



Camilla Tabarrini

**COMPRENDERE LA "BIGMIND".
IL GDPR SANA IL DIVARIO DI
INTELLIGIBILITÀ
UOMO-MACCHINA?**

Estratto



GIUFFRÈ FRANCIS LEFEBVRE

CAMILLA TABARRINI

COMPRENDERE LA “*BIG MIND*”: IL GDPR SANA IL DIVARIO DI INTELLIGIBILITÀ UOMO-MACCHINA?

SOMMARIO: 1. Una introduzione empirica. — 2. La sfida del GDPR: il diritto ad una spiegazione. — 3. L’opacità intellettuale: la trappola della trasparenza. — 4. L’opacità intenzionale: uno scontro con la proprietà intellettuale? — 5. L’opacità intrinseca: un costo che vale la pena pagare?

1. UNA INTRODUZIONE EMPIRICA.

Il termine industria 4.0 evoca tutte le nuove frontiere imprenditoriali offerte dalle recenti innovazioni tecnologiche volte alla raccolta, all’archiviazione e all’analisi di dati strutturati e non strutturati caratterizzati da un volume e una varietà senza precedenti, e generati ad un ritmo in costante crescita: i *Big Data*¹. Allo stato attuale, è stimato che ogni giorno vengono creati 2,5 quintilioni di *bites* di dati e che questo ritmo è principalmente imputabile a due nuove fonti di informazioni: (i) gli *user-generated contents* prodotti automaticamente dagli utenti con il loro comportamento *online* (e.g. interazioni nei *social network* o ricerche effettuate a mezzo *browser*)² e (ii) *l’Internet of Things*, termine con il quale si indica il *network* globale di dispositivi fisici

* Il presente lavoro è stato preventivamente sottoposto a referaggio anonimo affidato a un componente il Comitato Scientifico dei Referenti della Rivista secondo le prassi correnti nella comunità dei giuristi.

¹ Il parametro delle cc.dd. tre V (Volume, Velocità e Varietà) è generalmente integrato da altre due V: Veridicità e Variabilità, ad indicare il grado di accuratezza e di pressoché quotidiano aggiornamento del flusso informativo. Vedi GRUPPO DI LAVORO ARTICOLO 29, Parere 03/2013 sulla *Purpose Limitation*, 00569/13/EN WP 203, 2 aprile 2013. Circa i caratteri distintivi dei *Big Data* si veda anche M.N. HELVESTON, *Consumer Protection in the Age of Big Data*, 93, *Wash. U. L. Rev.*, 2016, 859, 867; F.

MATTASSOGLIO, *Big Data*, in M. T. PARACAMPO (a cura di), *FinTech: Introduzione ai profili giuridici di un mercato unico tecnologico dei servizi finanziari*, Giappichelli Editore, Torino, 2017, 69 e ss. Per un puntuale inquadramento del fenomeno si veda anche V. MAYER-SCHONBERGER e T. RAMGE, *Reinventing Capitalism in the Age of Big Data*, John Murray Publishers, Londra, 2018.

² Questa tendenza alla condivisione online quotidiana sta conducendo alla c.d. “*hyper-self-consciousness*” dal momento che, come acutamente osservato: “[n]ever before in the history of humanity have so many people monitored, recorded, and reported so many details about themselves to such a large audience”. F. PASQUALE, *The*

ad alta intensità tecnologica che, nel servire specifiche esigenze di vita quotidiana, sono contemporaneamente in grado di raccogliere e trasmettere dati (e.g. sensori telematici inseriti nelle macchine, *wearables* o dispositivi di domotica) ³.

Maggiore quantità di dati, tuttavia, non implica necessariamente dati più grandi, poiché ciò che rende un dato “grande” è la sua sfera di influenza, ovvero il numero di attributi che descrivono quella particolare informazione. Maggiore è il numero dei descrittori, più probabile sarà trovare sovrapposizioni tra le sfere di influenza di dati diversi, così tracciando *pattern* non individuabili in un contesto *small data*. Per questo motivo, i motori di ricerca (e.g. Google) si stanno progressivamente trasformando in “motori di scoperta” che, sfruttando tali tecniche di estrazione e collegamento di attributi, offrono una nuova forma di conoscenza altamente individualizzata: è questo il processo di c.d. *knowledge discovery in databases* ⁴. Sebbene siffatta attitudine alla “scoperta” sia ancora piuttosto latente nei motori di ricerca, lasciando all’utente finale la scelta delle informazioni che ritengono più utili tra i risultati prodotti dalla ricerca *web*, la rapida evoluzione delle tecnologie di elaborazione dei dati sta fortemente incrementando l’abilità degli *internet service providers* di profilare gli utenti. Per questo motivo è verosimile immaginare che nel prossimo futuro i risultati indicizzati dai motori di ricerca mostreranno non solo un grado di accuratezza accentuato dai *pattern* sempre più profondi che li legano, ma anche un livello di personalizzazione tale da filtrare preventivamente i possibili *output* in base a ciò che il

Fourth Revolution. How the Infosphere is Reshaping Human Reality, Oxford, Oxford University Press, 2014, 62 ss.

³ Google, Amazon, Apple, Facebook e Alibaba (anche noti come i GAFAA) controllano gran parte delle nuove categorie di dati, favorendo così ipotesi di abuso di mercato derivanti dai cc.dd. “*data lock-ups*”, intesi come l’esercizio di atteggiamenti dominicali su tali imponenti patrimoni informativi, così compromettendo l’assetto concorrenziale del mercato attraverso lo sfruttamento unilaterale del potenziale di *Data Analytics* insito nelle accurate analisi predittive di mercato inferibili dai *Big Data* e sfruttabili al fine di indirizzare le strategie imprenditoriali verso i settori più remunerativi, relegando le imprese meno “*data-driven*” ai margini del mercato (c.d. *cream-skimming*). Tra queste ultime sono annoverabili anche i principali operatori tradizionali del settore interessato (*incumbents*) dotati di grandi moli di dati di alta qualità circa la propria clientela, ma privi dell’avanzato *know-how* tecnologico attualmente ne-

cessario per monetizzare tali conoscenze. Cfr. R. VAN LOO, *Making Innovation More Competitive: The Case of Fintech*, 6, *UCLA L. Rev.*, 2018, 232, 242; GENEVA ASSOCIATION, *Big Data and Insurance: Implications for Innovation, Competition and Privacy*, Zurigo, 2018, 22 ss.

⁴ Il processo generalmente si compone di diversi passaggi: (i) la selezione dei dati; (ii) il pre-trattamento dei dati, consistente nell’aumentarne la qualità rimuovendo qualsiasi interferenza suscettibile di incidere negativamente sulla precisione delle successive analisi; (iii) la trasformazione dei dati, intesa come riduzione della varietà dei dati ad una dimensione c.d. “estraibile o *minable*”; (iv) l’estrazione dei dati (o *data mining*), ovvero l’analisi algoritmica deputata alla individuazione di *pattern* tra i dati e (v) la valutazione dei dati, volta ad inferire conoscenza monetizzabile dalle informazioni fin qui ricavate. Vd U. FAYYAD, G. PIATETSKY-SHAPIRO, and PADHRAIC SMYTH, *From Data Mining to Knowledge Discovery in Databases*, 17, *AI Magazine*, 3, 1996, 37, 39 ss.

motore di scoperta ha ritenuto più rilevante per il profilo attribuito allo specifico utente ⁵.

Le tecniche per il trattamento dei dati volte ad allargarne la sfera di influenza si distinguono in attive e passive. Le prime richiedono un qualche intervento umano teso ad aggiungere attributi ai dati, per esempio, il *crowdsourcing* ingrandisce le informazioni tramite *feedbacks*, commenti, Captcha, *hashtags* di Instragram, e molte altre forme di interazione *web*. Dall'altro lato, la principale tecnica passiva di trattamento dei dati è il *Machine Learning* (ML), una sottocategoria *deep learning* di sistemi di Intelligenza Artificiale in grado di migliorare in modo automatico, così potendo raccogliere, archiviare ed ingrandire dati in modo autonomo ⁶.

Ciononostante, il notorio valore economico dei *Big Data* non è imputabile alla peculiare e crescente estensione della loro sfera di influenza, bensì alla menzionata e altamente redditizia forma di conoscenza che quest'ultima offre alle imprese disposte ad investire. In questo senso, ciò che gioca un ruolo realmente fondamentale nella *Data-Driven Economy* è l'abilità di "minare" nuove informazioni da *Big Data* allo stato grezzo (i c.d. metadati) e utilizzarle per sviluppare prodotti, servizi e strategie di *marketing* modulate sull'utente finale. Personalizzazione e cura della clientela sono, infatti, le nuove frontiere dell'economia *consumer-centric*, nonché obiettivo primario di qualsiasi imprenditore determinato a preservare la produttività della sua attività ⁷.

Tali prospettive di guadagno, tuttavia, vanno ben oltre il campo d'azione degli inserzionisti, poiché la possibilità di creare dei "*digital searchable files*" contenenti tutte le informazioni *online* e *offline* disponibili su un determinato cittadino è, per ovvie ragioni, estremamente attraente in molti settori pubblici e privati.

Il principale esempio delle potenzialità insite nella profilazione a mezzo *Big Data* è rappresentato dal Sistema di Credito Sociale cinese: un frammentato ecosistema di iniziative pubbliche e private finalizzate a rafforzare la *compliance* giuridica e regolamen-

⁵ G. D'ACQUISTO e M. NALDI, *Big Data e Privacy by Design. Anonimizzazione, Pseudonimizzazione, Sicurezza*, Giappichelli Editore, Torino, 2017, 12-14. Per una descrizione più generica del ruolo delle correlazioni vd V. MAYER-SCHÖNBERGER e K. CUKIER, *Big Data. A Revolution that Will Transform how We Live, Work and Think*, Houghton Mifflin Harcourt Publishing, New York, 53 ss.

⁶ G. D'ACQUISTO e M. NALDI, *Big Data e Privacy by Design. Anonimizzazione, Pseudonimizzazione, Sicurezza*, Giappichelli Editore, Torino, 2017, 9 ss.

⁷ Il *Data Mining* è anche definito come

"the automated or convenient extraction of patterns representing knowledge implicitly stored or catchable in large databases, data warehouses, the Web, other massive information repositories, or data streams." T. CALDERS e B. CUSTERS, *What is Data Mining and How Does it Work?* in B. CUSTERS, T. CALDERS, B. SCHERMER, e T. ZARSKY (a cura di), *Discrimination and Privacy in the Information Society. Data Mining and Profiling in Large Databases*, Springer, New York, 2013, 28; vedi anche V. ZENOVICH, *Ten Legal Perspectives on the "Big Data Revolution"*, 1, *Concorrenza e mercato*, 29, 2016, 29.

tare attraverso strumenti di controllo sociale ad alta intensità tecnologica⁸. In ragione delle risalenti problematiche di *enforcement* storicamente affrontate dalla Cina, infatti, lo sviluppo tecnologico è stato da subito percepito come uno strumento per rafforzare il controllo sociale attraverso l'implementazione di programmi mossi dalla c.d. "cultura della sincerità". La credenza del mandato divino, infatti, che investe di potere un sovrano giusto e responsabile anche della promozione di valori morali all'interno di una società concepita come un tutto armonico, ha progressivamente indebolito il confine tra la sfera pubblica e la sfera privata, favorendo l'emersione di una nozione di credito che linguisticamente ricomprende tanto la meritevolezza creditizia quanto l'onestà del singolo. A partire dagli anni 2000, tale approccio paternalistico è poi sfociato in un rinnovato appello a comportamenti affidabili nel mercato, concepito come il principale mezzo di rafforzamento dell'economia cinese⁹.

Nel 2006, successivamente alla creazione del "Registro di Credito Bancario e Sistema di Consulenza", la Banca Popolare Cinese ha istituito un'agenzia unica a livello nazione per il *credit-scoring*: il Centro di Referenze Creditizie, al quale confluiscono tanto i dati relativi al merito creditizio della clientela in possesso di imprese bancarie, quanto informazioni non finanziarie aggiuntive fornite dalle corti, dalle autorità pubbliche, nonché dalle compagnie di telecomunicazioni. In alcuni casi, come quello della città di Suining, questo fenomeno è sfociato in un programma di credito di massa in cui i mille punti di credito iniziale assegnato a ogni cittadino veniva ridotto in modo proporzionale alla gravità delle infrazioni eventualmente commesse. Il punteggio residuo avrebbe poi determinato la misura dei benefici riconosciuti (e.g. accesso privilegiato all'impiego), ovvero del trattamento deteriore riservatogli nel prendere specifiche decisioni nei suoi confronti, come in caso di arruolamento militare, alloggi popolari o la concessione di licenze. Nel 2002 un sistema analogo è emerso in Hangzhou, la capitale della provincia di Zhejiang, dove nel 2007 è stato implementato anche un piano provinciale. Questi programmi locali sono stati poi replicati in più larga scala nel 2014 con il Piano di Credito Sociale, volto al raggiungimento di tre

⁸ In altre parole, l'obiettivo principale perseguito tramite tali iniziative è quello di implementare "meccanismi cibernetici di controllo sociale, nell'ambito dei quali gli individui e le organizzazioni sono monitorati al fine di porli di fronte alle conseguenze delle loro azioni in modo automatizzato." (traduzione dell'autore) R. CREEMERS, *China's Social Credit System: An Evolving Practice of Control* (9 maggio 2018) disponibile all'indirizzo <https://ssrn.com/abstract=3175792>.

Cfr. V. SHUE e P. THORNTON, *Beyond Implicit Political Dichotomies and Linear Models of Change in China*, in Id. (a cura di), *To Govern China: Evolving Practices of Power*, Cambridge University Press, Cambridge, 2017, 4 ss.

⁹ Sull'origine della versione in cinese mandarino del termine "credito" (*xinyong*) si veda R. CREEMERS, *China's Social Credit System: An Evolving Practice of Control*, *op. cit.*, spec. nt. 13.

obiettivi principali: aumentare la trasparenza dell'azione governativa, rafforzare la fiducia nell'economia di mercato e incrementare lo scrutinio pubblico sulle condotte *online* ¹⁰.

Il Sistema di Pene Congiunto del 2016 è il principale componente del piano sessennale di Credito Sociale. Quest'ultimo, ispirato al principio della "sanzione sproporzionata", era originariamente concepito come uno strumento per assicurare una più effettiva ottemperanza di sentenze vincolanti ai sensi della emendata legge sul contenzioso civile. In base ad un regolamento del 2013 della Corte Popolare Suprema, infatti, chiunque violi un ordine della corte è inserito in una lista nera. Dal 2016, tuttavia, tale lista è gestita congiuntamente da quarantacinque organismi partecipanti, tenuti a prendere in considerazione lo *status* risultante dalla *black list* per applicare diverse categorie di restrizioni, quali ad esempio il divieto di acquistare azioni in borsa o di avvalersi di altre forme di accesso al settore finanziario, di sussidi di Stato, ovvero limitazioni al consumo di lusso (e.g. viaggi in prima classe, pernottamenti in grand hotel o acquisti di lusso), così come il divieto di sfruttamento di risorse naturali o l'esclusione dalla possibilità di ricevere titoli onorari ¹¹.

Con riferimento al settore privato cinese, invece, la Ant Financial Group (controllata di Alibaba) ha adottato un suo proprio sistema di credito chiamato Credito Sesame, con il quale vengono assegnati punteggi dai 350 ai 950 ad ogni cliente in base alla propria storia creditizia, allo storico dei propri acquisti, alla stabilità di reddito, alla qualità e quantità di dati personali verificabili e perfino alle interazioni in *social media* ¹². Sebbene il sistema di calcolo mostri una componente di elaborazione algoritmica dei dati ancora piuttosto residuale, l'ampliamento della pletora di categorie di informazioni disponibili e l'evoluzione delle opportunità di archiviazione, organizzazione e trattamento offerte dai *Big Data*, rendono difficile escludere che il contestuale

¹⁰ *China invents the digital totalitarian state. The worrying implications of its social-credit project*, The Economist (Pechino, 17 dicembre 2016); C. NGUYEN, *China might use data to create a score for each citizen based on how trustworthy they are* (Business Insider, 26 ottobre 2016). Per una dettagliata descrizione del paradigmatico caso del giornalista cinese Liu Hu che, condannato per diffamazione, non aveva potuto acquistare biglietti aerei in quanto "non qualificato" in ragione del suo basso credito sociale, si veda N. VANDERKLIPPE, *Chinese blacklist an early glimpse of sweeping new social-credit control* (The Globe and Mail, Pechino, 18 gennaio 2018).

¹¹ I soggetti che contattano telefonicamente individui inseriti in *black lists* sono

peraltro generalmente preventivamente avvisati della circostanza per cui "la persona che stai chiamando è inserita nella lista nera del della centrale del credito", vd *supra* nt. 9; Cfr. A. XU, *Chinese Judicial Justice on the Cloud: A Future Call or a Pandora's Box? An Analysis of the 'Intelligent Court System' of China*, 26, *Information & Communications Technology Law*, 1, 59, 2017, 62 ss.

¹² Ad esempio, il Canada non richiede più gli estratti bancari ai fini della elaborazione delle richieste di visti di ingresso nel Paese presentate da clienti di Ant Financial dotati di un credito sesame superiore a 750. C. UDEMANS, *Alipay's Sesame Credit now accepted by Canada for visa applications* (Technode, 23 novembre 2018).

sviluppo di sempre più sofisticati sistemi di Intelligenza Artificiale possa finire col condurre ad un “sistema automatizzato di controllo sociale basato sui dati”, così inasprendo la già montante preoccupazione per l'arbitrarietà e l'intrusività ingiustificata di questi procedimenti di calcolo della “affidabilità” dei cittadini ¹³. Nonostante il peso di ogni variabile sia pubblicamente specificato, infatti, il funzionamento dei punteggi Sesame rimane perlopiù secretato e opaco, rendendo così sostanzialmente impossibile per i singoli utenti prevedere l'impatto che ogni comportamento avrà sul loro punteggio complessivo.

Al di là di tali scenari orwelliani ¹⁴, fortunatamente ancora rari, analoghi processi computazionali sono adottati da un crescente numero di imprese finanziarie tradizionali (i.e. *Incumbents*), di *Bigtechs* e *Start up* del settore *Fintech*, per svolgere valutazioni del merito creditizio della clientela più complete e dettagliate ¹⁵.

A differenza del sistema tradizionale di *credit scoring*, generalmente supportato da soluzioni algoritmiche semplici in parte riconducibili ai limiti tecnologici all'utilizzazione di categorie non tradizionali di dati solo recentemente superati, molti progetti pilota stanno attualmente promuovendo l'utilizzo di metodi *Machine Learning* per trovare correlazioni tra una pletora di dati *online* e *offline* relativi alla propria clientela molto più ampia, in quanto disponibile pubblicamente e/o acquistata presso *data brokers* in ossequio al c.d. “*all data is credit data approach*” ¹⁶.

Dal punto di vista informatico, il processo di sviluppo di sistemi

¹³ R. CREEMERS, *China's Social Credit System: An Evolving Practice of Control*, op. cit., vd. *supra* nt. 9. Cfr. N. KOBIE, *The complicated truth about China's social credit system*, in *Wired*, 21 gennaio 2019; L. YILUN CHEN, *Jack Ma's Ant Apologizes for Baiting Users Into Credit System*, in *Bloomerang*, 4 gennaio 2018; A. VISWANATHA e KATE O'KEEFE, *Before It Was Hacked, Equifax Had a Different Fear: Chinese Spying*, in *The Wall Street Journal*, 12 settembre 2018.

¹⁴ Vedi S. GRAFANAKI, *Autonomy Challenges in the Age of Big Data*, 27, *Fordham Intell. Prop. Media & Ent. L.J.*, 803, 2017, 817 ss.

¹⁵ A tal proposito, l'esempio più calzante è il prodotto Decision Hub offerto da Big Data Scoring che, nelle parole stesse della società è un “*powerful cloud-based decision engine that takes care of the entire underwriting process: data collection, pre-processing, risk management, business rules, predictive algorithms and monitoring. All of that happens in real time and fully automatically, eliminating the possibility of human errors and personal judgement.*” [en-

fasi aggiunta] (<http://www.bigdatascore.com/decision-hub/>). Nessun accenno, tuttavia, viene fatto alla possibilità di errori o effetti discriminatori legati alla errata programmazione del *software*, rispetto alla quale si dirà meglio nei prossimi paragrafi. Cfr. EBA, ESMA, EIOPA, *Joint Committee Final Report on Big Data*, 15 marzo 2018 (JC/2018/04); FINANCIAL STABILITY BOARD, *FinTech credit Market structure, business models and financial stability implications*, 22 maggio 2017 e COMITATO DI BASILEA PER LA VIGILANZA BANCARIA, *Sound Practices Implications of fintech developments for banks and bank supervisors*, febbraio 2018.

¹⁶ M. HURLEY - J. ADEBAYO, *Credit Scoring in the Era of Big Data*, 18 *Yale J.L. & Tech.*, 2016, 148, 164. Per una approfondita analisi del funzione di ZestFinance (il principale sistema di *credit scoring* attualmente rispondente al principio del “*all data is credit data*”) vedi S. AZIZ - M. DOWLING, *Machine Learning and AI for Risk Management*, in T. LYNN, J.G. MOONEY, P. ROSATI e M. CUMMINS (a cura di), *Disrupting Finance: FinTech and Strategy in the 21st*

di *scoring* è composto da tre passaggi principali, ognuno dei quali caratterizzato da peculiari problematiche giuridiche¹⁷. In primo luogo è necessario definire il problema che, nel caso in questione, consiste nello specificare le variabili rilevanti ai fini della definizione della specifica formula di meritevolezza creditizia da applicare alla clientela. La mancanza di una nozione universalmente accettata di merito creditizio, tuttavia, presenta il rischio latente di discriminazione ogni qualvolta il programmatore ometta di ponderare adeguatamente ogni variabile, integrando una emblematica espressione dell'attuale dominio tecnocratico sulla consapevolezza giuridica¹⁸.

La seconda fase consiste nella raccolta e trasformazione dei dati, volta a tradurli in una forma analizzabile, quale quella di punteggi percentuali o variabili ordinarie. Sebbene la quantità e varietà potenzialmente illimitata di dati personali disponibili nell'era dei Big Data possa indubbiamente migliorare la precisione dei modelli di accertamento della meritevolezza creditizia, il descritto ampliamento della sfera di influenza dei dati rende altamente probabile l'individuazione di correlazioni spurie tra gli attributi dei dati da parte di algoritmi ML, così pregiudicandone l'esattezza del risultato. Inoltre, se nella fase di raccolta non si rispetta un ferreo criterio di neutralità, raggruppando un campione non pienamente rappresentativo di dati, il modello algoritmico così prodotto potrebbe perpetuare la posizione deteriorata di gruppi sociali già sotto-rappresentati (e.g. privi di una connessione internet)¹⁹.

L'ultimo passaggio implica invece l'utilizzo di processi di ma-

Century, Palgrave Macmillan, Dublino, 2019, 35.

¹⁷ In linea generale, è stato correttamente affermato che “distorsioni ed errori possono manifestarsi nel contesto di decisioni robotiche in tre ambiti: i dati, la programmazione dell'algoritmo, e nel risultato.” G. MALGIERI e G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7, *International Data Privacy Law*, 4, 2017, 243, 248. Cfr. S. BARO-CAS - A.D. SELBST, ‘Big Data’s Disparate Impact’, 104, *California Law Review*, 2016, 671, 680; F. FERRETTI, *Consumer Access to Capital in the Age of FinTech and Big Data: the Limits of EU Law*, 25, *Maastricht Journal of European and Comparative Law*, 4, 2018, 476-499.

¹⁸ A. OTTOLIA, *Big Data e Innovazione Computazionale*, in *Quaderni di Aida* n. 28, Torino, Giappichelli Editore, 2017, 6 ss. Circa la “dittatura dell'algoritmo” di veda anche S. RODOTÀ, *Il mondo nella rete. Quali*

diritti, quali vincoli, Editori Laterza, Roma, 2014, 37 ss. Più genericamente in materia di discriminazione algoritmica vedi M. RHOEN - Q. YI FENG, *Why the ‘Computer says no’: illustrating big data’s discrimination risk through complex systems science*, 8, *International Data Privacy Law*, 2, 2018, 140.

¹⁹ In particolare, si possono distinguere due forme di pregiudizio algoritmico: (i) il pregiudizio cognitivo, inteso come una inesatta rappresentazione della realtà imputabile a difetti connaturati alla fase di raccolta ovvero a discriminazioni implicite nel raggruppamento degli stessi; e (ii) il pregiudizio statistico, ricorrente ogni qualvolta i dati siano estrapolati da contesti affetti da discriminazione strutturale. G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit., 243, 249 ss. Per ulteriori osservazioni inerenti il *digital divide* vedi A. MASERA - G. SCORZA, *Internet, I Nostri Diritti*, Editori Laterza, Bari, 2016, 10 e ss.

chine learning per ottimizzare e analizzare i dati. Con il primo termine ci si riferisce alla circostanza per cui i *software* di *deep learning* generalmente elaborano i dati immessi come *input* per determinare quali variabili sono più rilevanti e che peso dovrà essere assegnato ad ognuna nel corso del processo di *scoring*. Questa fase, tuttavia, per ovvie ragioni pone importanti ostacoli alla trasparenza del punteggio finale, compromettendo così la capacità dell'utente finale di valutarne la correttezza, l'equità, la legittimità ed eventualmente l'opportunità di contestarlo. In linea teorica, questa affermazione potrebbe essere valida per qualsiasi decisione tecnicamente complessa, ma nel caso del *machine learning* “non supervisionato” i programmatori non predeterminano le categorie attraverso le quali gli attributi dei dati verranno raccolti, organizzati e processati, conferendo alla macchina stessa la capacità di compiere queste valutazioni autonomamente imparando dall'esperienza (i.e. i dati utilizzati per programmarla e quelli successivamente acquisiti). Per questa ragione è conseguentemente impossibile per gli stessi programmatori prevedere il risultato finale del processo decisionale automatizzato. Peraltro, anche la precedente fase di trasformazione, attraverso la quale i dati sono tradotti in forme intelligibili ad un computer, non è idonea a rendere gli stessi altrettanto intelligibili ad operatori umani, indipendentemente dal grado di *expertise* maturato nell'ambito del diritto bancario o in economia. Infine, poiché determinare la capacità di restituzione di un debito costituisce un c.d. “problema non strutturato”, questo viene risolto attraverso una procedura intrinsecamente incerta, in quanto l'algoritmo utilizzabile varia a seconda degli specifici dati di input inseriti, cosicché per lo stesso problema possono essere fornite più risposte corrette ²⁰.

2. LA SFIDA DEL GDPR: IL DIRITTO AD UNA SPIEGAZIONE.

Alla luce di quanto detto, si può facilmente comprendere perché autorevoli studiosi parlino della nostra come della “*black box society*”, per evidenziare tanto il rafforzato potenziale tecnologico di archiviazione e registrazione delle informazioni, quanto l'intrinseca oscurità del funzionamento dei sistemi di decisione automatizzata ²¹. La tendenza a convertire “*quantified selves*” ²² in punteggi utilizzati per prendere decisioni sempre più frequentemente delegate a sistemi di Intelligenza Artificiale, infatti, implica

²⁰ Al contrario, i problemi strutturati sono risolvibili attraverso algoritmi fissi che rimangono immutati al mutare delle variabili di *input* inserite. M. HURLEY - J. ADEBAYO, *Credit Scoring in the Era of Big Data*, 18 *Yale J.L. & Tech.*, 2016, 148, 159.

²¹ In linea di massima “si ha una vera scatola nera ogni qualvolta un'affermazione sia presentata come una mera constatazione di fatto senza alcun riferimento alla sua genesi o al suo autore” (traduzione dell'autore) come affermato da G. HARMAN, *Prince*

l'esercizio di un potere computazionale sulle persone "unredeemably opaque"²³ poiché, come saggiamente sottolineato, l'abilità di "scrutinare gli altri evitando di essere scrutinati a propria volta è una delle più importanti forme di potere"²⁴. Peraltro, che le imprese *Tech* siano pienamente consapevoli del problema è ulteriormente corroborato dalla recente *partnership* tra Microsoft e Zestfinance finalizzata alla progettazione di una tecnologia che possa consentire alle istituzioni finanziarie di comprendere le motivazioni alla base della specifica decisione circa la sottoscrizione di un determinato finanziamento presa dal sistema di ML²⁵.

Di conseguenza, non si può più prescindere dalla necessità di assicurare l'effettività di un quadro etico e giuridico appropriato e, dal momento in cui tutti i dati processati per valutare il merito creditizio devono essere, per definizione, riferibili allo specifico cliente, sono qualificabili come dati personali e ricadono nell'ambito di applicazione del Regolamento Generale sulla Protezione dei Dati Personali (d'ora in avanti GDPR)²⁶. In particolare, ai sensi dell'articolo 22 GDPR quando una decisione basata "unicamente" sul trattamento automatizzato dei dati produce effetti giuridici ovvero incide in modo analogo significativamente sui soggetti interessati ("come il rifiuto automatizzato di una richiesta

of Networks: Bruno Latour and Metaphysics, re.press, Melbourne, 2009, 37 e menzionato da F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, 2015, 4 e 222, spec. nt. 7.

²² A. DEMBOSKY, *Invasion of the Body Hackers*, in *Financial Times*, 10 giugno 2011. Cfr. N. DOW SCHÜLL, *Self in the Loop: Bits, Patterns, and Pathways*, in *the Quantified Self*, in Z. PAPACHARISSI (a cura di), *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, Routledge, New York, 2019, 26-30.

²³ F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, op. cit., 2 spec. nt. 5. Cfr. D. KEATS CITRON - F. PASQUALE, *The Scored Society: Due Process for Automated Predictions*, 89, *Washington Law Review*, 2014, 1, 10 e ss..

²⁴ F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, op. cit., 3. Cfr. J. VAN DIJCK, *Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology*, 12, *Surveillance & Society*, 2, 2014, 197-208.

²⁵ Come scritto dal giornalista di Forbes A. Knapp nel riportare le parole di D.

Berglund (vicepresidente senior della sezione di intelligenza artificiale della U.S. Bank): "Il fatto è che non abbiamo avuto la possibilità di utilizzare alcuni dei più avanzati sistemi di deep learning a causa del problema della spiegazione [...]. La risposta ottenuta con sistemi di deep learning può essere precisa, ma nel caso di richieste di finanziamento, questo non è sufficiente." Id., *Microsoft Partners With Fintech Startup ZestFinance To Bring Transparency To AI-Powered Financial Models*, in *Forbes*, 19 dicembre 2018.

²⁶ Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati) [d'ora in avanti GDPR], Articolo 4 para 1(1). Per una riflessione esauritiva sulle tendenze di cambiamento registrate nell'ambito della disciplina sulla protezione dei dati personali vedi C. HODGES, *Delivering Data Protection: Trust and Ethical Culture*, 1, *EDPL*, 2018, 1, 6 ss.; V. MAYER-SCHÖNBERGER - Y. PADOVA, *Regime Change? Enabling Big Data Through Europe's New Data Protection Regulation*, 17, *The Columbia Science & Technology Law Review*, 2016, 315, 321 ss.

di credito *online* [...] senza l'intervento umano")²⁷, questi ultimi hanno il diritto di non essere sottoposti a tale decisione.

Con riferimento al termine "unicamente", parte della dottrina ha sostenuto che tale requisito non precluda l'applicazione dell'articolo 22 GDPR a decisioni che ammettono forme marginali di coinvolgimento umano. Un intervento umano meramente passivo e volto alla meccanica applicazione di linee guida prefissate, infatti, non sarebbe idoneo a raggiungere "il livello minimo di interazione umana" richiesto dalla norma, nella misura in cui quest'ultimo impone l'esercizio di un potere di effettiva influenza sull'esito del processo decisionale automatizzato²⁸.

Il Gruppo di Lavoro Articolo 29 ha confermato questa interpretazione, sottolineando come un mero gesto simbolico non sia idoneo a qualificarsi come intervento umano ai sensi della norma che, al contrario, presuppone una significativa revisione umana "condotta da qualcuno che abbia l'autorità e la competenza per cambiare la decisione"²⁹.

In tal senso, nel caso in cui il processo decisionale automatizzato non presenti questo "sostanziale intervento umano": (i) se la decisione automatizzata è autorizzata dalla legge, è necessaria per concludere o eseguire il contratto, ovvero è basata sul consenso esplicito del soggetto interessato, ai sensi dell'articolo 22 co. 3 GDPR, quest'ultimo ha il diritto di ottenere una valutazione umana proattiva del "materiale automatizzato prodotto dall'algoritmo di profilazione"³⁰; in caso contrario, (ii) sebbene i soggetti interessati possano rifiutare la decisione automatizzata, ai sensi degli articoli 13 co. 2 lett. *f*, 14 co. 2 lett. *g* e 15 co. 1 lett. *h* GDPR, ove i dati personali siano raccolti rispettivamente presso loro stessi, presso terzi ovvero ogni qualvolta tali dati vengano trattati, il titolare del trattamento deve fornire agli interessati informazioni relative "all'esistenza di un processo decisionale automatiz-

²⁷ GDPR, Considerando 71.

²⁸ Il linguaggio utilizzato dall'articolo sembrerebbe riferirsi al caso in cui tanto i lavori preparatori quanto il risultato finale siano condotti dalla macchina, relegando il pur presente coinvolgimento umano ad una mera formalità. Questa è l'opinione fatta propria dal UK INFORMATION COMMISSIONER'S OFFICE (ICO), *Feedback request- profiling and automated decision-making*, 2017, 19.

²⁹ GRUPPO DI LAVORO ARTICOLO 19, *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679 adottate il 3 ottobre 2017* (versione emendata e adottata in data 6 febbraio 2018) (WP 251 rev.01).

³⁰ G. MALGIERI - G. COMANDÈ, *Why a*

Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation, op. cit., 243, 252. Per una interessante classificazione dei processi decisionali sostanziali e processuali, così come della distinzione tra quelli algoritmici e non, ovvero tra decisioni *cc.dd. rule-based* e *law-based* si veda M. BRKAN, *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond*, 11, *International Journal of Law and Information Technology*, 2019, 1-31. Cfr. T. ZARSKY, *The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making*, 41, *Science, Technology, & Human Values*, 2016, 118-32.

zato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato”³¹.

Di conseguenza, alcuni giuristi hanno sostenuto che l'articolo 22 co. 1 e 4 GDPR, se letto in combinato disposto con gli articoli 13 co. 1 lett. *f*, 14 co. 2 lett. *g* e 15 co. 1 lett. *h* GDPR, fissi il diritto ad una spiegazione della decisione che incida sulla sfera giuridica dei soggetti interessati ovvero leda in modo significativo i loro interessi legittimi, nel caso in cui tale decisione sia stata presa senza l'intervento umano³².

Questa interpretazione sistematica è ulteriormente corroborata dal considerando 71, in base al quale “tale 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”³³.

Nonostante l'indiscussa mancanza di autonomo valore normativo dei considerando, è comunque condivisibile, come ribadito dalla Corte di Giustizia dell'Unione Europea e dalla Commissione Europea, la considerazione per cui i primi svolgono un ruolo ermeneutico supplementare “facendo luce sull'interpretazione da attribuire ad una norma di legge”³⁴.

Riconosciuta l'esistenza del diritto, quindi, il successivo dibattito dottrinale si è spostato sul suo contenuto ed estensione. A tal proposito, utili indicazioni vengono dal requisito del carattere “significativo” delle informazioni introdotto dal GDPR, così specificando la natura dei doveri informativi già posti in capo ai responsabili del trattamento dalla previgente Direttiva 1996/45/CE. Alcuni studiosi, infatti, hanno sostenuto che affinché le informazioni sulla “logica utilizzata” nel processo decisionale automatizzato siano qualificabili come “significative”, queste dovrebbero essere “parametrata alle specifiche capacità di compren-

³¹ Dal disposto letterale della norma sembrerebbe, peraltro, che se un diritto ad una spiegazione esiste, dovrebbe essere considerato vincolante solo nei casi di cui ai commi 1 e 4 dell'articolo 22 GDPR. Per una disamina diacronica dello sviluppo storico della disposizione, in prospettiva comparatistica, si veda I. MENDOZA - L.A. BYGRAVE, *The Right Not to be Subject to Automated Decisions Based on Profiling*, in T.E. SYNOUDOU, P. JOUGLEUX, C. MARKOU, T. PRASTITOU (a cura di), *EU Internet Law. Regulation and Enforcement*, Springer, Berlino, 2017, 79.

³² Diversa l'autorevole opione di S.

WACHTER, B. MITTELSTADT e L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7, *International Data Privacy Law*, 2, 76 ss. Nello stesso senso, A.D. SELBST - J. POWLES, *Meaningful information and the right to explanation*, 7, *International Data Privacy Law*, 4, 2017, 233-242.

³³ Corsivo aggiunto.

³⁴ G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, *op. cit.*, 243, 254.

sione umana”³⁵. In altri termini, l’informativa dovrebbe raggiungere una soglia di intellegibilità tale da ricomprendere la comprensibilità, l’accessibilità e la trasparenza delle informazioni. Ciò discende dalla considerazione per cui i soggetti interessati sono effettivamente messi nella condizione di poter esercitare il loro diritto di contestare la decisione automatizzata solo nella misura in cui sono sì resi consapevoli del funzionamento matematico dell’algoritmo (trasparenza), ma sono nondimeno informati circa gli obiettivi e il potenziale impatto della sua applicazione al caso specifico.

Sebbene, come ribadito dal Parlamento Europeo, “dovrebbe essere sempre possibile fornire la logica sottostante la decisione presa con l’aiuto di IA che può avere un impatto sostanziale sulla vita di una o più persone; [e] [...] che dovrebbe essere sempre possibile ridurre i calcoli del sistema di AI ad una forma comprensibile dagli umani,”³⁶ le decisioni automatizzate si qualificano come “super decisioni” in quanto seguono inevitabilmente un flusso cognitivo che è quasi impossibile da replicare per la mente umana³⁷. Ne consegue che una descrizione del funzionamento dell’algoritmo meramente trasparente potrebbe essere controproducente, relegando il diritto ad una spiegazione ad un mero adempimento formale³⁸.

La difficoltà a rimuovere in modo efficace gli ostacoli all’intellegibilità è principalmente ascrivibile a quelle che Burrell ha definito come opacità intellettuale, intrinseca e intenzionale³⁹. La prima si riferisce alla frequente mancanza nei soggetti interessati delle competenze informatiche minime necessarie alla compren-

³⁵ IVI, 245.

³⁶ Par. 12, sezione “principi etici” della Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

³⁷ Come sottolineato da G. GITTI in occasione del suo intervento al convegno intitolato “La funzione delle norme generali sui contratti e sugli atti di autonomia privata. Prospettive di riforma del codice civile” svoltosi a Pisa il 29-30 novembre 2018.

³⁸ Come efficacemente messo in luce da F. Pasquale “quando la complessità è troppo elevata per essere monitorata o compresa, la disclosure diventa un gesto vuoto”. Perciò “dobbiamo porre fine al gioco ricorrente della “disclosure” e dei “trucchi per sconfiggere la disclosure” che hanno flagellato i legislatori. Operazioni che sono troppo complesse per essere spiegate a soggetti esterni, potrebbero ben essere troppo complesse perché ne sia consentita l’esistenza.” (traduzione dell’autore) F. PAQUALE, *The Black Box Society. The Secret Algorithms*

That Control Money and Information, op. cit., 16. Cfr. O. BEN-SHAHAR - C.E. SCHNEIDER, *More Than You Wanted to Know. The Failure of Mandated Disclosure*, Princeton University Press, Princeton, 2014, 55 ss.

³⁹ Più specificamente Burrell sostiene che “[t]hree distinct forms of opacity include: (1) opacity as intentional corporate or institutional self-protection 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.” J. BURRELL, *How the machine ‘thinks’: Understanding opacity in machine learning algorithms*, 1, *Big Data & Society*, 2016, 1-2. Cfr. B. LEPRI - J. STAIANO - D. SANGOKOYA - E. LETOUZÉ - N. OLIVER, *The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-*

sione dei meccanismi di funzionamento dei sistemi di *machine learning* utilizzati per prendere decisioni che li riguardano. L'opacità intenzionale, invece, indica le barriere regolamentari volontariamente erette (i.e. i diritti di proprietà intellettuale) dai responsabili del trattamento e generalmente addotte a giustificazione del rifiuto opposto a richieste di accesso al codice sorgente di un sistema *machine learning*, così come di informazioni di dettaglio sul funzionamento del *software* o persino dei *metadata* di profilazione ottenuti dalla elaborazione dei cc.dd. *raw Big Data*.

Da ultimo, l'opacità intrinseca, esprime l'impossibilità tecnologica di predire e tradurre in una forma umanamente intellegibile il complessivo processo decisionale algoritmico come avviene, per esempio, nel menzionato caso del *software* di *deep learning* rispetto al quale nemmeno il programmatore può prevedere le variabili che la macchina prenderà in considerazione e, *a fortiori*, l'esito del processo in ogni dato momento.

Alla luce di quanto detto, la trattazione proseguirà con l'analisi di ognuno di questi tre livelli di opacità.

3. L'OPACITÀ INTELLETTIVA: LA TRAPPOLA DELLA TRASPARENZA.

La principale strategia legislativa storicamente adottata per bilanciare l'asimmetria informativa tra le parti contrattuali (di cui il *gap* di intelligibilità uomo-macchina rappresenta una peculiare e recente declinazione) fu l'introduzione di diritti di accesso riconosciuti ai soggetti interessati per rafforzare la trasparenza. Quest'ultima, infatti, è stata tradizionalmente concepita come ontologicamente legata alla realizzazione di un effettivo controllo su un dato fenomeno.

Come efficacemente affermato “[t]ransparency is seen as one of the bastions of democracy, liberal government, accountability and restraint on arbitrary or self-interested exercise of power”⁴⁰, originariamente sviluppata nella sfera pubblica come espressione della libertà di informazione⁴¹ e recentemente declinata in ciò che Bryce Goodman e Seth Flexman per primi hanno definito come il diritto ad una spiegazione⁴². Come precedente-

Making for Social Good, in T. CERQUITELLI, D. QUERCIA, F. PASQUALE (a cura di), *Transparent Data Mining for Big and Small Data*, *Studies in Big Data*, Vol. 11, Berlino, Springer, 2017, 12 ss.

⁴⁰ L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16, *Duke Law & Technology Review*, 18, 2017, 39.

⁴¹ L.D. BRANDEIS, *Other People's Money, and How Bankers Use It*, 62, *National*

Home Library Foundation, 1933 *Ivi* citato nella parte in cui afferma “[s]unlight is said to be the best of disinfectants; electric light the most efficient policeman”.

⁴² B. GOODMAN - S. FLAXMAN, *European Union regulations on algorithmic decision-making and a “right to explanation”*, 38, *AI Magazine*, 3, 2017, presentato per la prima volta nel 2016 al ICML Workshop relative al tema della *Human Interpretability in Machine Learning* (WHI 2016), New York, NY.

mente dimostrato, infatti, il sempre più intensivo sfruttamento automatizzato dei *Big Data*, supportato dalla recente implementazione di applicazioni di intelligenza artificiale altamente avanzate e finalizzate allo svolgimento di processi decisionali sempre più delicati, ha condotto all'emersione di quella che è stata definita come "la tirannia dei dati"⁴³. Questo, a sua volta, sta alimentando una sempre più diffusa richiesta di *accountability* dei processi decisionali algoritmici, che (come altre decisioni imprenditoriali tecniche) fronteggiano il ben noto ostacolo dell'asimmetria informativa tra esperti e controparti non professionali. Ciò che tradizionalmente si configurava come una relazione contrattuale B2C, tuttavia, si sta progressivamente trasformando in una relazione trilaterale (esperti di ICT-Professionisti-Consumatori) caratterizzata da due diversi livelli di asimmetria informativa. Si pensi, ad esempio, al settore bancario. Nella misura in cui i banchieri si affidano sempre più intensamente all'esattezza dei risultati di sistemi algoritmici di *credit scoring* per prendere decisioni relative alla concessione di finanziamenti, e considerata la già menzionata natura non strutturata di questo tipo di problemi computazionali frequentemente risolti attraverso sistemi ML "non supervisionati", il problema dell'intelligibilità del risultato finale è duplice: da un lato, i consumatori non hanno le competenze finanziarie necessarie per comprendere i criteri economici e regolamentari seguiti dalla banca per valutare la meritevolezza creditizia della clientela; dall'altro, nel caso di applicazioni *Fintech* di questo tipo, i banchieri stessi non sono in grado di conoscere appieno le variabili acquisite dalla macchina (tantomeno l'intera pletora di dati personali strutturati e non strutturati utilizzati per programmare e poi alimentare il *software*) al fine di prendere la specifica decisione di finanziamento.

Di conseguenza, il già compromesso rimedio della trasparenza viene ad essere ulteriormente indebolito dalla necessità di tradurre processi decisionali altamente tecnici, nonché complessi dal punto di vista algoritmico, in termini comprensibili per soggetti interessati privi delle competenze tanto finanziarie quanto informatiche necessarie.

⁴³ In particolare, "*we refer to as the "tyranny of data", that is the adoption of data-driven decision-making under the technocratic and top-down approaches highlighted by Easterly. We elaborate on the need for social good decision-making algorithms to provide transparency and accountability, to only use personal information — owned and controlled by individuals — with explicit consent, to ensure that privacy is preserved when data is analyzed in aggregated and anonymized form, and to be*

tested and evaluated in context, that is by means of living lab approaches involving citizens." B. LEPRI - J. STAIANO - D. SANGOKOYA - E. LETOUZÉ, - N. OLIVER, *The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good*, cit., 5. Cfr. W. EASTERLY, *The Tyranny of Experts. Economists, Dictators, and the Forgotten rights of the Poor*, Basic Books, New York, 2014, 286 e S. RODOTÀ, *Il mondo nella rete. Quali i diritti, quali i vincoli*, op. cit., 37-40.

In questo contesto, si accentua il rischio di cadere ancora più a fondo nel ben noto problema nella fallacia della trasparenza ⁴⁴. Il nostro tradizionale approccio consumeristico, infatti, indurrebbe ad integrare i già prolissi prospetti informativi con ulteriori informazioni tecniche concernenti il funzionamento algoritmico dei *software* di *credit scoring* utilizzati dalle imprese finanziarie. Questo, tuttavia, non farebbe altro che aggravare il fraintendimento circa la natura umana alla base del cosiddetto “*disclosure approach*” che, secondo un orientamento dottrinale, non considerando il carattere fisiologicamente avverso alle decisioni dell'uomo medio, ha finito per determinare il fallimento per il regime della *disclosure* obbligatoria per tre principali concause: (i) la generalizzata percezione delle richieste di consenso come fardelli burocratici esclusivamente finalizzati a proteggere i professionisti in caso di contenzioso; (ii) la diffusa frustrazione provocata dal linguaggio altamente tecnico utilizzato tanto in moduli contrattuali quanto in prospetti informativi; (iii) la persistente indifferenza verso la richiesta di chiarezza, efficacia e sinteticità ⁴⁵.

Il rischio così creato è quello di favorire l'emersione di un paradigma di trasparenza privo di significato degno della già nota trappola del consenso ⁴⁶, nella misura in cui gli sforzi normativi non riescono a raggiungere il cuore del problema dell'asimmetria: la persistente mancanza di comprensione. I soggetti interessati, pur essendo generalmente ben consapevoli della quantità di dettagliate informazioni tecniche a loro disposizione, non sono disposti ad affrontare il consistente investimento di tempo ed energie necessario a decifrarne il concreto significato, ciò principalmente a causa della sensazione di disconnessione dal potere di incidere attivamente sullo stesso, rendendo così l'intero processo informativo privo di scopo ⁴⁷. Ciò che realmente manca, quindi, non riguarda la quantità di informazione ma la sua qualità.

⁴⁴ M. ANNANY - K. CRAWFORD, *Seeing Without knowing: limitations of the transparency ideal and its application to algorithmic accountability*, 20, *New Media & Society*, 3, 973-989.

⁴⁵ Questo è il fenomeno generalmente noto come *information overload*, in virtù del quale, come correttamente affermato: “le persone vogliono sapere, non di più, e l'informazione porta ansia e non sicurezza.” (traduzione dell'autore) Inoltre, i soggetti cc.dd. *overloaded* tendono a cercare alternative, rimandare o evitare decisioni (attenendosi al c.d. *path of least resistance*). O.I. BEN-SHAHAR - C.E. SCHNEIDER, *More Than You Wanted to Know. The Failure of Mandated Disclosure*, op. cit., 70-76. Cfr. F.H. CATE, *Big Data consent and the future of data protection* in C.R. SUGIMOTO,

H.R. EBIKA, M.MATTIOLI (a cura di), *Big Data is not a Monolith*, MIT Press, Cambridge (MA), 2016, 3.

⁴⁶ Come sottolineato dagli Autori: “*there is some danger of research and legislative efforts being devoted to creating rights to form a transparency that may not be feasible and may not match user needs. As the history of industries like finance and credit show us, rights to transparency do not necessarily secure substantive justice or effective remedies*”. L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, op. cit. 22-23.

⁴⁷ “Le informative sono generalmente complesse. [...] Di conseguenza, gli individui sono sovrastati da lunghe e dettagliate normative sulla privacy. Nel marzo 2012, il

Molti studiosi hanno sostenuto che proprio questo è lo scopo sottostante la riformulazione dell'articolo 12 della previgente direttiva 1995/46/EC, in base al quale “gli Stati membri [devono] garantire ad ogni soggetto interessato il diritto di ottenere dal responsabile del trattamento: [...] la *conoscenza della logica* applicata nei trattamenti automatizzati dei dati che lo interessano, per lo meno nel caso delle decisioni automatizzate”. Diversamente, il menzionato e attualmente in forza articolo 15 GDPR afferma che “i soggetti interessati hanno il diritto di ottenere dal responsabile del trattamento *informazioni significative sulla logica utilizzata*”⁴⁸.

Questa specificazione ha indotto eminenti giuristi a ritenere che l'articolo 15 GDPR richieda qualcosa di più e di diverso dalla mera trasparenza, introducendo uno standard di intelligibilità che implica non soltanto informazioni dettagliate ed esaustive (i.e. trasparenti), ma anche e soprattutto comprensibili e significative (*rectius* utili) per lo specifico soggetto interessato⁴⁹. A dispetto delle indubbiamente nobili intenzioni alla base di questa interpretazione, come messo in luce da Stefano Pagliantini: “a fare della trasparenza un mero sinonimo di non opacità (delle clausole), si finisce per declassarla ad una nozione evanescente e, tutto sommato, frusta. [...] Tutto cambia, e sta qui il *novum* saliente della disciplina comunitaria, se la trasparenza invece viene pure rappresentata come una regola concernente un modo di essere del contratto diseguale”⁵⁰. Ai sensi del diritto europeo dei contratti, quindi, la trasparenza assurge a informazione qualificata, includendo: (i) la completezza, intesa come il tasso di esaustività che incide sulla soglia massima e minima dello standard quantitativo di informazioni da fornire al consumatore; (ii) la chiarezza, tradizionalmente finalizzata ad assicurare la leggibilità dei segni

comitato britannico di controllo sui consumatori Which? ha calcolato che la *policy* sulla privacy di PayPal, aggiunta agli altri termini e condizioni del servizio sottoposte ai consumatori ammontava ad un totale di 36,275 parole, più lungo di Amleto (30,066 parole), mentre quelle di iTunes arrivano a 19,972 parole, più lungo di Macbeth (18,110 parole). Uno studio ha calcolato che leggere le *policies* di privacy dei siti *web* più popolari richiederebbe ad un singolo individuo 244 ore — o più di 30 intere giornate lavorative — ogni anno.” (traduzione dell'autore) F.H. CATE, *Big Data consent and the future of data protection* in C.R. SUGIMOTO, H.R. EBIKA, M.MATTIOLI (a cura di), *Big Data is not a Monolith*, op. cit. 7.

⁴⁸ Corsivo aggiunto.

⁴⁹ Questa è la definizione data dal dizionario Cambridge della lingua inglese per

la parola “*meaningful*”. Gli autori più precisamente distinguono tra: (i) *readability*, Intesa come la capacità del destinatario di comprendere il contenuto di un determinato documento; (ii) *legibility*, ovvero il concetto per primo introdotto da Mortier e finalizzato a rendere “*data and analytics algorithms both transparent [i.e. complete and exhaustive] about the commercial use and comprehensible in its functionality*”; ed infine (iii) *explanation*, il quale sottintende un approccio proattivo, essendo tarata all'educazione e capacità di comprensione del singolo individuo. G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit., 245.

⁵⁰ S. PAGLIANTINI, *Trasparenza contrattuale* (voce), in *Enc. Dir.*, Annali VI, 2012, par. 12.

grafici utilizzati nel testo, impedendo così alle imprese di utilizzare formati illeggibili o idonei ad ingenerare confusione nel lettore; (iii) l'intellegibilità, concepita come il tasso di difficoltà per la comprensione del significato e delle implicazioni di un testo utilizzando uno standard di ordinaria diligenza, generalmente parametrato a quella del consumatore medio; (iv) l'equità e l'accessibilità, rispettivamente riferite alla attitudine di buona fede mostrata nella redazione del testo e alla possibilità di leggere i termini e le condizioni in ogni dato momento e per più di una volta ⁵¹.

In altre parole, la trasparenza non solo opera come regola di condotta che conforma il *drafting* dei termini contrattuali (così come dei prospetti informativi), ma anche come scudo normativo a salvaguardia della legittima aspettativa del consumatore a che il contratto sia eseguito in modo sì conforme al suo contenuto, ma così come espresso e ragionevolmente percepito attraverso l'adempimento da parte del professionista dei suoi doveri informativi. Il principale obiettivo sottostante la nozione di trasparenza, quindi, è quello di costringere i professionisti a rispettare uno standard informativo strumentale ad un'effettiva protezione di un'autodeterminazione pienamente consapevole del consumatore, in tal modo indirettamente salvaguardando la competitività del mercato unico europeo ⁵².

Come autorevolmente sottolineato da De Franceschi, la stessa Corte di Giustizia dell'Unione Europea, nel caso Kásler ha ribadito che la *ratio* sottostante l'articolo 5 della Direttiva concernente le clausole abusive nei contratti stipulati con i consumatori (per menzionare una delle molte disposizioni normative europee che impongono simili doveri) in base al quale i termini contrattuali "devono sempre essere redatti con un linguaggio chiaro e intelligibile", è quella di controbilanciare la fisiologica debolezza dei consumatori al fine di renderli in grado di "prevedere quando uno specifico evento potrà determinare l'applicazione di una clausola contrattuale, nonché gli effetti di una tale applicazione." Il requisito della trasparenza ivi fissato, quindi, non può essere soddisfatto da clausole intellegibili da un punto di vista meramente formale e grammaticale, ma deve consentire ai consumatori di comprendere sul piano sostanziale le condizioni e le conseguenze dell'applicazione delle stesse ⁵³.

Non sorprende, quindi, che il linguaggio utilizzato richiami

⁵¹ Ivi, para 6. Cfr. M. DELL'UTRI, *Principi generali e condizioni di liceità del trattamento dei dati personali*, in V. CUFFARO, R. D'ORAZIO, V. RICCIUTO (a cura di), *I dati personali nel diritto europeo*, Giappichelli Editore, Torino, 2019, 206.

⁵² S. PAGLIANTINI, voce *Trasparenza contrattuale*, op. cit., par. 11.

⁵³ Árpád Kásler e Hajnalka Káslerné Rábai c. OTP Jelzálogbank Zrt., C-26/13 (30 aprile 2014) ECLI:EU:C:2014:282. Cfr. A. DE FRANCESCHI, *European Contract Law*

proprio la specificazione introdotta dall'articolo 15 (1)(h) GDPR nella parte in cui aggiunge che i soggetti interessati devono anche poter ottenere informazioni riguardanti "l'importanza e le conseguenze previste di tale trattamento per l'interessato." Alla luce di quanto sin qui osservato, appare ancora più chiaro come, introducendo il requisito della significatività delle informazioni offerte ai soggetti interessati, il legislatore europeo non abbia fatto altro che tradurre in termini legislativi più espliciti la menzionata interpretazione del principio della trasparenza, trasponendola nella peculiare relazione B2C che intercorre tra soggetti interessati e responsabile del trattamento dei loro dati personali. L'obiettivo perseguito con l'introduzione del requisito della significatività, quindi, è quello di indurre i responsabili del trattamento a focalizzarsi di più su una valutazione individualizzata della qualità delle informazioni offerte ai soggetti interessati, al fine ultimo di promuoverne una più efficace comprensione del modo in cui dati personali sono acquisiti e processati per prendere decisioni che incidono significativamente sulla loro sfera giuridica (c.d. *smart disclosure*)⁵⁴.

In altre parole, ciò che la disposizione normativa cerca di salvaguardare è la cosiddetta libertà negativa dei soggetti interessati, intesa come la possibilità di ritirarsi in uno spazio privato e imperscrutabile di auto-riflessione dove ciascuno possa valutare le proprie opzioni e decidere il proprio comportamento futuro⁵⁵. Tuttavia, il c.d. "capitalismo di sorveglianza", risultato della crescente economia dei *Big Data*, sta inesorabilmente conducendo al superamento dell'inviolabilità di questa ultima frontiera di intimità mentale lasciando i soggetti interessati nella mera illusione di autodeterminazione⁵⁶.

Si prenda ad esempio il sistema FICO. Questo sistema di *credit*

and the Digital Single Market. The Implications of the Digital Revolution, Intersentia, Cambridge, 2016, 201.

⁵⁴ "Smart disclosure involves providing consumers access to data in user friendly electronic formats, in order to fuel the creation of products and tools that benefit consumers, including tools that will help them make important marketplace decisions." *Smart Disclosure and Consumer Decision Making: Report of the Task Force on Smart Disclosure* (Executive Office of the President National Science and Technology Council, maggio 2013), 7. Cfr. O.I. BEN-SHAHAR - C.E. SCHNEIDER, *More Than You Wanted to Know. The Failure of Mandated Disclosure*, op. cit., 133.

⁵⁵ M. HILDEBRANDT, *The Dawn of a Critical Transparency Right for the Profiling Era*, in J. BUS, M. CROMPTON, M. HIL-

DEBRANDT, G. METAKIDES (a cura di), *Digital Enlightenment Yearbook 2012*, IOS Press, Amsterdam, 47.

⁵⁶ Il capitalismo di sorveglianza viene generalmente definito come "l'abilità di raccogliere informazioni riguardanti i consumatori, di classificare tali soggetti, di predire il loro comportamento e di personalizzare le offerte e i prodotti in base alle diverse categorie di soggetti" (traduzione dell'autore), R. TAYLOR, *No Privacy Without Transparency*, in R. LEENES, R. VAN BRAKEL, S. GUTWIRTH, P. DE HERT (a cura di) *Data Protection and Privacy: The Age of Intelligent Machines*, Hart Publishing, Oxford, 2017, 77. Utilizzando le parole di Shoshana Zuboff "Surveillance capitalism is the puppet master that imposes its will through the medium of the ubiquitous digital apparatus. I now name the apparatus Big Other: it

scoring, utilizzato dal “90% dei *top lenders*” considera le ricerche *online* relative ai tassi di interesse o alle offerte di finanziamento come aventi un’incidenza negativa sul punteggio finale di chi le ha effettuate, tuttavia, il modo nel quale è calcolato il peso di questa variabile cambia a seconda della specifica versione della formula di *scoring* scelta dal finanziatore⁵⁷. I soggetti interessati sono perciò generalmente inconsapevoli che: (i) i percorsi mentali generalmente seguiti per arrivare ad una decisione e originariamente impenetrabili sono oggi per lo più presi a mezzo di ricerche *web* non solo generalmente registrate, ma sempre di più tracciate, i.e. collegate tra loro al fine di inferire il profilo del soggetto a cui sono riferibili; (ii) le opzioni (dalle proposte di mutuo fino quale pubblicità) che gli vengono presentate cambiano a seconda del profilo; (iii) ogni variabile che confluisce in un profilo incide sullo stesso in modo differente a seconda della specifica formula algoritmica utilizzata dal responsabile del trattamento.

Ne consegue che le prassi indiscriminate di *brain-hacking* attualmente praticate dalle imprese di *data analytics* non possono che sfociare in forme sempre più aggressive di manipolazione cognitiva, in tal modo rinviando la richiesta di ciò che viene recentemente definito come il diritto alla privacy mentale o perfino alla libertà cognitiva⁵⁸. Fin quando le imprese di *data analytics* di tipo predittivo continueranno ad evitare la *disclosure* dello specifico peso attribuito ad ogni variabile ritenuta rilevante ai fini della specifica decisione automatizzata finale prima che tale decisione sia presa, i soggetti interessati saranno inevitabilmente indotti a cadere nella così detta trappola dell’autonomia “perché [se] una persona non è consapevole dei profili che le sono attribuiti potrebbe essere indotta ad agire in modo diverso da ciò che avrebbe scelto qualora li avesse conosciuti. [...] La mia autonomia è elusa e la mia volontà indebolita ogni volta in cui non sono consapevole delle informazioni che vengono utilizzate”⁵⁹.

Ciononostante, due principali ostacoli normativi continuano a

is the sensate, computational, connected puppet that renders, monitors, computes, and modifies human behavior. Big Other combines these functions of knowing and doing to achieve a pervasive and unprecedented means of behavioral modification. Surveillance capitalism's economic logic is directed through Big Other's vast capabilities to produce instrumental power, replacing the engineering of souls with the engineering of behavior.” Id., *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, New York, 2019, 353.

⁵⁷ MyFico, *Credit Education*, disponibile all’indirizzo [www.myfico.com/credit-](http://www.myfico.com/credit-education/credit-reports/credit-checks-and-inquiries)

[education/credit-reports/credit-checks-and-inquiries](http://www.myfico.com/credit-education/credit-reports/credit-checks-and-inquiries).

⁵⁸ G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, *op. cit.*, 253.

⁵⁹ (traduzione propria) M. HILDEBRANDT, *The Dawn of a Critical Transparency Right for the Profiling Era*, in J. BUS, M. CROMPTON, M. HILDEBRANDT, G. METAKIDES (a cura di), *Digital Enlightenment Yearbook 2012*, *op. cit.*, 49; Cfr. F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, *op. cit.*, 5; A. OTTOLIA, *Big Data e Innovazione Computazionale*, *op. cit.*, 2017,

compromettere l'efficacia dei diritti di trasparenza fissati nel terzo capitolo (sezione prima) del GDPR.

Innanzitutto, la dottrina maggioritaria sostiene che, nonostante il linguaggio utilizzato negli articoli 13, 14 e 15 del GDPR sia sostanzialmente identico, soltanto l'ultimo possa essere interpretato come fonte del diritto dei soggetti interessati ad una spiegazione caso per caso della decisione individuale, in ragione della sua peculiare formulazione parzialmente orientata al passato. Al contrario, gli articoli 13 e 14 GDPR, prevedendo esclusivamente un dovere di notificazione *ex ante* in capo ai responsabili del trattamento, non potrebbero dettare che un mero diritto ad essere informati sul generale ed astratto funzionamento del processo decisionale automatizzato. In virtù di questa interpretazione, quindi, i soggetti interessati non potrebbero vedersi riconosciuto un diritto ad una spiegazione individualizzata *ex ante* del significato e delle conseguenze previste per il trattamento dei loro dati. Ciò implica anche che, per esempio, in tutti i casi menzionati di *credit scoring*, nel caso in cui il responsabile del trattamento (i.e. agenzie di *scoring* o imprese finanziarie) ottiene i dati personali dei suoi clienti, questi ultimi possono essere informati dei loro punteggi personali solo dopo che la specifica valutazione del loro merito creditizio (basata su quel punteggio) è stata effettuata, e solo in virtù dell'articolo 15 GDPR (i.e. diritto di accesso).

Questa prospettiva, tuttavia, trascura la circostanza per cui i clienti possono avere una reale consapevolezza delle conseguenze del loro scelte (e quindi reale autonomia) solo nella misura in cui sono informati dello specifico peso che ogni azione può avere sul loro punteggio di credito *prima* che questo venga trasmesso e utilizzato dai finanziatori. Al contrario, nel descrivere il funzionamento generale del sistema di *scoring* alcune imprese utilizzano espressioni quali "l'impatto delle ricerche di offerte di finanziamento potrà *variare da persona a persona* in ragione della loro *unica* storia di credito. *In generale*, ricerche di finanziamento [...] hanno un ridotto impatto. Per la *maggior parte* delle persone, una ricerca aggiuntiva ridurrà il punteggio di meno di 5 punti"⁶⁰. Sebbene questa formulazione sia indubbiamente sufficiente a rispettare i doveri di notificazione come sopra interpretati, è altrettanto innegabile che questo tipo di descrizione sia insignificante (*rectius* non significativa) per gli interessati che stanno cercando di esercitare in modo consapevole la propria

202; M. MATTIOLI, Disclosing Big Data, in *Minnesota Law Review*, 99, 2014, 2, 535 ss.; T.Z. ZARSKY, "Mine your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion, 5,

Yale Journal of Law and Technology, 1, 1-56.

⁶⁰ MyFico, Credit Education, disponibile all'indirizzo www.myfico.com/credit-education/credit-reports/credit-checks-and-inquiries.

autodeterminazione, essendogli preclusa la possibilità di prevedere le conseguenze delle proprie azioni. È qui che di nuovo si insinua la menzionata minaccia della fallacia della trasparenza ⁶¹.

Per di più, questo non sembra neppure essere l'intento del legislatore europeo, dal momento che il considerando 60 del GDPR afferma che "il titolare del trattamento dovrebbe fornire all'interessato eventuali ulteriori informazioni necessarie ad assicurare un trattamento corretto e trasparente, *prendendo in considerazione le circostanze e del contesto specifici* in cui i dati personali sono trattati" ⁶². Per questo motivo si sostiene che la cosiddetta spiegazione modello-centrica non costituirebbe un'informazione significativa ai sensi degli articoli 13 e 14 GDPR ⁶³.

Questa conclusione preliminare, tuttavia, conduce alla seconda delle due barriere legislative precedentemente menzionate. Anche ammettendo il diritto di una spiegazione individualizzata (indipendentemente dalla sua natura *ex ante* o *ex post*), infatti, la conoscenza estratta dal responsabile del trattamento dai cc.dd. *raw Big Data*, così come il *software* utilizzato, rientrano nell'ambito di applicazione della disciplina in materia di proprietà intellettuale e *trade secret*. Sono proprio questi, infatti, gli elementi che costituiscono il secondo livello di opacità delle decisioni automatizzate: l'opacità intenzionale, all'analisi della quale sarà dedicato il successivo paragrafo.

4. L'OPACITÀ INTENZIONALE: UNO SCONTRO CON LA PROPRIETÀ INTELLETTUALE?

Una volta riconosciuta, da un lato, la necessità di una cosiddetta spiegazione soggetto-centrica, e il valore economico della conoscenza di carattere predittivo e di profilazione estrapolata dai dati

⁶¹ L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, op. cit., 23.

⁶² Che il Considerando 60 GDPR si riferisca esclusivamente agli articoli 13 e 14 GDPR è chiaramente dimostrato dalle somiglianze nel linguaggio utilizzato dal considerando e dagli articoli, i quali menzionano "informazioni necessarie ad assicurare un trattamento equo e trasparente", nonché dal fatto che il considerando colleghi espressamente questi principi con il diritto ad essere informato sull'esistenza del trattamento automatizzato, così sottintendendo un dovere di notificazione *ex ante* ai sensi degli articoli 13 e 14 GDPR. Ciò è peraltro confermato dalla versione italiana del GDPR

come redatta dal Garante della Privacy, che espressamente collega i considerando 60-62 ai soli articoli 13 e 14 GDPR.

⁶³ Una spiegazione modello-centrica è definita come quella spiegazione che, rimanendo indifferente agli *input* e agli *output* specifici del caso concreto, si limita a descrivere l'obiettivo generale sottostante l'intero procedimento algoritmico, la generica classificazione dei dati utilizzati per programmare l'algoritmo, l'attitudine predittiva del modello, nonché la logica globale dello stesso, intesa come una spiegazione semplificata e umanamente comprensibile del come *input* casuali producano *output* casuali. L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, op. cit., 22.

personali⁶⁴, dall'altro, il problema si trasferisce dall'*an* al *quantum* della *disclosure* necessaria ai fini della spiegazione "significativa"⁶⁵. Più chiaramente, il considerando 63 GDPR specifica che il diritto di accesso ai dati personali (e *a fortiori* il diritto a conoscere la logica utilizzata in qualsiasi trattamento automatizzato di dati personali) "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"⁶⁶.

Per questa ragione le seguenti considerazioni saranno volte ad individuare il metodo più appropriato per bilanciare questi diritti confliggenti. Come messo in luce da Andrea Ottolia⁶⁷, infatti, è possibile distinguere tra due diversi approcci: quando gli interpreti bilanciano i diritti *per* il sistema, muovono dal presupposto della natura fondamentale di uno dei due, ne consegue che il risultato sarà inevitabilmente condizionato da questa prevalenza assiologica. Al contrario, se prodotto di un bilanciamento *nel* sistema, l'esito interpretativo sarà fondato su una ponderazione paritaria di diritti confliggenti ma di uguale peso⁶⁸.

A dispetto della prospettiva di Ottolia, in base alla quale il conflitto tra il diritto ad una spiegazione e la protezione del *trade secret* si qualificherebbe come un bilanciamento nel sistema⁶⁹, il linguaggio utilizzato dalla direttiva 2016/ 943/EU suggerisce diversamente. Infatti, come già messo in luce da Malgieri e Comandè⁷⁰, mentre il considerando 63 GDPR afferma che la protezione dei dati "*should not adversely affect*" il segreto industriale

⁶⁴ Come efficacemente affermato: "è importante comprendere le decisioni prese dai motori di ricerca attraverso i loro algoritmi, in quanto l'approccio algoritmico alla ricerca di informazioni può essere considerato come "una particolare forma di conoscenza logica, costruita attorno ad una specifica concezione di cosa sia la conoscenza e di come se ne debbano individuare le componenti più rilevanti. Che stiamo passando agli algoritmi per individuare ciò che abbiamo bisogno di conoscere è tanto importante quanto essersi affidati a esperti accreditati, al metodo scientifico, al buon senso o alla parola di Dio". D. LEWANDOWSKI, *Is Google Responsible for Providing Fair and Unbiased Results?*, in M. TADDEO, L. FLORIDI (a cura di), *The Responsibilities of Online Service Providers, Law, Governance and Technology Series*, Vol. 31, Springer, New York, 2017, 62 dove viene citato T. GILLESPIE, *The relevance of algorithms*, in T. GILLESPIE, P.J. BOCZKOWSKI, & K.A. FOOT (a cura di), *Media Technologies: Essays on Communication, Materiality, and Society*, MIT Press, Cambridge (MA), 168. G. MAL-

GIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit., 262; M. HILDEBRANDT, *Defining Profiling: A New Type of Knowledge?*, in M. HILDEBRANDT, S. GUTWIRTH (a cura di), *Profiling the European Citizen. Cross-Disciplinary Perspectives*, Springer, Berlin, 2008, 17-45.

⁶⁵ A. OTTOLIA, *Big Data e Innovazione Computazionale*, op. cit., 103.

⁶⁶ Considerando 63 GDPR.

⁶⁷ A. OTTOLIA, *Big Data e Innovazione Computazionale*, op. cit., 180.

⁶⁸ V. ZENO-ZENCOVICH, *Diritto d'autore e libertà di espressione: una relazione ambigua*, in AIDA, 2005, 151 e ss.; A. OTTOLIA, *Proprietà intellettuale e trattamento dei dati personali: riflessioni su privacy "per il sistema" e "nel sistema"*, in AIDA, Giappichelli Editore, Torino, 2010, 319 ss.

⁶⁹ A. OTTOLIA, *Big Data e Innovazione Computazionale*, op. cit., 182 e 202.

⁷⁰ G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit., 247.

e aziendale così come i diritti d'autore che tutelano il *software*, il considerando 34 della Direttiva sui segreti commerciali più radicalmente afferma che i diritti ivi riconosciuti “*should not affect*” i diritti e gli obblighi fissati dalla direttiva 1995/46/EC, in particolare il diritto dei soggetti interessati di avere accesso ai propri dati personali ⁷¹.

Ciò che rimane da analizzare, quindi, è la misura in cui il perseguimento di una effettiva implementazione dei diritti alla protezione dei dati personali fissati nel GDPR possa derogare le prerogative legislative fatte proprie dalla disciplina sul *trade secret*. Da un lato, sappiamo che il primo non dovrebbe incidere negativamente sul secondo ⁷², in tal modo richiamando il bisogno di evitare che esercitando i propri diritti, i soggetti interessati causino un danno al detentore del segreto commerciale. Dall'altro, il considerando 63 GDPR specifica che “tali considerazioni non dovrebbero condurre a un diniego a fornire all'interessato tutte le informazioni.”

Il successivo passo interpretativo non potrà quindi che consistere nell'individuazione del modello di spiegazione soggetto-centrico più adeguato a tale scopo. Riassumendo le conclusioni fin qui raggiunte, affinché una spiegazione rispetti il requisito del carattere significativo delle informazioni fornite, questa dovrebbe: (i) consentire ai soggetti interessati di comprendere la logica seguita dal *software* che ha assunto la decisione automatizzata; (ii) consentire agli stessi la possibilità di contestare la decisione (i.e. verificare la legittimità e la veridicità dei dati utilizzati e vagliarne il carattere discriminatorio); (iii) attribuire ai destinatari della decisione la possibilità di prevedere gli effetti che le loro azioni produrranno sul risultato automatizzato, consentendogli così di esercitare in modo consapevole la propria autodeterminazione ⁷³.

Per queste ragioni, una cosiddetta spiegazione controfattuale non condizionata (anche definita spiegazione “*sensitivity-based*”) ⁷⁴ non è sufficiente. In ossequio a questo approccio, basato sulla nozione del più vicino dei mondi possibili, i soggetti

⁷¹ 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. Da notare, tuttavia, che la menzionata differenza linguistica non appare riprodotta nella traduzione italiana della normativa.

⁷² La versione italiana del Regolamento utilizza la parola “ledere”, che analogamente alla versione inglese richiama l'idea di causare un danno o di agire in modo non favorevole. OXFORD ENGLISH DICTIONARY

ONLINE, Oxford University Press, Oxford, 2019.

⁷³ Questo si discosta parzialmente da quanto sostenuto da S. WACHTER - B. MITTELSTADT - C. RUSSELL, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, 31, *Harvard Journal of Law & Technology*, 2, 2018, 25.

⁷⁴ L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, *op. cit.*, 58.

interessati dovrebbero poter ottenere informazioni relative al più piccolo cambiamento che avrebbe dovuto essere apportato alla situazione reale per ottenere l'esito desiderato. Nelle parole degli autori, questo assumerebbe per esempio la seguente formulazione: "ti è stato negato un prestito perché il tuo reddito annuale era di 30.000£. Se avessi avuto un reddito di 45.000£ ti sarebbe stato offerto il finanziamento" ⁷⁵.

In altri termini, focalizzandosi sulla relazione tra i fattori esterni e la specifica decisione automatizzata, i controfattuali indicano il più piccolo cambiamento che il soggetto interessato avrebbe dovuto apportare alle variabili iniziali per ottenere il risultato che si aspettava o che, più semplicemente, avrebbe preferito, senza tuttavia dover aprire la scatola nera e rischiare di violare diritti d'autore sul *software*. In questo modo, si ritiene che il titolare del trattamento sarebbe in grado di rispettare il principio di trasparenza di cui al GDPR, senza dover affrontare il notorio e complesso problema dell'analisi (e della spiegazione) dei processi algoritmici interni che hanno prodotto la decisione automatizzata. Questo è, peraltro, esattamente il modo in cui funzionano i tre più comuni metodi di spiegazione dei sistemi ML esistenti: LOCO ("*Leave One Column Out*"), PI (*Permutation impact*) e LIME ("*Local Interpretable Model-Agnostic Explanations*") ⁷⁶.

Tutte queste tecniche condividono la caratteristica di consentire un'analisi meramente esterna del processo decisionale automatizzato. In aggiunta, dal momento che solo LOCO utilizza dati reali per testare il modello e individuare le correlazioni *input-output*, siffatto limite tecnologico rende possibili tre diversi tipi di spiegazione soggetto-centrica, tutti però indifferenti alle specifiche circostanze del caso poiché alimentati dai dati ipotetici creati appositamente per spiegare il concreto esito algoritmico prodotto. Il primo è stato definito come un modello *demographic-based* e appare caratterizzato da una forte dipendenza dal concetto di privacy di gruppo o privacy di categoria ⁷⁷. Attraverso questo metodo, infatti, i soggetti interessati dovrebbero ottenere una

⁷⁵ S. WACHTER - B. MITTELSTADT - C. RUSSELL, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, *op. cit.*, 5-6.

⁷⁶ Il metodo LOCO consiste nell'eliminare una variabile alla volta, ricalcolando poi il risultato. Maggiore sarà la variazione di questo nuovo esito rispetto a quello iniziale, maggiore sarà il peso esercitato dalla variabile eliminata sulla decisione automatizzata. Analogamente, il metodo PI implica la sostituzione di una variabile con una diversa e casuale, ricalcolando poi il risultato. Infine, il metodo LIME utilizza un'ap-

prossimazione lineare del modello per tradurre in termini più semplici il funzionamento reale del modello, avvalendosi peraltro di variabili sintetiche. In altri termini si tratta di una simulazione semplificata con dati fittizi del modello reale. J. BUDZIK, *Most AI Explainability is Snake Oil. Ours isn't* (ZestFinance, 2019) disponibile all'indirizzo <https://www.zestfinance.com/snake-oil-explainability>.

⁷⁷ M. HILDEBRANDT, *The Dawn of a Critical Transparency Right for the Profiling Era*, in J. BUS, M. CROMPTON, M. HILDEBRANDT, G. METAKIDES (a cura di), *Digital*

spiegazione per associazione, ovvero tale da indicare i fattori rilevanti che hanno portato alla specifica decisione nei loro confronti attraverso la divulgazione delle caratteristiche che essi stessi condividono con altre persone che hanno ricevuto un esito analogo. Il secondo, definito *performance-based*, garantisce uno scorcio sulla logica algoritmica seguita dalla macchina ancora più ristretto, poiché offre ai soggetti interessati una mera forma di tasso di errore dell'algoritmo personalizzato, indicando la percentuale di persone appartenenti alla medesima categoria che hanno ricevuto un risultato errato. Il terzo modello soggetto-centrico, definito *case-based*, avvalendosi di un approccio casistico offrirebbe ai soggetti interessati una spiegazione consistente nella descrizione dei dati utilizzati per programmare il sistema ML che più si avvicinano alla situazione specifica del destinatario ⁷⁸.

È piuttosto evidente, quindi, come tutti questi metodi di spiegazione, chiaramente ideati per essere il meno gravosi possibile per i titolari del trattamento ⁷⁹, non possono raggiungere la soglia della significatività così come descritta precedentemente. Nonostante presenti l'innegabile pregio di consentire ai destinatari di orientare in modo più consapevole il proprio comportamento futuro, una spiegazione meramente esterna (*rectius* concernente la logica esterna dell'algoritmo) lascerebbe inevitabilmente irrisolte questioni imprescindibili per un effettivo esercizio del diritto dei soggetti interessati di verificare la legittimità del trattamento ⁸⁰. Ad esempio, essendo accomunati dal limite di prendere in considerazione una sola variabile per volta ⁸¹, questi metodi non riescono a fornire ai soggetti interessati una visione organica di tutti i fattori che hanno inciso sulla specifica decisione. I controfattuali, inoltre, specialmente nel caso in cui si utilizzino modelli fittizi o dati astratti utilizzati per la programmazione del

Enlightenment Yearbook 2012, op. cit., 49; R. TAYLOR, *No Privacy Without Transparency*, in R. LEENES, R. VAN BRAKEL, S. GUTWIRTH, P. DE HERT (a cura di), *Data Protection and Privacy: The Age of Intelligent Machines*, op. cit., 96. Cfr. A. MANTELLERO, *Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*, 32, *Computer Law & Security*, 2016, 238; L. KAMMOURIEH - T. BAAR -, JOS BERENS - E. LETOUZÉ - J. MANSKE - J. PALMER - D. SANGOKOYA - P. VINCK, *Group Privacy in the Age of Big Data*, in L. TAYLOR, L. FLORIDI, B. SLOOT (a cura di), *Group Privacy. New Challenges of Data Technologies*, Philosophical Studies Series, Vol. 126, Springer, Berlino, 43.

⁷⁸ L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explana-*

tion' Is Probably Not the Remedy You Are Looking For, op. cit., 55.

⁷⁹ S. WACHTER - B. MITTELSTADT - C. RUSSELL, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, op. cit., 35.

⁸⁰ Come espressamente richiesto dal considerando 63 GDPR. Cfr. T. CALDERS - I. ŽILIOAITĖL, *Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures*, in B. CUSTERS, T. CALDERS, B. SCHERMER, e T. ZARSKY (a cura di), *Discrimination and Privacy in the Information Society. Data Mining and Profiling in Large Databases*, op. cit., 43 ss.

⁸¹ A tal proposito si veda anche M. RHOEN - Q. YI FENG, *Why the 'Computer says no': illustrating big data's discrimination risk through complex systems science*, op. cit., 140-159.

*software*⁸², impedirebbero ai soggetti interessati di esercitare qualsivoglia forma di controllo sul peso attribuito ad ogni variabile, sulla correttezza dei dati utilizzati, ovvero su di un eventuale effetto discriminatorio della decisione⁸³.

Contrariamente a quanto sostenuto da Watcher, infatti, i titolari del trattamento hanno un obbligo di “aiutare i soggetti interessati ad esercitare i loro diritti”⁸⁴, altrimenti la previsione dell’articolo 12 comma 2 GDPR (in base al quale i titolari del trattamento devono “agevola[re] l’esercizio dei diritti dell’interessato ai sensi degli articoli da 15 a 22”) verrebbe declassato a lettera morta legittimante un mero adempimento formale, rafforzando così i già menzionati timori di ineffettività della trasparenza. Nonostante l’indiscutibile complessità, quindi, i titolari del trattamento non dovrebbero poter evitare una qualche forma di apertura sulla logica algoritmica interna seguita dal sistema di ML per arrivare alla specifica decisione.

Questa consapevolezza, tuttavia, non rende più agevole la ricerca di una soddisfacente versione di spiegazione. In base a quanto fin qui dimostrato, infatti, affinché una spiegazione soggetto-centrica si possa configurare come significativa questa dovrebbe: (i) prendere in considerazione le specifiche circostanze del caso; (ii) focalizzarsi sullo scoprire il funzionamento interno del *software* utilizzato per la decisione; (iii) consentire ai soggetti interessati di avere un’idea chiara di tutte le variabili concretamente prese in considerazione e il peso che ciascuna ha esercitato sullo specifico esito. Allo stesso tempo, tuttavia, una tale spiegazione non dovrebbe essere così dettagliata da diventare troppo tecnica per essere intellegibile da un soggetto interessato medio, né troppo intrusiva per non incidere negativamente sui diritti di autore che proteggono il *software* utilizzato.

Alla luce di quanto detto, potrebbero rimanere dubbi circa la misura in cui il diritto dei soggetti interessati ad avere accesso ai propri profili (i cc.dd. metadati creati algoritmicamente dai tito-

⁸² Ciò si pone in contrasto al dettato inequivocabile del considerando 60, in base al quale “il titolare del trattamento dovrebbe fornire all’interessato eventuali ulteriori informazioni necessarie ad assicurare un trattamento corretto e trasparente, prendendo in considerazione le circostanze e del contesto specifici in cui i dati personali sono trattati”.

⁸³ Al contrario, Watcher sostiene che “comprendere la logica interna del sistema che ha portato a considerare il reddito come una variabile rilevante ai fini della decisione, che richiederebbe una spiegazione tecnica diversa da spiegazione con-

trofattuale [...] seppure di per sé desiderabile, non è assolutamente necessaria per contestare la decisione basata su quella variabile.” (traduzione dell’autore) S. WACHER - B. MITTELSTADT - C. RUSSELL, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, op. cit., 41. Tuttavia, come ammesso alla stessa Watcher il concetto di spiegazione tipicamente si riferisce alla logica interna, che è ciononostante trascurata dai controfattuali.

⁸⁴ S. WACHER - B. MITTELSTADT - C. RUSSELL, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, op. cit., 39-40.

lari del trattamento) sia giuridicamente percorribile ed economicamente sostenibile⁸⁵, ma non ci dovrebbero essere discussioni sul fatto che divulgare il codice sorgente del *software* che ha preso la decisione non presenta le menzionate caratteristiche, necessarie per bilanciare i diritti fissati nel GDPR e i diritti d'autore in modo adeguato.

A prescindere dall'ovvia violazione del diritto di autore che tale soluzione imporrebbe, è abbastanza evidente che avere accesso al codice sorgente utilizzato dal titolare del trattamento per arrivare alla specifica decisione difficilmente aiuta i soggetti interessati ad esercitare il proprio diritto di contestare la decisione. Per quanto dettagliato possa essere il codice sorgente divulgato, infatti, questo non potrebbe che indurre a trascurare ancora una volta il principio di trasparenza, in base al quale "le informazioni e le comunicazioni relative al trattamento di tali dati personali [devono essere] facilmente accessibili e comprensibili e [...] [fornite in] un linguaggio semplice e chiaro"⁸⁶. Questo sacrificio della posizione del titolare del diritto d'autore sul *software*, peraltro, non appare neppure necessario dal momento in cui l'articolo 1 comma 2 della direttiva 2009/ 24/EC sulla protezione dei programmi per elaboratore stabilisce che "le idee e i principi alla base di qualsiasi elemento di un programma per elaboratore, compresi quelli alla base delle sue interfacce, non sono tutelati dal diritto d'autore a norma della presente direttiva"⁸⁷. In questo senso, un'analisi umanamente comprensibile e specifica al caso concreto del funzionamento interno del *software*, da un lato, costituirebbe un'informazione significativa ai sensi del GDPR, dall'altro supererebbe il rischio di entrare in collisione con il diritto d'autore che legislatore europeo riconosce sul *software*⁸⁸.

Questa non è, tuttavia, la conclusione raggiunta (seppure in un contesto legislativo leggermente diverso) dal Tribunale Amministrativo Regionale del Lazio in una sentenza molto particolare concernente il diniego opposto dal Ministero Italiano dell'Istruzione alla richiesta di accesso al codice sorgente del *software* utilizzato per prendere decisioni interamente automatizzate riguardanti il collocamento sul territorio del personale docente, avanzata da una associazione sindacale rappresentativa di inse-

⁸⁵ R. TAYLOR, *No Privacy Without Transparency*, in R. LEENES, R. VAN BRAKEL, S. GUTWIRTH, P. DE HERT (a cura di), *Data Protection and Privacy: The Age of Intelligent Machines*, op. cit., 108; N. DIAKOPOULOS, *Accountability in algorithmic decision making*, 59, *Communications of the ACM*, 2, 60.

⁸⁶ Considerando 39 GDPR.

⁸⁷ Direttiva 2009/24/CE del 23 aprile 2009 del Parlamento Europeo e del Consiglio relativa alla tutela giuridica dei programmi per elaboratore.

⁸⁸ B. DANIEL MITTELSTADT - P. ALLO - M. TADDEO - S. WACHTER - L. FLORIDI, *The Ethics of Algorithms: Mapping the Debate*, 2, *Big Data & Society*, 2016, 1-21.

gnanti⁸⁹. In particolare, il MIUR ha ritenuto sufficiente l'accesso consentito ad un documento denominato "descrizione dell'algoritmo" nel quale veniva riassunto il quadro regolamentare di riferimento, procedendo poi a fornire una breve descrizione dell'algoritmo. Ciononostante, l'associazione sindacale ha agito in giudizio sostenendo che tale documento non fosse idoneo a sostituire la divulgazione del codice sorgente così come dell'algoritmo utilizzato, nella misura in cui tale informazione ometteva di fornire l'indicazione del peso e del ruolo esercitato ai fini della decisione da *tutte* le variabili prese in considerazione dalla normativa vigente. Il Ministero, dal canto suo, ha resistito in giudizio obiettando come l'algoritmo non possa essere considerato un documento amministrativo ai sensi dell'articolo 22 lett. d L. n. 241/1990, poiché tale norma si riferisce esclusivamente alla forma di un atto che, a differenza del mero algoritmo, può essere qualificato come amministrativo. Il diritto individuale di accesso ivi fissato, quindi, non può essere applicato estensivamente al codice sorgente o all'algoritmo. Peraltro, la protezione riconosciuta al diritto d'autore sul *software*, continua il MIUR, impedisce anche l'esercizio del differente diritto di accesso civico generalizzato, di cui all'articolo 6 D.lgs. n. 97/2016.

La Corte, tuttavia, ha ritenuto che nella misura in cui l'algoritmo vada sostanzialmente a sostituire l'intero procedimento amministrativo di trasferimento, possa essere qualificato come un c.d. atto amministrativo a elaborazione elettronica. Quest'ultimo si distingue dall'atto amministrativo informatico in senso stretto di cui all'articolo 22 lett. d L. n. 241/1990, in quanto il primo (rientrando nell'interpretazione estensiva dell'ambito di applicazione della norma fatta propria dalla Corte) non solo potrebbe presentare un formato elettronico, ma sarebbe caratterizzato dall'essere il prodotto di un procedimento amministrativo interamente computazionale, attraverso il quale il *software* determina il contenuto dell'atto individuando i principi normativi rilevanti al caso di specie, così come i dati necessari, in modo completamente automatizzato. In questo senso, ciò che contraddistingue un atto amministrativo ad elaborazione elettronica è la circostanza per cui l'intero percorso logico che conduce alla decisione finale (*rectius* la motivazione dell'atto) è esclusivamente imputabile al *software*.

Al di là della riflessione operata dalla Corte circa l'ammissibilità dell'utilizzo di questa soluzione informatica per l'adozione di atti amministrativi diversi da quelli a discrezionalità vincolata, la sentenza è particolarmente rilevante ai fini dell'analisi fin qui

⁸⁹ TAR Lazio, sez. III-bis, n. 3769/2017.

condotta in virtù di due ulteriori conclusioni raggiunte dal giudice amministrativo.

In primo luogo, dal momento in cui il caso evidentemente richiama delle decisioni interamente automatizzate al pari di quelle di cui all'articolo 22 GDPR, è interessante notare come nonostante il TAR abbia confermato la natura creativa del *software* in questione (entrando così nell'ambito di applicazione della specifica protezione apprestata al diritto d'autore sullo stesso), il giudice abbia ciononostante escluso che l'esercizio di un diritto di accesso individuale al codice sorgente potesse compromettere l'interesse economico del titolare di diritti di proprietà intellettuale sullo stesso. Più specificamente, la Corte ha sottolineato che la disciplina sul diritto d'autore non preclude qualsivoglia riproduzione dell'opera di ingegno, ma esclusivamente quelle idonee a consentirne uno sfruttamento economico da parte di terzi. Questo, tuttavia, non è il caso del diritto individuale di accesso ai sensi dell'articolo 24 L. 241/1990, dato il limitato numero di soggetti ammessi dalla norma a presentare una richiesta giuridicamente motivata ⁹⁰.

Passando alla seconda rilevante conclusione raggiunta dalla Corte, quest'ultima ha ritenuto che nonostante l'indiscussa intrusività della richiesta di accesso al codice sorgente, la decisione organizzativa assunta dal ministero "di adottare delle soluzioni tecnologiche innovative per meglio eseguire i propri compiti non può incidere negativamente sui diritti dei titolari dei diritti individuali di accesso". È necessario, infatti, consentire ai destinatari di queste decisioni di vagliare il concreto funzionamento dell'algoritmo al fine di poter verificare l'esistenza di errori nel trattamento dei dati. Questa valutazione, tuttavia, nell'ottica del TAR, può essere effettivamente operata dai destinatari della decisione soltanto nella misura in cui siano dotati di quella conoscenza realmente completa assicurabile solo ed esclusivamente dal pervasivo accesso al codice sorgente. La Corte conclude, quindi, affermando che l'interesse sotteso alla richiesta di accesso sollevata dal ricorrente non possa essere considerata adeguatamente soddisfatta dalla menzionata descrizione del funzionamento dell'algoritmo redatta dal programmatore del *software* "in quanto, *evidentemente e intuitivamente*, la descrizione della modalità di

⁹⁰ Il TAR ha inoltre specificato che la conclusione raggiunta sarebbe stata opposta se il diritto di accesso fosse stato concepito dal legislatore come uno strumento di controllo generalizzato sull'operato della Pubblica Amministrazione a disposizione dei cittadini per il caso in cui quest'ultima si avvalga di processi decisionali automatizzati. TAR Lazio, sez. III-bis, n. 3769/2017. Circa la differenza tra la pubblicazione diffusa di

interne categorie di informazioni operata dalla PA spontaneamente, ovvero a fronte della richiesta immotivata di qualsiasi cittadino, ex articolo 5-bis D.Lgs. n. 97/2016 e la "disclosure reattiva" innescata dalla richiesta individuale e motivata di accesso a specifici documenti, si veda S. Foà, *La nuova trasparenza amministrativa*, in *Diritto Amministrativo*, 1, 2017, 65.

funzionamento dell'algoritmo assicura una conoscenza assolutamente non paragonabile a quella che deriverebbe dall'acquisizione del richiesto linguaggio sorgente, atteso che, se non altro, la predetta descrizione è, comunque, atto di parte"⁹¹.

Trasponendo questo ragionamento al simile ambito di applicazione del diritto di accesso fissato dal GDPR, potrebbe essere sostenuto che affinché un soggetto interessato sia abilitato ad esercitare un controllo effettivo sulla decisione esclusivamente automatizzata, dovrebbe essergli garantito un accesso individuale (motivato) al codice sorgente del *software* utilizzato, nella misura in cui questo sarebbe l'unico modo per fornirgli una descrizione imparziale ed esaustiva nel funzionamento interno dell'algoritmo.

Questa soluzione, tuttavia, non potrebbe che trascurare completamente il più volte menzionato principio di trasparenza fissato nel GDPR, come peraltro recentemente ribadito dalla Carta etica europea sull'uso dell'intelligenza artificiale nel sistema giurisdizionale e dintorni (d'ora in avanti ECAI) adottata dalla Commissione Europea per l'Efficienza della Giustizia del Consiglio d'Europa (CEPEJ) nella sua trentunesima seduta plenaria svoltasi a Strasburgo nel dicembre 2018⁹².

Il quarto dei cinque principi etici ivi fissati (il principio della trasparenza, dell'imparzialità e dell'equità), stabilisce la necessità di "rendere il trattamento dei dati accessibile e comprensibile"⁹³. Il CEPEJ stesso, nei successivi paragrafi esplicativi enfatizza la difficoltà di trovare un equilibrio tra la necessità di garantire una qualche forma di accesso al processo di programmazione e la necessità di garantire la protezione dei diritti di proprietà intellettuale che gravano su gran parte dei metodi di trattamento dei dati. L'opzione della completa trasparenza tecnica, tuttavia, non sembra convincere neppure la commissione poiché, come riconosciuto anche in altri studi del Consiglio d'Europa e di altri organismi nazionali dal primo richiamati⁹⁴, la diffusione dell'algoritmo e del codice sorgente finirebbe inevita-

⁹¹ Corsivo aggiunto e traduzione non ufficiale.

⁹² La COMMISSIONE PER L'EFFICIENZA DELLA GIUSTIZIA DEL CONSIGLIO D'EUROPA (CEPEJ) è stata istituita nel 2002 dai ministri della giustizia degli Stati membri del Consiglio d'Europa. È composta da esperti provenienti da tutti i 47 Stati membri e si pone l'obiettivo di migliorare la qualità dei sistemi giurisdizionali europei, rafforzando la fiducia dei cittadini negli stessi. Di conseguenza, il ruolo del CEPEJ è quello di fornire ai legislatori europei, così come ai professionisti del settore, le conoscenze tecniche necessarie a creare un effettivo foro di discussione. Maggiori dettagli sono disponi-

bili all'indirizzo www.coe.int/en/web/cepej/about-cepej.

⁹³ CEPEJ, *European Ethical Charter on the use of artificial intelligence in judicial systems and their environment*, adottata in occasione della 31esima sessione plenaria del CEPEJ (Strasburgo, 3-4 dicembre 2018), 12.

⁹⁴ Vedi Ivi, nt. 3 che richiama lo studio del Consiglio d'Europa e del MSI-NET circa "Algorithms and Human Rights. Studies on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications", DGI (2017)12 (CONSIGLIO D'EUROPA, marzo 2018), così come il report di Villani "AI for huma-

bilmente per scontrarsi con l'opposizione dei detentori dei diritti di proprietà intellettuale, compromettendo in questo modo non solo le già scarse possibilità di garantire una effettiva trasparenza⁹⁵, ma anche l'interesse economico che muove gli investimenti in sistemi di intelligenza artificiale sempre più innovativi.

Una volta esclusa l'opportunità legislativa ed economica di un approccio esplicativo fondato sul menzionato accesso all'algoritmo, così come della possibilità di rispettare il dovere di fornire informazioni significative semplicemente offrendo una spiegazione esterna, rimane l'esigenza di determinare la misura e la qualità dell'apertura della scatola nera necessaria, e che una cosiddetta spiegazione destrutturata⁹⁶ dovrebbe presentare per consentire ai soggetti interessati di avere una comprensione utile della logica algoritmica seguita dal *software* per arrivare alla decisione unicamente automatizzata.

In questo senso, si è sostenuto che ci sono forme di informazioni concernenti il funzionamento interno di un *software* proprietario di tipo decisionale "che possono essere divulgati senza il rischio di violare alcun diritto di proprietà intellettuale"⁹⁷, è questo, ad esempio, il caso delle variabili realmente utilizzate, dei generali obiettivi operativi perseguiti attraverso il processo decisionale automatizzato, della tipologia e quantità di dati utilizzati per programmare l'algoritmo, così come di quelli concretamente trattati nel caso individuale⁹⁸.

Al di là dell'evidente fluidità della linea di confine tra una spiegazione trasparente del funzionamento interno del *software* e una violazione del segreto commerciale, appare estremamente

nity" elaborato in esecuzione del compito affidatogli dal primo ministro francese, così come il report della House of Lords, "AI in the UK: ready, willing and able?", spec. para 92, 96-99.

⁹⁵ Vedi § 3 *supra*.

⁹⁶ Come chiaramente affermato, questo tipo di spiegazione, per definizione "attempts to open the black box, and understand how the structures within, such as the weights, neurons, decision trees and architecture, can be used to shed light on the patterns that they encode." L. EDWARDS - M. VEALE, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, op. cit., 64.

⁹⁷ R. TAYLOR, *No Privacy Without Transparency*, in R. LEENES, R. VAN BRAKEL, S. GUTWIRTH, P. DE HERT (a cura di), *Data Protection and Privacy: The Age of Intelligent Machines*, op. cit., 103.

⁹⁸ Vedi lo studio del Consiglio d'Europa e del MSI-NET circa "Algorithms and Human Rights. Studies on the Human

Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications", DGI (2017)12 (CONSIGLIO D'EUROPA, marzo 2018), 38. Alla luce del report esplicativo della Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale (Consiglio d'Europa, Strasburgo, 28 gennaio 1981), anche nota come Convenzione 108, "I soggetti interessati hanno il diritto di conoscere il ragionamento sottostante il trattamento dei dati. [...] [P]er esempio nel caso di sistemi di rating creditizio, i soggetti finanziati hanno il diritto di conoscere la logica del trattamento dei propri dati che ha condotto alla decisione di concedere o rifiutare il finanziamento, piuttosto che essere semplicemente informati della decisione in sé". Si veda anche CEPEJ, *European Ethical Charter on the use of artificial intelligence in judicial systems and their environment*, adottata in occasione della 31esima sessione plenaria del CEPEJ (Strasburgo, 3-4 dicembre 2018), 46 spec. nt. 58.

difficile definire in modo aprioristico un regime di *disclosure* unico in grado di assicurare ai soggetti interessati, per ogni tipologia di decisione automatizzata, la possibilità di “provare l’accuratezza dei dati personali prima del loro utilizzo, l’adeguatezza dei profili applicati alla loro particolare situazione ovvero dei fattori che hanno avuto un impatto sull’esito del processo decisionale automatizzato”⁹⁹. Inoltre, anche ammettendo la possibilità astratta di fissare un *minimum* nel livello di *disclosure* necessario per rispettare il “principio del controllo dell’utente” come stabilito nel principio 5 della ECAI, la crescente complessità dei sistemi di *deep machine learning*, rende questa operazione non sempre tecnicamente possibile.

In altre parole, al di là della menzionata difficoltà di tradurre in termini umanamente comprensibili i ragionamenti algoritmici, la complessità intrinseca delle forme *deep learning* di intelligenza artificiale rende impossibile agli stessi programmatori predire l’esito del trattamento automatizzato, così come identificare *ex post* le categorie e le variabili elaborate ed utilizzate in modo autonomo dal software per arrivare alla specifica decisione. È precisamente questa, infatti, la terza e ultima frontiera di opacità che compromette l’effettività del diritto ad una spiegazione significativa della logica seguita nel trattamento automatizzato: l’opacità intrinseca.

5. L’OPACITÀ INTRINSECA: UN COSTO CHE VALE LA PENA PAGARE?

Come osservato¹⁰⁰, la *Knowledge Discovery in Databases* (KDD) è caratterizzata dall’attitudine al cosiddetto “*repurposing*”, in virtù della quale le ipotesi scientifiche, le opzioni per lo sfruttamento economico o, più coerentemente all’analisi qui condotta, le regole decisionali seguite dalla macchina, tendono ad emergere nel corso del procedimento automatizzato in modo dinamico, totalmente imputabile alla macchina e non completamente prevedibile dai programmatori¹⁰¹.

⁹⁹ Questa è l’espressione efficacemente utilizzata nel *report* esplicativo della Convenzione 108, come modificata dal Protocollo adottato a maggio 2018, para 75 e articolo 9(1)(c) — vd anche nota precedente —. Il documento peraltro prosegue specificando che “[q]uesto è precisamente il caso in cui l’adozione di un ragionamento algoritmico, importando la limitazione di un diritto, il rifiuto di una prestazione sociale ovvero la valutazione della loro capacità di restituzione di un finanziamento sulla sola base del software, abbia l’effetto di stigmatizzare gli individui”.

¹⁰⁰ M. HILDEBRANDT, *The Dawn of a Critical Transparency Right for the Profiling Era*, in J. BUS, M. CROMPTON, M. HILDEBRANDT, G. METAKIDES (a cura di), *Digital Enlightenment Yearbook 2012*, op. cit., 44; B. CUSTERS, *The Power of Knowledge Ethical, Legal and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Wolf Legal Publishers, Nimega, 2004, 160.

¹⁰¹ M. HILDEBRANDT, *The Dawn of a Critical Transparency Right for the Profiling Era*, in J. BUS, M. CROMPTON, M. HILDEBRANDT, G. METAKIDES (a cura di), *Digital*

Tale caratteristica non solo confligge fisiologicamente con il principio della limitazione delle finalità e della trasparenza di cui all'articolo 5 co. 1 lett. (b) e (a) GDPR ¹⁰², ma pone anche gravi minacce all'efficacia dei doveri del responsabile del trattamento di dimostrare il rispetto del principio di *accountability* stabilito nell'articolo 5 co. 2 GDPR. Infatti, "a prescindere dal fatto che questa informazione [i.e. una spiegazione nella logica algoritmica] possa essere significativamente fornita a non esperti" ¹⁰³, l'oscurità matematica caratterizzante i sistemi di *deep machine learning* rende tecnologicamente impossibile condurre una valutazione *ex post* delle variabili estratte dalla macchina dai dati acquisiti e utilizzate per prendere la decisione automatizzata, così come offrire un'analisi *ex ante* esaustiva delle conseguenze previste di tale trattamento ¹⁰⁴.

La natura altamente problematica di questa considerazione è ulteriormente corroborata dai rari progetti pilota attualmente esistenti volti a sviluppare dei metodi esplicativi proprietari veloci precisi e coerenti suscettibili di superare il cosiddetto "problema della scatola nera", in tal modo focalizzandosi sullo sviluppo di tecniche in grado di svelare i funzionamenti interni del sistema di *machine learning* attraverso una verifica *ex post* dei passaggi algoritmici autonomamente condotti dal *software*. A differenza dei menzionati sistemi LOCO, LIME e PI, *Zest Automated Machine Learning* (ZAML) è presentato come "l'unica soluzione per lo sviluppo di sistemi di *machine learning* per il credito e per la gestione del rischio con una "spiegabilità" *end-to-end*." Più specificamente, Zestfinance sostiene di offrire l'unico sistema per la concessione di prestiti basato su un sistema di *machine learning* in grado di consentire ai finanziatori di: (i) analizzare il funzionamento interno del sistema; (ii) analizzare il modello concretamente utilizzato per ogni singola decisione di credito piuttosto che avvalersi di modelli astratti; (iii) analizzare i dati individuali e non "dati sintetici o meramente randomizzati" ¹⁰⁵.

Enlightenment Yearbook 2012, op. cit., 44; R. TAYLOR, *No Privacy Without Transparency*, in R. LEENES, R. VAN BRAKEL, S. GUTWIRTH, P. DE HERT (a cura di), *Data Protection and Privacy: The Age of Intelligent Machines*, op. cit., 110.

¹⁰² Ai sensi dell'articolo 5 co. 1 lett. b GDPR, i dati personali sono "raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità; un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici non è, conformemente all'articolo 89, paragrafo 1, considerato incompatibile con le

finalità iniziali (« limitazione della finalità »)".

¹⁰³ S. WACHTER - B. MITTELSTADT - C. RUSSELL, *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, op. cit., 22.

¹⁰⁴ G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit., 258; T.Z. ZARSKY, *Mine your Own Business!*, op. cit., 56.

¹⁰⁵ "It's Zest Automated Machine Learning, and it will transform your underwriting through the power of machine learning." (ZestFinance) disponibile all'indirizzo www.zestfinance.com/zaml.

Prescindendo dalla veridicità di queste affermazioni pubblicitarie, lo stato sperimentale di tali pionieristici approcci a quella che può essere definita come una tecnica di *black unboxing* mette in luce la sfida tecnologica attualmente sottostante l'implementazione di un effettivo controllo su questi processi decisionali automatizzati.

Particolarmente rilevante a tal proposito è il programma *Explainable Artificial Intelligence* (XAI) recentemente lanciato dell'Agenzia per i progetti di ricerca avanzati sulla difesa (DARPA) con l'obiettivo di promuovere la “terza onda intelligenza artificiale” ideata per: (i) “offrire modelli più spiegabili, mantenendo allo stesso tempo un alto livello di *learning performance*”, e (ii) “consentire agli operatori umani di comprendere, riporre adeguata fiducia, e di gestire efficacemente l'emergente generazione di *partner* di intelligenza artificiale”¹⁰⁶.

L'obiettivo ultimo del DARPA, coerente con la missione militare affidatagli dal dipartimento della difesa statunitense a partire dal 1957 con il lancio dello Sputnik, è quello di alimentare una *partnership* sempre più paritaria tra gli operatori umani e i *software* di *machine learning* attraverso la quale i primi possano trarre effettivo vantaggio dai processi neurali e dalle capacità predittive dei sistemi di *deep learning*, così aumentando le probabilità dei militari di prendere complesse decisioni belliche in ridotti intervalli temporali.

Scopo simile è quello perseguito dal Data Transparency Lab, una comunità inter-istituzionale con il dichiarato obiettivo di offrire supporto a qualsivoglia progetto finalizzato ad offrire soluzioni tecnologiche volte a rafforzare una effettiva protezione dei dati nel contesto online¹⁰⁷.

La breve rassegna di tali sforzi pionieristici mostra come lo standard legislativo della *accountability* fissato dal GDPR rimanga attualmente ineguagliato dalle ancora embrioniche tecnologie *glass-box*¹⁰⁸, alla base del funzionamento del *big data due process*¹⁰⁹. In particolare, a causa dell'attuale impossibilità tecnologica di accertare il funzionamento interno (così come l'intera

¹⁰⁶ D. GUNNING, *Explainable Artificial Intelligence* (XAI) (Defense Advanced Research Projects Agency) disponibile all'indirizzo www.darpa.mil/program/explainable-artificial-intelligence.

¹⁰⁷ DATA TRANSPARENCY LAW, disponibile all'indirizzo www.datatransparencylab.org/technical-program/.

¹⁰⁸ N. BAIRD, *Retail Has Three Big AI Dilemmas*, in *Forbes*, 13 agosto 2018, disponibile all'indirizzo www.forbes.com/sites/nikkibaird/2018/08/13/retail-has-three-big-ai-dilemma.

¹⁰⁹ D. KEATS CITROW, *Technological Due Process*, 85, *Wash UL Rev.*, 2008, 1256; K. CRAWFORD - J. SCHULZ, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55, *Boston College Law Rev.*, 93, 2014, 119; K. DE VRIES - M. HILDEBRANDT, *Introducing Privacy, Due Process and the Computational Turn at a Glance: Pointer for the Hurried Reader*, in Id. (a cura di), *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Routledge, New York, 2013, 9.

gamma di scopi del trattamento) che caratterizza la variante di *deep learning* dei sistemi di Intelligenza Artificiale, nessuna delle due forme di *auditing* suggerite da Malgieri e Comandè sembra essere attualmente tecnologicamente sostenibile, seppure desiderabile da un punto di vista legislativo. Tanto *l'auditing* sull'architettura, quindi, quanto quello sull'implementazione del trattamento dei dati, rimarrebbero altamente inefficaci ¹¹⁰.

Di conseguenza, rimane controversa la misura in cui un'ipotesica *trusted third party*, quali i meccanismi di certificazione della *compliance* del trattamento dei dati ex articolo 42 GDPR, potrebbe individuare violazioni o rischi. Lo stesso, *a fortiori*, non può che essere altrettanto vero con riguardo al tasso di effettività dei codici di condotta o degli *impact assessment* sulla protezione dei dati che, ai sensi dell'articolo 35 GDPR, i responsabili del trattamento che conducono attività di elaborazione dei dati suscettibili di esporre i diritti e delle libertà i soggetti interessati a rischi (e.g. utilizzando dati personali per prendere decisioni "riguardanti determinate persone fisiche in seguito a una valutazione sistematica e globale di aspetti personali relativi alle persone fisiche, basata sulla profilazione di tali dati, o in seguito al trattamento di categorie particolari di dati personali, dati biometrici o dati relativi a condanne penali e reati o a connesse misure di sicurezza") ¹¹¹ dovrebbero condurre ¹¹².

Ciononostante, come già affermato ed esplicitamente ribadito dalla menzionata sentenza del TAR, l'opacità strutturale dei sistemi di *machine learning* (di ogni tipo) non può giustificare un diniego dei diritti di accesso dei destinatari delle decisioni, inclusi quelli concernenti il *due process*, nella misura in cui questo è finalizzato a offrire ai soggetti interessati una descrizione intelligibile della logica sottostante il processo decisionale automatizzato. In caso contrario, tale spiegazione rimarrebbe inevitabilmente vaga, così favorendo decisioni cc.dd. "*unaccountable*" ¹¹³. Inoltre, le sole disposizioni del GDPR che disciplinano il pro-

¹¹⁰ G. MALGIERI - G. COMANDÈ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, op. cit., 258 ss. Per un interessante *case study* del *Digital Due Process* si veda D. C. NUNZIATO, *With Great Power Comes Great Responsibility: Proposed Principles of Digital Due Process for ICT Companies*, in L. FLORIDI (a cura di), *Protection of Information and the Right to Privacy. A New Equilibrium?*, Law, Governance and Technology Series, Vol. 17, Springer, New York, 2014, 80-84.

¹¹¹ Considerando 91 GDPR.

¹¹² Questo obbligo, peraltro, dovrebbe essere adempiuto prima del tratta-

mento, prendendone in considerazione la funzione, il contesto e gli obiettivi in modo tale da poter utilizzare i risultati di questa preliminare valutazione del rischio per individuare le misure più adatte a dimostrare il rispetto della normativa del GDPR, al fine ultimo di implementarle. Vd Considerando 74, 84 e 90 GDPR.

¹¹³ K. BRENNAN-MARQUEZ, "*Plausible Cause*": *Explanatory Standards in the Age of Powerful Machines*, 70, *Vanderbilt Law Review*, 2017, 7; J.A. KROLL - J. HUEY - S. BAROCAS - E.W. FELTEN - J.R. REIDENBERG - D.G. ROBINSON - H.YU, *Accountable Algorithms*, 165, *University of Pennsylvania Law Review*, 2017, 633.

blema della impossibilità ad implementare i diritti dei soggetti interessati nel perseguire gli obiettivi del trattamento dei dati, ammettono deroghe solo con riferimento ai diritti fissati negli articoli 15, 16 e 21 GDPR (diritti di accesso, rettifica, limitazione del trattamento e di opposizione) e solo nel caso in cui tali restrizioni siano necessarie ad effettuare trattamenti mossi da scopi scientifici, di ricerca storica, statistici ovvero di archiviazione nel pubblico interesse. Inoltre, l'articolo 89 co. 4 GDPR specifica che “qualora il trattamento di cui ai paragrafi 2 e 3 funga allo stesso tempo a un altro scopo, le deroghe si applicano solo al trattamento per le finalità di cui ai medesimi paragrafi”.

Un'esenzione leggermente più ampia può essere individuata nell'articolo 19 GDPR dove viene fatto genericamente riferimento all'impossibilità di rispettare i doveri di rettifica, cancellazione, e notificazione della limitazione del trattamento, a prescindere dallo scopo perseguito dai responsabili del trattamento. Tuttavia, nessuna disposizione fissa all'esenzione al diritto di “non essere sottoposto ad una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona” ex articolo 22 GDPR. Perciò, qualora il responsabile del trattamento volesse negare una spiegazione intellegibile *ex post* (ai sensi dell'articolo 15 GDPR) di una decisione automatizzata resa ai sensi dell'articolo 22 co. 2 lett. (a) o (c) GDPR in ragione della eccezione di impossibilità legittimata dagli obiettivi statistici perseguiti, il soggetto interessato dovrebbe comunque avere il diritto ad ottenere l'intervento umano, a contestare la decisione e ad avere informazioni significative *ex ante* sulla logica seguita dal processo decisionale automatizzato¹¹⁴. Di conseguenza, in ogni altro caso i responsabili del trattamento non potrebbero adottare delle tecnologie che, pur perseguendo gli obiettivi del trattamento in modo più efficace, renderebbero impossibile rispettare il diritto dei soggetti interessati ad ottenere un'informazione significativa.

Sebbene sia innegabile, infatti, che passare a *decision trees* o altre tecniche di *clear-box* ridurrebbe il livello di precisione predittiva dei risultati di *machine learning*, vale la pena notare che optare per la tecnologia più performante non assicura necessariamente un'adeguata ponderazione del giudizio automatiz-

¹¹⁴ Più specificamente, il Considerando 62 e gli articoli 14 (5) (b), 17 (3) (d), 89 (2) e (3) GDPR si riferiscono al caso in cui l'adempimento di tali doveri di prova si dimostri “impossibile o [tale da] pregiudicare gravemente il conseguimento delle finalità di tale trattamento”. Come menzionato nel testo, una previsione leggermente

diversa la si può invece rinvenire nel disposto dell'articolo 19 GDPR, nella misura in cui quest'ultimo fissa un'eccezione ai doveri di notificazione circa la rettifica/cancellazione dei dati personali, ovvero circa limitazioni al trattamento, nel caso in cui l'adempimento “risult[i] impossibile o im-plich[i] uno sforzo sproporzionato”.

zato ¹¹⁵. Come efficacemente sottolineato: “in alcuni contesti, un giudizio è necessario non perché questo garantisca la perfezione statistica, ma perché rende l’esercizio del potere intellegibile” ¹¹⁶. Questo è particolarmente vero nel caso degli algoritmi di apprendimento che, in un contesto di crescente quantità e qualità di dati personali altamente interconnessi, rendono estremamente difficoltoso individuare variabili discriminatorie indirette (ad esempio, il codice postale potrebbe essere una variabile indiretta per l’etnia) ¹¹⁷. Perciò, nella misura in cui gli standard esplicativi perseguono un equilibrio caso per caso di una pluralità di valori confliggenti, l’utilizzo di decisioni automatiche non solo è inammissibile da un punto di vista normativo ma è anche inefficace da quello tecnologico.

Questa è, per esempio, la ragione per cui studiosi di diritto amministrativo nel segnalare il passaggio da una cosiddetta burocrazia *street-level* (nella quale i processi decisionali rientrano completamente nell’ambito di competenza dell’agente umano) ad una burocrazia *screen-level* (dove il ruolo umano è confinato alla programmazione *ex ante* e ad una valutazione passiva *ex post* del processo automatizzato), ritengono tale transizione ammissibile solo con riferimento agli atti a discrezionalità vincolata ¹¹⁸.

A tal proposito, il Parlamento Europeo ha chiaramente ribadito che sebbene i sistemi di *machine learning* apportino indubbi benefici economici ed è verosimile che supereranno le capacità intellettive umane, lo sviluppo dell’intelligenza artificiale dovrebbe essere finalizzata a complementare e non sostituire quest’ultima, garantendo agli umani “un controllo sulle macchine

¹¹⁵ Per una apologia della semplicità si veda anche I. VOTSIS, *Philosophy of Science and Information*, in L. FLORIDI (a cura di), *The Routledge Handbook of Philosophy of Information*, Routledge, New York, 2016, 255, e nt. 8, il quale citando Occam scrive: “per esempio, nella *Summa Logicae*, egli afferma: ‘*Pluralitas non est ponenda sine necessitate*’ (1974, p. 185). Tale espressione si traduce, approssimativamente, nella frase ‘la pluralità non deve essere adottata senza necessità’. La pluralità in questione sembrerebbe riferirsi ad una tipo ontologico. In altri termini, l’enfasi viene posta sulla riduzione della complessità ontologica. Non c’è nessun riferimento diretto a ipotesi più semplici, in questa o altre citazioni.” (traduzione dell’autore).

¹¹⁶ K. BRENNAN-MARQUEZ, “*Plausible Cause*”: *Explanatory Standards in the Age of Powerful Machines*, 70, *Vanderbilt Law Review*, 2017, 14. Come magistralmente affermato da Taruffo: “*at present the pro-*

blem is neither of making a priori acts of faith in the all-encompassing virtues of AI nor of making a priori acts of disbelief in any possibility of using AI as a means to elicit, interpret and represent the judge’s reasoning. Both attitudes, in fact, would be misplaced and unjustified on the basis of the current “state of the art” of AI applied to the problem of judicial decision making. Rather, one should take into account the emerging trends in this area and consider whether they appear well oriented and possibly fruitful.” M. TARUFFO, *Judicial Decisions and Artificial Intelligence*, 6, *Artificial Intelligence and Law*, 1998, 317.

¹¹⁷ T. MILLER - P. HOWE - L. SONENBERG, *Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences*, in ArXiv, 2017, 3.

¹¹⁸ L. VIOLA, *L’Intelligenza Artificiale nel procedimento e nel processo amministrativo: lo stato dell’arte*, 2, *Foro Amministrativo*, 9, 2018, 1598.

intelligenti in ogni momento”¹¹⁹. A tal fine, maggiore enfasi dovrebbe essere posta su due principi cardine ripresi dal Parlamento europeo.

Il primo è il principio di inclusività, cuore dell’intero articolo, in base al quale viene sottolineata l’importanza di assicurare a tutti i portatori di interesse la partecipazione nel processo decisionale. Il secondo è il principio di precauzione, in base al quale i ricercatori dovrebbero attentamente anticipare i potenziali impatti sulla sicurezza delle applicazioni di intelligenza artificiale¹²⁰. Questo è esattamente il motivo per cui l’ultimo paragrafo è stato specificamente dedicato ai doveri di *Data Protection Impact Assessment* introdotti dal GDPR che, sebbene rappresentino la più emblematica espressione di tale principio, hanno suscitato una insufficiente attenzione con riferimento ai problemi riguardanti la loro efficacia¹²¹.

Sebbene ciò esuli dallo scopo della presente analisi, peraltro, è bene sottolineare che indipendentemente dal “carattere apparentemente obiettivo ed incontrovertibile”¹²² dei processi automatizzati e delle loro crescenti capacità di autoapprendimento, i decisori umani non dovrebbero essere indotti ad abdicare la loro responsabilità, specialmente fin quando la “spiegabilità” dell’intelligenza artificiale (dove tecnologicamente possibile) continui ad essere progettata da esperti informatici per esperti informatici (noto come il problema degli “*inmates running the asylum*”)¹²³ e il quadro normativo rimanga inadeguato ad affrontare il tema della responsabilità di applicazioni imprevedibili di intelligenza artificiale, quali sono le menzionate applicazioni di *deep machine learning*. Seppure in via del tutto incidentale, quindi, si ritiene indispensabile in questa sede sottolineare come riconoscere lo *status* di persone elettroniche ai sistemi di intelligenza artificiale, da un lato, non sembrerebbe giuridicamente percorribile, dal momento che i *software* di *machine learning* non presentano necessariamente il minimo supporto fisico richiesto dalla risolu-

¹¹⁹ Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)), para 3,12 e considerando O, N, P, Q.

¹²⁰ CODICE ETICO-DEONTOLOGICO DEGLI INGEGNERI ROBOTICI allegato alla Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

¹²¹ S. RODOTÀ, *Il mondo nella rete. Quali i diritti, quali i vincoli*, op. cit., 40 e 201.

¹²² COMMISSIONE EUROPEA, Proposta mo-

dificata di direttiva del Consiglio relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, COM(92) 422 DEF- SYN 287, Bruxelles, 15 ottobre 1992. Cfr. D. KAMARINOU - C. MILLARD, - J. SINGH, *Machine Learning with Personal Data*, Queen Mary School of Law Legal Studies Research Paper No. 247/2016, disponibile all’indirizzo <https://ssrn.com/abstract=2865811> last accessed 25 April 2019.

¹²³ A. COOPER, *The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*, Sams Publishing, Indianapolis, 2004.

zione europea ¹²⁴. Dall'altro, concepire siffatte applicazioni di AI come autonomi centri di imputazione di diritti e di doveri, rischierebbe di accentuare forme di deresponsabilizzazione degli operatori umani coinvolti (in modo comunque sempre più passivo) nei descritti processi decisionali automatizzati, così fornendo *humus* normativo alla già incontrollata attitudine delle *Information Technologies* a declinarsi in forme di *human-out-of-loop* ¹²⁵.

In conclusione, come recentemente affermato dal Consiglio di Stato nel giudizio n. 2270/2019, concernente l'appello presentato da insegnanti neoassunti contro la decisione algoritmica presa dal MIUR con riferimento alla loro assegnazione: "il meccanismo attraverso il quale si concretizza la decisione robotizzata (ovvero l'algoritmo) deve essere "conoscibile", secondo una versione rafforzata del principio trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio diverso da quello giuridico. [...] In altri termini, la "caratterizzazione multidisciplinare" dell'algoritmo [...] 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, sia per

¹²⁴ Il Parlamento europeo ha ribadito la necessità di introdurre "definizioni europee comuni di sistemi ciberfisici, di sistemi autonomi, di robot autonomi intelligenti e delle loro sottocategorie, prendendo in considerazione le seguenti caratteristiche di un robot intelligente: (i) l'ottenimento di autonomia grazie a sensori e/o mediante lo scambio di dati con il suo ambiente (interconnettività) e lo scambio e l'analisi di tali dati; (ii) l'autoapprendimento dall'esperienza e attraverso l'interazione (criterio facoltativo); (iii) almeno un supporto fisico minore; (iv) l'adattamento del proprio comportamento e delle proprie azioni all'ambiente; (v) l'assenza di vita in termini biologici". Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

¹²⁵ Nella consapevolezza dell'ampiezza del risalete e autorevole dibattito sviluppato attorno all'opportunità normativa di riconoscere forme di personalità giuridica ai sistemi di AI, in questa sede ci si ripropone, più modestamente, di destare attenzione alle conseguenze psicologiche, prima ancora che giuridiche, che siffatto status potrebbe provocare, specialmente alla luce dei recenti sviluppi applicativi dell'Intelligenza Artificiale fin qui analizzati. Con riguardo al tema dello *status* di persona elettronica, a titolo meramente esemplifica-

tivo, si menzionano i fondamentali contributi di C.E. A. KARNOW, *The Encrypted Self: Fleshing Out the Rights of Electronic Personalities*, 13, *Journal of Computer & Information Law*, 1994, 1; L.B. SOLUM, *Legal Personhood for Artificial Intelligences*, 70, *N.C. L. Rev.*, 1992, 1231, e C.D. STONE, *Should Trees Have Standing? Law, Morality, and the Environment*, III ed., Oxford University Press, Oxford, 2010, 8 ss. Per una prospettiva più organica e recente si veda invece G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, Edizioni Scientifiche Italiane, Napoli, 2019, 36 ss.; nonché W. BARFIELD, *Towards a Law of Artificial Intelligence*, in W. BARFIELD e U. PAGALLO (a cura di), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, Northampton, 2018, 36 ss. Pertinenti anche le riflessioni di Miller *et al. supra* nt 115 a supporto dell'"enfasi posta dalle recenti sollecitazioni del DARPA circa la necessità di creare tecniche "human-in-the-loop" utilizzabili dagli sviluppatori [...] per consentire valutazioni umane più intensive" e condivide l'opinione di Doshi-Velez e Kim per cui affinché si possa produrre un impatto sul mondo reale "è fondamentale che noi come comunità rispettiamo il tempo e lo sforzo necessari a compiere tali valutazioni." (traduzione dell'autore) Cfr. F. DOSHI-VELEZ, B. KIM, *Towards A Rigorous Science of Interpretable Machine Learning*, in *arXiv*, 2017.

i cittadini che per il giudice.” Sebbene le circostanze fattuali fossero molto simili a quelle della sentenza del TAR sopra descritta, in questo caso il giudice non ha preso in considerazione alcun diritto di proprietà intellettuale, ma si è limitato a stabilire che l'impossibilità di comprendere il funzionamento interno dell'algoritmo integri una violazione del principio di imparzialità, pubblicità e trasparenza, sufficiente a compromettere la legittimità del procedimento amministrativo ¹²⁶.

Alla luce di quanto detto, se si vuole che il mondo cessi di simulare un cosiddetto “specchio unidirezionale” ¹²⁷ in cui la tirannia di un potere algoritmico irreversibilmente opaco si sostituisce progressivamente alla discrezionalità umana, nel nome di una maggiore precisione predittiva e della crescita economica, così trasformando i soggetti interessati in imperscrutabili punteggi, è necessario assicurare l'effettività di un adeguato livello di trasparenza ¹²⁸. Di conseguenza, i responsabili del trattamento dovrebbero astenersi dall'implementare tecniche di elaborazione dei dati in cui i procedimenti interni non possano essere completamente controllati o tradotti in forme umanamente comprensibili e rispettose di uno standard di intelligibilità parametrato al livello medio di un soggetto interessato non esperto, anche se ciò potrebbe comportare la necessità di pagare il prezzo di un rallentamento dello sviluppo tecnologico. Come correttamente affermato nel considerando 4 GDPR, infatti, “il trattamento dei dati personali dovrebbe essere al servizio dell'uomo”, e non il contrario.

¹²⁶ Consiglio di Stato, sez. VI, sentenza dell'8 aprile 2019, n. 2270.

¹²⁷ F. PASQUALE *The Black Box Society. The Secret Algorithms That Control Money and Information*, op. cit., 9.

¹²⁸ Come acutamente osservato, infatti, “[t]he decline in personal privacy might be worthwhile if it were matched by comparable levels of transparency [...]. But

when every move we make is subject to inspection by entities whose procedures and personnel are exempt from even remotely similar treatment, the promise of democracy and free markets rings hollow. Secrecy is approaching critical mass, and we are in the dark about crucial decisions. Greater openness is imperative.” Ivi, 4.