Cause": *Explanatory Standards in the Age of Powerful Machines*, 70 *Vanderbilt Law Review*, 1249, 1258 (2017), ove si osserva che «probability is about predictive likelihood. Past observations can be extrapolated to new data (...) machine learning has intensified the process. Today, prediction is more powerful—and possible in more domains—than ever before».

²¹ In questo senso valgono le illuminanti osservazioni in B. Bodo, N. Helberger, K. Irion, F. Zuiderveen Borgesius, J. Moller, B. van Es, C. de Vreese (eds.), *Tackling the Algorithmic Control Crisis - The Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents*, in 19 *Yale Journal of Law and Technology*, 133, 171 (2018), ove si osserva in modo assai pertinente che le attuali forme di intelligenza artificiale «are not the huge, shiny, metal robots, or standalone machines with superhuman strength who have the potential to cause physical harm to humans in various local contexts» e che gli agenti algoritmici operano in modo immateriale ed interconnesso e «rather than being engaged in local, one-on-one interactions with individuals, they operate on planetary scale, simultaneously affecting the lives of billions».

²² Cfr. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi *The ethics of algorithms*, cit., 1; M. Durante, *Potere computazionale*, cit., 17 ss.

²³ Peraltro sul punto cfr. già J.R. Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society*, Cambridge, 1989, e P. Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, in 43 *Hastings Law Journal*, 1321 (1992).

²⁴ *Ex multis* cfr. M. Ziewitz, *Governing algorithms: Myth, mess, and methods*, in 41 *Science, Technology, & Human Values*, 1 (2015); T. Gillespie, *The politics of 'platforms'*, in 12 *New Media & Society*, 347 (2010); T. Zarsky, *The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making*, in 41 *Science, Technology, & Human Values*, 118 (2015); K. Crawford, J. Schultz, *Big data and due process: Toward a framework to redress predictive privacy harms*, in 55 *Boston College Law Review*, 93 (2014).

²⁵ Le problematiche connesse a quella che viene talvolta definita come vera e propria 'algcrazia' (così J. Danaher, *The threat of algocracy: Reality, resistance and accommodation*, in 29

L'idea che ha preso corpo è che l'utilizzo di algoritmi non possa limitarsi a garantire una maggiore efficienza del processo decisionale ma, piuttosto, che sia al tempo stesso necessario garantire che l'esito del processo decisionale sia corretto, legittimo ed effettivamente accurato²⁶. Questo anche e soprattutto sotto il profilo del rispetto della dimensione individuale e dei diritti dei soggetti che sono formali destinatari o che comunque si trovano a subire gli effetti di tali decisioni²⁷.

Sebbene sia pertanto generalmente riconosciuta la necessità di una sorta di 'accountability by design'²⁸ nella programmazione e messa in opera degli algoritmi predittivi, tale risultato non è facilmente ottenibile nella pratica, per via della menzionata "differenza ontologica" fra ragionamento umano e processo decisionale algoritmico, che, come si è visto, ha come essenza caratterizzante proprio l'attitudine a trascendere la comprensione umana²⁹.

204

È specificamente da questo punto di vista che rivelano notevoli limiti i tentativi di regolare il potere algoritmico facendo ricorso ad un dovere di trasparenza o al concetto di spiegabilità. Si tratta di due prospettive, cui in questa sede si può solo fare cenno, che operando per così dire dall'interno della logica algoritmica scontano proprio il limite della differenza ontologica di cui abbiamo riferito.

La trasparenza algoritmica appare per questo in certa misura come una soluzione ingenua³⁰, che solleva problemi già su un piano intuitivo, ossia quello della piena accessibilità dei codici algoritmici. Questo ancor più in ragione del fatto che spesso detti codici sono oggetto di una qualche forma di privativa intellettuale, e ciò di per sé implica un ulteriore limite strutturale all'obbligo di svelarne pienamente il contenuto³¹.

Philosophy and Technology, 245 (2016)) rappresentano peraltro un fenomeno che deve necessariamente essere studiato e affrontato in prospettiva multidisciplinare, in cui i profili giuridici si colleghino non solo con quelli tecnologici ed economici ma anche etici e filosofici.

²⁶ L'efficienza degli algoritmi non può evidentemente essere disgiunta da una più ampia prospettiva di trasparenza e correttezza. Così B. Goodman, S. Flaxman, *EU regulations on algorithmic decision-making and a "right to explanation"*, in 38 *AI Magazine*, 7 (2017).

²⁷ In tal senso cfr. J. Danaher, M.J. Hogan et al., *Algorithmic governance: Developing a research agenda through the power of collective intelligence*, in 1 *Big Data & Society*, 1, 2 (2017) e T. Zarsky, *The trouble with algorithmic decisions* cit., 119.

²⁸ Cfr. J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, *Accountable Algorithms*, in 165 *University of Pennsylvania Law Review*, 633, 639 (2017).

²⁹ Cfr. A.D. Selbst, J. Powles, *Meaningful information and the right to explanation*, in 7 *International Data Privacy Law*, 233 (2017).

³⁰ Così J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, *Accountable Algorithms*, cit., 657 ove si legge: «to demand transparency of the source code as well as inputs and outputs for the relevant decisions (...) is a naive solution to the problem of verifying procedural regularity».

³¹ Sul punto cfr. Bodo, N. Helberger, K. Irion, F. Zuiderveen Borgesius, J. Moller, B. van Es, C. de Vreese (eds.), *Tackling the Algorithmic Control Crisis*, cit., 142. Cfr. altresì G. Noto La Diega, *Against the Dehumanisation of Decision-Making: Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection and Freedom of Information*, in 9 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 3 (2018), in cui è ben sottolineato il ruolo necessariamente concorrente dei tre plessi normativi (proprietà intellettuale, protezione dei dati e libertà di informazione) nel dare forma, e porre limiti, al fenomeno delle decisioni algoritmiche.

A ben vedere poi, e anche a prescindere da questi profili, la ragione principale e per così dire strutturale per cui tale soluzione non sembra adeguata risiede nelle modalità di funzionamento degli algoritmi predittivi, basati su logiche di autoapprendimento che prevedono la modifica della propria struttura dopo ogni decisione o analisi di dati³². Ciò rende arduo non solo per i soggetti che subiscono gli effetti delle decisioni algoritmiche, ma anche per gli stessi soggetti che tali algoritmi hanno originariamente creato, affidarsi ad un efficace funzionamento del principio di trasparenza.

Limiti non dissimili si ritrovano in relazione alla possibilità di un controllo a posteriori del contenuto delle decisioni algoritmiche dalla prospettiva della 'spiegabilità' o 'comprensibilità', rimedio che, pur volto a superare le aporie dell'opzione appena esposta relativa alla trasparenza, risulta in certa misura in continuità con essa e ad essa complementare. La comprensibilità si propone infatti di reindirizzare l'attenzione dalle modalità mediante cui si svolge il processo decisionale algoritmico verso gli esiti di quest'ultimo³³, che rappresentano del resto il profilo autenticamente problematico delle decisioni automatizzate in relazione ai diritti individuali³⁴.

Peraltro lo stesso diritto ad una spiegazione presenta rilevanti limiti che, in essenza, risiedono nelle variabili nascoste che possono minare la dimostrazione di accuratezza che sia raggiunta, anche in sede probatoria, allorché l'algoritmo sia testato attraverso scenari controfattuali volti a corroborarne, in modo comprensibile, l'affidabilità³⁵.

Questo quadro dà ben conto di come la stessa nozione di discriminazione algoritmica sfugga ai classici strumenti del diritto antidiscriminatorio. Per questo, come è stato ben argomentato, tali strumenti sono spesso almeno in parte inefficaci e richiedono di essere integrati con tutti gli strumenti del plesso normativo dedicato alla *data protection*.³⁶

In un simile quadro, caratterizzato da quella che si è definita 'ontologica differenza' fra intelligenza umana ed intelligenza artificiale, appare insomma necessario introdurre un ribaltamento di prospettiva e con esso un volano di

³² Cfr J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, *Accountable Algorithms*, cit., 659 ove si sottolinea che i sistemi più avanzati di *machine learning* «can update their model for predictions after each decision, incorporating each new observation as part of their training data».

³³ Si può considerare la trasparenza come una caratteristica a priori mentre la comprensibilità è elemento valutato a posteriori, sulla base di un momento applicativo. In proposito, per una prospettiva maggiormente focalizzata sul risultato ma anche, al contempo, sulla fase di immissione dei dati posti alla base della decisione, cfr. A. Chander, *The Racist Algorithm?*, in 115 *Michigan Law Review*, 1023, 1039 (2017), ove si osserva che «instead of transparency in the design of the algorithm, what we need is a transparency of inputs and outputs».

³⁴ Come si osserva in G. Malgieri, G. Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in 7 *International Data Privacy Law*, 243, 244 (2017): «the problem of these algorithms is the lack of 'readability' and 'legibility' from individuals directly concerned by them».

³⁵ Sul tema cfr. L. Edwards, M. Veale, *Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for*, in 16 *Duke Law & Technology Review*, 18 (2017).

³⁶ In tema cfr. P. Hacker, *Teaching Fairness to Artificial Intelligence*, cit., in particolare 24 ss.

riequilibrio che si ponga al di fuori della logica dell'algoritmo: tale prospettiva, cui ora ci si deve dedicare, si lega alla valorizzazione dell'intervento umano.

3. Il diritto all'intervento umano con particolare riferimento all'art 22 GDPR

Come è noto l'Unione europea, con l'adozione del Regolamento generale sulla protezione dei dati personali³⁷ (GDPR), ha inteso consolidare, sostituendo la precedente Direttiva 95/46³⁸ e in forza di uno strumento normativo con vocazione alla uniforme applicazione, la coerenza normativa in materia di protezione dei dati personali tra gli Stati membri. La tutela dei cittadini europei nei confronti dei soggetti che raccolgono e utilizzano dati personali è assicurata nel Regolamento indentificando un *numerus clausus* di basi di legittimazione del trattamento dei dati personali, e conseguentemente ponendo limiti stringenti ed uniformi al trattamento³⁹.

Per quanto l'individuazione di tali limiti sia sicuramente significativa, devono rilevarsi questioni più sottili e di non evidente soluzione, con particolare riferimento alle problematiche sottese al concetto di profilazione, di cui già si diceva nel precedente paragrafo e che trova definizione all'articolo 4, n. 4) del GDPR⁴⁰.

È in questo quadro che assume particolare rilevanza l'articolo 22 del GDPR, rubricato *Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione*⁴¹. Da una lettura sistematica dell'articolo si può dedurre come il

³⁷ Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE, in GU L 119/1. Sul Regolamento cfr. in generale, F. Pizzetti, *Privacy e il diritto europeo alla protezione dei dati personali. Dalla Direttiva 95/46 al nuovo Regolamento europeo*, Torino, 2016; Id., *Privacy e il diritto europeo alla protezione dei dati personali. Il Regolamento europeo 2016/679*, Torino, 2016; P. Voigt, A. von dem Bussche, *The EU General Data Protection Regulation (GDPR). A Practical Guide*, Berlin, 2017 e L. Feiler, N. Forgó, M. Weigl, *The EU General Data Protection Regulation (GDPR): A Commentary*, Woking, 2018; G.M. Riccio, G. Scorza, E. Belisario (a cura di) *GDPR e Normativa Privacy. Commentario*, Milano, 2018.

³⁸ Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, in GU L 281/31.

³⁹ In tema cfr. G. Finocchiaro, *Il quadro d'insieme sul Regolamento europeo sulla protezione dei dati personali*, in G. Finocchiaro (a cura di) *La protezione dei dati personali in Italia. Regolamento UE 2016/679 e d.lgs. 2018 n. 101*, Bologna, 2019, 1 ss.

Per un'analisi dei contenuti del Regolamento nella prospettiva del rapporto fra decisioni algoritmiche, diritti e principio di uguaglianza cfr. il recente e ampio studio di G. Resta, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 2019, 199 ss.

⁴⁰ Essa è ivi definita quale 'trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali 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, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica'. In tema cfr. M. Hildebrandt, S. Gutwirth (eds), *Profiling the European Citizen. Cross-Disciplinary Perspectives*, Berlin, 2008.

⁴¹ L'art. 22 GDPR prevede al comma 1 che 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

principio di cui al comma 1, ossia il diritto a non essere sottoposti a una decisione basata esclusivamente sul trattamento automatizzato, sia posto come principio e sia pertanto concepito come dotato di valore generale, mentre il comma 2 ne rappresenti una eccezione, basata su tre circostanze tipiche. Il comma 3 a sua volta introduce delle garanzie nel caso ci si trovi in due delle tre circostanze tipizzate al comma 2. In particolare risulta qualificante la possibilità, per un soggetto su cui abbia inciso significativamente una decisione adottata in deroga al comma 1, di instaurare un contraddittorio che prevede il recupero dell'intervento umano già garantito, a livello di principio, proprio dal comma 1. Infine il comma 4 esclude, come base per le decisioni eccezionalmente previste al comma 2, l'utilizzo di categorie particolari di dati personali (e segnatamente quelli di cui all'art. 9 GDPR, ossia dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona). Ciò a meno che non sia applicabile l'articolo 9, paragrafo 2, lettere a) o g), che si riferisce ad un consenso esplicito dell'interessato o alla necessità di trattamento per motivi di interesse pubblico, e sempre che siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato.

La lettura dell'articolo rivela pertanto una sistematica complessa di regole ed eccezioni, dalla quale emerge un principio generale, enunciato al comma 1, temperato da svariate deroghe che appaiono interagire fra loro e di cui sarà cruciale, chiarire, anche e soprattutto in chiave giurisprudenziale, i reciproci rapporti⁴².

In relazione al principio generale emergono, da un punto di vista ermeneutico, almeno tre rilevanti questioni⁴³. In primo luogo, vi è la portata del riferimento alla 'incidenza significativa sulla sfera individuale', in relazione al quale

sulla sua persona. Il comma 2 dispone che il paragrafo 1 non si applica nel caso in cui la decisione: a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato. Il comma 3 ulteriormente precisa che nei casi di cui al paragrafo 2, lettere a) e c), 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. Il comma 4 infine dispone che le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato.

⁴² In tema cfr. E. Gil Gonzalez, P. de Hert, *Understanding the legal provision that allow processing and profiling of personal data – an analysis of GDPR provisions and principles*, in *Era Forum* February 2019, 597 ss.

⁴³ In tema cfr. L.A. Bygrave, *Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision Making*, in K. Yeung, M. Lodge (a cura di) *Algorithmic Regulation*, Oxford, 2019, 248 ss.; E. Pehrsson, *The Meaning of the GDPR Article 22*, EU Law Working Papers, n. 31/2018, Stanford-Vienna Transatlantic Technology Law Forum.

assumono rilievo in Consideranda⁴⁴ 70 e 71 del GDPR⁴⁵. Al riguardo, l'Article 29 Data Protection Working Party (WP29)⁴⁶ ha avuto modo di chiarire che, affinché la previsione sia applicabile, è necessaria, ma al contempo sufficiente, una qualche forma di interferenza, ossia l'utilizzo di dati personali al fine di valutare o predire qualcosa riguardo al comportamento di un individuo⁴⁷. Dunque ricadono sotto la disposizione certamente le conseguenze che incidano sui diritti di un soggetto (quali, ad esempio, libertà di associazione, diritto di voto, diritto a una tutela giurisdizionale effettiva) o siano relative al suo *status*, ma anche decisioni che senza incidere direttamente sui diritti, siano idonee ad incidere sul comportamento e sulle scelte di un individuo, avendo un impatto duraturo o permanente e potendo potenzialmente dare luogo a una qualche forma di esclusione o di discriminazione⁴⁸.

In secondo luogo, si pone la questione del significato da attribuire al riferimento ad una decisione che dovrebbe essere "basata *unicamente* sul trattamento automatizzato". Secondo alcuni commentatori tale previsione implicherebbe che la presenza di un qualsivoglia intervento umano, anche su un piano meramente formale, osterebbe all'applicazione della garanzia di cui all'articolo 22⁴⁹. Anche un'analisi delle diverse versioni della stessa previsione succedutesi nei vari passaggi della procedura legislativa sembrerebbero avvalorare

⁴⁴ Al riguardo, giova ricordare quanto affermato da parte della Corte di Giustizia dell'UE quanto al fatto che il considerando di un regolamento, pur non essendo giuridicamente vincolante, «può [...] consentire di chiarire l'interpretazione di una *regula juris*» (sentenza della Corte di giustizia del 13 luglio 1989, causa C-215/88, *Casa Fleischhandel / BALM*, punto 31).

⁴⁵ Al primo si può leggere che qualora i dati personali siano trattati per finalità di *marketing* diretto, l'interessato dovrebbe avere il diritto, in qualsiasi momento e gratuitamente, di opporsi a tale trattamento, sia con riguardo a quello iniziale o ulteriore, compresa la profilazione, e che tale diritto dovrebbe essere esplicitamente portato all'attenzione dell'interessato e presentato chiaramente e separatamente da qualsiasi altra informazione. Al secondo si afferma che l'interessato dovrebbe avere il diritto di non essere sottoposto a un processo decisionale automatizzato nei casi – evidentemente, meramente esemplificativi – relativi a domande di credito online o pratiche di assunzione elettronica senza interventi umani.

⁴⁶ Si tratta del gruppo di lavoro per la tutela delle persone con riguardo alla tutela dei dati personali, istituito sulla base della direttiva 1995/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati. È un organo consultivo composto da un rappresentante della o delle autorità di controllo designate da ciascuno Stato membro e da un rappresentante della o delle autorità create per le istituzioni e gli organismi comunitari, nonché da un rappresentante della Commissione europea che ha il compito, tra l'altro, di esaminare questioni attinenti all'applicazione della normativa in materia di protezione dei dati personali e di consigliare la Commissione in materia.

⁴⁷ Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 17/EN WP251rev.01, 6 febbraio 2018, p. 7.

⁴⁸ Cfr. Article 29 Data Protection Working Party *Guidelines* cit., p. 21-22, ove si considerano tali anche le decisioni relative alla sfera economica di un soggetto, ad esempio per quanto attiene all'accesso al credito, all'accesso ai servizi educativi o sanitari e alle opportunità di lavoro.

⁴⁹ S. Wachter, B. Mittelstadt, L. Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in 7 *International Data Privacy Law*, 76 (2017).

una simile idea⁵⁰. Tuttavia, le indicazioni fornite dal WP29 sono idonee, se non a contestare interamente tale linea argomentativa, quanto meno a stemperarla in un'ottica di maggiore garanzia sostanziale. A ben vedere, infatti, il riferimento è proprio volto ad evitare una artificiosa e non sostanziale costruzione dell'intervento umano con finalità elusive della norma⁵¹. In relazione ad entrambe le questioni, il rischio di una interpretazione restrittiva della disposizione è evidente perché porterebbe a neutralizzare il rimedio in discussione, la cui applicazione sarebbe esclusa in tutte quelle fattispecie in cui le decisioni automatizzate siano adottate con riguardo a circostanze che non raggiungano la soglia degli effetti giuridici e in quelli in cui si rilevi un passaggio, anche solo meramente formale, attraverso l'attività umana⁵².

Una terza e assai rilevante questione riguarda, infine, il modo in cui andrebbe interpretata la previsione dell'articolo 22, paragrafo 1, in relazione alla alternativa fra una sua ricostruzione come riconoscimento in capo del titolare dei dati personali di un diritto a contestare una decisione risultante dal processo decisionale automatizzato o invece come generale divieto di adottare una tale decisione. La formulazione, sul punto, non è inequivoca e vi sono invero opinioni divergenti circa il contenuto⁵³, la portata e la reale efficacia del rimedio in discussione⁵⁴. Certamente, le conseguenze di una interpretazione rispetto all'altra non sono di poco rilievo anche in chiave applicativa. Un divieto sarebbe tale da imporre un onere in capo ai soggetti che svolgono l'attività di profilazione, o che si avvalgono ad altro titolo di decisioni algoritmiche automatizzate, di contemplare

⁵⁰ E. Pehrsson, *The Meaning of the GDPR Article 22*, cit., p. 9-10.

⁵¹ Cfr. Article 29 Data Protection Working Party, *Guidelines* cit., p. 21, ove si legge: «The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing. To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data».

⁵² Cfr. M Hildebrandt, *The Dawn of a Critical Transparency Right for the Profiling Era*, in J. Bus et al. (eds.) *Digital Enlightenment Yearbook 2012*, IOS Press, 2012, 41, spec. 51.

⁵³ Ad esempio, «it is not clear whether data controllers are prevented from performing these kinds of processing or if there is an active option for data subject to 'block' these processing» (cfr. G. Malgieri, G. Comandé, *Why a Right to Legibility of Automated Decision-Making Exists* cit. 247).

⁵⁴ Nel senso di un divieto generale pare orientata la prevalente dottrina di lingua tedesca, e in particolare S. Simitis, *Bundesdatenschutzgesetz. Kommentar*, Baden Baden, 2014; H.A. Wolff, S. Brink, *Beck'scher Online-Kommentar Datenschutzrecht*, München, 2017; B.P. Paal, D.A. Pauly, *Datenschutz-Grundverordnung. Bundesdatenschutzgesetz*, München, 2018). Nello stesso senso anche I. Mendoza, L.A. Bygrave, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, in T. Synodinou, P. Jogleux, C. Markou, T. Prastitou (eds.), *EU Internet Law: Regulation and Enforcement*, Berlin, 2017, 77 ss.

Fra le ricostruzioni in termini di diritto cfr. I.S. Döhmman, O. Tambou, P. Bernal, M. Hu, C.A. Molinaro, E. Negre, I.W. Sarlet, L. Schertel Mendes, N. Witzleb, F. Yger, *The Regulation of Commercial Profiling – A Comparative Analysis*, in 2 *European Data Protection Law Review*, 535, (2016); B. Berendt, S. Preibusch, *Toward Accountable Discrimination-Aware Data Mining: The Importance of Keeping the Human in the Loop-and Under the Looking Glass*, in 5 *Big Data*, 135 (2017); A. Vedder, L. Naudts, *Accountability for the use of algorithms in a big data environment*, in 31 *International Review of Law, Computers & Technology*, 206 (2017).

ex ante, e in certa misura *by design* un intervento umano significativo. Tale vincolo sarebbe superabile solamente nel caso in cui, come si è visto, si facessero valere le eccezioni di cui al paragrafo 2, ma comunque con le salvaguardie e le precisazioni di cui ai commi 3 e 4.

D'altro canto, ove si ricostruisse la previsione come orientata a riconoscere al singolo il diritto di contestare le decisioni automatizzate, esse potrebbero essere adottate senza previ limiti o condizionamenti ma occorrerebbe prevedere meccanismi adeguati per rispondere alle contestazioni dei soggetti interessati. La portata effettiva della disposizione sarebbe, in certo senso, subordinata alla proposizione di un ricorso, ma certo la costruzione di un diritto soggettivo avrebbe un valore, anche simbolico, non trascurabile e si collegherebbe armonicamente con il diritto di prendere parte al procedimento decisionale e di influenzarne i risultati di cui si dispone al comma 3⁵⁵.

Sul punto soccorre, ancora una volta, il WP29, che ha avuto modo di chiarire la propria posizione come favorevole alla ricostruzione di un generale divieto, affermando che «a general prohibition on this type of processing exists to reflect the potential risks to individuals' rights and freedoms»⁵⁶ e che «this prohibition applies whether or not the data subject takes an action regarding the processing of their personal data»⁵⁷.

A ben vedere, peraltro, si può ben sostenere che le due ricostruzioni non si escludano vicendevolmente, ma piuttosto possano mutualmente rinforzarsi. Si tratta di una lettura che darebbe conto efficacemente sia della logica intrinseca dell'art. 22, con particolare riferimento al rapporto fra il comma 1 e i commi 2 e 3, e segnatamente il diritto ivi contemplato di ottenere anche *ex post* l'intervento umano da parte del soggetto su cui la decisione incida significativamente, senza rinunciare alla portata tendenzialmente generale, e per questo applicabile a priori, del divieto di sottoporre l'individuo a decisioni totalmente automatizzate.

Questa lettura, che qui si propone, appare un tassello qualificante ai fini della piena valorizzazione della disposizione e sul punto è auspicabile che gli interventi giurisprudenziali a venire possano interpretare in questo senso, valorizzandolo pienamente, un dettato normativo non inequivoco.

⁵⁵ Cfr. M. Hildebrandt, K. De Vries *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Abingdon, 2013.

⁵⁶ Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 17/EN WP251rev.01, 6 febbraio 2018, p. 9. Sotto questo punto di vista, può notarsi la conformità tra la nozione di *profiling* rilevante ai fini del diritto dell'Unione e quella fatta propria dal Consiglio d'Europa (cfr. la *Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling*, 3 novembre 2010, p. 6, ove si legge: «Profiling means an automatic data processing technique that consists of applying a “profile” to an individual, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes»).

⁵⁷ *Ibidem*, p. 19.

4. L'intervento umano come possibile *gold standard* internazionale e la sua valorizzazione a tutela dell'autonomia umana nel quadro di regolazione internazionale dell'intelligenza artificiale.

Dall'analisi svolta sull'articolo 22 GDPR si evince che la logica di valorizzazione dell'intervento umano è quella di concretizzare sistematicamente l'interazione sostanziale con un interlocutore umano vuoi nel momento della adozione di una decisione automatizzata vuoi nel momento in cui si agisce per ottenere la modifica di una decisione automatizzata già assunta. L'alveo in cui tale principio è normativizzato è quello specifico della disciplina relativa alla *Data Protection* e il principio stesso presenta rilevanti collegamenti sia con la sistematica generale del GDPR, improntata al principio di *accountability*,⁵⁸ sia con specifiche disposizioni del medesimo che ne costituiscono in certo senso il presupposto, quali in particolare gli artt. 13 e 14 relativi alle informazioni da fornire all'interessato⁵⁹.

È peraltro indubbio che, pur muovendo da una ermeneutica testuale dell'art. 22 GDPR, si perviene facilmente ad avvalorare, su un piano più generale e anche al di fuori della logica intrinseca allo strumento, una concezione teorica che vorrebbe il processo decisionale algoritmico finalizzato a supportare il ragionamento umano⁶⁰ e non già a sostituirlo⁶¹. In questo senso, e a ben vedere, l'intervento umano appare necessario da una duplice prospettiva, che è ancor prima tecnica che giuridica.

Sotto il primo profilo, infatti, tale intervento si collega a talune caratteristiche del processo decisionale algoritmico, il quale, in particolare in relazione alla c.d. randomizzazione predittiva che caratterizza le forme più avanzate di *machine learning*⁶², richiede sistematici aggiustamenti in relazione al caso applicativo concreto.

⁵⁸ Il nesso con la nozione di *accountability* di cui all'art. 5 del GDPR e con la responsabilità del *data controller* è ben evidenziato in B. Berendt, S. Preibusch, *Toward Accountable Discrimination-Aware Data Mining: The Importance of Keeping the Human in the Loop* cit. in particolare ai parr. 4.3 e 5.1.

⁵⁹ Sul collegamento con il tema della comunicazione delle 'informazioni significative sulla logica utilizzata' in un processo decisionale automatizzato, e in generale sulla sistematica delle due disposizioni, cfr. M. Durante, *Commento all'art. 13 GDPR. Informazioni da fornire qualora i dati personali siano raccolti presso l'interessato e Commento all'art. 14 GDPR. Informazioni da fornire qualora i dati personali non siano stati ottenuti presso l'interessato* in A. Barba, S. Pagliantini (a cura di), *Commentario del Codice Civile Utet. Modulo delle Persone*, Vol. II, Utet Giuridica, Torino, 2019, pp. 218-234, e 235-245.

⁶⁰ Come si osserva opportunamente in K. Brennan Marquez, "Plausible Cause", cit., 1254, una volta precisata la necessità che il risultato di un processo decisionale algoritmico sia analizzato sotto il profilo della comprensibilità e non solo sotto quello della probabilità dei risultati forniti, «[i]t means, rather, that their role is to aid human reasoning, not to supplant it».

⁶¹ Cfr., anche per ulteriori riferimenti, P. Domingos, *The Master Algorithm: How the Quest for Ultimate Machine Learning will Remake Our World*, New York, 2015.

⁶² In effetti, l'intervento umano appare necessario anche perché «[m]any machine learning systems use randomization as part of their normal operation. It turns out that guessing randomly and adjusting the probability of each class of output often leads to much better performance than trying to determine the absolute best decision at any point» (cfr. J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, *Accountable Algorithms*, cit., 655).

Sotto il secondo profilo, non sfugge poi come sul piano giuridico il diritto all'intervento umano riporti alla intrinseca e sostanziale differenza fra l'attività di scrittura del codice di un algoritmo e l'elaborazione di norme giuridiche⁶³, in particolare per il diverso grado di rigidità delle due tipologie di regole, tanto nella fase di formulazione quanto, e specialmente, nella fase applicativa⁶⁴.

Queste indicazioni impongono di riconsiderare il tema delle decisioni algoritmiche, e della loro ontologica differenza rispetto al ragionamento umano, in una più ampia cornice che consideri la tutela dei diritti fondamentali e con essa il tema stesso, fondamentale, della tutela dell'umano.

Sono infatti evidenti i pregiudizi che la soggezione a decisioni automatizzate sono suscettibili di comportare per alcuni diritti fondamentali degli individui⁶⁵, e in particolare per lo stesso diritto alla dignità e all'autodeterminazione della persona umana⁶⁶. Questo, giova sottolineare, anche a prescindere da ogni valutazione rispetto alla correttezza o accuratezza degli esiti della decisione, nonché dalla prova di un preteso effetto discriminatorio, che come si è visto può essere difficile da dimostrare con i tradizionali strumenti propri della tutela antidiscriminatoria.

È infatti lo stesso meccanismo di profilazione a fini decisionali a presentare criticità rispetto alla piena conformità con tale novero di diritti fondamentali⁶⁷. Come noto, infatti, la profilazione ha assunto una dimensione collettiva in forza della quale gli individui vengono classificati e valutati non tanto in base alle proprie azioni, quanto piuttosto in base a quelle che gli algoritmi ipotizzano potrebbero essere le loro azioni future⁶⁸; e soprattutto, tali ipotesi circa le possibili condotte future non sono elaborate sulla base di dati direttamente ed esclusivamente riferibili ai soggetti destinatari della decisione, quanto piuttosto

⁶³ Questo dualismo è ben sottolineato, come nascente da una tensione fra predittività e prescrittività, in G. Della Morte, *Big Data e protezione internazionale dei diritti umani*, cit. in particolare p. 275 ss.

⁶⁴ Sotto tale profilo, si osserva che «[s]oftware code is, ultimately, a rigid and exact description of itself: the code both describes and causes the computer's behavior when it runs. In contrast, public policies and laws are characteristically imprecise, often deliberately so. Thus, even when a well-designed piece of software does assure certain properties, there will always remain some room to debate whether those assurances match the requirements of public policy» (cfr. J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, H. Yu, *Accountable Algorithms*, cit. 646).

⁶⁵ Cfr. Consultative Committee of the Convention for the Protection of Individuals with regard to automatic processing of personal data, *Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data*, 2017, spec. p. 3

⁶⁶ Cfr. L. Moerel, *Big Data Protection. How to Make the Draft EU Regulation on Data Protection Future Proof*, Tilburg, 2014, spec. 11; cfr. più in generale anche V. Zeno Zencovich, *Il concetto di "autonomia privata" ai tempi dei "Big Data"*, in P. Passaglia, D. Poletti (a cura di), *Nodi virtuali, legami informali: Internet alla ricerca di regole*, Pisa, 2017, p. 31.

⁶⁷ Il nesso è ben evidenziato in M. Leta Jones, *The right to a human in the loop: Political constructions of computer automation and personhood*, in *47 Social Studies of Science*, 216, 220 (2017), ove si osserva che «automated data processing systems have been associated with the abuse and indignity of individuals and groups throughout the twentieth century».

⁶⁸ Gli individui, in altri termini, «are no longer judged on the basis of [their] actions, but on the basis of what all the data about [them] indicate [their] probable actions may be» (cfr. il Report dell'Autorità della Privacy Norvegese, *Big data – privacy principles under pressure*, 2013, p. 7).

sulla base di dati e informazioni che riguardano un numero indeterminato di altri soggetti che lo stesso algoritmo, autonomamente, ritiene per qualche ragione assimilabili al destinatario.

Questo schema di elaborazione di *pattern* comportamentali, operante anche a fini di processi decisionali automatizzati, rimanda seriamente non solo al tema della tutela della *privacy* in dimensione collettiva⁶⁹, ma anche a quello, essenziale, della dignità umana come elemento fondativo dello stesso concetto di *privacy*⁷⁰. Come è noto, una lettura piena e sostanziale di tale concetto si pone in relazione diretta con la prospettiva dello sviluppo umano: proteggere la vita privata e familiare significa in questo senso tutelare dalla sindrome di un moderno *panopticon* quello che è stato efficacemente definito un santuario in cui si conservano gli arcani dell'uomo⁷¹, nell'intento di evitare all'uomo stesso una 'nudità' assoluta che risulterebbe disumanizzante.

È proprio in quest'ultima prospettiva di tutela che la nostra considerazione del rapporto fra decisioni automatizzate e intervento umano si apparenta con il concetto di dignità, in ragione della considerazione dell'umano come per sua natura irriducibile a un numero chiuso e determinato di attività computazionali su un numero a sua volta definito di dati⁷².

Questa prospettiva, definibile come 'incomputabilità' dell'umano, è stata sviluppata, in una prospettiva più generale, in relazione alla funzione della *privacy* ma appare altresì particolarmente calzante in relazione al tema dell'intervento umano come argine alla generalizzazione di decisioni automatizzate⁷³. Essa consente una più fondamentale comprensione di cosa dovrebbe significare mantenere l'uomo parte sostanziale della decisione, o, come è stato efficacemente espresso, "*keeping the human in the loop*"⁷⁴. La presenza dell'umano diviene, in questa prospettiva, momento qualificante di validazione della decisione algoritmica⁷⁵, assolvendo al contempo l'essenziale funzione di garantire sempre un nesso di attribuzione, che è presupposto per poter coltivare, anche nel contesto della società algoritmica, una autentica prospettiva di responsabilità⁷⁶. Si tratta di un punto

⁶⁹ In argomento cfr. L. Kammourieh, T. Baar, J. Berens, E. Letouzé, J. Manske, J. Palmer, D. Sangokoya, P. Vinck, *Group privacy in the age of big data*, in L. Taylor, L. Floridi, B. van der Sloot (eds.), *Group Privacy: new challenges of data technologies*, Dordrecht, 2017, spec. p. 38).

⁷⁰ Sul punto cfr. L. Floridi, *On Human Dignity as a Foundation for the Right to Privacy*, in 29 *Philosophy and Technology*, 307 (2016).

⁷¹ Il riferimento va a S. Zuboff, *The Age of Surveillance Capitalism. The Fight for a human future at the new frontier of power*, cit. in particolare 475 ss.

⁷² Si è visto nei paragrafi precedenti come questo elemento di determinatezza dei passaggi di calcolo sia coesistente alla logica algoritmica, e come esso costituisca in certo modo il fulcro della stessa definizione di algoritmo.

⁷³ Sul punto cfr. M. Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, in 19 *Theoretical Inquiries of Law*, 83 (2019).

⁷⁴ Il riferimento va ancora a M. Leta Jones, *The right to a human in the loop*, cit.

⁷⁵ In questa stessa prospettiva vi è chi sviluppa, intorno alla garanzia dell'intervento umano, sia esso preventivo o successivo, ma sempre necessariamente basato sulla disponibilità di informazioni significative sul processo decisionale automatizzato, un vero e proprio principio di contestabilità *by design*. Sul punto cfr. M. Almada, *Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems*, International Conference on Artificial Intelligence and Law (ICAIL) Papers, June 2019, 2 ss.

⁷⁶ Il requisito della attribuzione è per questo valorizzato al punto da essere proposto, a

essenziale in cui la reviviscenza dell'umano nel processo decisionale algoritmico assolve alla necessità di individuare un fulcro certo di responsabilità che è, in certo modo, legato alla stessa nozione di fallibilità umana, la cui reintroduzione può, solo in modo apparentemente paradossale, servire proprio la tutela dell'umano. Come si è visto, infatti, non sempre il legame con l'umano può essere tenuto vivo con riferimento alla concezione o al funzionamento automatico dell'algoritmo, specie se ci riferiamo ad algoritmi che hanno la caratteristica dell'autoapprendimento di cui si è analizzata la diversità 'ontologica'.

Su questo punto si evidenzia quindi il rischio di una lacuna di attribuzione, e conseguente responsabilità, che è potenzialmente assai grave e che potrebbe essere arginata mediante il recupero di un intervento umano sostanziale. Non è facile, certamente, immaginare che l'intervento umano possa sovvertire le risultanze di una elaborazione algoritmica, sostituendosi totalmente ad essa. Tuttavia, un ruolo di validazione che implichi una assunzione di responsabilità, potenzialmente aggiuntiva a quella già contemplata sul piano dell'impianto di *data protection*, potrebbe essere una significativa novità, configurabile come requisito da realizzarsi anche nella forma di un 'intervento umano *by design*': un meccanismo che condurrebbe a una piena valorizzazione, come divieto generale, proprio del principio di cui all'art. 22 GDPR, che esclude, salve specifiche eccezioni che potrebbero essere mantenute a questo punto senza rischio di abuso, la sottoposizione a decisioni che discendano da processi totalmente automatizzati.

In una simile prospettiva, il principio di cui all'art. 22 GDPR, quale aggancio normativo concreto, potrebbe superare il proprio campo di applicazione naturale, per assurgere ad autentico *gold standard* a vocazione globale. Se infatti è vero che il GDPR è uno strumento che esprime la duplice volontà della UE di garantire alti standard di *data protection* ai propri cittadini, servendo al contempo l'obiettivo di rafforzare la fiducia nella circolazione dei dati, che è coesistente allo sviluppo del mercato unico digitale, si deve riconoscere che, proprio per questa sua seconda anima, esso rivela una notevole *vis expansiva* e una attitudine a proiettarsi ben al di là della dimensione territoriale della UE⁷⁷. Come noto, infatti l'art. 3 GDPR detta una precisa vocazione alla applicazione extraterritoriale⁷⁸ e vi sono chiari indizi della crescente tendenza a trattare i contenuti del GDPR come standard generale e globale in materia di tutela dei dati personali⁷⁹. Da questo punto di vista, i contenuti del GDPR, e in particolare il riferimento in esso contenuto all'intervento umano, incarnano indubbiamente una specificità europea caratterizzata da una sensibilità di protezione particolarmente elevata, che si accredita come modello

complemento delle tre leggi di Balkin, come quarta legge della Robotica in F. Pasquale, *Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility and Explainability in an Algorithmic Society*, University of Maryland Francis King Carey School of Law, Legal Studies Research Paper n. 2017-21.

⁷⁷ Sul punto e sulle specificità dell'approccio della UE alla società algoritmica cfr. U. Pagallo, *Algo-Rhythms and the Beat of the Legal Drum*, in 31 *Philosophy and Technology*, 507 (2018).

⁷⁸ Sul punto cfr. fra i molti R. Vatanparast, *Designed to serve mankind. The politics of General Data Protection Regulation*, in *Völkerrechtsblog*, 27 May 2019.

⁷⁹ In questo senso cfr. anche G. Buttarelli, *The EU GDPR as a clarion call for a new global digital gold standard*, in 6 *International Data Privacy Law*, 78 (2016).

avanzato in tema di *data protection*, per questo destinato a fungere da modello anche per le normative extraeuropee⁸⁰.

L'allargamento di prospettiva che si è proposto conduce, a questo punto, a collegare l'analisi ad un più ampio contesto di riferimento, ossia il quadro dei tentativi di regolazione, mediante strumenti internazionali che assumono perlopiù la forma di *soft law*, della intelligenza artificiale.

Si tratta di un processo che muove dal riconoscimento di quanto i processi decisionali algoritmici, e particolarmente quelli basati sul *machine learning*, richiedano lo sviluppo di un approccio e di standard internazionali, per assicurare la promozione e la protezione dei diritti umani di fronte di uno sviluppo digitale sempre più suscettibile di impattare con incisività sulle sfere individuali⁸¹.

In tema, si deve innanzitutto riscontrare l'interazione fra tre plessi di regole di natura diversa, e in particolare tecniche, etiche e giuridiche. È infatti la risultante di queste tre forze a determinare la complessa *governance* della intelligenza artificiale⁸². In proposito, è assai attuale il dibattito relativo al bilanciamento reciproco fra i menzionati plessi di regole, e segnatamente a quanto peso specifico dovrebbero avere le regole giuridiche in un contesto pervaso dalla forza per così dire intrinseca degli standard tecnologici e, al contempo, tradizionalmente improntato a valorizzare prevalentemente i precetti di *digital ethics*⁸³.

Assai vario è il panorama degli strumenti volti a fissare internazionalmente principi per l'intelligenza artificiale e di esso non si può dare conto compiutamente, stante la complessità che lo caratterizza⁸⁴. Ciò che rileva in questa sede è riscontrare come negli strumenti citati perlopiù manchi un riferimento esplicito all'intervento umano. Essi sono per certo orientati a riconoscere un ruolo dell'intelligenza artificiale che sia servente rispetto all'umano, e in questo senso

⁸⁰ Sulla diversa sensibilità che, sul punto dell'intervento umano, sussiste fra le due sponde dell'Atlantico cfr. M. Leta Jones, *The right to a human in the loop: Political constructions of computer automation and personhood*, cit. in particolare 219, anche per l'esplicito collegamento fra intervento umano e tutela della dignità come specificità europea.

⁸¹ Cfr. in questo senso la *Declaration on Ethics and Data Protection in Artificial Intelligence*, della International Conference of Data Protection and Privacy Commissioners, 2018, spec. 3.

⁸² Il punto è assai ben evidenziato in L. Floridi, *Soft Ethics and the Governance of the Digital*, in 31 *Philosophy and Technology*, 4 (2018) ove si indaga precisamente l'interazione fra *Digital Ethics*, *Digital Governance* e *Digital Regulation* rispetto all'obiettivo di dare forma attuale e futura all'intelligenza artificiale.

⁸³ Sul punto, emerge come significativa proprio la prospettiva di arricchire il modello di etica digitale attraverso un sistema di regolazione digitale che sia via via reso più incisivo (G. Resta, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, cit. in particolare 220).

⁸⁴ Per un interessante tentativo di analisi comparativa delle diverse iniziative in corso, cfr. L. Floridi, *A Unified Framework of Five Principles for AI in Society*, in 1 *Harvard Data Science Law Review*, 1 (2019), in cui si analizza la declinazione dei 5 principi di *Beneficence*, *Nonmaleficence*, *Autonomy*, *Justice* e *Explicability* all'interno di alcuni rilevanti strumenti di codificazione dei principi. Fra questi si devono innanzitutto menzionare gli *Ethical Principles* contenuti dello *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, pubblicato dallo *European Commission's European Group on Ethics in Science and New Technologies*, Marzo 2018, e si possono altresì ricordare gli *Asilomar AI Principles*, sviluppati sotto l'egida del *Future of Life Institute*, in occasione della Conferenza di Asilomar del gennaio 2017 e la *Montreal Declaration for Responsible AI*, sviluppata sotto l'egida dell'Università di Montreal a seguito del *Forum on the Socially Responsible Development of AI* del Novembre 2017.

depongono i frequenti riferimenti ad una intelligenza artificiale *'human centered'* che si possono ritrovare ad esempio nella recente Raccomandazione del Consiglio dell'OECD sull'intelligenza artificiale⁸⁵, o nelle *Draft Ethics Guidelines for Trustworthy AI* elaborate da Gruppo di esperti della Commissione europea sull'intelligenza artificiale⁸⁶.

Si deve d'altronde rimarcare l'assenza di un riferimento esplicito e diretto all'intervento umano e occorre ricollegare questo concetto all'area di operatività del principio di autonomia, che è frequentemente posto come uno dei principi cardine per governare l'intelligenza artificiale⁸⁷. Tale principio si regge sull'importante presupposto che la decisione di delegare all'intelligenza artificiale non possa non essere, sempre, revocabile e reversibile, e pertanto non sia adottabile una volta per tutte. Questa caratura del principio è essenziale per la tutela dell'umano nel senso anche etimologico del termine 'autonomia', che rimanda alla attitudine di un soggetto a darsi regole proprie, e con ciò all'implicita costante possibilità di modificare le stesse.

Se, come è vieppiù riconosciuto, gran parte del dibattito sull'intelligenza artificiale dei prossimi anni insisterà sul delicato bilanciamento fra attività delegate ad agenti autonomi e attività conservate in capo all'uomo, si rivela appieno la assoluta centralità di questo concetto.

È proprio in questo perimetro di azione del concetto di autonomia che si ritrova l'esigenza di uno strumento atto rinforzare il portato sostanziale della stessa, al fine di conservare all'uomo la decisione sul se e quando delegare all'algoritmo. Ed è in questa prospettiva che l'intervento umano può entrare in gioco come elemento, non solo etico ma giuridico, che risulta al contempo concreto e qualificante.

Come si è visto, la previsione giuridica dell'intervento umano attualmente formalizzata all'art. 22 GDPR appare dotata di un potenziale espansivo che potrebbe anche travalicare l'ambito strettamente legato alla *data protection*, nel quale peraltro è già stato chiarito che il riferimento all'umano non possa essere superficiale, solo formale o cosmetico ma richieda invece una giuridicizzazione in senso sostanziale.

In questo quadro, si ritiene che la codificazione, e il successivo pieno riconoscimento giurisprudenziale, dell'intervento umano nel GDPR potrebbe precisamente fungere da base per un più pieno riconoscimento del principio dell'intervento umano anche nel quadro dei nascenti principi regolatori dell'intelligenza artificiale, ove anche i principi dell'etica dell'intelligenza artificiale potrebbero prendere più esplicita posizione nel senso di favorire forme sempre più sostanziali di 'intervento umano *by design*'.

⁸⁵ OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019, punto 1.2

⁸⁶ Cfr. in particolare il riferimento al Principio di Autonomia di cui a p. 9 del *Draft* pubblicato il 18 Dicembre 2018 e rubricato: *The Principle of Autonomy: Preserve Human Agency*.

⁸⁷ Sull'importanza del tema nella prospettiva delle decisioni algoritmiche cfr. M.S. Gal, *Algorithmic Challenges to Autonomous Choice*, in *25 Michigan Telecommunication and Technology Law Review*, 60 (2018).

Una tale impostazione sembrerebbe coerente con l'idea, già introdotta nei paragrafi precedenti, di un superamento della alternativa, in materia di decisioni completamente automatizzate, fra divieto generale operante *ex ante* e diritto soggettivo azionabile *ex post*. In un simile ribaltamento di prospettiva, basato sulla logica *by design*, si suffragherebbe la complementarietà delle due prospettive e la necessità di una concezione degli algoritmi, e delle correlate forme di intelligenza artificiale, tale da escludere strutturalmente l'assenza di un significativo coinvolgimento dell'umano come interfaccia necessaria proprio per la tutela dell'umano⁸⁸.

Questa caratura di quello che si è definito 'intervento umano *by design*' avrebbe per molti versi la capacità di riequilibrare il rapporto fra sfera umana e tecnologica, e darebbe corpo ben più sostanziale allo stesso principio di autonomia, attualmente declinato in una dimensione assai 'morbida', nonché a quello, parimenti essenziale, di attribuzione, che è imprescindibile in una logica generale di responsabilità che, anche su un piano astratto, pare sempre necessario contraltare di ogni esercizio del potere, sia esso anche declinato nella sua dimensione algoritmica.

Ciò proietterebbe altresì al di fuori della dimensione europea l'intervento umano come standard globale, esportando e generalizzando una sensibilità 'umanista' europea che da qualche tempo cerca di affermarsi, anche in chiave di riequilibrio geopolitico⁸⁹. Essa pare meritevole, ben al di là di considerazioni di sola opportunità politica, di assurgere essa stessa a *gold standard*, come in certa misura sta avvenendo in generale per il plesso normativo a tutela della *data protection* basato sul GDPR.

In definitiva, un riconoscimento più esplicito della ineludibile necessità di una cooperazione uomo-macchina, della natura necessariamente servente dell'intelligenza artificiale rispetto all'umano e, in ultimo, della dimensione della dignità umana come irriducibile a elementi puramente computazionali.

Alberto Oddenino
Dip.to di Giurisprudenza
Università degli Studi di Torino,
alberto.oddenino@unito.it

⁸⁸ Come si osserva assai pertinentemente in L. Floridi *On Human Dignity* cit. 310, l'essenza dell'umano, nel quadro della cd. Quarta rivoluzione dell'Infosfera, risiede specificamente nella sua natura aperta e interrelazionale.

⁸⁹ In tema cfr. la significativa recente pubblicazione del Libro Bianco della Commissione europea (Commissione europea, *Libro Bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*, Documento COM(2020) 65 final del 19.2.2020) in cui, in particolare nella parte 5 (pp. 10-12) si evidenzia, ancorché larvamente, un legame fra il concetto di fiducia, l'edificazione di un quadro normativo per l'intelligenza artificiale e il margine riservato all'intervento umano.