JUS CIVILE

EMILIANO TROISI

PhD in Law – Università degli Studi Suor Orsola Benincasa di Napoli

DECISIONE ALGORITMICA, BLACK-BOX E AI ETICA: IL DIRITTO DI ACCESSO COME DIRITTO A OTTENERE UNA SPIEGAZIONE

1. – Decisione algoritmica, trattamento automatizzato dei dati, intelligenza(e) artificiale(i); sono termini (e fenomeni) spesso correlati, talvolta – nella narrazione – anche in modo confusivo. Pertanto, pur senza pretese di nitidezza – abbisognando, ogni discorso che si proponga di contribuire alla generale trattazione di un problema, riferirsi anzitutto ad una certa sistemazione dello stesso – qualche premessa classificatoria risulta imperativa.

Quella che chiamiamo, in questa sede, decisione algoritmica è il risultato – l'output – di un cosiddetto processo decisionale automatizzato (spesso individuato dall'acronimo inglese A.D.M., *Automated Decision Making*), espressione che si riferisce – in senso ampio – a null'altro che qualunque procedimento che consenta, attraverso l'impiego di strumenti tecnologici, di prendere decisioni senza, o comunque con irrilevante coinvolgimento umano. Una tale definizione non implica, dunque, ma chiaramente include, l'utilizzo delle tecnologie cd. di Intelligenza Artificiale, come più in generale qualsiasi tecnica informatica che, basandosi su algoritmi – ovvero una sequenza di operazioni eseguibili da un processore [1] – consenta di eseguire compiti ripetitivi con i dati senza la necessità di una costante guida umana [2].

Tra questi, i sistemi di IA – volendoli definire brevemente – e accogliendo una delle definizioni tra le più accreditate – si distinguono in quanto, essenzialmente, dei sistemi razionali [3] capaci di '*agire e pensare uma-*

[1] *Treccani online*, voce *Algoritmo*, *https://www.treccani.it/vocabolario/algoritmo/.*

[2] Nell'ambito dello *State-of-the-Art Report* su *Algorithmic decision-making* di *Algo:aware*, studio del dicembre 2018 commissionato dal *DG Connect (Directorate General for Communications Networks, Content and Technology)*, Dipartimento della Commissione Europea, l'agoritmo decisionale (*decision making algorithm*) viene definito più in generale come "*A software system – including its testing, training and input data, as well as associated governance processes – that, autonomously or with human involvement, takes decisions or applies measures relating to social or physical systems on the basis of personal or non-personal data, with impacts either at the individual or collective level.*" e di conseguenza "*the definition of algorithmic decision-making is to be interpreted as a decision taken by a decision-making algorithm*".

[3] S. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2009. Per approfondire, quanto all'ambito più strettamente giuridico, appare utile fare rinvio anche a M. DURANTE, voce *Intelligenza artificiale. Applicazioni giuridiche*, in *Dig. it. Terza appendice di aggiornamento della IV edizione. Disc. Privatistiche*, vol. II, II Agg., 714 ss.

namente'[4]; in grado, cioè, di risolvere problemi imitando quello che sarebbe il comportamento umano in circostanze analoghe. Un sistema intelligente è capace di raccogliere dati da un certo *data-base* o ambiente di riferimento, interpretarli[5] e – alla luce dell'obiettivo da raggiungere – decidere quale sia l'azione o la decisone migliore, quindi agire di conseguenza[6] in modo pressoché automatico[7]. Questo processo decisionale è condotto dalla macchina – a seconda delle tecniche di AI implementate – applicando schemi ragionativi statici ovvero ricorrendo a tecniche di apprendimento automatico (e.g. *machine learning*, *deep learning, reti neurali, decision trees* e altre). Senza entrare nel dettaglio, basti sapere – ai nostri fini – che in quest'ultimo caso la macchina, anziché eseguire regole di comportamento pre-impartite in modo definito, elabora 'da sola' e dinamicamente – in applicazione di algoritmi di apprendimento e adattivi – la regola decisionale; in taluni casi risultando, perciò, anche in grado di rispondere e adattarsi meglio ai cambiamenti dell'ambiente o affinare, con l'esperienza di utilizzo, la capacità di generare un *output* adeguato[8].

Queste tecniche fanno dell'IA uno strumento incredibilmente utile, capace di giungere a decisioni e fare predizioni molto prima ed in modo più accurato di quanto farebbe l'essere umano, per di più in contesti che richiedono l'analisi di enormi quantità di dati – spesso anche non strutturati[9]– altrimenti proibitiva per l'agente umano. I dati sono infatti la materia prima dei processi algoritmici. La diffusione capillare di reti di dati e infrastrutture ICT come *Internet*, l'ascesa di social media e piattaforme digitali, rendono possibile una raccolta massiva di dati da varie fonti – esseri umani, macchine, organizzazioni – forniti spontaneamente, frutto di osservazione o tracciamento (si pensi al diffondersi esponenziale di sensori e del cd. *Internet of Things*), frutto di inferenza: è il caso dei cd. dati derivati e della *Big Data Analytics*[10]. è questa enorme disponibilità di dati[11], unita alla capacità dell'Intelligenza Artificiale di ricavarne 'valore' individuandovi correlazioni – facendo *clustering*, estraendo quindi dai dati, in maniera automatica, informazioni che possono mettere in evidenza tendenze, fenomeni, prevedere esiti e comportamenti[12], profilare utenti[13] – che ha dato e da slancio ad investimenti e applicazioni dell'Intelligenza artificiale in un numero sempre crescente di settori, tanto privati che pubblici, da Banche, Finanza e settore assicurativo, fino al Fisco, l'assistenza sanitaria, l'Agricoltura, l'Ambiente, il Marketing. In tutti questi ambiti, enorme appare il potenziale dell'IA quale *boo-*

---

[4] M. SOMALVICO, F. AMIGONI, V. SCHIAFFONATI, *Intelligenza Artificiale*, in S. PETRUCCIOLI (a cura di), *Storia della scienza*, vol. IX, Roma, Istituto della Enciclopedia Italiana, 2003, 615-624; AA.VV., *Artificial Intelligence and life in 2030*, *One hundred year study on Artificial Intelligence*, Stanford University Press, 2016, p. 5.

[5] Ragionando su di essi o elaborando le informazioni desunte dai dati.

[6] Modificando eventualmente l'ambiente (naturale o virtuale) in cui opera o comunque proponendo un certo output, la soluzione ad uno specifico problema

[7] Cfr. la definizione elaborata dal Gruppo di esperti istituito dalla Commissione europea in High-Level Expert Group on Artificial Intelligence (AI HLEG), *A definition of AI: Main capabilities and scientific disciplines*, Brussels, 18 dicembre 2018.

[8] *Amplius*, v. Y. BATHAEE, *The Artificial Intelligence Black Blox And The Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, Vol. 31, n. 2, 2018, 890 ss.

[9] È il caso degli algoritmi di auto-apprendimento.

[10] Si veda, anche con riferimento alle rinnovate esigenze di tutela che ciò comporta, F. PIZZETTI, *La protezione dei dati personali e le sfide dell'intelligenza artificiale*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino 2018 (*I diritti nella 'rete' della rete*), 3 ss. (in particolare da p. 40).

[11] Si è parlato infatti della nostra come di una *data-driven society*; PENTLAND, A., *The data-driven society*, in *Scientific American*, Vol. 309, n. 4, 2013, 78-83, stable URL: *https://www.jstor.org/stable/10.2307/26018109.*

[12] M.C. CARROZZA et al., *Automazione e Autonomia: dalla definizione alle possibili applicazioni dell'Intelligenza Artificiale*, in FONDAZIONE LEONARDO, *Statuto Etico e Giuridico dell'IA*, 2019.

[13] Per una definizione e un approfondimento sul tema dell'*automated profiling*, M. HILDEBRANDT, *Defining Profiling: A New Type of Knowledge?*, in M. HILDEBRANDT, S. GUTWIRTH (eds.) *Profiling the European Citizen*, Springer, 2008. *https://doi.org/10.1007/978-1-4020-6914-7_2.*

ster della prosperità umana; come strumento promettente per migliorare il benessere individuale e sociale e il bene comune nonché favorire progresso e innovazione [14].

Qual è, dunque, il problema? L'IA può essere usata, com'è ovvio, per scopi criminali; può essere intrusiva, violare la riservatezza, la vita privata delle persone; soprattutto, può sbagliare [15]. L'uso di sistemi di *Automated Decision Making* può portare a decisioni distorte, dannose per l'essere umano, discriminatorie e lesive dei diritti fondamentali della persona; mette a rischio il diritto dell'uomo ad avere un libero ed equo accesso a beni e servizi,all'in formazione; a non essere manipolato a sua insaputa da forme di *marketing* scorrette, aggressive e invadenti; minaccia, come vedremo, la legittima pretesa a contestare le decisioni lesive assunte da una macchina o, soprattutto, a che queste non siano assunte in modo poco trasparente e sulla base, magari, di dati errati e incompleti [16].

Tutto ciò lo si è compreso bene, forse, dallo scandalo di Cambridge Analytica quando è stato chiaro che i nostri 'Mi Piace' su *Facebook* – meglio, i profili psicometrici che potevano ricavarvisi – oltre che per 'invadere' la nostra privacy, potevano essere utilizzati per 'orientare' le opinioni politiche degli utenti [17] attraverso un *microtargeting* comportamentale. Ma i casi sono innumerevoli; arcinoto quello di COMPAS – acronimo di *Correctional Offender Management Profiling for Alternative Sanctions* – sistema di *machine learning* usato da diversi stati americani fin dal 2001 quale *tool* a servizio del giudice per valutare il rischio di recidiva dei condannati e che ha dimostrato, a parità di altre condizioni, di essere discriminatorio nei confronti dei criminali neri afroamericani. Ancora due, curiosi: il sistema di raccomandazione di Amazon [18], nel 2017, che agli avventori che cercavano di acquistare un certo agente chimico sulla piattaforma, suggeriva tra gli articoli 'spesso comprati insieme' altri prodotti che, combinati, avrebbero potuto essere utilizzati per fabbricare un esplosivo artigianale: addirittura un set di pallini per massimizzare la letalità dell'ordigno [19]; il caso di Tay, il *chatter bot* capace di auto-apprendimento sviluppato da Microsoft e lanciato su Twitter nel 2016, "diventato" razzista e antisemita in meno di un giorno [20].

---

[14] sul possibile valore aggiunto di una Intelligenza Artificiale al servizio dell'uomo e della società, si veda lo Studio dell'European Parliamentary Research Service (EPRS), *European framework on ethical aspects of artificial intelligence, robotics and related technologies. European added value assessment*, del Settembre 2020, per cui, in sintesi, "*The analyses of this European added value assessment suggest that a common EU framework on ethics has the potential to bring the European Union €294.9 billion in additional GDP and 4.6 million additional jobs by 2030*".

[15] Cosí si apre il *Libro Bianco sull'Intelligenza Artificiale* della Commissione Europea – COM(2020) 65 final del febbraio 2020: "*L'intelligenza artificiale* […] *cambierà le nostre vite migliorando l'assistenza sanitaria (ad esempio rendendo le diagnosi più precise e consentendo una migliore prevenzione delle malattie), aumentando l'efficienza dell'agricoltura, contribuendo alla mitigazione dei cambiamenti climatici e all'adattamento ai medesimi, migliorando l'efficienza dei sistemi di produzione mediante la manutenzione predittiva, aumentando la sicurezza dei cittadini europei e in molti altri modi che possiamo solo iniziare a immaginare. Al tempo stesso, l'intelligenza artificiale (IA) comporta una serie di rischi potenziali, quali meccanismi decisionali opachi, discriminazioni basate sul genere o di altro tipo, intrusioni nelle nostre vite private o utilizzi per scopi criminali*.".

[16] Tra gli altri, v. B. CASEY, *Title 2.0: Discrimination Law in a Data-Driven Society*, in *J. L. & MOB.*, 2019, p. 36, *https://doi.org/10.36635/jlm.2019.title*; K. CRAWFORD, *The Hidden Biases in Big Data*, in *Harv. Bus. Rev.*, 2013, *https://perma.cc/E95C-TUQU*; si veda anche il nostro E. TROISI, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità' dell'algoritmo*, in *European Journal of Privacy Law & Technologies*, 1/2019, online: *https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1027/276*.

[17] Per approfondire la vicenda: *https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump; https://www.nytimes.com/2018/03/18/us/cambridge-analytica-facebook-privacy-data.html?*.

[18] Si veda *www.channel4.com/news/potentially-deadly-bomb-ingredients-on-amazon.*

[19] L'algoritmo in questione funziona secondo una logica nota come *collaborative filtering*: se un numero di persone considerevole compra sia X che Y, i due prodotti saranno con grande probabilità correlati; cfr. M. AIROLDI, D. GAMBETTA, *Sul mito della neutralità algoritmica*, in *The Lab's Quarterly*, XX, 4, 2018, 25.

[20] *https://www.theguardian.com/world/2016/mar/29/microsoft-tay-tweets-antisemitic-racism*.

Si è ormai infranto l'iniziale mito della neutralità algoritmica [21]; le macchine intelligenti sono create, programmate e addestrate ('*trained*') da operatori umani e perciò come l'uomo sono fallibili, possono 'ragionare', o 'imparare a ragionare' in modo distorto [22]. Le regole date alla macchina dal suo programmatore – con cui poi questa interpreta la conoscenza e si muove nel suo ambiente di riferimento – possono rifletterne pregiudizi culturali, anche benevoli e non intenzionali [23]. Nei sistemi basati sui dati, come i sistemi di apprendimento automatico, poi, distorsioni possono derivare anche dalla raccolta dei dati, che possono essere insufficienti, incompleti o viziati; dall'addestramento, per *bias* indotte dall'interpretazione umana dei dati, in caso di algoritmi ad apprendimento supervisionato, o – quando *un-supervised* – a causa dell'apprendimento online e dell'auto-adattamento tramite l'interazione con gli utenti (si pensi ai comuni algoritmi di *ranking*). Queste distorsioni possono condurre a decisioni inique [24].

2. – Altro – generale e più grande problema, e che qui maggiormente interessa – è quello dell'opacità, che si riflette sulla possibilità di sindacare la decisione assunta con mezzi automatici [25].

I set di dati, i processi che determinano la decisione degli algoritmi, il perché di una certa decisione che incida la sfera giuridica di una persona dovrebbero essere tracciabili, trasparenti, spiegati; ciò anche per mettere in condizione l'interessato di contestarne i contenuti. Non sempre invece lo sono, o lo sono adeguatamente: vuoi per scelta [26] – per ragioni competitive, di tutela del *know-how* – o per limiti tecnologici: è il caso di quegli algoritmi che propriamente chiamiamo '*black-box*', sistemi i cui meccanismi inferenziali non sono (completamente) prevedibili *ex ante* [27] o che, comunque, non rendono sempre possibile spiegare perché un modello abbia generato un particolare risultato o decisione (e quale combinazione di fattori vi abbia contribuito) [28].

---

[21] M. AIROLDI, D. GAMBETTA, *op. cit*., 29.

[22] J. BURRELL, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, 1/2016, 1-12, per cui l'affermazione che gli algoritmi classifichino le informazioni, e quindi decidano, in modo più "oggettivo", non può essere presa alla lettera dato il grado di giudizio umano coinvolto nella progettazione degli algoritmi stessi, in particolare sotto il profilo della definizione dei criteri di *clustering*, la pre-classificazione dei dati di allenamento e la regolazione di soglie e parametri decisionali.

[23] Su come i sistemi di classificazione possano essere e siano concretamente influenzati, anche con conseguenze di notevole portata, dal 'punto di vista' di chi li costruisce, si veda l'interessante lavoro di G.C. BOWKER, S.L. STAR, *Sorting Things Out: Classification and Its Consequences*, Cambridge, MA.

[24] Si veda, sul punto, High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 8 April 2019; FONDAZIONE LEONARDO, *Statuto Etico e Giuridico dell'IA*, 2019.

[25] Tra gli altri, J. BURRELL, *op. cit.*, per cui "*opacity seems to be at the very heart of new concerns about 'algorithms'*(operating on data) *among legal scholars and social scientists*".

[26] Il fortunato appellativo di 'Black Box Society' si deve a Frank Pasquale, che magistralmente ne delinea i tratti con espressioni altrettanto evocative quali '*the Secret Judgments of Software*' e '*the Secrecy of Business and the Business of Secrecy*'; cfr. F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard, 2015.

[27] Ma è talvolta possibile – va sottolineato – studiarne il comportamento analizzando le risposte che esso produce a fronte delle sollecitazioni che riceve. Cosiddetti 'explanatory tools'(o post-hoc explanation techniques, c.f.r. J. ZHONG, E. NEGRE, *AI: To interpret or to explain?,* INFORSID, 2021) sono in grado di ricostruire ex-post il funzionamento di taluni modelli decisionali 'opachi'; in particolare risultati del modello esaminato verrebbero spiegati trovando i collegamenti tra le caratteristiche dei dati di input e i risultati o costruendo un modello più semplice per approssimare il modello originale (*ivi*, p. 6); la precisione e affidabilità di queste 'spiegazioni' viene contestata, ad esempio, da C. RUDIN, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, Vol. 1, n. 5, 2019, 206-215.

[28] J. BURRELL, *op. cit.*, p. 3, con cui si può facilmente concordare, distingue 3 tipologie di opacità degli algoritmi decisionali, che così descrive: "*Three distinct forms of opacity include: (1) opacity as intentional corporate or institutional self-protection*

L'insorgenza di questi problemi e la crescente preoccupazione, anche dovuta al timore che la diffidenza verso i nuovi strumenti tecnologici potesse limitarne la diffusione nel mercato, ha portato all'adozione di una serie di Carte Etiche – di derivazione pubblica o anche privata – a tutti i livelli: internazionale, europeo, nazionale [29].

Muovendo dall'idea secondo cui – poiché la tecnologia digitale diventa una parte sempre più centrale di tutti gli aspetti della vita umana – le persone dovrebbero potersi fidare di tale tecnologia [30] e questa debba essere sviluppata al servizio dell'uomo, essere etica e rispettare i diritti fondamentali, si sono individuati una serie di principi e requisiti cui i sistemi intelligenti, le rispettive applicazioni, i produttori, programmatori ed utilizzatori dovrebbero attenersi. Leva, o una delle principali leve, di questa strategia [31] per una IA '*human-centric'* e '*trustworthy*', per usare le parole del Gruppo di Esperti sull'intelligenza artificiale nominato dalla Commissione europea [32], è proprio la trasparenza [33]. Spesso, presupposto fondamentale per garantire che i diritti umani fondamentali e i principi etici siano rispettati, protetti e promossi [34].

Se l'I.A. deve essere sviluppata al servizio dell'uomo, l'uomo dev'essere messo in grado di servirsene

---

and concealment and, along with it, the possibility for knowing deception; (2) opacity stemming from the current state of affairs where writing (and reading) code is a specialist skill and; (3) an opacity that stems from the mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation". V.si anche Y. BATHAEE, *op. cit.*; D. CASTELVECCHI, *Can We Open the Black Box of AI*, in *Nature*, Vol. 538, 2016, 20 ss.

[29] Un tentativo, ben riuscito ma già non completo, di mappare le varie Carte etiche, Dichiarazioni di principi o Linee Guida, suddivise per contesto geo-politico e analizzate per contenuti, lo si deve a A. JOBIN, M. IENCA, E. VAYENA, *Artificial Intelligence: the global landscape of ethics guidelines*, in *Nat. Mach. Intell*., 2019. Per citarne qualcuno tra i più rilevanti: Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, United Nations Educational, Scientific and Cultural Organization (UNESCO), *First Draft Of The Recommendation On The Ethics Of Artificial Intelligence*, SHS/BIO/AHEG-AI/2020/4 REV.2, 7 September 2020; Council of Europe, *Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*, CM/Rec(2020)1; più risalente e limitata al settore Giustizia: European Commission For The Efficiency Of Justice (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, December 2018; nell'ambito dell'Unione Europea: High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 8 April 2019; in Italia si segnala il documento a cura della Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale, *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino*, 2018. Tra le dichiarazioni di principi di matrice privata, celebre quella redatta in seno al Future of Life Institute e sottoscritta da circa 1800 ricercatori e quasi 4000 altri *endorsers*, alcuni celebri come Stephen Hawking o colossi del digital market come Elon Musk e Jaan Tallinn: *The Asilomar AI Principles*, 2017, *https://futureoflife.org/ai-principles/*.

[30] Così la Commissione Europea, *White Paper on Artificial Intelligence. A European approach to excellence and trust*, COM(2020) 65, 19 febbraio 2020.

[31] Europea; si veda Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions, *Building Trust in Human-Centric Artificial Intelligence*, COM(2019) 168 final, 8 april 2019.

[32] Si veda il già citato High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 8 April 2019.

[33] Così nel già citato documento dello Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, United Nations Educational, Scientific and Cultural Organization (UNESCO), *First Draft of The Recommendation on the Ethics of Artificial Intelligence*, SHS/BIO/AHEG-AI/2020/4 REV.2, 7 September 2020, III (2) §39, secondo cui cui "*transparency may contribute to trust from humans for AI systems".* Si veda anche il documento stilato dal National Institute of Standards and Technology (NIST) statunitense: AA.VV., *Four Principles of Explainable Artificial Intelligence*, disponibile online, nella versione di bozza, col seguente doi: *https://doi.org/10.6028/NIST.IR.8312-draft*.

[34] Cosí, testualmente, Ad Hoc Expert Group (AHEG) UNESCO, Ibid., III (2) §37. Sul rapporto tra trasparenza e fiducia v.si H. FELZMANN, E. FOSCH-VILLARONGA, C. LUTZ, A. TAMÓ-LARRIEUX, *Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns*, in *Big Data & Society*, 1/2019, 1-14; J. SCHOEFFERA, Y. MACHOWSKIA, N. KUEHLA, *A Study on Fairness and Trust Perceptions in Automated Decision Making*, 2021, online su *arXiv*: *arXiv:2103.04757v1*.

consapevolmente [35]: ha il diritto di essere sempre a conoscenza del fatto che sta interagendo con un sistema di IA [36]; deve poterne capire lo scopo, le capacità e le modalità di funzionamento (si parla di '*explicability*') e le decisioni, per quanto possibile, devono poter essere spiegate a coloro che ne sono direttamente o indirettamente interessati [37] (si parla di '*explainability*' [38]) affinché possano (c.f.r. '*contestability*') sindacarne le modalità e il contenuto, potendo a questo scopo ricorrere all'intervento umano.

3. – Posto, dunque, un pacifico dovere etico di trasparenza dell'algoritmo [39] e di spiegazione della (singola) decisione raggiunta con mezzi automatizzati, occorre chiedersi se esista – e in che termini (cioè con quale ampiezza) – un corrispondente diritto sul piano dello *ius positum* – e quali siano i suoi limiti, di natura giuridica ma anche tecnologica.

Quadro giuridico generale [40] di riferimento, almeno nei casi in cui il trattamento decisionale automatizzato riguardi dati personali è dato dal Regolamento (EU) 2016/679 (in sigla, GDPR) [41]. L'articolo 22 del Regola-

---

[35] AA.VV., *Paper sui Principi etici*, in FONDAZIONE LEONARDO, *Statuto Etico e Giuridico dell'IA*, 2019.

[36] Così §78, High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI, 8 April 2019..

[37] Così High-Level Expert Group on Artificial Intelligence (AI HLEG), IBID., §53.

[38] Alla explainability si riferisce il punto III (2) §40 della richiamata Proposta di Raccomandazione UNESCO sulla'etica dell'IA., Cit., per cui *"Explainability refers to making intelligible and providing insight into the outcome of AI systems. The explainability of AI systems also refers to the understandability of the input, output and behaviour of each algorithmic building block and how it contributes to the outcome of the systems. Thus, explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should be understandable and traceable, appropriate to the use context".*

[39] Secondo quanto risulta dallo studio condotto nel 2019 da A. JOBIN, M. IENCA, E. VAYENA, *op. cit.*, 7 ss., 'Transparency' è il principio più diffuso nella letteratura attuale, presente in 73 degli 84 documenti analizzati, sebbene con diverse sfumature di significato, riassunte dalle locuzioni: 'transparency, explainability, explicability, understandability, interpretability, communication, *disclosure, showing'.*

[40] Disposizioni specifiche possono trovarsi, invece, ad esempio, nell'ambito della normativa a tutela dei consumatori; per una panoramica della situazione e dei problemi specifici, si v, tra gli altri, M. GROCHOWSKI, A. JABLONOWSKA, F. LAGIOA, G. SARTOR, *Algorithmic transparency and explainability for EU Consumer Protection: unwrapping the regulatory premises*, in *Critical Analysis of Law*, Vol. 8, n. 1, 2021, 43-63.

[41] Il 21 aprile 2021 la Commissione UE ha pubblicato una proposta di Regolamento per l'intelligenza artificiale, il c.d. AI Act, con cui si intende definire il quadro regolatorio per lo sviluppo e immissione in commercio di sistemi di intelligenza artificiale. Secondo un approccio regolatorio progressivo basato sul rischio – che può essere inaccettabile (per cui il sistema viene vietato), alto, oppure basso o minimo – i sistemi IA considerati ad alto rischio saranno consentiti sul mercato europeo subordinatamente al rispetto di determinati requisiti obbligatori e previa certificazione di conformità. Tra questi requisiti c'è la trasparenza. L'art.13 del Proposal dispone infatti che i sistemi 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 forniti con precise istruzioni per un uso appropriato. Per gli altri sistemi, reputati a rischio minore, sono imposti soltanto obblighi di trasparenza molto limitati, ad esempio in termini di fornitura di informazioni per segnalare l'utilizzo di un sistema di IA nelle interazioni con esseri umani; peraltro con deroghe ampie e già oggetto di diverse critiche (si pensi ad esempio alla non applicabilità dell'obbligo di trasparenza ai sistemi di riconoscimento biometrico usati a fini di prevenzione e accertamento di reati). Non è questa la sede adatta per estendere il discorso all'analisi dell'impianto normativo dell'AI Act; tuttavia, non si possono trascurare alcuni brevi rilievi. Innanzitutto, posto che l'Intelligenza artificiale richiede un massivo trattamento di dati, molto spesso personali, è di tutta evidenza che il futuro Regolamento sull'AI dovrà funzionare in modo complementare al GDPR, armonizzandosi con le previsioni e gli obblighi di quest'ultimo, eventualmente specificandolo e/o integrandolo con riferimento alle peculiarità dei sistemi 'intelligenti'; tuttavia – proprio in un punto di un chiaro raccordo tra le due discipline (anche di principio) – il Proposal sembra carente, col rischio di ingenerare vuoti normativi (anche gravi) o incertezze, finanche in ordine ai ruoli e alle competenze delle Autorità di controllo coinvolte. Oltre però alle perplessità connesse alla mancanza di un'adeguata armonizzazione col Regolamento generale per la protezione dei dati (per le censure si rinvia al Parere congiunto n.5/2021 di EDPB e GEPD sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale; online qui: *https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en*), ciò che preoccupa della proposta in commento è l'ina-

mento [42] espressamente sancisce il diritto dell'interessato a "*non essere sottoposto ad una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione* [43]*, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona*". Il divieto, ai sensi del paragrafo 2 dello stesso articolo, non si applica solo se, e nella misura in cui i trattamenti decisionali automatizzati sono necessari per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento oppure siano basati sul consenso dell'interessato [44].

Sorvolando sulla portata di questa deroga – peraltro non l'unica [45] – che ammette, sostanzialmente,

---

deguatezza, sotto il profilo della previsione di meccanismi di tutela (anche per le molte deroghe) e redress – cioè meccanismi che consentano agli utenti che subiscono decisioni automatizzate pregiudizievoli di ottenere un rimedio diretto – (assenti), a superare la disciplina dell'art. 22 GDPR, che non copre l'intero ambito di applicazione dei sistemi decisionali automatizzati potenzialmente dannosi ed è – come si vedrà –per molti versi precaria, perciò oggetto di annosa critica in dottrina. Anche sotto lo specifico profilo della trasparenza, non essendo sempre possibile – anche per i limiti tecnologici di cui diremo *infra* – ottenere una previa spiegazione comprensibile dei risultati prodotti dalla macchina, l'EDPB e il GEPD non mancano di segnalare (v. parere congiunto n.5/2021, cit., spec. §72) come il regolamento sull'Intelligenza artificiale dovrebbe promuovere modalità nuove, più proattive e tempestive per informare gli utenti dei sistemi di IA in merito allo status decisionale in cui si trova il sistema in ogni momento, predisponendo allarmi rapidi su risultati potenzialmente nocivi affinché le persone i cui diritti e le cui libertà potrebbero essere compromessi dalle autonome decisioni della macchina siano in grado di reagire o impugnare le decisioni in questione.

[42] Non si tratta, peraltro, di un'assoluta novità legislativa, posto che il problema del trattamento decisionale automatizzato era già affrontato, con una previsione di divieto meno ampia di quella attuale, dall'art. 15 e dal considerando 41 della Direttiva 95/46/CE. *Amplius*, tra gli altri, L.A. BYGRAVE, *Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, in *Comput. Law Secur. Rev.*, 17, 2001, 17-24.

[43] Di trattamento automatizzato il Regolamento non dà una precisa definizione, l'art. 22 (1) vi include però espressamente la profilazione. Ai sensi dell'art. 4, per profilazione, nell'ambito del GDPR, si intende qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di questi per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare e 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". La profilazione è quindi una specie di trattamento automatizzato che consiste nell'esercizio, da parte di un software, un motore inferenziale, di un'operazione deduttiva capace di ricavare cioè da una certa quantità di dati inseriti in un data-set – analizzati e individuate tra loro, automaticamente, delle correlazioni statistiche – altri dati relativi a caratteristiche o schemi comportamentali attribuibili ad un determinato individuo o gruppi di essi al fine di classificarlo/i in precisi gruppi o categorie e/o predirne probabili comportamenti futuri.

[44] Va sottolineato che sono in ogni caso vietati i trattamenti automatizzati e le profilazioni che coinvolgano i dati sensibili di cui all'art. 9 del GDPR, salvo che l'interessato abbia prestato il proprio consenso esplicito o questo sia consentito da apposita base normativa nel diritto dell'Unione o degli Stati Membri e purché, anche in questo caso, il trattamento sia necessario per motivi di interesse pubblico e proporzionato alla finalità perseguita. I dati trattati poi, ai sensi dell'art. 5 del Regolamento, perché il trattamento (automatizzato) sia lecito, devono, com'è noto, essere raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità; come è stato opportunamente rilevato, l'apprendimento automatico alla base di sistemi ML coinvolge *set* di dati esistenti ed altri flussi di dati in tempo reale in processi dinamici complessi e dagli esiti a volte non completamente prevedibili, difficili da conciliare con l'esigenza di una compiuta individuazione anticipata delle finalità del trattamento cui acconsentirsi *ex ante*, al punto che taluno arriva a revocare in dubbio la stessa possibilità di esprimere un consenso realmente informato a trattamenti affidati a sistemi *black-box* (v. C. KUNER, D.J.B. SVANTESSON et al., *Machine learning with personal data: is data protection law smart enough to meet the challenge?*, in *International Data Privacy Law*, vol. 7, n. 1, 2017, 1 ss.).

[45] il GDPR ammette pure, infatti, che possa essere consentito – ai sensi del paragrafo 2, lettera b) dell'articolo in esame – adottare decisioni sulla base di un trattamento automatizzato in altre singole ipotesi previste da apposita norma autorizzativa del diritto dell'Unione o dello Stato Membro cui è soggetto il titolare del trattamento; ciò eccezionalmente pur in assenza del consenso dell'interessato e sempre che siano, anche qui, adottate misure adeguate alla tutela delle persone fisiche coinvolte. Secondo il Considerando 71 del Regolamento queste ipotesi possono includere, ad esempio, l'utilizzo di trattamenti informatizzati a fini di monitoraggio o prevenzione di frodi od evasione fiscale, ma in ogni caso deve ritenersi necessario – a parere di chi scrive – che il ricorso a tali meccanismi risulti giustificato alla luce del perseguimento di scopi di rilevante interesse pubblico e la relativa compressione dei diritti e delle libertà dei soggetti coinvolti sia proporzionata. In questi casi il Regolamento si limita a prescrivere l'obbligo per il legislatore di prevedere misure adeguate a tutela dei singoli ma il diritto a richiedere l'intervento umano e a contestare eventualmente la decisione non risulta assicurato dal paragrafo 3 dell'art. 22 (come invece negli altri casi esaminati), seppur nulla toglie, ovviamente, che le norme autorizzative di tali trattamenti, nel prevedere le salvaguardie richieste, possano predisporre forme di tutela anche più ampie.

l'*Automated Decision Making* in presenza di tutte quelle condizioni autorizzative per così dire "consensuali" [46] – che sarebbero idonee, almeno in teoria, ad assicurare una più ampia consapevolezza dell'interessato [47] – si prevede che nelle ipotesi in cui il trattamento decisionale automatizzato sia consentito, il titolare del trattamento ha il dovere – ai sensi del paragrafo 3 dell'esaminando art. 22 GDPR – di adottare misure appropriate per tutelare i diritti, le libertà e i legittimi interessi del *data subject*, e tra questi, in particolare, è tenuto in ogni caso a garantirgli il diritto ad ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestarne la decisione.

Ulteriori prescrizioni per il caso di trattamenti decisionali automatizzati così come definiti dall'art. 22 sono contenute negli articoli 13 (2)(f),14 (2)(g) e 15 (1)(h) del Regolamento.

Queste disposizioni, con formulazione identica, prevedono – nel caso degli articoli 13 e 14 nell'ambito dell'informativa, nel caso dell'art. 15 quale portato del diritto di accesso – che l'interessato ha il diritto di essere informato circa "*l'esistenza di un processo decisionale automatizzato, compresa la profilazione, di cui all'art. 22, paragrafi 1 e 4, e, almeno in tali casi,* [a ricevere, *n.d.a.*] *informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato*".

Sulla base di queste norme, il titolare del trattamento (decisionale automatizzato, s'intende) è tenuto dunque a uno specifico dovere di informazione, e quindi di trasparenza (che coinvolga: esistenza di un processo decisionale automatizzato/informazioni significative sulla logica dell'algoritmo/importanza e conseguenze previste di tale trattamento per l'interessato); e se è vero – come è stato sostenuto – che nell'intero discorso prescrittivo del GDPR, l'obiettivo della sicurezza della circolazione dei dati personali risulta affidato a una strategia incentrata, anzitutto, sulla trasparenza dei processi di trattamento dei dati [48], è evidente che il dovere di trasparenza in questione deve interpretarsi, in ogni caso, come un *quid pluris* – qualcosa in più – rispetto al 'contenuto minimo' dell'Informativa già dovuta per qualsivoglia trattamento. Ogni altra interpretazione, di senso cioè 'riduttivistico', renderebbe, infatti, del tutto superflua la specificazione del legislatore europeo; senza contare che una norma di maggiore tutela ben si giustificherebbe per la maggiore lesività (potenziale) dell'ADM [49], che il legislatore decide per l'appunto di regolare con una previsione di divieto.

Percorrendo questa direttrice, l'informazione di cui agli articoli 13 (2)(f),14 (2)(g) e 15 (1)(h) del Rego-

---

Amplius, v. MALGIERI, G., *Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations*, in *Computer Law & Security Review*, 35(5), 2019, *http://dx.doi.org/10.1016/j.clsr.2019.05.002*.

[46] Si v. il nostro E. TROISI, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità' dell'algoritmo*, cit.

[47] Con la conseguenza che deve ritenersi illegittimo in tutti i casi in cui l'interessato non vi abbia acconsentito espressamente attraverso una propria consapevole manifestazione di volontà, o direttamente, o nell'ambito di un più complesso rapporto contrattuale, sul presupposto però, in quest'ultimo caso, che il trattamento automatizzato sia da considerarsi necessario alla conclusione o all'esecuzione dell'accordo. Questo in punto di diritto; per chi voglia approfondire invece il tema dell'efficacia del meccanismo del consenso informato in ambiente digitale ad assicurare la consapevolezza, e quindi la tutela dell'interessato, si fa rinvio a I.A. CAGGIANO, *Il consenso al trattamento dei dati personali nel nuovo Regolamento europeo. Analisi giuridica e studi comportamentali*, in *Osservatorio del diritto civile e commerciale*, 2018, 67-106; e al recente L. GATT, I.A. CAGGIANO, R. MONTANARI (eds.), *Privacy and Consent. A Legal and UX&HMI Approach for Data Protection*, Suor Orsola University Press, 2021, che affrontano il tema con approccio empirico e dignità statistica, accompagnando alla lettura giuridica l'analisi comportamentale.

[48] Cfr. R. MESSINETTI, *La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata*, in *Contratto e Impresa*, 3/2019, p. 861.

[49] Quando il trattamento automatizzato è anche idoneo a produrre effetti giuridici o comunque a incidere in modo analogo significativamente sulla persona i cui dati sono trattati, non è più soltanto la privacy del data-subject ad essere in pericolo ma il suo diritto a non essere discriminato, ad avere un libero ed equo accesso a beni e servizi, a non essere manipolato a sua insaputa da forme di marketing scorrette, aggressive e invadenti, ma anche la legittima pretesa a contestare le decisioni lesive di una macchina o, soprattutto, a che queste non siano assunte in modo poco trasparente e sulla base, magari, di dati errati e incompleti.

lamento (e quindi la 'logica dell'algoritmo' e le 'conseguenze previste') deve consistere certamente in qualcosa di più della indicazione in forma coincisa, chiara, facilmente accessibile ed intellegibile [50] delle categorie di dati trattati e delle finalità del trattamento, oltreché della modalità automatizzata dello stesso [51]. Il dubbio è se un tale dovere di trasparenza si spinga però fino a poter considerare giuridicamente sancito, sul piano del diritto positivo, un diritto del soggetto interessato dal trattamento alla '*disclosure*' dell'algoritmo – cioè quantomeno a qualcosa in più della generale logica funzionale implicata dal sistema, un diritto ad 'aprire la scatola nera' [52] e guardarci dentro – e, di più ancora, un diritto alla 'spiegazione' della specifica decisione che lo riguarda e che incida sulla sua sfera giuridica: le ragioni, le circostanze individuali per cui una precisa decisione, con un preciso contenuto, sia stata adottata dall'algoritmo nei suoi riguardi [53].

Com'è stato acutamente osservato [54] le prescrizioni di cui agli artt. 13(2)(f) e 14(2)(g), che impongono un dovere di trasparenza da assolversi, appunto, nell'ambito dell'informativa privacy – in un momento, quindi, logicamente precedente al trattamento decisionale – non possono che riferirsi alla generale funzionalità dell'algoritmo e al tipo di decisioni normalmente attese dal suo funzionamento con valutazione astratta ed *ex ante* ('conseguenze *previste*' dice infatti il testo normativo). Il contenuto del diritto conoscitivo dell'interessato – riferito a queste basi giuridiche – è quindi necessariamente limitato e sicuramente inidoneo a fondare un dovere di 'motivazione' della decisione automatica. Tuttavia il riferimento alle "informazioni significative", alla "importanza" e le "conseguenze previste" per l'interessato impongono al titolare del trattamento, se non certo una *disclosure* dell'algoritmo [55], quantomeno di fare luce, in modo accessibile a tutti, sui principali criteri di funzionamento dello stesso [56], fornendo anche "informazioni sul trattamento previsto o futuro,

---

[50] Considerando 39 GDPR.

[51] Per B. GOODMAN, S. FLAXMAN, *European Union Regulations on algorithmic decision-making and a "right to explanation"*, in *AI Magazine*, vol. 38, n. 3, 2017, 50-57 oppure online: *https://arxiv.org/abs/1606.08813v3*, "*It is reasonable to suppose that any adequate explanation would, at a minimum, provide an account of how input features relate to predictions, allowing one to answer questions such as: Is the model more or less likely to recommend a loan if the applicant is a minority? Which features play the largest role in prediction?*".

[52] l'espressione è di Messinetti; v. R. MESSINETTI, *op. cit.*

[53] Si tenga presente che il Considerando 71 del GDPR menziona espressamente il diritto alla "spiegazione" dell'interessato, in tali termini: "Il trattamento dovrebbe essere subordinato a garanzie adeguate, che dovrebbero comprendere la *specifica informazione all'interessato* e il diritto di ottenere l'intervento umano, di esprimere la propria opinione, di ottenere una *spiegazione della decisione* conseguita dopo tale valutazione e di contestare la decisione (il corsivo è nostro)." Come fanno notare però 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 *International Data Privacy Law*, vol. 7, n. 2, 2017, 76-99, dopo un'attenta analisi di tutti i lavori preparatori, tale formulazione, originariamente inserita nella proposta di Regolamento, è in seguito espunta dal testo infine sottoposto ad approvazione legislativa. Ricostruisce la genesi della disposizione anche C. DJEFFAL, *The normative potential of the European rule on Automated Decisions: a new reading for Art. 22 GDPR*, in *ZaöRV* 80/2020, 847-879.

[54] S. WACHTER, B. MITTELSTADT, L. FLORIDI, *op. cit.*

[55] nel senso della non necessarietà della *divulgazione dell'algoritmo*, per adeguarsi al livello minimo di trasparenza richiesto dal Regolamento, v.si Article 29 Working Party's *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, n. 251 del 3 Ottobre 2017, per cui, nel versione italiana del testo: "Il regolamento impone al titolare del trattamento di fornire informazioni significative sulla logica utilizzata, ma *non necessariamente una spiegazione complessa degli algoritmi utilizzati o la divulgazione dell'algoritmo completo* (il corsivo è nostro).".

[56] La complessità non è una scusa per non fornire informazioni all'interessato. Il Considerando 58 al GDPR afferma che il *principio di trasparenza* è "particolarmente utile in situazioni in cui la molteplicità degli operatori coinvolti e la *complessità tecnologica* dell'operazione fanno sì che sia difficile per l'interessato comprendere se, da chi e per quali finalità sono raccolti dati personali che lo riguardano, quali la pubblicità online (*corsivo* nostro)". Alla luce di un'interpretazione che guardi al contesto delle disposizioni regolamentari – in particolare ai requisiti di chiarezza, intelligibilità, accessibilità, semplicità di linguaggio che presidiano all'informazione cui ha diritto in via generale l'interessato soggetto a un trattamento dei propri dati personali ex art. 12 (1) GDPR – deve escludersi che il legislatore europeo voglia riferirsi ad un dovere di *mera 'disclosure' dell'algoritmo*, soprattutto se inteso nella sua veste mate-

nonché sulle possibili conseguenze del processo decisionale automatizzato sull'interessato"[57] Le informazioni fornite dovrebbero tuttavia essere sufficientemente complete perché l'interessato possa comprendere i motivi alla base della decisione[58] ed esprimere un valido consenso ai sensi dell'art. 4 GDPR[59]. Si pensi ad una pratica di finanziamento interamente gestita online per cui una banca si affidi ad algoritmi di profilazione per individuare il punteggio di affidabilità creditizia della persona richiedente (*credit scoring*) e decidere quindi se concedere o respingere il prestito e quale tasso d'interesse proporre[60]. Laddove la decisione si basi *unicamente* (art. 22 GDPR) o comunque *sostanzialmente* su tale punteggio, il titolare del trattamento dovrà spiegare all'interessato che sarà valutato da un algoritmo e come questo funziona – non la formula matematica, bensì in che modo il sistema 'ragiona' – quindi, quali informazioni sono prese in considerazione (es. *situazione lavorativa e reddituale; grado di indebitamento; storia creditizia; ecc*.); quali sono le fonti (es. *dati forniti dall'interessato nel modulo di domanda o ottenuti da terzi; dati già raccolti in occasione di precedenti rapporti; informazioni derivanti da pubblici registri; ecc*.); quali circostanze incidono maggiormente sulla decisione e in che modo (es. individuando una possibile casistica; facendo ricorso a simulazioni); come sono classificati gli utenti (quanti e quali sono i *ranks* considerati, in base al punteggio); come e quanto lo *score* incide sulla domanda di credito. Per rendere queste informazioni (*cfr*. i*mportanza* e *conseguenze previste* di tale trattamento) *significative* e comprensibili, dovrebbero essere forniti esempi reali e concreti del tipo di possibili decisioni ed effetti per l'interessato.[61] Il titolare del trattamento dovrà anche spiegare brevemente all'interessato quali misure adotta per evitare, ad esempio, malfunzionamenti e distorsioni dell'algoritmo, e informarlo che ha diritto a chiedere il riesame della decisione invocando l'intervento umano (*cfr.* art. 22(3) GDPR). Oltre a questo[62], dovrebbe anche chiarire che l'utente ha diritto a conoscere (avere *accesso* a) il *pun-*

---

matica come cioè *architettura di sistema*, dovendosi piuttosto ritenere che quello che qui interessa, e compete al titolare del trattamento esplicare – magari in aggiunta alla funzione matematica che sostanzia l'algoritmo in questione – è la sua *concreta implementazione*: il *contesto* in cui è applicato, gli *scopi perseguiti*, le *tecniche applicate* e magari anche degli *esempi circa il suo concreto funzionamento*. Solo allegando tali circostanze all'informativa dovrebbe potersi ritenere assolto quel *dovere "esplicativo"* che il legislatore sembrerebbe richiedere.

[57] In tal senso si esprime Article 29 Working Party, *op. cit.*, aggiungendo anche che "*per rendere queste informazioni significative e comprensibili, dovrebbero essere forniti esempi reali e concreti del tipo di possibili effetti*".

[58] Cosí, testualmente, Article 29 Working Party, *op. cit.*

[59] A tal proposito, si veda, ad esempio, la recente Cass., sez. I, 25 maggio 2021 n. 14381 (ord.) sulla validità del consenso reso dall'utente che non è venuto preventivamente a conoscenza dello schema esecutivo di un algoritmo (nel caso di specie un algoritmo di *rating*, impiegato da parte resistente per l'elaborazione di profili reputazionali degli utenti); nel cassare la pronuncia di merito la Suprema Corte, parlando proprio di "*scarsa trasparenza dell'algoritmo*" ha infatti affermato che "in tema di trattamento di dati personali, il consenso è validamente prestato solo se espresso liberamente e specificamente in riferimento a un trattamento chiaramente individuato" con la conseguenza che "nel caso di una piattaforma web (con annesso archivio informatico) preordinata all'elaborazione di profili reputazionali di singole persone fisiche o giuridiche, incentrata su un sistema di calcolo con alla base un algoritmo finalizzato a stabilire i punteggi di affidabilità, *il requisito di consapevolezza non può considerarsi soddisfatto ove lo schema esecutivo dell'algoritmo e gli elementi di cui si compone restino ignoti o non conoscibili da parte degli interessati*" (il corsivo è nostro).

[60] La fattispecie è evocata nel Considerando 71 del Regolamento.

[61] I corsivi si riferiscono alla lettera degli artt. 13(2)(f) e 14(2)(g) GDPR.

[62] Se fin qui si resta in area – per così dire – speculativa, oltre che generale, quando si tratti di attività procedimentale algoritmizzata della Pubblica Amministrazione, quando cioè sistemi decisionali automatizzati siano usati in sede decisoria pubblica, il contenuto minimo del dovere di trasparenza imposto al titolare del trattamento (in questo caso, un soggetto pubblico nell'esercizio di un potere autoritativo) è ben più definito (e – deve ritenersi – marcato), grazie al già intenso contributo interpretativo della giurisprudenza. Il Consiglio di Stato, in particolare, nella pronuncia n. 8472/2019 (v.si commenti di A. MASCOLO, *Gli algoritmi amministrativi: la sfida della comprensibilità*, in *Giorn. dir. amm.*, 3/2020, 366 ss.; M. TIMO, *Algoritmo. Il Procedimento di assunzione del personale scolastico al vaglio del Consiglio Di Stato*, in *Giur. it*., 2020, 5, 1190 ss.) – in parte richiamando ma superando il precedente n. 2270/2019 – con statuizione di principio, afferma (para 15.1), nell'interpretare quello che definisce *principio di conoscibilità*, ovvero il principio – introdotto dal GDPR – per cui per cui ognuno ha diritto a conoscere l'esistenza di processi decisionali automatizzati che

teggio assegnatogli dall'algoritmo e i motivi – cioè gli antecedenti logici – di questa classificazione; ma ci arriviamo subito.

Se, come detto, si può facilmente convenire che gli obblighi di trasparenza di cui agli artt. 13(2)(f) e 14(2)(g) non paiono idonei a fondare un vero e proprio diritto del *data subject* alla spiegazione *ex post* della decisione automatizzata, non convince invece un'interpretazione altrettanto restrittiva dell'art. 15(1)(h) del Regolamento, variamente fondata, in dottrina, sul dato letterale della disposizione [63] – identica alle preceden-

---

lo riguardino ed in questo caso a ricevere informazioni significative sulla logica utilizzata, che "*tale diritto* […] *va accompagnato da meccanismi in grado di decifrarne la logica. In tale ottica, il principio di conoscibilità si completa con il principio di comprensibilità* […]". Più in particolare, chiarisce come (para 13.1) "*il meccanismo attraverso il quale si concretizza la decisione robotizzata (ovvero l'algoritmo) debba essere 'conoscibile', secondo una declinazione rafforzata del principio di trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico. Tale conoscibilità dell'algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti";* la "formula tecnica" deve essere cioè corredata "*da spiegazioni che la traducano nella "regola giuridica" ad essa sottesa e che la rendano leggibile e comprensibile*". Ciò "*al fine di poter verificare che i criteri, i presupposti e gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione a monte di tale procedimento e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato*." L'assenza di un tale livello minimo di *comprensibilità* (possiamo dire, spiegazione) dell'algoritmo costituisce, di per sé, a parere del Consiglio di Stato, un vizio tale da inficiare l'intera procedura (e i suoi esiti provvedimentali) per esserne l'intero metodo censurabile per difetto di trasparenza. È chiaro tuttavia, e lo rende esplicito l'Estensore del citato provvedimento, che una tale interpretazione delle norme del Regolamento sulla protezione dei dati, di taglio generale e perciò applicabili sia a decisioni assunte da soggetti privati che, come in questo caso, da soggetti pubblici, risente però, in quest'ultimo caso, della concorrente efficacia dei principi già informanti l'agire pubblico: in particolare, il *principio di conoscibilità* di cui al GDPR, interpretato nei termini suddetti, costituisce nell'ipotesi di specie, più che altro, "*diretta applicazione specifica dell'art. 42 della Carta Europea dei Diritti Fondamentali ("Right to a good administration"), laddove afferma che quando la Pubblica Amministrazione intende adottare una decisione che può avere effetti avversi su di una persona, essa ha l'obbligo di sentirla prima di agire, di consentirle l'accesso ai suoi archivi e documenti, ed, infine, […] di 'dare le ragioni della propria decisione*'". Più diffusamente, sulla trasparenza e la necessaria 'comprensibilità dell'algoritmo decisionale usato nell'esercizio dell'attività amministrativa, si veda G. FASANO, *L'intelligenza artificiale nella cura dell'interesse generale*, in *Giorn. dir. amm.*, 6/2020, 715 ss.

In Francia è il legislatore che con la LOI n. 2016-1321 du 7 octobre 2016 pour une République numérique (*https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte*) ha modificato il Code des relations entre le public et l'administration prevedendo l'introduzione dell'art. L. 311-3-1 per cui"*Sous réserve de l'application du 2° de l'article L. 311-5, une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées par l'administration à l'intéressé s'il en fait la demande*". Norma ulteriormente specificata con il Décret n. 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique (*https://www.legifrance.gouv.fr/eli/decret/2017/3/14/PRMJ1632786D/jo/texte*) che ha introdotto i seguenti Art. R. 311-3-1-1 ("*La mention explicite prévue à l'article L. 311-3-1 indique la finalité poursuivie par le traitement algorithmique. Elle rappelle le droit, garanti par cet article, d'obtenir la communication des règles définissant ce traitement et des principales caractéristiques de sa mise en œuvre, ainsi que les modalités d'exercice de ce droit à communication et de saisine, le cas échéant, de la commission d'accès aux documents administratifs, définies par le présent livre*") e Art. R. 311-3-1-2 ("*L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes: 1°) Le degré et le mode de contribution du traitement algorithmique à la prise de décision; 2°) Les données traitées et leurs sources; 3°) Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé; 4°) Les opérations effectuées par le traitement*). I principi sono peraltro oggetto della Décision n. 2018-765 DC du 12 juin 2018 del Conseil Constitutionnel che pare darne una lettura restrittiva. Disponibile al link: *https://www.conseil-constitutionnel.fr/decision/2018/2018765DC.htm*. Sulla decisione algoritmica nell'ordinamento francese si sofferma anche G. MALGIERI, *Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations*, cit., 13 ss. Si veda anche nota 102 *infra*.

[63] Cosi S. WACHTER, B. MITTELSTADT, L. FLORIDI, *op. cit.*; G. FINOCCHIARO, *Intelligenza artificiale e diritto. Intelligenza artificiale e protezione dei dati personali*, in *Giur. it*., 2019, 7, 1657, per cui non esisterebbe un pieno diritto alla spiegazione nell'art. 15 del Regolamento, ostandovi il dato letterale della disposizione, ipotesi di "mancata sintonia" fra la norma e il Considerando 71 del Regolamento.

ti [64] – o sull'assunto che il diritto d'accesso abbia il medesimo contenuto e la medesima finalità del diritto all'informativa [65]. Nel quadro delle norme analizzate – che il GDPR prevede a tutela della persona fisica in caso di *Automated Decision Making* – il potere di controllo attribuito all'interessato si declina, oltre che nel diritto all'informazione, in quello di ottenere l'intervento umano e soprattutto di contestare la decisione prodotta dal sistema automatizzato [66]. In questo sistema di salvaguardia dell'interessato, che fonda sulla volontà – sul consenso in senso lato – la liceità del trattamento decisionale automatizzato, l'informativa – come l'abbiamo delineata – assolve l'importante funzione di permettere all'interessato di esercitare, con cognizione, il suo diritto di gestione dei consensi [67], mentre il diritto d'accesso assommerebbe a questa medesima funzione anche quella, ulteriore (in questo specifico contesto), di consentire una contestazione effettiva – perché debitamente informata – della decisione nociva.

A dispetto di una formulazione pressoché identica delle disposizioni di cui agli articoli 13 (2)(f) e14 (2)(g), da un lato, e dell'art. 15 (1)(h), dall'altro, non è possibile farne un'interpretazione simmetrica: mentre nell'ambito dell'informativa le informazioni dovute non possono che riferirsi alla *generale funzionalità dell'algoritmo* e al tipo di *decisioni normalmente attese dal suo funzionamento*, nel caso del diritto d'accesso è più opportuno interpretare estensivamente la norma, traendone un dovere di informazione, in capo al titolare del trattamento, più pregnante e dettagliato e che attenga alla *specifica decisione* – eventualmente già adottata dal sistema nei confronti dell'interessato – e ai passaggi inferenziali che hanno portato (la macchina) a *quel* dato output: [68] una spiegazione *ex post*.

Riprendendo l'esempio del prestito online, è ragionevole pensare – su queste premesse – che il diritto d'accesso riconosciuto dal GDPR all'interessato – il nostro richiedente – si estenda allo *score*, e cioè al profilo, assegnatogli dall'algoritmo reputazionale che ne definirebbe l'affidabilità creditizia; questo è infatti, a tutti gli effetti – a ben vedere – un suo dato personale, *prodotto* di un processo di profilazione e dunque decisivo ai fini della domanda di credito cui ha interesse. Conoscere questo dato, oltre che un diritto dell'interessato (art. 15 GDPR), è anche essenziale perché possa, ad esempio, valutarne l'esattezza, chiederne magari la rettifica (o in generale esercitare gli ulteriori diritti sui propri dati di cui agli artt. da 16 a 20 del Regolamento) [69]; contestare

---

[64] Per S. WACHTER, B. MITTELSTADT, L. FLORIDI, *op. cit.*, la disposizione, cennando anche in questo caso alle "conseguenze previste" del trattamento decisionale farebbe pacificamente riferimento, anche qui, ad una 'spiegazione' necessariamente ex ante, e pertanto limitata.

[65] In questo senso, Article 29 Working Party, *op. cit.*, che, nella versione italiana del documento, conclude come segue: "*L'articolo 15, paragrafo 1, lettera h), afferma che il titolare del trattamento deve fornire all'interessato informazioni sulle conseguenze previste del trattamento, piuttosto che una spiegazione di una particolare decisione*".

[66] R. MESSINETTI, *op. cit.*; C. MALGIERI, G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, Vol. 7, n. 4; M.E. KAMINSKI, *The Right to Explanation, Explained*, in *Berkeley Technology Law Journal*, 34, 2019, 189-218.

[67] Inclusa la revoca, se del caso, ovvero l'esercizio del diritto di opposizione di cui all'art. 21 del GDPR, tanto per fare degli esempi; quest'ultimo, l'opposizione, particolarmente rilevante nella configurazione di cui al paragrafo 2 dell'art. 21 del Regolamento, quale diritto *incondizionato* dell'interessato ad opporsi al trattamento dei suoi dati personali per finalità di marketing diretto e quindi, in buona sostanza, alla profilazione (nella misura in cui sia connessa a tale marketing diretto), peraltro a prescindere dal fatto che abbia o meno luogo un processo decisionale *unicamente* automatizzato (e cioè che si rientri nell'applicazione dell'art. 22 GDPR). Di un ruolo dell'Informativa quale strumento volto ad avviare un controllo dell'interessato sul procedimento in cui si concreta l'attività di trattamento, parla I.A. CAGGIANO, *Il consenso al trattamento dei dati personali nel nuovo Regolamento europeo. Analisi giuridica e studi comportamentali*, *op. cit.*, che cita S. MAZZAMUTO, *Il principio del consenso e il potere della revoca* in AA.VV., *Libera circolazione e protezione dei dati personali*, a cura di R. PANETTA, t. I, Milano, 2006, 1004.

[68] Di questo avviso, tra gli altri: G. MALGIERI, G. COMANDÈ, *Right to legibility of automated decision-making*, cit.; T.W. KIM, B.R. ROUTLEDGE, *Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach,* in *Business Ethics Quarterly*, Vol. 32, Issue 1, 2022, 75-102, doi: *https://doi.org/10.1017/beq.2021.3*.

[69] Vedi nota 80 *infra*.

propriamente la decisione automatica (ad esempio la mancata concessione del finanziamento o condizioni indesiderabili dello stesso). Certo, il profilo – il *rank* assegnato al punteggio individuato dall'algoritmo – esprime comportamenti o caratteristiche *probabili*, *previste*, ma che in quanto giuridicamente rilevanti (per la domanda o le condizioni del credito) devono pur essere *ragionevolmente* desunte dai dati dell'interessato, che deve perciò essere messo in condizione di valutare questa ragionevolezza ed '*esprimere la propria opinione'* (cfr. art. 22(3)). Perché ciò accada, l'informazione che gli è dovuta dal titolare del trattamento (*ex* art 15) non può avere lo stesso contenuto – già visto – dell'Informativa *ex ante*; non può limitarsi a spiegare in modo comprensibile il presumibile funzionamento dell'algoritmo, piuttosto *dovrebbe* – nel nostro esempio – indicare lo *score* assegnato; individuarne i motivi concreti (cioè perlomeno le circostanze specifiche che hanno pesato nel giudizio); spiegare (o quantomeno riportare in modo 'intellegibile') perché quel punteggio abbia comportato il rigetto dell'istanza e magari quale diverso punteggio o caratteristica avrebbe consentito un esito diverso [70].

4. – La necessità di un'interpretazione tal fatta è, per chi scrive, la naturale conseguenza di una lettura sistematica e coerente del paragrafo 1, lettera h) dell'art. 15 del Regolamento. Come visto, infatti, l'art. 22 del GDPR, al paragrafo 3, prevede espressamente il diritto dell'interessato destinatario di una misura di ADM a mettere in discussione la decisione automatizzata attraverso la possibilità di ottenere l'intervento umano e cioè di relazionarsi in modo 'dialettico' col titolare del trattamento, esprimendo la propria opinione, chiedendo di procedere ad una verifica della decisione e potendo anche, successivamente, contestarne gli assunti. Si tratta di un diritto, a ben vedere, che, salvo volerne frustrare il contenuto sostanziale, sottende necessariamente, da parte dell'interessato (e quindi richiede al titolare del trattamento) un'informazione specifica sul funzionamento dell'algoritmo decisionale; un'informazione non astratta e prognostica, ma concreta ed *ex post* [71], calata nell'applicazione specifica che lo riguarda; solo in tal modo infatti l'interessato sarebbe messo in condizione di esercitare appieno il suo diritto oppositivo, disponendo degli strumenti idonei a muovere una contestazione specifica e motivata [72]. Non è a caso, infatti – a parere di chi scrive – che il Considerando

[70] L'ultimo riferimento è alla tecnica delle cd. counterfactual explanations; si legga in proposito l'interessantissimo S. WATCHER, B. MITTELSTADTD, C. RUSSELL, *Counterfactual explanations without opening the Black box: automated decisions and the GDPR*, in *Harvard Journal of Law & Technology*, Vol. 31, n. 2, 2018.

[71] Nel senso inteso da S. WACHTER, B. MITTELSTADT, L. FLORIDI, *op. cit.*, p. 6, per cui "*an ex post explanation occurs after an automated decision has taken place*" and "[…] *can address both system functionality and the rationale of a specific decision*". Cfr. anche T.W. KIM, B.R. ROUTLEDGE, *op. cit.*

[72] Tra gli altri, v. A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation*, in *International Data Privacy Law*, Vol. 7, Issue 4, 2017, 233-242. *https://doi.org/10. 1093/idpl/ipx022*. Parla di una trasparenza 'qualificata', da definirsi proprio in ragione del suo nesso funzionale con l'esercizio degli altri diritti riconosciuti al *data subject*, non ultimo quello a impugnare la decisione automatica: M.E. KAMINSKI, *op. cit.*, per cui "*there is a clear relationship between the other individual rights theGDPR establishes—contestation, correction, and erasure—and the kind of individualized transparency it requires. This suggests something interesting about transparency: the substance of other underlying legal rights often determines transparency's substance. If one has a right of correction, one needs to see errors. If one has a right against discrimination, one needs to see what factors are used in a decision. Otherwise, information asymmetries render underlying rights effectively void.*" (*ivi*, p. 213). Si noti come pure in Cons. Stato 8472/2019 – v. *supra* nota 62– nel giustificare un concetto rafforzato di trasparenza, funzionale al controllo giurisdizionale della decisione algoritmica in ambito amministrativo, si è evidenziato come (para 13.3) "l'articolo 15, diversamente dagli articoli 13 e 14, abbia il pregio di prevedere un diritto azionabile dall'interessato e non un obbligo rivolto al titolare del trattamento, e *permette* inoltre *di superare i limiti temporali posti dagli articoli 13 e 14*, *consentendo al soggetto di acquisire informazioni anche qualora il trattamento* […] *abbia* addirittura *già prodotto una decisione*"; valgono naturalmente tutte le considerazioni di cui già *supra*. Riconosce l'esistenza di un diritto alla spiegazione *ex post* della *singola* decisione automatica anche T.W. KIM, B.R. ROUTLEDGE, *op. cit.*, che però motiva valorizzando l'istituto del consenso informato nei casi di trattamento algoritmico decisionale, che vuole basato su un dovere di informazione non istantaneo (ed *ex ante*) ma continuo e pregnante, un "*right to an updating explanation*".

71 del Regolamento [73] – il quale, a differenza del testo normativo, espressamente cita il diritto dell'interessato a "ottenere una spiegazione" – esplicitamente metta in relazione questa garanzia al diritto di "ottenere l'intervento umano" e di poter "esprimere la propria opinione" in merito alla decisione automatica. Il rapporto di funzionale implicazione tra informazione (si consenta, spiegazione) e contestazione è, a ben vedere, valorizzato anche dalla relazione esplicativa che accompagna la cd. Convenzione 108+ del Consiglio d'Europa – ovvero il Protocollo di modifica, sottoscritto dall'Italia, che nell'aggiornare la Convenzione di Strasburgo del 1981 sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale al mutato contesto tecnologico, espressamente aggiunge nel nuovo articolo 9 (rubricato: *Rights of the data subject*), lett.a, il diritto di ciascun individuo "*not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration*" – laddove, al § 75a, sottolinea come il diritto di accesso sia in realtà funzionale alla possibilita di sindacare in modo effettivo la decisione automatica: "*It is essential that an individual who may be subject to a purely automated decision has the right to challenge such a decision by putting forward, in a meaningful manner, his or her point of view and arguments*. […]" [74].

Il diritto d'accesso, del resto, come può evincersi anche dal Considerando 63 del GDPR, dà all'interessato il diritto a ottenere tutte le informazioni di cui all'art. 15 del Regolamento, non solo in un momento anteriore o coincidente con l'inizio del trattamento, ma anche nel corso di questo, e si estende necessariamente anche a quei dati (e processi) nel frattempo '*prodotti*' dal trattamento, ossia alle informazioni ricavate dai dati trattati ('*inferred*') quale prodotto (anche solo intermedio) del processo automatizzato (di estrapolazione, clusterizzazione etc.) [75] e che, se riferibili all'interessato nei termini dell'art. 4, ricadono nell'ambito oggettivo del GDPR [76]. Questi dati *output*, infatti, oltre che 'base' di una eventuale (successiva) decisione automatizzata idonea ad incidere la sfera giuridica dell'interessato, sono già a loro volta il 'risultato' di un trattamento automatizzato (e.g. *data analytics*, profilazione etc.) [77] rispetto al quale il diritto di accesso si connota, pertanto, come una

[73] Si ricordi che sebbene non vincolanti, i Considerando guidano l'interpretazione delle norme. Si veda, in giurisprudenza, CGUE, Judgment of the Court (Third Chamber) of 13 July 1989, *Casa Fleischhandels-GmbH v Bundesanstalt für landwirtschaftliche Marktordnung*, c. 215/88. ECLI:EU:C:1989:331. E anche EUROPA, '*Guide to the Approximation of EU Environmental Legislation ANNEX I* '(Environment, 2015), secondo cui: Recitals explain the background to the legislation and the aims and objectives of the legislation. They are, therefore, important to an understanding of the legislation which follows.".

[74] Sia utile richiamare qui anche lo High Level Expert Group on Artificial Intelligence (AI HLEG), che nel citato documento pubblicato l'8 aprile 2019, e significativamente intitolato "Orientamenti Etici per un'IA affidabile", dopo aver individuato una serie di principi fondamentali, tra cui l'equità, cui il nuovo quadro tecnologico dovrebbe informarsi per essere "trustworthy", significativamente afferma: "La dimensione procedurale dell'equità implica la capacità di impugnare le decisioni elaborate dai sistemi di AI […] e la possibilità di presentare un ricorso efficace contro di esse. A tal fine […] i processi decisionali devono essere spiegabili". Ivi, ancora, più specificatamente sull'esplicabilità, si aggiunge: "Tale principio implica che i processi devono essere trasparenti, le capacità e lo scopo dei sistemi di AI devono essere comunicati apertamente e le decisioni, per quanto possibile, devono poter essere spiegate a coloro che ne sono direttamente o indirettamente interessati. Senza tali informazioni una decisione non può essere debitamente impugnata".

[75] Il riferimento del considerando 63 a dati quali "diagnosi, risultati di esami, pareri di medici" sono di fatto esempi *ante-litteram* di 'dati dedotti', *inferences* appunto; il progresso tecnologico e l'intelligenza artificiale applicata a modelli di data analytics rende possibili ben maggiori e invasive '*deduzioni*'. V. B. MITTELSTADT, S. WACHTER, *A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI*, in *Columbia Business Law Review*, 1/2019.

[76] Cfr. Article 29 Working Party's *Opinion 4/2007 on the concept of personal data*, disponibile online qui: *https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm*; sulla natura di dato personale delle inferenze, v. anche B. MITTELSTADT, S. WACHTER, *op. ult. cit*., da 25.

[77] Operazioni di d*ata mining* e profilazione, anche se non preordinate ad un momento decisionale, sono di per sé già potenzialmente lesive di interessi individuali meritevoli di tutela, come in più occasioni rilevato dalla Corte Europea dei Diritti dell'Uomo (per una Guida aggiornata alla Case Law sul punto, si v. ECHR, Guide to the Case-Law of the of the European Court of Human Rights.

legittima richiesta di informazioni *ex post*. Se non altro in questa circostanza – peraltro non necessariamente coincidente col campo d'applicazione dell'art. 22 GDPR, ma riferibile ad un numero molto più ampio di casi che contemplino una profilazione o comunque un trattamento dei dati (anche non automatizzato) che importi elementi previsionali o *lato sensu* inferenziali [78] – il diritto d'accesso deve considerarsi, infatti, certamente riferibile ad un contenuto informativo molto più ampio di quello di cui agli artt. 13 e 14 del Regolamento, che abbracci non solo informazioni dettagliate sui dati personali utilizzati come *input* del trattamento [79] (ovvero, in caso di profilazione, i dati utilizzati per creare il profilo) ma anche tutte le informazioni possibili sui dati di *output* (e.g. il profilo o lo *score* assegnato all'interessato; tutte le informazioni 'ricavate' dall'analisi dei suoi dati personali e le circostanze dedotte) [80]. Quando poi il trattamento (cioè l'inferenza) sia automatizzato(a) e sia funzionale all'adozione di una decisione dotata di effetti giuridici per l'interessato (basata *unicamente* sulla bontà di tale inferenza) – e *almeno* [81] in questo caso – anche tutte le (ulteriori) 'informazioni *significative* sulla logica utilizzata' dall'algoritmo che tale *output* ha (già) generato; non, evidentemente, una informazione prognostica e astratta sul funzionamento dell'algoritmo *in genere*, ma più realisticamente le ragioni della decisione (l'*output* prodotto) *in concreto*; la logica utilizzata nel caso specifico, il come e il perché [82]: una spiegazione. Con il solo limite, nei contenuti, della tutela del segreto commerciale e della proprietà intellettuale del titolare del trattamento circa la (*fully disclosure* della) *formula* dell'algoritmo decisionale, che il Considerando 63 significativamente contempla proprio con riferimento al diritto d'accesso, peraltro sottolineandone in modo esplicito il carattere non di limite assoluto, ma anzi rimesso a ponderato bilanciamento, *case-by-case*, con l'avverso interesse (alla spiegazione e contestabilità della decisione) di chi sia soggetto al processo automatizzato (e mai idoneo, in ogni caso, a fondare il rifiuto di fornire informazioni all'interessato) [83].

Data protection., 30 April 2021, online at: *https://www.echr.coe.int/Documents/Guide_Data_protection_ENG.pdf*), oltre a sollevare specifici problemi in ordine alla protezione dei dati '*prodotti*', in particolare quando abbiano natura di dati *sensibili*. V. B. MITTELSTADT, S. WACHTER, *op. ult. cit*. Il problema si è posto con particolare evidenza con riguardo ai dati sanitari; *amplius*, W. SCHÄFKE-ZELL, *Revisiting the definition of health data in the age of digitalized health care*, in *International Data Privacy Law*, 2021; G. MALGIERI, G. COMANDÈ, S*ensitive-by-distance: quasi-health data in the algorithmic era*, in *Information & Communication Technology Law*, Vol. 26, n. 3, 2017, 229-249.

[78] *Amplius*, B. MITTELSTADT, S. WACHTER, *op. ult. cit*.

[79] la loro origine, la finalità per cui sono trattati e tutte le altre informazioni richieste *ex* art. 15 GDPR.

[80] Se così non fosse, risulterebbero, per paradosso, inevitabilmente frustrati i diritti di rettifica, cancellazione e limitazione di trattamento (riconosciuti rispettivamente *ex* artt. 16, 17 e 18 GDPR) proprio nei casi in cui ce ne sarebbe maggiormente bisogno: la profilazione e più in generale le tecniche di *data analytics*, comportando elementi di 'previsione' ed 'estrapolazione', aumentano grandemente il rischio intrinseco di inesattezza dei dati (*output*) e ampliano il novero dei dati trattati rendendone più difficile il controllo da parte dell'interessato. Nell'interpretazione che qui si vuole adottare, funzionale all'esercizio di questi diritti (come pure a quello dell'impugnazione della decisione basata unicamente su trattamento automatizzato dei dati, nel caso di ADM) sarebbe da intendersi l'informazione di cui all'art. 15 GDPR. Cfr. sul punto, con riferimento alla profilazione, Article 29 Working Party's *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, n. 251 del 3 ottobre 2017, WP 251 rev01, versione italiana, p. 19, secondo cui, testualmente, '*i diritti di rettifica e cancellazione si applicano tanto ai dati personali di input quanto ai dati di output*'. In particolare, sul contenuto (ampio) del diritto di accesso in caso di *inferenza*, v. G. MALGIERI, G. COMANDÈ, *op. ult. cit*., p. 246, che valorizza in tal senso alche il portato dell'art. 20 GDPR che riconosce all'interessato il diritto alla portabilità dei dati in un formato *leggibile*. All'articolo 20, sul punto, riconosce però una portata decisamente inferiore il Gruppo di Lavoro dell'articolo 29 che esclude dall'ambito di operatività dell'obbligo i dati prodotti da attività algoritmica del data controller: v. in particolare, Article 29 Working Party, *Guidelines on on the right to data portability,* n. 242 del 13 dicembre 2016, WP 242 rev.01, p.10 ss (recentemente richiamata da EDPB, *Guidelines on the targeting of social media users*, n. 8 del 2 settembre 2020).

[81] Il richiamo è alla lettera dell'art. 15(1)(h) GDPR, immediatamente allusivo ad un *quid pluris* rispetto al contenuto minimo di trasparenza e tutela previsto in via generale, e possibile nel caso concreto (*ex post*).

[82] Così vuole interpretarsi, in questa prospettiva e in questo contesto (cioè di una informazione destinata a pervenire all'interessato s*uccessivamente* alla decisione), l'espressione '*importanza* […] *di tale trattamento*' di cui all'art. 15(1)(h) GDPR.

[83] Se ne dirà più ampiamente *infra.* Si v. G. NOTO LA DIEGA, *Against the Dehumanisation of Decision-Making. Algorithmic Deci-*

È l'esigenza essenziale di una tutela effettiva della persona umana nei confronti del potere digitale [84] a imporre che il significato dell'art. 15(1)(h) GDPR sia ricostruito in modo da garantire che il titolare dei dati personali abbia comunicazione di tutti i dati che lo riguardano: anche di quelli c.d. derivati; possa valutarne l'esattezza, la completezza, la rilevanza (nei limiti che il Regolamento e le finalità del trattamento consentono) e, laddove posti alla base di una decisione (automatizzata) che leda (o possa ledere) suoi interessi giuridicamente tutelati, abbia accesso almeno a tutte le informazioni necessarie per valutarne, ed eventualmente sostenerne un'effettiva contestazione (ricorrendo ad un interlocutore umano o ad un'aula di giustizia) [85]. Se il digitale è diventato uno spazio 'sociale', vivo, entro cui si svolge e delinea l'esistenza "*onlife*" [86] di una persona, si determina, cioè, la sua identità, in modo, come s'è visto, giuridicamente rilevante (ed altrimenti incontrollato), allora la tutela della dignità umana – ancoraggio di tutta la legislazione sulla protezione dei dati – la libertà individuale, il diritto fondamentale di ciascuno a incidere la propria dimensione individuale e collettiva, passano anzitutto attraverso la conoscenza, la trasparenza intesa come giustificazione del potere digitale –come spiegazione – e, soltanto, quindi, attraverso un potere *attivo* di controllo sui propri dati – in particolare 'derivati' e dedotti – e la possibilità di arginare (e contestare) l'impatto delle tecnologie su di essi e a partire da essi, che senza un (previo) diritto alla spiegazione sarebbero svuotati. Negare *questa* trasparenza, che in ultima sponda vuol dire non già diritto a conoscere le ragioni di una certa decisione automatica che ci riguardi ma anche i dati (*inferiti*) rilevanti alla sua base (ovvero, *chi siamo* online), significa, in una certa misura, disumanizzare l'individuo [87].

5. – Postulata la possibile base giuridica di un diritto alla spiegazione della decisione algoritmica, veniamo però ai suoi limiti. Un primo attiene all'ambito di applicazione, come ha rilevato qualcuno [88] piuttosto limitato, in base alla formulazione letterale dell'art. 22 GDPR. I trattamenti automatizzati di cui stiamo parlando rilevano infatti, espressamente, solo in quanto siano 'decisionali', conducano cioè ad (e applichino) una decisione che influisca sull'interessato. Ciò comporta che siano esclusi dalla fattispecie tutti quei trattamenti di dati non tipicamente 'inferenziali', in cui cioè l'utilizzo di tecniche informatiche sia limitato, magari, alla conservazione e/od organizzazione di dati personali senza comportare alcun passaggio analitico e/o

---

sions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information, in *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* (JIPITEC), 9/2018, Vol. 3, n. 1, Available at SSRN: *https://ssrn.com/abstract=3188080* (v. in particolare p. 23, para 69).

[84] Oltre che il principio generale di correttezza del trattamento, cui certamente può dirsi contrario il rifiuto, quando ingiustificato, di fornire informazioni sui dati derivati riferibili all'interessato o circa i motivi di una certa decisione automatica.

[85] Per R. MESSINETTI, *op. cit.*, p. 867, "il sistema giuridico-tecnologico deve permettere alla persona di comprendere la comprensione che la macchina ha della persona medesima nell'ambito di processi decisionali diretti ad incidere sulla sua sfera giuridica e vitale"; sarebbe perciò l'esigenza di conservare il controllo sulla propria identità personale (e la sua formazione) a giustificare un'interpretazione del diritto di accesso come diritto a comprendere la logica e le ragioni giustificatrici della decisione automatizzata (avente ad oggetto proprio quell'identità) e dei propri dati personali "derivati" dagli *inputs* originari, che costituiscono gli *outputs* intermedi e/o conclusivi del trattamento.

[86] Il termine, evocativo, si deve a Luciano Floridi, che l'ha scelto per rappresentare l'esperienza che l'uomo vive nelle società iperstoriche in cui non "non distingue più tra online o offline" e addirittura diventa sempre più "non ragionevole chiedersi se si è online o offline"; Cfr. L. FLORIDI, *The Fourth Revolution. How the Infosphere is reshaping human reality,* Oxford University Press, 2014; ID. (ed.), *The Onlife Manifesto: Being Human in a Hyperconnected Era,* Springer, 2015.

[87] Una tale lettura della trasparenza si ritrova, ad esempio, in R. WAELEN, *Why AI Ethics Is a Critical Theory*, in Philosophy & Technology, 35:9, 2022, *https://doi.org/10.1007/s13347-022-00507-5*.

[88] Così, tra gli altri, S. WACHTER, B. MITTELSTADT, L. FLORIDI, *op. cit.*

valutativo degli stessi. Nei trattamenti decisionali automatizzati (ADM), invece, i dati sono, al contrario, raccolti (o sottoposti alla macchina) proprio perché questa, eseguendo un dato calcolo – applicando cioè al data-set regole deduttive algoritmicamente definite nel programma – li analizzi per arrivare ad una certa 'soluzione', un *output*' decisionale', appunto.

Questa decisione – perché il relativo trattamento sia rilevante per gli effetti della disposizione in esame – deve essere "*basata unicamente sul trattamento automatizzato*" [89]: interpretazione letterale della disposizione vorrebbe dunque che ricadano nell'ambito applicativo dell'esaminanda disciplina solo quelle decisioni *fully automated*, dovendosi invece ritenere escluse tutte quelle in cui sia possibile riscontrare un benché minimo coinvolgimento umano che possa variamente 'interferire' col processo decisionale automatizzato, potendo verificare o modificare la decisone ma anche, ad esempio, meramente ratificarla. Un'interpretazione siffatta, a ben vedere, rischierebbe di escludere dall'ambito applicativo della rigorosa disciplina di tutela in esame una buona parte di decisioni che sono in sostanza esito di trattamenti automatizzati ma, ad esempio, formalmente applicate per intervento umano [90]. Tutto ciò – pena vanificare gli scopi di tutela della disposizione che si analizza – dovrebbe perciò più opportunamente condurre ad adottare un'interpretazione larga dell'articolo 22, ritenendo incluse nella fattispecie tutte le decisioni 'automatizzate nella sostanza', anche quelle, cioè, che, benché implicanti nel processo decisionale un qualche intervento umano, questo si configuri come ininfluente rispetto al contenuto della decisione [91]. Del resto, come pure fa notare qualcuno [92], un indizio a conferma di tale interpretazione può in realtà intravedersi già nello stesso dato letterale del disposto dell'art. 22 (1) GDPR che si riferisce non a *decisioni 'unicamente' automatizzate*, bensì a *decisioni 'basate' unicamente su trattamenti automatizzati*: la sottile differenza conduce ad una conseguenza di non poco conto: potersi considerare 'letteralmente 'incluse anche quelle decisioni formalmente imputabili ad un agente umano ma in realtà mera applicazione passiva di valutazioni computerizzate.

Non pare sollevi invece particolari problemi il resto della disposizione. Procedendo nell'interpretazione,

---

[89] Art. 22 (1) GDPR.

[90] Molte di queste si prestano anche ad avere effetti potenzialmente molto incisivi sugli interessati; si pensi a tutti i meccanismi di *'scoring'*, ampiamente utilizzati nella prassi digitale, ad esempio, per selezionare gli aspiranti ad una posizione lavorativa o valutare la probabile solvibilità di chi miri ad ottenere un fido bancario: non è raro che in questi casi gli interessati siano valutati con strumenti informatici che, applicando ai dati personali funzioni statistiche, assegnino loro un certo valore di merito, uno 'score 'appunto, e che la decisione finale, formalmente spettante ad un funzionario umano sia in realtà meramente passiva, limitandosi magari questo ad abbinare, spesso anche seguendo rigide linee guida predefinite, un certo esito ad un determinato 'punteggio 'senza avere od esercitare alcun potere realmente 'valutativo'. Cfr. G. MALGIERI, G. COMANDÈ, *Right to legibility of automated decision-making*, in *International Data Privacy Law*, 2017, vol.7, n.4. In generale, invece, sullo *scoring*, si v. D.K. CITRON, F. PASQUALE, *The Scored Society: Due Process for Automated Predictions*, in *Washington Law Review*, Vol. 89, 2014, p. 1, anche in *University of Maryland Legal Studies*, Research Paper n. 2014-8. SSRN: *https://ssrn.com/abstract=2376209*.

[91] In tal senso, del resto, è di conforto lo stesso Article 29 Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, n. 251 del 3 Ottobre 2017*, p.21, dove, testualmente: "*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*". E prima ancora registrava un simile orientamento anche il Garante Privacy inglese, l'Information Commissioner's Office (ICO) per cui "*the interpretation of the word "solely" in the context of Article 22(1) [...]is intended to cover those automated decision-making processes in which humans exercise no real influence on the outcome of the decision, for example where the result of the profiling or process is not assessed by a person before being formalized as a decision*" Cfr. Information Commissioner's Office, *Feedback request – Profiling and automated decision-making*, 2017, 19, richiamato anche in G. MALGIERI, G. COMANDÈ, *Right to legibility of automated decision-making*, cit.

[92] G. MALGIERI, G. COMANDÈ, *op. ult. cit.*

queste decisioni (cui sia ricollegabile un diritto di spiegazione), a norma dell'art. 22 (1) del Regolamento, oltre che basate su trattamenti automatizzati, devono pure "*produrre effetti giuridici* [nei confronti dell'interessato] *o incidere in modo analogo significativamente sulla sua persona*". L'effetto di tali decisioni deve dunque influire, negandolo o limitandolo, sul libero esercizio di un diritto della persona riconosciuto dall'ordinamento giuridico. Si pensi, al diritto (meglio, ai diritti) previsti dalla disciplina giuslavoristica a non essere discriminati in una pratica di assunzione online che non preveda intervento umano. Anche in questo caso, però, può facilmente estendersi la portata della disposizione. Il catalogo di diritti potenzialmente afflitti da trattamenti decisionali automatizzati si allarga, infatti, notevolmente ove si tengano in considerazione anche tutti quei diritti fondamentali della persona di rango costituzionale – previsti dalla nostra Carta Fondamentale, come dalla Convenzione europea dei diritti umani e dalla Carta di Nizza – diritti, espressi in forma aperta, come la libertà di espressione, la libertà e segretezza della corrispondenza, la libertà di circolazione, di culto, il diritto di difesa e, più di tutti, il principio di non discriminazione: diritti e libertà naturalmente 'elastici', che non costituiscono cioè un *numerus clausus* ma si prestano a coprire tutta una gamma aperta di situazioni giuridiche individuali che possono essere esposte ad un concreto rischio, per chi ne sia titolare, a fronte di decisioni confliggenti, ed ingiuste, di un'Intelligenza Artificiale [93].

Questa tutela ampia dei diritti della persona, accordata dall'articolo in esame, è infine ulteriormente allargata dal prosieguo stesso della disposizione; il §1 dell'art. 22 estende infatti la sua portata anche a quei fenomeni di ADM che pur non intaccando situazioni giuridiche soggettive apertamente qualificate come diritto *"incidano* [comunque, ndr] *in modo analogo e significativamente sulla persona*", ovverosia – nell'interpretazione che pare preferibile – colpiscano, e gravino profondamente (questo il significato da attribuirsi all'espressione 'significativamente') su interessi legittimi o comunque legittime pretese della persona umana purché non di mero fatto, ossia situazioni giuridiche che l'ordinamento, benché non riconosca come diritti immediatamente azionabili dall'individuo, ritenga ugualmente degne di protezione o accordandovi una qualche forma di tutela o quantomeno dimostrando un complessivo giudizio di riprovevolezza verso pratiche e comportamenti che ne minaccino il libero e pieno godimento individuale [94].

---

[93] Si ricorda inoltre che la stessa 'libertà informatica', come è stata autorevolmente definita, intesa in senso attivo, e cioè come libertà di valersi senza limitazioni, neppure tecnologiche come può essere una decisione automatizzata, degli strumenti informatici al servizio del proprio diritto di comunicare, trasmettere e ricevere informazioni, accedere a servizi telematici, partecipare, insomma, alla società digitale trova un suo radicato fondamento costituzionale ed esprime, per via ermeneutica, nuovi diritti soggettivi, parimenti meritevoli di tutela, da quelli tradizionalmente statuiti opportunamente adattati dall'interprete al rinnovato contesto tecnologico in cui viviamo. Sul punto V. FROSINI, *L'Orizzonte giuridico di Internet*, in *Il diritto dell'informazione e dell'informatica*, n. 2, 2002, 275; T.E. FROSINI, *Tecnologie e libertà costituzionali* in *Il diritto dell'informazione e dell'informatica*, n. 3, 2003, nonché in *Libertè Egalitè Internet*, Napoli, 2016, 19 ss.; ID., *Il diritto costituzionale di accesso a Internet*, in M. PIETRANGELO (a cura di), *Il diritto costituzionale di accesso a Internet*, Napoli, 2011, nonché in *Libertè Egalitè Internet*, Napoli, 2016, 50 ss.

[94] Questo è il significato che a parere di chi scrive dovrebbe attribuirsi all'espressione *"incidano in modo analogo*" ("*similarly*" nella versione inglese del Regolamento), espressione che merita considerevole attenzione da parte dell'interprete perché assolutamente nuova: è infatti aggiunta *ex novo* dal GDPR accanto all'espressione *"incidano significativamente*" già presente invece nella pregressa formulazione della norma in esame, l'art. 15 dell'abrogata Direttiva 95/46/CE. Si lasci chiarire con un esempio afferente una circostanza frequentissima nel digital marketing. Non esiste nell'ordinamento un diritto a non ricevere *targeted advertising*, una forma cioè di marketing '*tailored*' basato sulla profilazione. Questa è una pratica perciò astrattamente lecita per quanto potenzialmente intrusiva della sfera personale; tuttavia quando la propria dimensione soggettiva risulti lesa, nel caso concreto, con un'incidenza definibile importante (sia riscontrabile cioè una certa gradazione dell'offesa – restando al caso esemplificativo, ad esempio, per l'eccessiva intrusività della raccolta dei dati o per l'invasività intollerabile dei *banner* pubblicitari, o ancora per l'implementazione di meccanismi di *pricing* che sfruttano le specifiche debolezze dell'*user* per proporgli prezzi d'acquisto elevati e fuori standard) è certamente plausibile pensare che tale pratica incida significativamente e in modo analogo sull'interessato rendendosi opportuna l'applicazione della speciale disciplina di cui all'art. 22 GDPR. Ciò in ragione non della violazione di un preciso diritto dell'interessato, ma per via dell'esistenza, comunque, nel caso in que-

6. – La *governance* degli algoritmi decisionali impone un'ulteriore, e non facile, operazione ermeneutica, connessa alla necessità di operare un bilanciamento tra il diritto di chi è soggetto alla decisione automatica ad essere adeguatamente informato sul funzionamento dell'algoritmo, e quello – opposto – di chi abbia sviluppato ed utilizzi a fini economici l'algoritmo a tenerne segreto il design per evitare di disperderne il valore imprenditoriale a vantaggio concorrenziale dei competitors. Ciò rappresenta un ulteriore limite al diritto alla spiegazione della decisione automatica; quest'ultimo, importante e più difficilmente aggirabile.

Secondo il Considerando 4 del GDPR, "*Il diritto alla protezione dei dati di carattere personale non è una prerogativa assoluta, ma va* […] *contemperato con altri diritti fondamentali, in ossequio al principio di proporzionalità*" e il Regolamento "*rispetta tutti i diritti fondamentali e osserva le libertà e i principi riconosciuti dalla Carta e sanciti dai trattati* […]" t*ra cui* "[…] *la libertà d'impresa* […]". Più specificamente, il Considerando 63, relativo al diritto di accesso, testualmente afferma che "*tale diritto non dovrebbe ledere i diritti e le libertà altrui, compreso il segreto industriale e aziendale e la proprietà intellettuale, segnatamente i diritti d'autore che tutelano il software*" con la precisazione che "*tuttavia, tali considerazioni non dovrebbero condurre a un diniego a fornire all'interessato tutte le informazioni*". In ciò facendo eco, tra l'altro, alla recente Raccomandazione del Consiglio d'Europa [95] secondo cui "*States should establish appropriate levels of transparency with regard to* […] *use, design and basic processing criteria and methods of algorithmic systems* […]' e che "*The legislative frameworks for intellectual property or trade secrets should not preclude such transparency, nor should States or private parties seek to exploit them for this purpose*" [96], che, peraltro, dà indubbia prevalenza, tra le situazioni tutelate, al diritto dei singoli ad essere opportunamente informati.

Sugli estremi di questo bilanciamento [97] ci illumineranno senz'altro la giurisprudenza europea e nazionale, la questione è però tutt'altro che irrilevante, considerato che la maggior parte degli algoritmi di decisione automatica, anche molto invasivi ed utilizzati – inclusi quelli di Google e Facebook, per intenderci – sono proprietari e tenuti riservati quanto al loro meccanismo. In generale, può dirsi – ma questo soprattutto con riguardo al settore della Concorrenza, per cui è data maggiore casistica – che la giurisprudenza della Corte di Giustizia [98] ammette limitazioni o deroghe al diritto d'autore e al segreto commerciale [99] in "*exceptional cir-*

---

stione, di una sua legittima pretesa (ricavabile dal complesso dell'ordinamento) a non essere vittima di pratiche commerciali scorrette o essere manipolato nell'atto di formarsi una propria adesione alla volontà di acquistare un determinato bene. *Amplius*, sul punto, G. MALGIERI, G. COMANDÈ, *Right to legibility of automated decision-making*, cit., ed altri, tra cui il nostro E. TROISI, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità' dell'algoritmo*, cit.

Volendo fare un ulteriore esempio, il generale principio di correttezza e non discriminazione può considerarsi alla base della legittima pretesa degli interessati a non vedersi rifiutata con processi automatizzati una domanda di credito online: quest'ultima è un'altra tipica ipotesi di ADM, rilevante per l'art. 22 GDPR, per di più autorevolmente citata esemplificativamente dallo stesso Considerando 71 del Regolamento. Per un'interpretazione restrittiva della disposizione in questione, si veda, invece, B. WONG, *Online personalised pricing as prohibited automated decision-making under Article 22 GDPR: a sceptical view*, in *Information & Communications Technology Law*, vol. 30, n. 2, 2021, 193-207, che sulla base di un'interpretazione stretta del primo paragrafo dell'art. 22 del Regolamento, in particolare delle locuzioni *'legal effects'* e '*similarly significant effects*', ritiene, ad esempio, che la nozione di trattamento decisionale automatizzato di cui all'art. 22 GDPR non includa buona parte delle pratiche di online personalised pricing, non avendo queste alcun effetto *diretto* giuridicamente rilevante sugli utenti salvo nell'ipotesi in cui la proposta di un prezzo eccessivamente fuori standard precluda a talun utente o categoria di utenti il completo accesso ad un prodotto o servizio (v. *ivi*, 198-200).

[95] Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.

[96] Articolo 4.1.

[97] V.si, tra gli altri, G. MALGIERI, *Trade Secrets v Personal Data: a possible solution for balancing rights*, in International Data Privacy Law, 6(2): ipv030, 2016. *http://dx.doi.org/10.1093/idpl/ipv030*; G. NOTO LA DIEGA, *op. cit.*; M.E. KAMINSKI, *op. cit.*

[98] Si vedano: CGUE, sentenza della Corte del 6 aprile 1995, Cause riunite C-241/91 P e C-242/91 P, ECLI:EU:C:1995:98 (cd. Magill); CGUE, Sentenza della Corte (Quinta Sezione) del 29 aprile 2004, C-418/01, ECLI:EU:C:2004:257 (cd. IMS Health);

cumstances" [100], quando siano cioè in gioco contrastanti ed altrettanto rilevanti interessi generali o diritti fondamentali, nei limiti di quanto sia proporzionato allo scopo. La questione non si è ancora posta, a livello europeo, con riferimento agli obblighi informativi di cui al GDPR; né pare, a una prima indagine, che elementi utili possano ricavarsi dalle poche decisioni dei giudici nazionali [101]; tra queste quelle più inclini a ritenere recedente il diritto alla riservatezza dell'algoritmo – considerando quindi giustificata (e dovuta) la *disclosure* – intervengono per lo più in merito a casi di utilizzo di algoritmi decisionali da parte dell'Autorità Pubblica, motivando, quindi, la debita trasparenza o la legittimità dell'accesso in quanto funzionali alla tutela di interessi pubblici prevalenti e al rispetto dei principi di trasparenza e motivazione dell'agire Pubblico [102].

Un esempio di contemperamento tra le due contrapposte esigenze qui specificamente in gioco ce lo da però lo stesso legislatore europeo. Nel cd. *Regolamento P2B*, il Regolamento (UE) 2019/1150 sui servizi di intermediazione online con cui l'Unione ha fissato le regole a garanzia della concorrenza nella *Platform economy*, l'obbligo, in capo ai fornitori di servizi, di indicare i parametri principali che determinano il posizionamento degli utenti commerciali attraverso metodi evidentemente automatizzati (cd. algoritmi di raccomandazione) [103] – *lex* ovviamente *specialis* rispetto ai generici doveri di trasparenza di cui al GDPR – è delimitato, al §6 dell'art. 5, dalla precisazione per cui i service providers nell'adempiere a tali prescrizioni"[…] *non*

CGUE, Sentenza del Tribunale di primo grado (grande sezione) del 17 settembre 2007, T-201/04, ECLI:EU:T:2007:289 (Microsoft c. Commissione).

[99] Direttiva (UE) 2016/943 del Parlamento europeo e del Consiglio dell'8 giugno 2016 sulla protezione del know-how riservato e delle informazioni commerciali riservate (segreti commerciali) contro l'acquisizione, l'utilizzo e la divulgazione illeciti.

[100] Si vedano le appena citate pronunce della Corte di Giustizia dell'Unione europea Magill, IMS Health e Microsoft c. Commissione.

[101] per l'Italia, sul carattere discriminatorio di un algoritmo decisionale (in particolare l'algoritmo di ranking dell'App Deliveroo) si segnala Trib. Bologna, 10 marzo 2021 (Ord.) in cui non è però oggetto del giudizio la *disclosure* dell'algoritmo decisionale, il cui carattere discriminatorio è presunto dalla descrizione del suo funzionamento nelle condizioni generali del contratto di lavoro sottoscritto dai *riders*.

[102] Per l'Italia si segnala, in particolare, tra le altre: TAR, Lazio-Roma, sez. III bis, sentenza 22 marzo 2017, n. 3769. Nel caso di specie il Tribunale amministrativo è chiamato valutare se possa essere riconosciuto il diritto di accesso agli atti della P.A. di cui all'agli artt. 22 ss. l. n. 241/1990 con riguardo all'algoritmo che gestisce il software relativo ai trasferimenti interprovinciali del personale docente ai sensi e per gli effetti del C.C.N.I. sulla mobilità 2016 di cui alla l. n. 107/2015, negato dal M.I.U.R.; il giudice, ritenendo che l'algoritmo possa configurarsi come atto amministrativo informatico e quindi che siano applicabili le norme sull'accesso agli atti ritiene che sebbene il software sia protetto quale opera dell'ingegno ne sia dovuta l'ostensione nel limite (visione ed estrazione di copia) di quanto strettamente necessario alla tutela dei diritti del ricorrente.

Si segnala, significativamente, sul punto, anche Cons. Stato n. 8472/2019, per cui (para 13.1) "*l'invocata riservatezza delle imprese produttrici dei meccanismi informatici utilizzati*" non esime dalla "*necessità che la formula tecnica che di fatto rappresenta l'algoritmo sia corredata da spiegazioni che la traducano nella regola giuridica ad essa sottesa e che la rendano leggibile e comprensibile*" per quanto quelle "*ponendo al servizio del potere autoritativo tali strumenti, all'evidenza ne accettano le relative conseguenze in termini di necessaria trasparenza*.".

Più recentemente torna sul punto Cons. Stato, sez. VI, sentenza n. 30 del 2 gennaio 2020, che afferma che la nozione di "riservatezza" *ex* art. 22, comma 1, lett. c), l. n. 241/1990 quale criterio di individuazione dei "controinteressati" va intesa in senso ampio, come comprensivo non soltanto della tutela dei "dati personali" della persona fisica, ma anche dei "segreti tecnici e commerciali" della persona giuridica (nel caso di specie si trattava del soggetto titolare del diritto d'autore sul codice sorgente di un software utilizzato per le prove di un pubblico concorso).

In Francia è il legislatore ad aver introdotto uno specifico dovere di trasparenza per la Pubblica Amministrazione che adotti, a certe condizioni, provvedimenti con strumenti decisionali automatici. In particolare, l'art. R. 311-3-1-2 del *Code des relations entre le public et l'administration*, nel prevedere che l'Amministrazione comunica al soggetto interessato da una decisione individuale assunta sulla base di un trattamento algoritmico, su richiesta di quest'ultimo, in forma intelligibile: 1) il grado e le modalità di contribuzione al processo decisionale dell'elaborazione algoritmica; 2) i dati trattati e le loro fonti; 3) i parametri del trattamento e, ove applicabile, il loro peso, applicati alla situazione dell'interessato; ecc.; fa espressamente salva la violazione dei segreti protetti dalla legge (*la traduzione è nostra*). Si veda anche nota 62 *supra*.

[103] Art. 5 (1-5) Reg. (UE) 2019/1150.

sono tenuti a rivelare algoritmi o informazioni che, con ragionevole certezza, si tradurrebbero nella possibilità di trarre in inganno i consumatori o di arrecare loro danno attraverso la manipolazione dei risultati di ricerca" e che in ogni caso è fatta salva a direttiva (UE) 2016/943 sulla protezione delle informazioni commerciali riservate. Nel Considerando 27 si specifica che a norma del Regolamento "*ai fornitori di servizi di intermediazione online o di motori di ricerca online non dovrebbe essere richiesto di divulgare il funzionamento dettagliato dei loro meccanismi di posizionamento, inclusi gli algoritmi* […]" *né dovrebbe compromettersi "la loro capacità di agire contro la manipolazione in mala fede del posizionamento da parte di terzi, anche nell'interesse dei consumatori* […]. *Una descrizione generale dei parametri principali di posizionamento dovrebbe salvaguardare tali interessi, fornendo nel contempo agli utenti commerciali e agli utenti titolari di siti web aziendali una adeguata comprensione del funzionamento del posizionamento nel contesto del loro utilizzo di servizi di intermediazione online o di motori di ricerca online specifici*". Per garantire che l'obiettivo del regolamento sia raggiunto, pertanto, "*la considerazione degli interessi commerciali dei fornitori di servizi di intermediazione online o di motori di ricerca online non dovrebbe* […] *mai portare ad un rifiuto di divulgare i parametri principali che determinano il posizionamento. A tale riguardo, mentre il presente regolamento non pregiudica la direttiva (UE) 2016/943* […], *la descrizione data dovrebbe perlomeno essere basata sui dati effettivi relativi alla rilevanza dei parametri di posizionamento utilizzati*" [104].

Tirando le somme, sebbene ampio rilievo sia accordato alla tutela delle ragioni competitive dei *providers*, un nucleo minimo di trasparenza sul funzionamento effettivo dell'algoritmo (a beneficio degli utenti commerciali che si servono del servizio), tale comunque da non svilire gli scopi del Regolamento, deve essere garantita; la fattispecie è molto diversa da quella qui in esame, ma il dato è ugualmente rilevante, tenuto conto che gli interessi commerciali dei fornitori di servizi non rappresentano l'unico contraltare all'obbligo di trasparenza ma vi concorre la necessità di evitare manipolazioni, a danno dei consumatori, dei servizi di ricerca o promozione online.

In ogni caso, salvo previsioni specifiche e posizioni settoriali anche molto diverse del legislatore eurounitario [105], la lezione che può apprendersi è che un *assessment case-by-case* dovrebbe essere di volta in volta richie-

---

[104] A "*come selezionare i parametri principali e prevenire comunque la manipolazione in cattiva fede del posizionamento*" è dedicato il Capitolo 4, §§ da 76 ad 84, degli "*Orientamenti sulla trasparenza del posizionamento a norma del regolamento (UE) 2019/1150 del Parlamento europeo e del Consiglio*" di cui alla Comunicazione della Commissione Europea (2020/C 424/01). Testualmente il §82: "*I fornitori non possono* […] *rifiutarsi di divulgare, ad esempio, i parametri principali per l'unico motivo che non hanno mai rivelato alcuno dei loro parametri in passato o che le informazioni in questione sono sensibili sotto il profilo commerciale*".

[105] Più restrittiva e severa sembra la posizione espressa dalla Commissione nella Proposta di Regolamento sui servizi digitali (cd. *Digital Services Act*): premesso che sono fatte salve le disposizioni del GDPR quanto alla tutela della riservatezza delle persone fisiche (Considerando 10), la Bozza introduce uno specifico diritto dei 'destinatari del servizio' – anche quindi persone fisiche – ad essere informati, nell'ambito delle condizioni generali del servizio, "*in modo chiaro, accessibile e facilmente comprensibile*" circa "*i principali parametri utilizzati nei loro sistemi di raccomandazione* [completamente o parzialmente automatizzati], *nonché qualunque opzione* […]" *a loro disposizione che possa consentirgli di "modificare o influenzare tali parametri principali* […]" (art. 29 del Proposal) riconoscendo paralleli poteri di accesso e controllo al Coordinatore dei servizi digitali del luogo di stabilimento ovvero alla Commissione per verificare la *compliance* (anche) agli obblighi di trasparenza. La stessa Proposta riconosce però al gestore della piattaforma di opporsi all'*audit* quando "*dare accesso ai dati comporterebbe notevoli vulnerabilità* […] *per la protezione delle informazioni riservate, in particolare dei segreti commerciali*" (art 31, §6 lett. b del Proposal).

Le scelte alla base del *Digital Services Act* sembrano confermare la tendenza del legislatore ad affrontare il problema della trasparenza degli algoritmi decisionali in modo settoriale e con strumenti orizzontali, dotati di specifici strumenti amministrativi di verifica ed enforcement delle disposizioni. Sul punto: HUSEINZADE N., *Algorithm Transparency: How to Eat the Cake and Have It Too*, in *europeanlawblog.eu*.

Sul contenuto del principio di trasparenza, nell'ambito della riforma dello Spazio Digitale europeo, si veda, invece, ad esempio: E. GARZONIO, *L'algoritmo trasparente: obiettivi ed implicazioni della Riforma dello Spazio Digitale europeo,* in *Riv. it. inf. e dir.*, 2/2021, 25 ss., in part. p.30.

sto tra tutti gli interessi in gioco, tenendo a mente, però, che quando i procedimenti decisionali riguardino persone fisiche, il grado di necessaria 'spiegabilità' dell'algoritmo (e della sua decisione), in base al contesto, dovrebbe aumentare con l'aumentare della gravità delle conseguenze (sui diritti umani e gli interessi in gioco) se tale *output* fosse errato o altrimenti impreciso. E che comunque, quando sia prevedibile un grave impatto sui diritti umani, la trasparenza possa anche richiedere la condivisione di codici o set di dati specifici.

7. – Se pure si fosse riusciti nell'intento di dimostrare l'esistenza, sul piano giuridico, di un diritto del singolo a ottenere una spiegazione della decisione algoritmica che lo riguardi, da intendersi – riepilogando – almeno come 'informazioni significative' sulla logica *concretamente* utilizzata nel(lo specifico) processo decisionale automatizzato (*ex post*), resterebbe ancora da porsi il più importante degli interrogativi: se il diritto suddetto non incontri, superati i limiti giuridici, limiti tecnologici in grado di ostacolarne l'attuazione.

Secondo quanto ampiamente sostenuto in letteratura [106] la regola inferenziale che anima i processi automatizzati dei sistemi di Machine Learning – come si è detto, adattiva, non definita a priori, in sede di programmazione, ma dinamicamente (e autonomamente) rielaborata dall'Intelligenza artificiale nell'esperienza d'utilizzo – non sarebbe (sempre) conoscibile [107], almeno relativamente alla ricostruzione – determinante, ai fini *de qua* – della logica *in concreto* alla base della elaborazione di uno specifico *output* [108].

Se questo non è sempre vero [109] – e seppur in attesa di nuovi promettenti elaborati della tecnica [110], ciò rappresenterebbe, pur tuttavia – allo stato attuale – un limite, di natura tecnologica, tale da rendere, in queste ipotesi,

---

[106] Tra gli altri, Y. BATHAEE, *op. cit.*; D. CASTELVECCHI, *op. cit.*

[107] Tra gli altri, v. P.J. LISBOA, *Interpretability in Machine Learning. Principles and Practice*, Springer, 2013. Sulla differenza tra *explainability* ed *interpretability* dei modelli decisionali e il rapporto tra loro, v. J. ZHONG, E. NEGRE, *op. cit.*; si veda anche D.A. BRONIATOWSKI, *Psychological Foundations of Explainability and Interpretability in Artificial Intelligence*, NISTIR 8367, 2021, doi: *https://doi.org/10.6028/NIST.IR.8367.*

[108] Appare opportuno rilevare, però, come non manchino in letteratura approcci differenti alla soluzione del problema mediato della 'equità' decisionale dei sistemi automatici, che, abbandonata la prospettiva della necessaria (ricostruibilità di una) 'spiegazione' della decisione 'opaca' della macchina, prevedano, a tutela dei singoli, l'implementazione di sistemi idonei a rilevare e prevenire trattamenti i discriminatori nell'ambito del processo decisionale automatizzato. Si veda a riguardo C. DWORK, M. HARDT et al., *Fairness Through Awareness*, in *ITCS 2012*, Cambridge, MA (USA), 2012 oppure online at *https://arxiv.org/abs/1104.3913v2*; citato anche in J. BURRELL, *op. cit.*

[109] Explanatory tools consentono in taluni casi, e solo ex post, di analizzare i passaggi ragionativi del sistema ricostruendone i momenti decisionali essenziali in modo da restituire, a richiesta, la logica decisionale concretamente utilizzata dall'algoritmo nel caso analizzato; v quanto *supra*, nota 27. La ricerca è del resto da tempo attiva in questo settore, con l'obiettivo di rendere gli algoritmi *ML* sempre più suscettibili di controllo *ex ante* ed *ex post*; v. *inetr alia* A. DATTA, S. SEN, Y. ZICK, *Algorithmic transparency via quantitative input influence*, in *37th IEEE Symposium on Security and Privacy*, 2016, online at: *https://ieeexplore.ieee.org /document/7546525*; S. GALHOTRA, R. PRADHAN, B. SALIMI, *Explaining black-box algorithms using probabilistic contrastive counterfactuals*, 2021, disponibile online su arXiv:2103.11972v1; S. WACHTER, B. MITTELSTADT, C. RUSSELL, *Counterfactual explanations without opening the Black Box: automated decisions and the GDPR*, in *Harvard Journal of Law & Technology*, Vol. 31, n. 2, 2018, 843-887; M. LOI, A. FERRARIO, E. VIGANÒ, *Transparency as design publicity: explaining and justifying inscrutable algorithms*, in *Ethics and Information Technology*, October 2020, doi: *https://doi.org/10.1007/s10676-020-09564-w*; sul concetto di *Algorithmic accountability*, e sul possibile ruolo, quale *audit* esterno al sistema decisionale automatizzato, di algoritmi di *reverse-engineering*, si veda anche N. DIAKOPOULOS, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, Tow Center for Digital Journalism, Columbia University, 2014, *https://doi.org/10.7916/D8ZK5TW2*.

[110] Il riferimento è a quel filone di ricerca che va sotto il nome di XAI, Explanatory Artificial Intelligence, volto alla definizione di metodi e strumenti di "deep explanation", in grado cioè di far luce sul funzionamento di algoritmi dal funzionamento 'opaco', ovvero alla costruzione di sistemi intelligenti maggiormente trasparenti ed interpretabili; v.si H.J. WATSON, *The Need for Explainable Processes and Algorithms*, in *Business Intelligence Journal*, Vol. 25, n. 2, p. 5; A.B. ARRIETA, N. DÌAZ-RODRÌGUEZ et al., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*, 2020, online in *ArXiv*, *abs/1910.10045*.

'inutile' il diritto alla spiegazione che qui si vuole prescritto dal GDPR; inutile, perché impossibile sarebbe, per il titolare del trattamento che si affidi all'utilizzo di sistemi '*black box*' ad apprendimento autonomo, ricostruire (e quindi condividere) la logica decisionale seguita dalla macchina. Sicuramente ex ante; in molti casi *ex post* [111].

Ciò posto, occorre chiedersi in che modo l'ipotesi della irrealizzabilità pratica del diritto alla spiegazione reagisca nel sistema del GDPR. Come argutamente rilevato [112], in questa prospettiva, nelle ipotesi in cui si assuma essere precluso il tipo di spiegazione richiesto, risultando impossibile verificare liceità e correttezza del trattamento, la conseguenza che si produce sul piano logico non sarebbe l'inapplicabilità della norma – quindi la sua inutilità – ma il ripristino del divieto – assoluto, mancandone le condizioni minime di derogabilità [113] – di sottoposizione dell'interessato al trattamento decisionale automatizzato di cui all'art. 22 (1) GDPR. A meno che il titolare del trattamento, dunque, non possa garantire, anche sul piano tecnico [114], che la decisione basata (unicamente) su trattamento automatizzato dei dati personali dell'interessato (e dotata, per questi, di effetti giuridicamente rilevanti) sia anche spiegabile, ne sia cioè rintracciabile la logica decisionale *concreta*, è tenuto ad astenersi dal mettere in pratica il trattamento decisionale [115].

---

[111] Scettico sul fatto che possano trovarsi, anche in futuro, soluzioni tecnologiche realmente in grado di assicurare standard elevati di trasparenza dei processi decisionali animati da complessi algoritmi di ML, e pertanto critico circa la possibilità ed opportunità di incidere sul design tecnologico attraverso strumenti di regolazione, è ad esempio, Y. BATHAEE, *op. cit.*, 929; sulla possibilità di approcci differenti a quello della trasparenza *ex ante* o *ex post* (*audit*) dell'algoritmo, v., tra gli altri, C. DWORK, M. HARDT et al., *op. cit.*; sui vantaggi di un approccio scalare *case-by-case*, lo stesso Y. BATHAEE, *op. cit.*, 936 ss.

[112] R. MESSINETTI, *op. cit.*, 889.

[113] In tal senso l'art. 22(3) GDPR.

[114] La sede per queste valutazioni è la DPIA, la valutazione d'impatto della protezione dei dati, prevista *ex* art. 35 (e Considerando 84, 89-93 e 95) GDPR, quale strumento essenziale di *accountability* per tutti i trattamenti di dati personali che presentino un rischio elevato per i diritti e le libertà delle persone fisiche. La valutazione d'impatto è *in particolare richiesta*, ai sensi dell'art. 35(3)(a) del Regolamento, laddove il trattamento sia funzionale a 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*". La fattispecie è ampia e pacificamente include i processi decisionali di cui all'art. 22(1) GDPR nell'area della DPIA obbligatoria, oltre a richiederla anche laddove la decisione non sia basata *unicamente* sul trattamento automatizzato, purché abbia un impatto elevato per le persone fisiche coinvolte. Si v. sul punto Article 29 Working Party, *op. cit.*, 33; piu in generale sulla DPIA, Article 29 Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, n. 248 del 4 Aprile 2017, WP 248 rev.01. Valorizza il ruolo dell'*accountability*, della valutazione d'impatto e delle Autorità garanti, in una lettura complessiva del GDPR orientata alla trasparenza e responsabilizzazione degli algoritmi decisionali, B. CASEY, A. FARHANGI, R. VOGL, *Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise*, in *Berkeley Technology Law Journal*, Vol. 34, No 1, 2019, doi: *https://doi.org/10.15779/Z38M32N986*.

[115] Naturalmente l'analisi anticipata e specifica di tutti i rischi connessi al sistema (v. nota precedente), che dovrebbe guidare la scelta e l'attuazione delle misure tecnico-organizzative previste per la loro *governance* una volta individuati, non è – e non va – limitata – com'é ovvio – all'obiettivo di garantire un trattamento trasparente e informato dell'interessato, o metterlo in condizione di accedere a strumenti di *redress* e invocare l'intervento umano; il titolare del trattamento è anche chiamato ad assicurare che il trattamento decisionale sia corretto, in quanto non viziato da *bias*, basato su dati completi ed esatti ed assistito da garanzie adeguate ad evitare *output* discriminatori e lesivi. A carico degli utilizzatori, ma anche degli ideatori e sviluppatori degli algoritmi decisionali che importino il trattamento di dati personali di persone fisiche dovrebbe infatti intendersi imposta, già ai soli sensi del GDPR, un'articolata rete di *obblighi di protezione*, che impongono l'adozione di strumenti e metodi di *privacy by design* e *by default* – vale a dire misure tecniche e organizzative – che assicurino la conformità ai principi normativi vigenti in materia di *data protection* e il rispetto dei diritti umani fondamentali; e ciò sin dalla fase iniziale dell'ideazione degli algoritmi e, dunque, per impostazione predefinita. Sul punto, v. A. ROIG, *Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)*, in *European Journal of Law and Technology*, Vol 8, No 3, 2017; C. DJEFFAL, *op. cit.*; in particolare sul dovere di garantire la correzione continua dei sistemi per arginare *bias* cognitivi e statistici, come portato naturale del Regolamento, C. MALGIERI, G. COMANDÈ, *op. ult. cit*. Sul concetto di Transparency-by-design si v., H. FELZMANN, E. FOSCH-VILLARONGA, C. LUTZ, A.TAMÒ-LARRIEUX, *Towards Transparency by Design for Artificial Intelligence, in Science and Engineering Ethics*, October 2020, online al doi: *https://doi.org/10.1007/s11948-020-00276-4*.