Algoritmi e decisioni automatizzate.
Tutele esistenti e linee evolutive della regolazione

Erica Palmerini

Sommario: 1. *Processi automatizzati di decisione: un primo inventario*. – 2. *Le classificazioni possibili. Dati personali, dati non personali, dati anonimi*. – 3. Segue. *Il soggetto che effettua il trattamento*. – 4. *Intelligenza artificiale e tutela dei dati*. – 5. *Le decisioni automatizzate nel GDPR*. – 6. *Dal controllo sulle informazioni "in entrata" alle tutele rispetto alla decisione. La "profilazione"*. – 7. *Il divieto di trattamenti esclusivamente automatizzati: natura e portata della prescrizione*. – 8. *Le eccezioni al divieto*. – 9. *Le garanzie applicabili: informazione, contestazione, revisione*. – 10. *La trasparenza nel mondo digitale e dei* big data. – 11. *I limiti di una regolazione dei processi algoritmici* data protection-driven. – 12. *Linee portanti e modelli di riferimento per una regolazione degli algoritmi*.

1. *Processi automatizzati di decisione: un primo inventario*

La varietà di contesti in cui sono oggi impiegati processi automatizzati di decisione, i diversi obiettivi che perseguono e le enormi proporzioni del loro impatto sulla vita delle persone possono essere efficacemente illustrati attraverso una carrellata di esempi e di casi reali, tratti, oltre che dalla cronaca e dai resoconti della stampa, da una letteratura fiorente, sia scientifica sia di divulgazione. Il livello di maturazione della riflessione al riguardo è a sua volta testimoniato dalla ricchezza delle analisi, ormai di rado ammaliate dalle potenzialità di Big Data e intelligenza artificiali combinati, e più spesso inclini a guardare cautamente, se non in senso apertamente critico, a una tecnologia in espansione.[1]

La rassegna è volutamente ricca, multiforme e caotica, proprio per restituire l'idea della proliferazione degli algoritmi in molteplici ambiti di vita; uno scomponimento e una classificazione, essenziali per individuare

[1] Fin da subito è opportuno menzionare, per la profondità dell'esplorazione e la competenza esperta da cui muove, C. O'NEIL, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*, New York, 2016.

le tecniche di inquadramento e regolazione del fenomeno, saranno presentati più avanti.

Nell'esperienza di Amazon i consigli di lettura proposti da algoritmi computerizzati sono risultati molto più efficaci – idonei cioè a trasformare la pubblicità relativa in un vero acquisto – di esperti impiegati *ad hoc* per decifrare le preferenze dei clienti.[2]

In un mercato sviluppato delle assicurazioni sanitarie come quello statunitense, al fine di determinare il prezzo della polizza sulla base del rischio individuale si è pensato di sperimentare la seguente alternativa: sottoporre i potenziali clienti a test sanitari, che implicano almeno un prelievo di sangue, con aggravio di costi per le assicurazioni e disagi personali per i clienti, oppure consentire l'uso di dati sullo stile di vita – estratti di carte di credito, acquisti effettuati, siti internet visitati, quantità di ore davanti alla televisione – per stabilire tramite un processo automatizzato di calcolo la stessa esposizione al rischio di patologie come pressione elevata, diabete o depressione.[3]

Sempre in ambito assicurativo, la polizza di responsabilità civile automobilistica può essere costruita sulla base di informazioni acquisite in tempo reale, grazie alla presenza di una sensoristica avanzata sul veicolo: in questo modo il rischio di incidenti, e quindi l'importo del canone, sono correlati al tipo di strade che l'utente percorre abitualmente, ai tempi e allo stile di guida, e non semplicemente a dati costanti come l'età e i record pregressi.

Farecast è un applicativo sviluppato da Oren Etzioni per predire la salita o la discesa dei prezzi dei biglietti aerei su moltissime tratte statunitensi. Lo aveva acquistato Microsoft, integrandolo nel proprio motore di ricerca Bing, ma ha dovuto cessarne l'operatività quando i costi dei dati necessari per alimentarlo sono divenuti insostenibili (più precisamente allorché ITA Software, un network per prenotazioni aeree che forniva i dati a

---

[2] M. BHASKAR, *In the age of the algorithm, the human gatekeeper is back*, *The Guardian*, 20 September 2016, <https://www.theguardian.com/technology/2016/sep/30/age-of-algorithm-human-gatekeeper>.

[3] V. MAYER-SCHÖNBERGER – K. CUKIER, *Big Data. The essential guide to work, life and learning in the age of insight*, London, 2013, 56 s.

Farecast, viene acquistato da Google).[4] Decide.com nasce come una piccola start-up con un obiettivo simile, predire l'andamento dei prezzi, applicato tuttavia non alle tariffe aeree ma a prodotti venduti *on line* e, in particolare, dispositivi digitali come cellulari, macchine fotografiche, televisori a schermo piatto. La diffusione di sistemi automatizzati per fissare i prezzi e la mole enorme di dati analizzati hanno condotto a previsioni accurate nel 77% dei casi e hanno permesso ai consumatori che le hanno seguite di risparmiare fino a 100 dollari per prodotto.[5] Sempre negli Stati uniti FlightCaster pronostica con almeno quattro ore di anticipo sulla partenza prevista la probabilità di ritardo di un aereo, analizzando ogni volo dei precedenti dieci anni in combinazione con le informazioni sulle condizioni metereologiche presenti e passate e con una miriade di altri dati.[6]

La diffusione di oggetti intelligenti amplifica le possibilità di raccogliere dati e di usarli per gli scopi più vari. La c.d. *Internet of things* non soltanto promette di semplificarci la vita, circondandoci di cose che svolgeranno compiti di routine, liberando il nostro tempo per attività più amene, ma costituisce altresì uno strumento potente di *datafication*: trasforma in dato ogni interazione con l'oggetto o addirittura – è il caso dei *personal digital assistants* – osserva e registra informazioni a prescindere dall'uso diretto che se ne faccia. Questi dati possono essere sfruttati in molti modi, e non necessariamente per assumere decisioni che ci riguardano. La bambola "Hello Barbie" della Mattel, prima che la produzione venisse interrotta, veniva venduta con un software integrato capace di registrare le conversazioni intorno a sé per trasmetterle ad un'azienda californiana specializzata in IA, al fine di migliorare la pertinenza e la qualità delle risposte del giocattolo.[7] Il sensore GPS posizionato su un certo tipo di inalatore per asmatici permette all'azienda produttrice di identificare i fattori scatenanti di una crisi, come la prossimità a certe coltivazioni;

---

[4] J. COOK, *Farewell Farecast: Microsoft kills airfare price predictor, to the dismay of its creator*, *GeekWire*, April 8, 2014, <https://www.geekwire.com/2014/farewell-farecast-microsoft-kills-airfare-price-predictor-dismay-creator/>.

[5] V. MAYER-SCHÖNBERGER – K. CUKIER, *op. cit.*, 123 s.

[6] GULLIVER, *An end to agony at the gate*, *The Economist*, December 3rd 2009, <https://www.economist.com/gulliver/2009/12/03/an-end-to-agony-at-the-gate>.

[7] Sui rischi di impiego a fini di marketing personalizzato v. I.D. MANTA - D.S. OLSON, *Hello Barbie: First They Will Monitor You, Then They Will Discriminate Against You. Perfectly*, 67 *Alabama Law Review* 2015, 135.

dispositivi indossabili per registrare la qualità del sonno evidenziano differenze nella fase REM per donne e uomini; i lettori per e-book rilevano circostanze – quanto ci si sofferma su una pagina, se si ritorna ripetutamente su altre, se si abbandona un testo prima della fine – di potenziale valore per autori ed editori intenzionati a migliorare struttura e contenuti dei libri.

Sensori posizionati in ambienti, pubblici o privati, che raccolgono un flusso continuo di informazioni, per la maggior parte di carattere non personale, dischiudono grandi possibilità per le municipalità, le pubbliche amministrazioni, le imprese che si dotino di sistemi di accesso e di trattamento automatizzato: al fine di razionalizzare la manutenzione delle strade, migliorare il traffico urbano ed extraurbano e con ciò contribuire alla riduzione delle emissioni nocive, consentire risparmio di carburante ed evitare incidenti; possono incorporare meccanismi di allerta e segnalare la necessità di interventi di riparazione o permettere di individuare la componentistica difettosa in molti prodotti.[8] La discussione che si sviluppa intorno al concetto di *smart cities* riferisce delle moltissime opportunità che le tecniche di *data analysis* e di *data mining* offrono per una trasformazione innovativa e intelligente delle realtà urbane.[9]

Il mondo del commercio *on line* è esposto a infinite forme di sfruttamento dei dati che ogni utente dissemina nella rete ogni qualvolta effettua una ricerca finalizzata ad un acquisto. Fenomeni come la *price discrimination*, in cui i prezzi degli acquisti sono differenziati, a livello individuale o più spesso di gruppo, in base ad elementi come la propensione agli acquisti, la fedeltà a un certo marchio, o ancora il tipo di computer o di sistema operativo dal quale si effettua la ricerca, sono tecnicamente realizzabili anche se al riguardo vi è solo una limitata evidenza aneddotica.

---

[8] Una recente mappatura della "*sensor society*" è nel rapporto del Ratheanu Institute: D. SNIJDERS *et al.*, *Citizens and sensors – Eight rules for using sensors to promote security and quality of life*, The Hague, 2020.

[9] K. FINCH , O. TENE, *Smart Cities: Privacy, Transparency, and Community*, in *Cambridge Handbook of Consumer Privacy*, edited by E. Selinger, J. Polonetsky, O. Tene, Cambridge, 2018, 125 ss.

Il mercato del credito rappresenta una vera e propria miniera di spunti sul funzionamento degli algoritmi e sull'elevato margine di errore che presentano.[10] Parallelamente allo sviluppo delle analisi computerizzate per asseverare l'affidabilità degli aspiranti a un finanziamento, muta anche la struttura stessa del mercato: accanto agli intermediari tradizionali che integrano le nuove tecnologie nella loro organizzazione, emergono nuovi operatori che possono sfruttarle per proporre ai clienti soluzioni altamente personalizzate e raggiungere soggetti finora esclusi dall'accesso al credito.[11]

Il sistema giudiziario statunitense fa ampio uso di algoritmi, per predire il rischio di recidiva e quindi stabilire se rilasciare sotto cauzione o tenere in carcere un imputato dopo l'arresto e durante il processo, come nel noto caso del software COMPAS, oppure per determinare la durata della pena. In Pennsylvania un software viene utilizzato per decidere quali chiamate a un *help center* che denunciano abusi su minori debbano essere considerate con maggiore urgenza perché presentano un più alto livello di rischio.[12] PredPol è una start-up californiana che ha sviluppato un software per determinare, attraverso l'analisi di dati storici sui crimini occorsi negli anni precedenti, dove sia più probabile che essi si verifichino in modo da evidenziare di ora in ora sui video in uso alle forze di polizia un'area grande come due campi da calcio da sottoporre a pattugliamento. Strumenti di questo tipo, in funzione presso numerosi dipartimenti negli Stati Uniti, sembrano possedere efficacia deterrente e scoraggiare la commissione di reati;[13] in certa misura si stanno diffondendo anche in Italia: ad esempio, un progetto congiunto tra la Questura di Pisa e il CNR è rivolto a mettere a punto un'applicazione da

---

[10] Si veda l'istruttivo cap. 8, dal titolo, "*Collateral Damage: Landing Credit*", del libro di C. O'NEIL citato alla nt. 1.

[11] Per una messa a punto dei rapporti tra il c.d. Fintech e il trattamento di dati personali sia consentito rinviare a E. PALMERINI *et al.*, *Il* Fintech *e l'economia dei dati: profili civilistici e penalistici*, Quaderni Consob Fintech n. 2, Roma-Milano, 2018.

[12] R. COURTLAND*, The bias detective*, in *Nature*, 2018 (558), 357 ss.

[13] Ne riferiscono ampiamente il cap. 5, "*Civilian Casualties: Justice in the Age of Big Data*", del volume di C. O'NEIL; V. MAYOR-SCHÖNBERGER, K. CUKIER, *op.cit.*, 158 s.

installare sui display delle auto di polizia, che segnalerebbe in tempo reale le zone più esposte a rischio per piccoli crimini.[14]

Molto discussa, anche tra gli operatori pratici del diritto,[15] è la possibilità di impiegare algoritmi nel contesto del processo civile;[16] mentre è già realtà l'utilizzo da parte di molti studi legali di software come Ross di IBM, che servono per determinare le probabilità di successo di una causa, consigliare i clienti e valutare se concedere il patrocinio.[17]

Ancora, le imprese private usano sistemi di punteggio basati su calcoli algoritmici per molteplici scopi: selezionare e ordinare le domande di impiego ricevute; valutare a chi concedere un prestito; individuare il target di clientela a cui rivolgersi con pubblicità personalizzata; predire la popolarità di un film e quindi il ritorno degli ingenti investimenti in programma.[18] Talvolta l'algoritmo interviene direttamente nella conclusione del contratto come nella negoziazione algoritmica diffusa sui mercati finanziari e nella sua variante dell'*high-frequency trading*.

Anche le amministrazioni pubbliche impiegano già algoritmi per assumere decisioni, come scegliere chi sottoporre un controllo fiscale,[19]

---

[14] Ne riporta la notizia l'articolo di M. NERI, *Ecco la sicurezza 4.0: un algoritmo per prevedere rapine, furti e aggressioni*, *Il Tirreno*, 13 aprile 2017 <http://iltirreno.gelocal.it/pisa/cronaca/2017/04/13/news/rapine-aggressioni-e-furti-un-algoritmo-per-prevederli-1.15192360>.

[15] Cfr. C. CASTELLI, D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Questione giustizia*, 4/2018, 153 ss.

[16] Si dedica all'esplorazione di questa ipotesi la prima sessione del convegno su "Decisione robotica", terzo dei Seminari 'Leibniz' per la teoria e la logica del diritto, tenutosi a Roma presso l'Accademia dei Lincei il 5 luglio 2018, i cui Atti sono ora raccolti nel volume omonimo, a cura di A. Carleo, Bologna, 2019. Più di recente la questione è stata oggetto del Convegno su "Prospettive algoritmiche della giustizia civile", organizzato dalla Corte di Appello di Firenze, in collaborazione con la Scuola Superiore della Magistratura, il Consiglio dell'Ordine degli Avvocati e la Camera civile di Firenze, 13 novembre 2019.

[17] Ne riferisce M. MAUGERI, *I robot e la possibile "prognosi" delle decisioni giudiziali*, in *Decisione robotica*, a cura di A. Carleo, Bologna, 2019, 159 ss.

[18] F.M. SIMON, R. SCHROEDER, *Big Data Goes to Hollywood: The Emergence of Big Data as a Tool in the American Film Industry*, in *Second International Handbook of Internet Research*, a cura di J. Hunsinger, M. Allen M., L. Klastrup, Dordrecht, 2019.

[19] M. GABANELLI, A. MARINELLI, *Tasse, ecco come un algoritmo difettoso ti fa pagare più di quanto devi*, in *Corriere della Sera*, Dataroom, 11 novembre 2019.

a chi concedere un visto, a quale sede destinare gli insegnanti[20] oppure quali docenti debbano essere licenziati, sulla base dei cattivi risultati agli esami dei loro allievi, per elevare lo standard qualitativo di una certa scuola.[21]

2. *Le classificazioni possibili. Dati personali, dati non personali, dati anonimi*

I processi appena descritti, in un elenco vario e disorganizzato, possono essere classificati secondo plurimi criteri: a seconda che si svolgano in ambito pubblico o privato; che impieghino dati personali oppure dati non personali (tali in origine ovvero resi anonimi successivamente); che conducano a vere e proprie decisioni, con un impatto sul singolo o su gruppi di individui, oppure si limitino ad effettuare predizioni (sul prezzo di un biglietto aereo o sul rischio che il volo ritardi, sul probabile sviluppo di un'epidemia di influenza, ecc.) senza incidenza sui singoli.

Un primo spartiacque riguarda dunque la natura delle informazioni che alimentano il processo algoritmico: se si tratta di dati personali, si rientra nell'ambito di applicazione della disciplina in materia di *privacy*. Tuttavia, larga parte dei dati impiegati nella casistica rappresentata nel paragrafo precedente non sono dati personali, ossia non sono riferibili a un individuo determinato o determinabile. Si tratta, ad esempio, di dati raccolti da sensori collocati in aree pubbliche, di dati acquistati o reperibili nella rete, di ogni informazione, ceduta o liberamente accessibile, che non riguarda il singolo ma circostanze ambientali, metereologiche, economiche.

La disciplina che interessa i dati non personali è allora un recente Regolamento europeo,[22] rivolto ad agevolare il flusso dei medesimi all'interno del mercato unico e a consentirne la conservazione e il trattamento da parte di imprese e amministrazioni pubbliche senza vincoli di localizzazione all'interno di uno Stato membro. Lo scopo di questo provvedimento, che costituisce una delle azioni incluse nella *Digital*

[20] TAR Lazio,10 settembre 2018, n. 9227; Cons. St., 8 aprile 2019, n. 2270.

[21] Questa discussa valutazione sperimentata dalla città di Washington DC è tra le esperienze descritte da C. O'NEIL, *op. cit.*, *passim*.

[22] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 o*n a framework for the free flow of non-personal data in the European Union*.

Single Market Strategy, è infatti quello di contribuire allo sviluppo di un'economia digitale competitiva, in larga misura basata sulla circolazione dei dati.

Questo primo criterio di classificazione dei processi automatizzati basato sulla natura dei dati trattati – criterio, come vedremo, decisivo – non è privo di ambiguità, che si appuntano anzitutto sulla possibilità di stabilire con certezza e in maniera definitiva l'appartenenza all'uno o all'altro ambito. Se infatti la differenza tra dato personale e dato non personale è chiara in astratto, in concreto è più difficile segnare un confine netto e, quindi, identificare con assoluta certezza il regime applicabile e le tutele attive. Ciò accade sia perché la possibilità tecnica di attingere una anonimizzazione autentica è sempre stata in dubbio,[23] sia per l'effetto combinato di *big data* e intelligenza artificiale, che accresce enormemente le probabilità di risalire da un dato anonimo all'attribuzione a un soggetto determinato. Non mancano gli esempi in questo senso, come dimostra l'esperimento compiuto da due ricercatori dell'Università di Austin nel Texas, [24] in occasione del rilascio da parte di Netflix dei dati, appositamente resi anonimi, relativi alle registrazioni effettuate per noleggio di film, corrispondenti a circa mezzo milione di utenti. L'operazione è legata a un concorso, con un ricco premio in denaro, destinato al team di sviluppatori che sia in grado di migliorare almeno del 10% il sistema di raccomandazioni di cui Netflix si serve. Comparando questi dati con altre informazioni presenti su siti aperti, i ricercatori dimostrano come sia possibile identificare alcuni clienti e certe loro attitudini private, come le opinioni politiche e l'orientamento sessuale, sulla base delle recensioni postate e delle date in cui erano state fatte. La vicenda ha originato anche un'azione di classe presso una District Court della California da parte di alcuni utenti di Netflix, tra i quali una madre, omosessuale non dichiarata, ritenutasi "tradita" dall'identificazione delle sue preferenze cinematografiche.[25]

---

[23] P. OHM, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 *UCLA Law Review* 2010, 1701 ss. In tema v. anche I.S. RUBINSTEIN, W. HARTZOG, *Anonymization and Risk*, 91 *Washington L. Rev.* 2016, 703 ss.

[24] A. NARAYANAN, V. SHMATIKOV, *Robust De-anonymization of Large Sparse Datasets*, *IEEE Symposium on Security and Privacy*, Oakland (CA), 2008, 111 ss.

[25] Per un breve resoconto v. B. SCHNEIER, *Why 'Anonymous' Data Sometimes Isn't*, Wired, 12.12.2017, <https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/#>.

Quando un Internet service provider come AOL (America on Line) mette a disposizione della comunità dei ricercatori migliaia di dati anonimi relativi alle ricerche compiute sul suo motore di ricerca,[26] il New York Times mostra come sia possibile associare il numero identificativo che connotava ricerche generiche come "uomini singoli sui 60 anni" e "giardinieri a Lilburn", ad una signora di quell'età che abitava nella stessa cittadina della Georgia.[27]
Più in generale, la tutela offerta dall'anonimato tende a sfumare in una realtà iperconnessa, e nel contesto delle *smart cities* o degli *smart environments* può risultare illusoria: tracce apparentemente di scarso significato come il consumo di energia monitorato da contatori automatici o dati di geolocalizzazione sono collegabili a individui determinati con un'approssimazione vicina alla certezza. In tal modo, possono rivelare abitudini consolidate od occasionali deviazioni dalla routine che hanno valore informativo e attenere alla vita intima delle persone.[28]
Infine, accanto alle difficoltà pratiche di ottenere un anonimato autentico e irreversibile, c'è il rischio di un impoverimento dell'utilità del dato che viene reso anonimo. Si tratta di un problema ben conosciuto nell'ambito della ricerca sanitaria, dove l'impossibilità di associare nuove informazioni ai dati già raccolti riduce le opportunità di riuso delle informazioni a scapito dello sviluppo ulteriore della conoscenza.

3. Segue. *Il soggetto che effettua il trattamento*

Quando il trattamento algoritmico si basa sull'impiego di dati personali, è applicabile la disciplina in materia, oggi contenuta pressoché interamente nel Regolamento europeo n. 679/2016 (GDPR).

---

[26] L'obiettivo era consentire analisi potenzialmente produttive di risultati interessanti. Proprio la digitazione su Google di domande relative a sintomi quali febbre e raffreddore ha successivamente consentito di mettere a punto un modello per predire l'espansione di epidemie di influenza: J. GINSBURG *et al.*, *Detecting Influenza Epidemics Using Search Engine Query Data*, in *Nature* 457 (2009), 1012 ss.
[27] M. BARBARO, T. ZELLER, *A Face Is Exposed for AOL Searcher No. 4417749*, The New York Times, 2006, <https://www.nytimes.com/2006/08/09/technology/09aol.html>.
[28] L. EDWARDS, M. VEALE, *Slave to the Algorithm? Why a Right to an Explanation is Probably Not the Remedy You Are Looking For*, in 16 *Duke Law & Tech. Rev.* 2017, 34 s.

Ulteriori articolazioni sono tuttavia necessarie. Se ad avvalersi del processo automatizzato di decisione è un soggetto pubblico, la disciplina subisce una torsione particolare: occorre, come avviene in generale per i trattamenti in ambito pubblico, una base normativa, nel diritto interno o nel diritto europeo, che legittimi il trattamento; in tal caso, sarà la stessa fonte a prevedere le misure di tutela e di garanzia degli interessati (art. 22, comma 2, lett. b), ferma restando l'operatività dei principi e delle regole che costituiscono l'architettura del Regolamento.

Viceversa, per quelle forme particolari di trattamento legate all'accertamento e al perseguimento di reati e, più in generale, che hanno scopi di tutela della sicurezza pubblica e di prevenzione delle minacce che la riguardano, opera un'esenzione esplicita, che rende inapplicabile il GDPR (art. 2, comma e, lett. d) e, in suo luogo, rinvia alla Direttiva 2016/680/UE che ha specifico riguardo a questa materia.[29]

4. *Intelligenza artificiale e tutela dei dati*

La regolazione giuridica delle applicazioni dell'intelligenza artificiale ha finora preso corpo esclusivamente nell'ambito della protezione dei dati personali. Il GDPR (come già in precedenza la direttiva 95/46) costituisce pertanto l'unico momento di emersione delle tutele rispetto alle tecniche di analisi dei dati basate sul calcolo algoritmico.
Invero, vi sono svariate iniziative in ambito accademico e in contesti istituzionali volte ad elaborare linee guida di ordine etico e principi di disciplina giuridica per l'IA. Tra di esse, numerosissime,[30] è opportuno ricordare almeno la strategia adottata dalla Commissione europea, che ambisce a integrare l'elaborazione degli esperti con la discussione pubblica, cui sottopone versioni provvisorie delle varie analisi, al fine di

[29] Sui rischi di decisioni automatizzate nell'ambito del processo penale (e, prima ancora, nell'attività di prevenzione dei reati), nonché sulle pallide difese erette al riguardo dalla disciplina europea e nazionale si veda tuttavia B. GALGANI, *Giudizio penale,* habeas data *e garanzie fondamentali*, in *Arch. Pen.*, 2019, 20 ss.
[30] Un lavoro recente conta, allo stato, almeno 84 documenti di "AI Ethics": B. MITTELSTADT, *Principles alone cannot guarantee ethical AI*, in *Nature Machine Intelligence*, 1(2019), 501 ss.

promuovere una regolazione attenta e non ridondante del fenomeno.[31]
Tutti questi documenti, tuttavia, costituiscono per il momento essenzialmente esempi di *soft law*, che invitano all'autodisciplina i vari soggetti implicati nello sviluppo teorico e applicativo dell'IA e propongono modelli di riferimento cui attenersi. Da un punto di vista formale, il valore di tali schemi è soltanto persuasivo e l'adesione da parte degli attori privati o istituzionali è libera e priva di incentivi che non siano quelli reputazionali. Lo stesso contenuto si ferma spesso ad una articolazione di principi generalissimi, che possono essere applicati in modi diversi e anche contraddittori, e risente talvolta della mancanza di un apporto di tipo prettamente regolatorio o, per converso, non coniuga la riflessione etica con la valutazione circa la fattibilità tecnica delle soluzioni accolte.

5. *Le decisioni automatizzate nel GDPR*

Le strategie di tutela rispetto al trattamento algoritmico di dati personali rintracciabili nel Regolamento privacy sono essenzialmente due: la prima è generalizzata e riguarda i dati "in entrata"; la seconda è invece specifica e concerne più precisamente le vere e proprie decisioni automatizzate. Quando dati personali sono processati in maniera automatica, la disciplina in questione si applica nella sua interezza. Si è infatti di fronte a un comune trattamento e la fattispecie richiama pertanto tutti i principi e le regole che ne disegnano i presupposti di liceità. Non è possibile entrare nel merito di ciascuno, ma è opportuno segnalare quelli che presentano maggiore rilevanza per il tema considerato ovvero che possono entrare in attrito con le modalità di svolgimento di analisi su grandi quantità di dati e che impiegano algoritmi c.d. black box.
Uno di essi è il principio di minimizzazione dei dati, secondo cui i dati sono "adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati" (art. 5, comma 1, lett. c), che può

[31] Avviata con la Communication *Artificial Intelligence for Europe*, COM(2018) 237, ha dato luogo all'emanazione di un *Coordinated Plan on Artificial Intelligence*, COM(2018) 795, e all'istituzione di un HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE che ha prodotto (dopo una prima versione provvisoria) le *Ethics Guidelines for Trustworthy AI*, 8.2.2019. Alla Communication *Building Trust in Human-Centric AI*, COM(2019) 168, ha fatto seguito l'adozione di uno White Paper *On Artificial Intelligence – A European approach to excellence and trust*, COM(2020) 65, 19.2.2020.

risultare incompatibile con la natura delle tecniche di analisi dei *big data*,[32] la cui stessa denominazione simboleggia la moltitudine di informazioni sottoposte a trattamento. Il principio ha invero una portata relativa, poiché la dimensione dei dati passibili di trattamento dipende dalle finalità di quest'ultimo, alla cui stregua si identificheranno pertinenza e necessità. Una contraddizione tuttavia permane, poiché dati in apparenza eccedenti rispetto alla finalità dichiarata dal titolare potrebbero *a posteriori* rivelarsi utili, se non essenziali, al raggiungimento degli obiettivi che il trattamento ha di mira. Se il vincolo rappresentato dal principio di minimizzazione viene inteso in termini molto rigidi, ciò potrebbe tradursi in una perdita di efficienza del processo a scapito dello stesso soggetto interessato dal trattamento. Lo dimostra, ancorché su base aneddotica, questo esempio: ZestFinance è una compagnia che concede piccoli prestiti, a breve termine, a persone che hanno un merito creditizio modesto che li escluderebbe dall'accesso ai canali tradizionali di finanziamento. La ragione dell'ampliamento della platea dei destinatari dipenderebbe dal fatto che l'impresa, anziché affidarsi alle tecniche consuete che analizzano una serie molto limitata di elementi per determinare la affidabilità del debitore, come precedenti ritardi di pagamento o episodi di insolvenza, dispone di una tecnologia che impiega un numero elevato di variabili, anche apparentemente inconferenti, che però hanno dimostrato in concreto una grande capacità predittiva.[33]

In casi simili, i principi di minimizzazione e di pertinenza si pongono in antitesi alla logica di funzionamento dell'algoritmo di calcolo, là dove proprio la vasta mole di dati presi in considerazione e l'apparente eccedenza rispetto ai fini del trattamento dischiudono vantaggi per l'interessato. La complessità e la numerosità delle variabili che un sistema basato sui *big data* può osservare permette infatti di affinare i processi di decisione, rendendo queste ultime più granulari. In tale guisa, sarebbero evitati anche quegli esiti discriminatori che dipendono dalle classificazioni più grossolane permesse da processi di stampo tradizionale: l'inserimento dell'interessato in una categoria "a rischio" (si

---

[32] O. TENE – J. POLONETSKY, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, 11 *J. on Telecomm. & High Tech. Law* 2013, 362; T.Z. ZARSKY, *Incompatible: The GDPR in the Age of Big Data*, 47 *Seton Hall L. Rev.* 2017, 1009 ss.

[33] V. MAYER-SCHÖNBERGER, K. CUKIER, *op. cit.*, 46 s.

tratti delle probabilità di recupero di un credito o della sicurezza di un viaggio aereo) non può essere vinta da un'indagine più sottile e sofisticata che fa uso di informazioni atte a sconfessare la prima impressione.[34]

Altro momento di attrito riguarda il principio di finalità, che richiede di indentificare in anticipo gli scopi del trattamento, portarli a conoscenza dell'interessato e attenervisi nelle successive operazioni (art. 5, comma 1, lett. b); principio solo in parte mitigato dall'art. 6, comma 4, che ammette l'impiego degli stessi dati già raccolti per una finalità diversa da quella iniziale, purché compatibile con questa. Nel mondo dei *big data*, tuttavia, la raccolta delle informazioni avviene spesso prima, o molto tempo prima, che queste siano destinate a uno specifico uso, il quale può anche rivelarsi *ex post*,[35] cioè in guisa di conferma di una mera ipotesi di lavoro, quando il trattamento algoritmico evidenzia un'utilità dell'opera di elaborazione dei dati medesimi. Mentre la *data protection* elegge il principio di finalità a canone centrale della liceità e correttezza di un trattamento, i sistemi di *machine learning* richiedono una continua ridefinizione degli scopi dell'analisi: ad esempio, dati raccolti in occasione dell'acquisto *on line* di libri sono usati per perfezionare gli algoritmi che generano consigli di lettura, poi impiegati da parte della stessa piattaforma di vendita o ceduti ad altri come indicatore di preferenze anche per altri settori di mercato.[36] Queste caratteristiche, a loro volta, precludono di avvalersi della legittimazione data dal consenso, che deve avere ad oggetto una finalità determinata in anticipo.

Altri principi, viceversa, mantengono rilevanza anche nell'ambiente digitale grazie alla loro intrinseca flessibilità. Così, la correttezza può svolgere una funzione di garanzia in tutte le ipotesi cui non si applica la disciplina specifica prevista per le decisioni interamente automatizzate.

---

[34] Evitando, ad esempio, che una persona con un passaporto di un Paese arabo che sale su un aereo con un biglietto di sola andata, pagato in contanti, sia necessariamente trattenuto ai controlli se altre informazioni rendono altamente improbabile che sia un terrorista: propongono questo esempio V. MAYER-SCHÖNBERGER, K. CUKIER, *op. cit.*, 160 s.

[35] "Chi colleziona dati con la rete a strascico non ci può dire per quale finalità li sta affastellando, perché lo ignora anche lui": G. DE MINICO, *Big Data e la debole resistenza delle categorie giuridiche. Privacy e lex mercatoria*, in *Diritto pubblico*, 2019, 92. Critiche molto puntuali alla "*purpose limitation*" sono mosse da T.Z. ZARSKY, *op. cit.*, 1005 ss.

[36] L. EDWARDS, M. VEALE, *op. cit.*, 32.

Questo ruolo è cruciale, poiché, come spiegheremo, la fattispecie riguardante le decisioni automatizzate è costruita in termini restrittivi e avrà necessariamente un orizzonte operativo limitato. Tuttavia, tutte le forme di trattamento, anche quelle che non conducono a decisioni (si pensi al caso del marketing personalizzato) ovvero che includono passaggi non automatici, saranno soggetti a tale principio che, unitamente a quello di *accountability*, dovrebbe informare il comportamento delle imprese a criteri di lealtà. Altrettanto vale per il requisito dell'esattezza dei dati (art. 5, comma 1, lett. d), che ne impone il continuo aggiornamento, e si pone in corrispondenza ideale con la logica di funzionamento degli algoritmi, che non possono produrre esiti validi se non a partire da dati in entrata veri.

6. *Dal controllo sulle informazioni "in entrata" alle tutele rispetto alla decisione. La "profilazione"*

La disciplina generale del GDPR investe anzitutto la fase della raccolta e della successiva conservazione dei dati, dettandone le condizioni di legittimità. Alcune previsioni puntuali sono invece dirette a regolare le decisioni automatizzate, ossia le ipotesi in cui le informazioni servono ad alimentare il processo di calcolo che avrà come esito una decisione riguardante l'interessato. Queste norme compongono una strategia di tutela più specifica, articolata essenzialmente in due momenti, che trovano collocazione rispettivamente negli artt. 13, 14, 15 e nell'art. 22.

Quest'ultima è la disposizione principale che riproduce, con alcune modifiche, il contenuto dell'art. 15 della Direttiva 95/56. Senonché, alla sostanziale continuità con il regime previgente sul piano normativo fa da contrappunto nella realtà la diffusione esponenziale dei processi automatizzati di decisione,[37] che trasforma l'unica forma di regolazione del fenomeno in un dispositivo da analizzare con estrema attenzione nella sua portata teorica e nella sua applicazione pratica.

L'art. 22, quale regola generale, contiene tre prescrizioni, che possiamo così semplificare: 1) un trattamento completamente automatizzato è vietato quando la decisione cui conduce ha un'importanza significativa

[37] M. BRKAN, *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond*, in *International Journal of Law and Information Technology*, 27(2), 2019, 95 s.

per l'interessato; 2) a tale divieto si fa eccezione in tre casi; 3) quando un trattamento completamente automatizzato può essere effettuato nelle tre situazioni eccezionali, devono essere applicati alcuni meccanismi di garanzia per l'interessato.

La complessità del testo è evidente già ad una prima sommaria disamina; addentrandosi nell'analisi ci si avvede come la norma, cruciale nella sua funzione di irreggimentare un fenomeno insidioso, sia densa di problemi interpretativi.

Un primo nodo da sciogliere riguarda il rapporto tra decisioni automatizzate e profilazione:[38] nella formulazione dell'art. 22, quest'ultima sembra costituire una *species* delle prime. Si dubita tuttavia che possano aversi attività di trattamento dei dati personali, diverse dalla profilazione, che diano luogo a decisioni automatizzate: in altre parole, se vi è un trattamento automatizzato di dati personali, questo implica sempre e comunque una profilazione, anche se l'obiettivo finale può tendere a un risultato ulteriore rispetto ad essa; ne costituisce insomma un passaggio necessario. Viceversa, processi automatizzati di decisione che non comportano il trattamento di dati personali e, pertanto, non includono nemmeno una profilazione si collocano al di fuori della portata dell'art. 22. Ad esempio, un applicativo, di sempre più comune utilizzo negli studi legali, che valuta il merito di una controversia e predice l'esito probabile in giudizio può indurre a negare assistenza legale a un cliente che l'ha richiesta, ma l'operazione stessa, in sé e per sé considerata, sfugge all'ambito di applicazione del GDPR poiché non tratta dati personali.

Se proprio vogliamo dare un significato alla formulazione della norma che distingue le due ipotesi, dobbiamo circoscrivere l'area dei trattamenti automatizzati che non costituiscono profilazione a quei trattamenti che si servono di dati personali per valutare una persona fisica, ma al fine di compiere analisi o previsioni su aspetti diversi da quelli espressamente menzionati dalla definizione dell'art. 4. Questi ultimi non sarebbero

[38] Di quest'ultima può essere utile riportare la definizione: "qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica" (art. 4, n. 4).

allora esemplificazioni di un insieme più ampio, ma un'elencazione tassativa di cosa costituisce profilazione.

7. *Il divieto di trattamenti esclusivamente automatizzati: natura e portata della prescrizione*

Secondo quanto appena detto, la prima prescrizione ricavabile dall'art. 22 è costituita dal divieto di trattamenti esclusivamente automatizzati. Questa proposizione richiede tuttavia una precisazione e un'integrazione.

Quanto alla precisazione, la norma non si esprime testualmente in termini di divieto, bensì statuisce che "l'interessato *ha il diritto* di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato…". Il significato distillato è, pertanto, una delle interpretazioni alternative aperte dal dettato letterale, quella più congruente con l'intero sistema di tutele e maggiormente protettiva per l'interessato. Se, infatti, la previsione attribuisse semplicemente il diritto di opporsi al trattamento (al pari delle facoltà di accesso, di cancellazione, di rettifica contenute nel GDPR), l'attivazione delle garanzie che presidiano i processi automatizzati di decisione sarebbe solo eventuale, in quanto rimessa all'iniziativa del singolo. Con la conseguenza che una decisione automatizzata potrebbe essere assunta non solo nei casi previsti dal comma 2 della norma, che già fanno eccezione alla regola generale, ma nei quali operano opportune salvaguardie, bensì ogni qualvolta l'interessato manchi di opporsi e, dunque, persino in assenza di tali garanzie.[39] L'opinione prevalente – sostenuta anche dall'Art. 29 Working Party –[40] è pertanto nel senso che la norma ponga un divieto generalizzato di decisioni esclusivamente automatizzate, a prescindere da una richiesta dell'interessato.

La regola deve essere integrata da alcune specifiche condizioni che ne delimitano il perimetro di operatività: anzitutto, a essere preclusi sono i processi interamente svolti tramite un sistema automatico senza l'intervento umano; in secondo luogo, il divieto concerne soltanto i trattamenti che conducano a una decisione individuale, riguardante cioè

---

[39] M. BRKAN, *op. cit.*, 98 s.

[40] ART. 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decisionmaking and Profiling for the purposes of Regulation 2016/679*, 6 February 2018, 19 s.

un singolo interessato; quest'ultima, infine, deve avere un impatto giuridico o comunque significativo sul soggetto.

Quanto all'intervento umano, esso deve essere sostanziale,[41] mentre non sarebbe sufficiente a superare il divieto un puro recepimento formale della decisione che non la sottoponga a un controllo di merito o una validazione meramente procedurale. Alcuni aspetti problematici nella lettura della norma emergono già a questo livello. Una prima questione attiene alla concreta praticabilità del limite che dispone: decisioni interamente automatizzate, come la fissazione di un prezzo diversificato per acquisti *online*, difficilmente potrebbero essere oggetto di un controllo umano, dato il processo pressoché istantaneo in cui si inseriscono. In questi casi – posto che si tratti di decisioni automatizzate riferite a un singolo individuo e, dunque, rientranti nel cono applicativo dell'art. 22 – la garanzia apprestata dalla norma non appare pienamente centrata sull'obiettivo di protezione: da un lato, considerarle vietate finirebbe per precludere l'uso di applicazioni che possono ridondare in un vantaggio per l'interessato (ad esempio, per incoraggiare l'acquisto di un soggetto normalmente poco propenso a comprare beni o servizi su piattaforme elettroniche, gli si offre un prezzo molto vantaggioso); da un altro lato, una protezione effettiva può derivare soltanto da una regolazione dell'algoritmo che eviti effetti discriminatori sul singolo.

Un secondo problema riguarda invece l'affidamento che offre all'interessato l'intervento umano: si tratta cioè di comprendere se la garanzia prevista abbia carattere puramente formale oppure sia in grado di operare come tutela di tipo sostanziale. Poiché la questione si lega essenzialmente all'intelligibilità dell'algoritmo e, dunque, ridonda in un profilo di trasparenza, su di essa torneremo più avanti.

Escluse dall'ambito di applicazione della norma, per quanto precisato, sono poi le decisioni che non hanno natura esclusivamente individuale, bensì si proiettano su un gruppo o una collettività.[42] Esse, invero, possono

---

[41] ART. 29 DATA PROTECTION WORKING PARTY, *op. cit.*, 20 s.

[42] Per una prima messa a punto del rapporto tra i processi di analisi e di decisione basati sui *big data* e la dimensione individuale della *data protection* cfr. A. MANTELERO, *Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*, in *Computer Law & Security Review*, 32 (2016), 238 ss.

avere un notevole potenziale discriminatorio, come ad esempio l'offerta di prezzi differenziati ad acquisti on line in ragione della zona in cui si deve effettuare la consegna; oppure le operazioni di polizia effettuate più di frequente in un determinato quartiere sulla base dei suggerimenti di intelligenze artificiali per la prevenzione dei crimini. Rispetto ad ipotesi del primo tipo, si propone di considerarle come un insieme di decisioni individuali, con il risultato di farle rientrare per questa via nell'ambito di applicazione dell'art. 22.[43] Lo stesso risultato potrebbe essere attinto semplicemente osservando che il processo automatizzato investe una moltitudine di interessati, tutti quelli che abitano nella zona contrassegnata da un determinato codice di avviamento postale; tuttavia, la decisione finale che su questo risultato si basa riguarda un singolo individuo, ed è raggiunta per mezzo del dato personale consistente nel suo indirizzo di residenza. Non si è pertanto al di fuori del regime operante per le decisioni automatizzate neppure intendendo queste ultime in senso stretto; ovviamente, per alcune ragioni già accennate, sarebbe comunque difficile attuare in pratica le tutele previste dalla norma.

Viceversa, sfuggono senz'altro al raggio applicativo dell'art. 22 decisioni del secondo tipo, che producono i loro effetti su tutta la popolazione che vive nel comprensorio indicato come frequente epicentro di reati. Ciò accade poiché la mobilitazione di maggiori risorse in quella zona porterà fatalmente ad una più elevata identificazione dei responsabili, con la conseguenza che la medesima sarà considerata più insicura e malfamata rispetto ad altre. La decisione è in questo caso autenticamente collettiva e la *data protection law* non offre nessun rimedio.

Infine, il terzo criterio da integrare perché si attivi la tutela principale prevista dall'art. 22 è che le decisioni abbiano un impatto giuridico o altrimenti significativo sulla persona. L'identificazione di questo secondo presupposto non è agevole, poiché si compone attraverso l'espressione doppiamente indeterminata di "*similarly significant*". Se al primo assegniamo il senso di incidenza sulle libertà e i diritti – fondamentali, sociali, che scaturiscono da un contratto – o sullo status della persona, si può individuare tipologicamente l'ambito delle decisioni che similmente colpiscono l'individuo: esse devono riguardare situazioni che pur non conferendo veri e propri diritti, rappresentano un interesse rilevante

[43] M. BRKAN, *op. cit.*, 101.

nell'ambito della vita personale, familiare o lavorativa. Seguendo anche l'indicazione del *considerando* 71, si giunge a comprendervi le decisioni riguardanti la concessione di un prestito o di una carta di credito/debito ovvero una selezione lavorativa che siano amministrate secondo criteri automatici.

La flessibilità intrinseca della seconda indicazione può consentire un'interpretazione piuttosto lata volta a estendere l'ombrello di protezione della norma. Rimangono tuttavia alcune ipotesi di collocazione incerta, come l'*advertising* personalizzato o l'invio di comunicazioni, anch'esse personalizzate, di tipo politico in occasione ad esempio di competizioni elettorali. In questi casi, nonostante il potere di manipolazione elevato che tali tecniche possiedono, è dubbio che ricorra una "decisione" vera e propria, poiché siamo di fronte a processi automatizzati che hanno come destinatario un individuo, ma non si estrinsecano in un atto di volontà. L'impatto effettivo, inoltre, difficilmente è determinabile in astratto: forme aggressive di marketing possono lasciare indifferenti la maggior parte di noi e condizionare invece pesantemente alcuni, soprattutto qualora l'informazione pubblicitaria sia specialmente diretta a sfruttare la conoscenza di vulnerabilità o dipendenze e indurre di conseguenza certe scelte di acquisto. Invero, l'effetto reale può essere stabilito soltanto *ex post*, e ciò propone due questioni ulteriori: se il divieto sia comunque applicabile di per sé oppure debba essere necessariamente attivato dall'interessato, cui spetterà allora una prima valutazione circa l'entità dell'impatto che il processo produce su di lui; in questo secondo caso, se sia sufficiente a inibire il trattamento l'esercizio del diritto di opposizione oppure occorra provare in concreto gli effetti significativi dell'attività che si vuole far cessare.

8. *Le eccezioni al divieto*

Il divieto generale di trattamenti esclusivamente automatizzati – da intendersi secondo la nozione e i limiti appena accennati – contempla tre eccezioni nelle quali un simile trattamento può essere realizzato, adottando all'uopo (almeno) le garanzie previste dalla stessa norma dell'art. 22. A questo riguardo il primo profilo problematico investe lo stesso rapporto tra regola ed eccezioni: l'estensione di queste ultime è infatti tale da portare quasi a un rovesciamento della prima.

Anche esaminate singolarmente, le tre fattispecie evidenziano alcune ambiguità interpretative. La prima riguarda l'ipotesi in cui il trattamento sia *necessario* per la conclusione di un contratto con l'interessato. Vi sono due letture "estreme" di questa previsione, in astratto plausibili, ma che devono essere escluse per ragioni logiche e sistematiche: la prima è quella che intende in termini rigidi il presupposto della necessità, con ciò praticamente svuotando pressoché integralmente lo spazio operativo dell'eccezione. Sarebbe cioè inibita la possibilità di usare un trattamento automatizzato ogni qualvolta, cioè pressoché sempre, questo possa essere sostituito da un trattamento manuale o comunque guidato dall'uomo. Ad esempio, accertare il merito creditizio ai fini della concessione di un finanziamento è certamente possibile anche senza fare ricorso a strumenti automatizzati di calcolo, quindi non può dirsi che il processo automatizzato sia strettamente necessario. La necessità non deve pertanto essere intesa come un bisogno altrimenti insuperabile, a pena di riservare ai trattamenti eseguibili in virtù di tale deroga uno spazio decisamente esiguo, se non inconsistente. All'opposto, l'eccezione potrebbe essere intesa in senso estremamente lato, cosicché la prospettiva, anche remota, di instaurare una relazione contrattuale attiverebbe l'eccezione alla regola generale. Ad esempio, il *targeting* pubblicitario ha precisamente lo scopo di preludere alla conclusione di un contratto con il destinatario, ma l'eventualità è troppo remota perché possa dirsi presente il presupposto della necessità. L'eccezione si applicherà invece quando il processo automatizzato si collochi in una fase prodromica all'instaurazione di una relazione contrattuale, rispetto alla quale si pone come essenziale.

In alcune situazioni permangono margini di incertezza, come ad esempio per le operazioni automatizzate volte a selezionare le domande per una posizione lavorativa, finalizzate unicamente a ridurre il numero dei curricula sottoposti a una rassegna più accurata oppure quello delle persone convocate per un colloquio.[44] La flessibilità, pur ridotta, del presupposto della "necessità" dell'impiego dell'automazione autorizza probabilmente a sceverare in concreto i casi in cui essa è ineludibile (perché le domande sono nell'ordine delle migliaia o è ridotto il numero

[44] Rientrano nella sfera dell'eccezione secondo ART. 29 DATA PROTECTION WORKING PARTY, *op. cit.*, 23; M. BRKAN, *op. cit.*, 104, ritiene viceversa eccessiva questa estensione.

dei valutatori) oppure costituisce un ausilio importante anche se non indispensabile o, ancora, si rivela quasi superflua.

La seconda eccezione riguarda una previsione di legge nazionale o europea che ammette la decisione automatizzata e, al contempo, è tenuta a individuare le misure di tutela applicabili.

La terza eccezione riguarda il consenso esplicito dell'interessato. In questo caso le criticità non si appuntano sulla specifica previsione che introduce tale deroga, bensì sulla valenza generale del consenso quale dispositivo di salvaguardia che dovrebbe permettere di realizzare la propria autodeterminazione informativa. I limiti teorici e pratici del modello di tutela basato sul consenso sono ormai noti per dovercisi troppo soffermare. Mi limito a richiamare le analisi, anche empiriche, che ne dimostrano la scarsa efficacia nel garantire un controllo effettivo sui propri dati personali.[45] Da tempo è stato evidenziato come sia puramente illusoria l'immagine di un individuo che, informato delle implicazioni della propria manifestazione di volontà al trattamento dei dati, la esprime consapevolmente magari selezionando tra le alternative possibili quelle che ritiene compatibili con i propri scopi. Per la mole di informazioni che sono fatte oggetto di comunicazione, per il tecnicismo e la complessità che le connota, per i *bias* cognitivi che ci affliggono e, infine, per la frequente presenza di dispositivi di semplificazione della scelta (ad esempio, la spunta di una casella che equivale ad accettazione), è stato dimostrato come il consenso non dia luogo a un reale *empowerment* del singolo.[46] Una garanzia, ancorché inadeguata a superare tutte queste criticità, è tuttavia costituita, nella specifica ipotesi che stiamo esaminando, dal predicato del consenso, che deve essere esplicito, con ciò escludendo quelle forme di manifestazione che lasciano spazio a una più debole partecipazione volitiva.

---

[45] La letteratura al riguardo è molto vasta: basti qui richiamare D. SOLOVE, *Introduction: Privacy Self-management and The Consent Dilemma*, in 126 *Harvard Law Review* 2013, 1880 ss. Tra gli studi che attingono ad evidenze anche empiriche, si segnala di recente L. GATT, R. MONTANARI e I.A. CAGGIANO, *Consenso al trattamento dei dati personali e analisi giuridico-comportamentale*, in *Nodi virtuali, legami informali: internet alla ricerca di regole*, a cura di P. Passaglia e D. Poletti, Pisa, 2017, 57 ss.

[46] Cfr. EUROPEAN DATA PROTECTION SUPERVISOR, *Privacy and Competitiveness in the Age of Big Data: The Interplay Between Data Protection, Competition Law and Consumer Protection in the Digital Economy* (Preliminary Opinion), marzo 2014, 34 s.

Conclusivamente, è opportuno svolgere una considerazione ulteriore sullo spazio operativo disegnato dalle tre eccezioni alla regola generale del divieto di trattamento esclusivamente automatizzato. Esse si compongono intorno alla dicotomia pubblico/privato: quando il trattamento in questione avviene sulla base di una previsione di legge, le garanzie per l'individuo sono disposte dalla medesima; in considerazione della fonte da cui promanano, si può ritenere che non solo si attesteranno sulla soglia minima già individuata a livello generale dal GDPR, e semmai saranno più intense, ma avranno anche una maggiore aderenza alla fattispecie da regolare, di cui terranno conto per contenere i rischi delle decisioni assunte in quello specifico campo. La situazione più verosimile è che questo tipo di trattamenti autorizzati dalla legge saranno effettuati da soggetti pubblici o comunque adiuveranno lo svolgimento di servizi di interessi pubblico. Un regime più attento e rigoroso è dunque applicabile in ambito pubblico, nel solco dell'impostazione più tradizionale, risalente già agli anni '70, di assicurare protezione ai cittadini contro forme di controllo autoritario e antidemocratico.

Viceversa, nel settore privato, dove ragionevolmente trovano spazio le due deroghe alternative, basate sulla strumentalità rispetto all'instaurazione di una relazione contrattuale o sul consenso, si fa più blanda la protezione assicurata dalla disciplina, in quanto l'uso del trattamento completamento automatizzato resta vincolato a presupposti in sé privi di valenza difensiva, mentre le misure di tutela che lo circondano sono circoscritte a quelle minime e aspecifiche disposte dalla normativa generale. Ciò accade in un contesto socio-economico in cui i soggetti che hanno le maggiori potenzialità di accesso a grandi quantità di dati e, altresì, le capacità economiche e le competenze per analizzarli in forma automatica anche a fini predittivi, sono proprio operatori privati come le grandi compagnie attive su Internet.

Il perpetuarsi della cesura tra i trattamenti in ambito pubblico, sottoposti a limiti più stringenti, e quelli in ambito privato, la cui principale fonte di legittimazione rimane il consenso, trova un punto di massima tensione proprio con riferimento all'analisi algoritmica delle informazioni personali e dei *big data*,[47] confermando anche a questo riguardo

[47] I.S. RUBINSTEIN, *Big Data: The End of Privacy or a New Beginning*?, *International Data Privacy Law*, 2013 3(2), 74 ss.

un'impressione di obsolescenza del Regolamento rispetto al progresso tecnologico.[48]

9. *Le garanzie applicabili: informazione, contestazione, revisione*

Fermo restando quanto detto circa il rapporto tra la regola generale sui trattamenti esclusivamente automatizzati (art. 22, comma 1) e lo spazio operativo disegnato dalle tre eccezioni (comma 2), che sembra quasi invertire la proporzione apparente tra ambito di illiceità e, viceversa, di liceità, quest'ultimo è presidiato dalle garanzie già individuate dalla norma. In particolare, tranne quando un simile trattamento sia autorizzato da una norma di legge, che allora dovrà anche prevedere le condizioni e le forme di tutela degli interessati, nelle altre due ipotesi tali misure minime consistono: nel diritto di richiedere l'intervento umano, di esprimere il proprio punto di vista e di contestare la decisione (art. 22, comma 3).

Benché articolate separatamente, queste facoltà sono da intendersi probabilmente in guisa sintetica: si può cioè contestare la decisione automatizzata, indicando le ragioni per le quali la si ritiene impropria, e al contempo richiedere un intervento di verifica ed, eventualmente, di revisione.

Ad abilitare, nella pratica, la possibilità di una contestazione sono le regole contenute negli artt. 13 (comma 2, lett. f), 14 (comma 2, lett. g) e 15 (comma 1, lett. h), GDPR: in quanto attengono essenzialmente a doveri di informazione, esse sono logicamente preliminari alla realizzazione di un trattamento automatizzato e i relativi adempimenti devono collocarsi anteriormente nel tempo. Sono tuttavia altresì propedeutiche all'esercizio dei diritti riconosciuti dall'art. 22, cosicché la loro stretta relazione con l'attivazione delle tutele rende necessario trattarle contestualmente a queste ultime. Esse, invero, costituiscono la prima forma di garanzia per l'interessato, in quanto servono a renderlo edotto dell'esistenza di un trattamento che lo riguarda. Più precisamente, le tre disposizioni – rispettivamente applicabili quando la raccolta dei dati avviene presso lo stesso interessato ovvero presso terzi o, ancora, a seguito dell'esercizio del diritto di accesso – richiedono che questi sia

[48] A. MANTELERO, *Responsabilità e rischio nel nuovo Reg. UE 2016/679*, in *Nuove leggi civ. comm.*, 2017, 145.

informato dell'esistenza di un processo decisionale automatizzato e della logica che il medesimo utilizza.

A quest'ultimo riguardo, è dubbio se l'informazione debba vertere sulla logica generale di funzionamento dell'algoritmo oppure se sia necessario individuare la specifica sequenza di calcolo che potrà condurre o ha già condotto ad una decisione con riguardo al singolo interessato, nonché la base di dati che ne costituisce l'input. Ora, poiché l'informazione deve essere resa al momento in cui i dati sono raccolti (art. 13, comma 1) o comunque entro breve tempo da quando sono stati ottenuti (art. 14, comma 3), è verosimile che non siano ancora stati usati per una decisione vera e propria, ancorché vi sia una predisposizione a farlo. Di conseguenza, l'informazione dovuta, almeno a norma degli artt. 13 e 14, ha probabilmente carattere generale; tuttavia, l'aggettivo "significativ(a)" che la connota, nonché l'oggetto su cui si appunta – oltre alla logica, anche "l'importanza e le conseguenze previste" per l'interessato – dischiudono ampi margini di flessibilità, che sottendono la necessità sia di una rilevanza sostanziale sia di una personalizzazione della comunicazione medesima. In modo particolare, come esprime meglio il termine inglese "*meaningful*", il predicato assegnato all'informazione implica di misurare tale sua qualità rispetto al singolo interessato:[49] si tratta perciò di una nozione che ha meno il senso dell'importanza "quantitativa" e più quello della capacità di trasmettere all'individuo elementi utili alla sua comprensione e all'eventuale esercizio delle scelte che ne conseguono.

Quando invece l'informazione sia resa a seguito dell'esercizio del diritto di accesso, essa potrà essere più compiuta: in particolare, se, com'è verosimile, il processo automatizzato ha già dato luogo a una decisione, all'interessato possono essere messi a disposizione la base di dati che hanno "nutrito" il sistema e una spiegazione del modo con cui la loro elaborazione ha potuto condurre all'esito definitivo.

L'informazione avrebbe in questo caso una valenza effettiva e permetterebbe all'interessato di ricorrere consapevolmente alle forme di tutela successive: contestare una decisione, esprimendo la propria opinione al riguardo, è infatti possibile soltanto se si hanno gli elementi

[49] A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation*, in *Int. Data Privacy Law*, 2017 (7), 236.

per capire come sia stata prodotta. Diversamente, questa rimane impenetrabile e insuscettibile di una valutazione, nonché di una critica circostanziata.

Invero, che tale tipo di spiegazione, pur tecnicamente praticabile, sia oggetto di un obbligo imposto al titolare del trattamento è controverso: il parere reso dall'Art. 29 Working Party non sembra istituire particolari differenze tra l'informazione che è oggetto delle notifiche preliminari a carico del titolare del trattamento e quella da fornire a seguito dell'esercizio del diritto di accesso; più in generale, conferma che non sia necessario dare la spiegazione di una decisione specifica.[50] La formulazione delle regole è pressoché identica negli artt. 13, 14 e 15, nonostante questi siano destinati a trovare applicazione in momenti diversi, e soltanto il *considerando* 71 contiene un riferimento al diritto "di ottenere una spiegazione della decisione conseguita dopo tale valutazione", ossia dopo che vi è stato l'intervento di un revisore umano. Da questi ed altri argomenti testuali e sistematici è stata persino dedotta l'inesistenza nel GDPR di un diritto a una (vera e propria) spiegazione.[51]

Tale conclusione – quasi provocatoriamente veicolata dal titolo del saggio, ma poi ridimensionata nella sua portata dall'argomentazione sviluppata nel testo – è confutata da altri studi,[52] e si presenta come un'interpretazione pressoché antiletterale: l'esistenza giuridica di una simile prerogativa non dovrebbe essere dubbia alla luce dell'impostazione complessiva del GDPR e della congerie di norme che, semmai con una certa ridondanza, fissano le garanzie applicabili al trattamento automatizzato di dati.

10. *La trasparenza nel mondo digitale e dei* big data

Complessivamente le norme analizzate perseguono un obiettivo di trasparenza: sia in ordine alla circostanza che un algoritmo sia impiegato a fini decisionali sia, internamente, con riguardo al suo modo di

---

[50] ART. 29 DATA PROTECTION WORKING PARTY, *op. cit.*, 27.

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

[52] A.D. SELBST, J. POWLES, *op. cit.*, 233 ss.; G. MALGIERI, G. COMANDÉ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *Int. Data Privacy Law*, 2017 (7), 243 ss.

funzionamento. Le interpretazioni possono divergere, per quanto appena detto, in ordine alle caratteristiche, di qualità e di quantità, della conoscenza sufficiente a soddisfare lo scopo; nondimeno, la base giuridica del diritto a ricevere informazioni è solida, rimanendo più ambiguo soltanto il contenuto specifico della nozione di trasparenza accolta.

I principali ostacoli all'affermazione di un simile principio non si rinvengono pertanto a livello giuspositivo, ma si delineano piuttosto sul piano delle scelte di *policy*, anche in vista della strategia europea di sostegno all'economia dei dati, oppure si configurano come limiti di carattere tecnico.

Dal primo punto di vista, le possibili obiezioni richiedono solo una rapida messa a punto. Vi è chi ritiene che la trasparenza sia inutile, se non addirittura controproducente, poiché gli algoritmi c.d. *black-box*, che tengono in considerazione una quantità più elevata di variabili e processano grandi moli di dati, per loro stessa natura sono meno interpretabili e, tuttavia, più affidabili nei risultati, dunque maggiormente efficaci. A questa stregua, l'obbligo di fornire una descrizione della logica su cui si basano i processi avrebbe un impatto non desiderabile sulle attività economiche che si avvalgono di tecniche di *data analysis*: gli operatori sarebbero infatti obbligati a scegliere algoritmi più semplici e, perciò, meno validi, con possibili esiti negativi sia per la buona conduzione dell'impresa sia per il soggetto sul quale la decisione finale, ad esempio ottenere un prestito, deve ricadere. In alternativa, si troverebbero costrette a impegnare risorse e competenze esperte per chiarire in termini comprensibili il modello decisionale impiegato. La necessità di rivedere i risultati del calcolo da parte di un essere umano finirebbe per vanificare il senso del ricorso all'IA, ossia automatizzare i processi per renderli più rapidi, meno costosi e più sicuri nei risultati. Tale visione sfocia in una critica radicale all'impostazione del GDPR in tema di decisioni automatizzate: gli obblighi imposti dal testo normativo al fine di tutelare più intensamente la persona si tradurrebbero in oneri

economici e organizzativi per le imprese senza apportare reali vantaggi ai consumatori.[53]

Una critica meno radicale investe il difficile contemperamento del principio di trasparenza rispetto ai diritti di privativa intellettuale o alla tutela del segreto commerciale che il titolare del trattamento potrebbe opporre. L'investimento nell'elaborazione di algoritmi innovativi ed efficienti è infatti senz'altro da proteggere, mentre una completa *disclosure* potrebbe non soltanto avvantaggiare i concorrenti, ma anche indurre i potenziali destinatari delle decisioni automatizzate ad attuare strategie di aggiramento che ugualmente possono diminuire il rendimento della tecnologia adottata.

Si tratta invero di rilievi che confermano, in certa misura, il non perfetto allineamento delle strategie di tutela implementate nel GDPR con l'evoluzione dell'ambiente digitale in cui sono chiamate ad operare. Tuttavia, per un verso, questa potenziale discrasia non può vanificare la portata normativa della nuova disciplina introdotta dal GDPR: occorre pertanto adeguarsi all'impostazione prescelta, cercando semmai di ovviare all'esistenza di una "*regulatory disconnection*"[54] per mezzo di un'interpretazione che esplori tutte le potenzialità insite nelle norme specifiche sulle decisioni automatizzate e nell'architettura complessiva del Regolamento 679. Ad esempio, la spiegazione richiesta dall'art. 22 deve essere tale da risultare fruibile per l'interessato non esperto: dunque, al contempo, leggibile per lui ma priva di informazioni coperte da privativa che non avrebbe senso dischiudere al profano.[55]

Per un altro verso, le opinioni scettiche sulla capacità di ambientare un principio giuridico di trasparenza nella realtà digitale non considerano che esso è condiviso dalla stessa scienza dei dati, da perseguire, in ipotesi, progettando algoritmi che siano in grado di dare una spiegazione del

---

[53] N. WALLACE, D. CASTRO, *The Impact of the EU's New Data Protection Regulation on AI*, Center for Data Innovation, March 27, 2018, <http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf>; T.L. ZARSKI, *op. cit.*, 1017.

[54] Su questo concetto v. R. BROWNSWORD, M. GOODWIN, *Law and the Technologies of the Twenty-First Century. Text and Materials*, Cambridge, 2012, 398 ss.

[55] L. EDWARDS, M. VEALE, *op. cit.*, 54 ss., propongono molteplici tipologie di spiegazione possibili della logica di un algoritmo compatibili con la tutela di eventuali diritti di privativa.

processo che ha condotto ad un certo risultato.[56] Benché, infatti, le decisioni automatizzate siano in molti casi probabilmente più affidabili, perché oggettive e depurate dell'errore umano, l'esperienza ha rivelato che non sono immuni da errori; e se i falsi dell'algoritmo sono innocui nel contesto di un'indagine sperimentale, possono avere invece esiti inaccettabili quando sono impiegati per assumere decisioni sulla vita delle persone.

La trasparenza rappresenta quindi una sfida prima ancora tecnologica che giuridica, e il precetto che la accoglie non è in contraddizione con le caratteristiche strutturali dei sistemi intelligenti, bensì, eventualmente, soltanto con i limiti contingenti all'attuale livello di progresso in materia di IA.

Questa osservazione conduce al secondo problema attinente agli ostacoli tecnici che, al momento, potrebbero opporsi alla attuazione pratica del principio di trasparenza. Gli algoritmi più complessi, in particolare se basati su meccanismi di *machine learning*, non sempre possono essere spiegati agevolmente, nemmeno dagli esperti. Ciò rende irrealistica la costruzione di un "diritto a una spiegazione" valevole in termini assoluti, a meno di spingerlo fino a significare un divieto di uso di algoritmi che abbiano una logica di funzionamento completamente oscura.

L'opacità dell'algoritmo di calcolo è in antitesi sia con l'obbligo di informare preventivamente l'interessato sulla sua logica di funzionamento, sia con la possibilità di contestazione sia, infine, di attuare una revisione. L'indecifrabilità del processo decisionale opera in effetti tanto nei confronti dell'interessato, che per contestare la decisione dovrebbe prima comprendere come sia stata presa, tanto nei confronti del revisore umano chiamato a controllare l'esito del processo.

A questo proposito, occorre tuttavia ribadire che il principio di trasparenza deve essere preso sul serio dal controllore dei dati. Esso impone, nel quadro del più generale obiettivo di *accountability*, che questi sia in grado di provvedere una spiegazione delle proprie attività e dei propri comportamenti, incluso l'uso di processi automatizzati, e degli

---

[56] Cfr. il programma Explainable Artificial Intelligence (XAI) della Defense Advanced Research Projects Agency (DARPA), <https://www.darpa.mil/program/explainable-artificial-intelligence>.

esiti che producono.[57] Il titolare dovrà perciò applicare un criterio di leggibilità degli algoritmi che impiega all'interno della sua attività, al fine di soddisfare le richieste di spiegazione di ogni soggetto interessato e di poter attuare un controllo delle decisioni assunte.

11. *I limiti di una regolazione dei processi algoritmici* data protection-driven

Il GDPR è l'unica fonte legislativa in cui trovano disciplina i processi algoritmici di decisione e costituisce pertanto un riferimento essenziale. Tuttavia, come evidenziato nei paragrafi precedenti, le regole che vi dedica presentano limiti, sia per l'ambiguità della formulazione sia per il ristretto raggio applicativo, che ne indeboliscono la portata effettiva e il valore di garanzia.

Dal punto di vista dell'orizzonte di applicazione, si consideri come il divieto posto dall'art. 22, comma 1, e le eccezioni al medesimo (presidiate allora da opportune misure di tutela) riguardino soltanto i processi completamente automatizzati, mentre non si occupano dei sistemi usati in funzione di mero sussidio tecnico alla decisione da assumere. Quest'ultimo, invero, è attualmente il caso più frequente, che presenta rischi non meno elevati per il destinatario della decisione. Infatti, benché il passaggio finale del procedimento sia ascrivibile a un responsabile umano, è verosimile che questi abbia difficoltà, di ordine piscologico e di ordine pratico, a discostarsi dall'esito proposto. Se non è in grado di controllare il percorso razionale che conduce al risultato, il quale a sua volta appare il frutto di un calcolo sofisticato e oggettivo, tenderà a sovrastimare il valore predittivo del medesimo; specialmente in situazioni di scarsità di tempo e risorse, è improbabile una difformità di giudizio (e nella risposta applicata) tra l'uomo e la macchina. L'interessato avrà tuttavia a disposizione soltanto i rimedi generali a garanzia di un corretto trattamento dei suoi dati personali, ma non quello che gli consente di richiedere una spiegazione della logica di funzionamento del processo e

[57] Sulle strategie esistenti per progettare "algoritmi responsabili" nel senso sia della regolarità procedurale, sia dell'inclusione di meccanismi interni per somministrare decisioni corrette e non discriminatorie, sia, infine, nell'apertura alla possibilità di ispezione e di controllo successivo cfr. J.A. KROLL *et al.*, *Accountable Algorithms*, 165 *Univ. Penn. Law Rev.* 2017, 633 ss.

di contestare immediatamente la decisione presso lo stesso titolare del trattamento.

Nel noto caso dell'impiego da parte di alcune corti nordamericane di un algoritmo per determinare il rischio di recidiva degli imputati, si è potuto apprezzare come un'inferenza arbitraria preluda a esiti discriminatori sia quando il risultato prodotto dal calcolo è applicato meccanicamente sia quando il giudice si limiti a tenerlo in considerazione come elemento di una valutazione più complessa. Tuttavia, proprio questa circostanza – essere soltanto uno tra i fattori oggetto di scrutinio, non dotato in sé di un peso determinante – avrebbe reso il software compatibile con i principi del giusto processo.[58]

Altra debolezza dell'impianto di tutela consiste nella sua applicabilità soltanto a vere e proprie "decisioni", il che lascia impregiudicate pratiche come il *targeted advertising* che pure hanno una capacità di manipolazione dei comportamenti molto elevata; a decisioni, inoltre, che riguardino l'interessato i cui dati personali sono stati trattati, con esclusione di quelle che incidono su un gruppo o una categoria, che aggrega i singoli intorno a un criterio distillato dalla stessa analisi automatizzata.

Alla portata limitata delle regole di protezione, si aggiungono le opacità interpretative in merito ai presupposti per l'azionabilità delle misure di garanzia e al loro contenuto; nonché le difficoltà pratiche, per un verso, di attuare il diritto alla spiegazione della logica sottesa all'algoritmo di decisione, per un altro verso, di dare corso al controllo e alla eventuale revisione della decisione assunta. Non soltanto, infatti, il processo algoritmico può essere scarsamente leggibile anche per un revisore esperto, ma la regressione degli effetti della decisione presa sarebbe in alcune ipotesi particolarmente macchinosa.[59]

Queste criticità sono intrinseche alla formulazione testuale delle norme e possono essere in parte superate in via interpretativa. Il problema principale deriva invece dalla scelta di fondo di regolare un fenomeno che

---

[58] Supreme Court of Wisconsin, State of Wisconsin v. Eric L. Loomis, July 13, 2016.

[59] Ad esempio, nel caso della *price discrimination* una simile possibilità dovrebbe essere implementata nel sito di acquisti *on line* e prevedere le modalità di reintegrazione del soggetto che abbia ricevuto un trattamento ingiusto.

ha una portata generale soltanto conferendo diritti al singolo.[60] Manca, cioè, una regolazione verticale indirizzata ai soggetti, pubblici o privati, che progettano e impiegano nella loro sfera di attività sistemi computerizzati di decisione, con l'eccezione relativa alle fasi di raccolta dei dati e di trattamento in senso stretto, soggette ai principi generali accolti dal GDPR.

Ora, l'iniziativa individuale è già di per sé scoraggiata dalla configurazione dei presupposti per l'attivazione dei rimedi previsti; ma anche ove effettivamente vi sia, può portare a rimuovere la decisione sfavorevole nel caso singolo, ma non necessariamente riesce a incidere sulla struttura del processo decisionale che conduca a esiti discriminatori, falsi, errati. La singola decisione viene riformata, ma l'algoritmo rimane impregiudicato.

A cascata, se le contestazioni sollevate sono rare e i loro effetti rimangono confinati nella sfera individuale, difficilmente si origineranno prassi virtuose di controllo dell'algoritmo allo scopo di contenere il possibile contenzioso.

Ciò induce a ritenere necessari, in chiave di regolazione futura, degli interventi normativi capaci di influenzare il design dell'algoritmo, la qualità dei dati con cui viene alimentato, e di istituire dei meccanismi di monitoraggio del loro funzionamento.

12. *Linee portanti e modelli di riferimento per una regolazione degli algoritmi*

La regolazione degli algoritmi non può essere affidata esclusivamente alla prospettiva della *data protection*, benché quest'ultima sia logicamente essenziale. Molte ricadute negative del fenomeno evidenziate dalle esperienze riferite non dipendono infatti dall'assenza di una base legittimante il trattamento delle informazioni o dalla violazione di una delle specifiche regole che lo disciplinano. Anche l'adeguatezza dei dati che alimentano l'algoritmo, governata dal GDPR, spesso non è apprezzabile in sé, ma in rapporto all'uso che deve farsene: dati esatti e

[60] Rilevano la scarsa efficacia di misure di tutela basate sull'*enforcement* individuale e da qui muovono per identificare altri dispositivi di tutela L. EDWARDS, M. VEALE, *op. cit.*, 74 ss.

aggiornati possono interagire e combinarsi in modi inaspettati, o semplicemente essere, per quanto di buona qualità, di quantità insufficiente perché le inferenze che se ne traggono siano valide.

Con approccio pragmatico, occorre partire dai problemi emersi dalla casistica per individuare le contromisure adatte a contrastarli, basando su di esse un quadro di regolazione dell'architettura dei sistemi automatizzati di decisione. [61]

L'algoritmo, anzitutto, deve essere solido – o robusto, secondo un lessico sviluppato in questo contesto – dal punto di vista scientifico e statistico. Deve evitarsi, cioè, di costruire un modello per giudicare la *performance* degli insegnanti sulla base dei risultati che una manciata di studenti (i pochi che compongono una classe) ha riportato annualmente, quando le ragioni dell'esito negativo dei test possono essere tantissime, e avere poco o nulla a che fare con la qualità dell'insegnamento impartito.[62]

Questa solidità, o robustezza, deve estendersi alla base di dati usati per il training dell'algoritmo. In particolare, occorre evitare gli errori che dipendono dall'inevitabile proiezione dell'analisi verso il passato fotografato attraverso i dati: le inferenze sono tratte da una situazione apparentemente registrata in maniera neutrale, ma in realtà intrisa di differenze che possono così perpetuarsi e aggravarsi, fino al paradosso della profezia che si autoavvera. La scarsa presenza di una certa caratteristica nel set di dati usati (ad esempio, il genere femminile nelle esperienze pregresse di assunzione per una certa posizione lavorativa), determina un funzionamento imperfetto del modello e rende meno affidabile la valutazione per il gruppo poco rappresentato.[63] È noto il circolo vizioso che può innestarsi anche con riguardo al rischio di recidivismo: la condanna più pesante, a più anni di reclusione, sulla base della predizione di una maggiore pericolosità sociale può far avverare il

---

[61] G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Pol. dir.*, 2019, 218 ss., usa l'espressione "legality by design" per indicare la necessità di una simile forma di regolazione.

[62] Questo è il caso che ha portato negli Stati Uniti al licenziamento degli insegnanti che avevano ricevuto un punteggio molto basso nel sistema di valutazione IMPACT, sviluppato da una società privata: è un esempio delle "*weapons of math destruction*" secondo C. O'NEIL, che ne riferisce nell'Introduzione al suo libro.

[63] L'esempio è tratto da J. KROLL *et al., op. cit.*, 44.

pronostico proprio perché le condizioni carcerarie e il prolungato contatto con altri detenuti favoriscono la commissione di nuovi crimini.

Ugualmente importante, ancora con riferimento alla qualità dei dati, è la pertinenza delle informazioni impiegate e la loro autentica capacità predittiva: si immagini un datore di lavoro che seleziona i dipendenti da assumere in base, tra l'altro, alla puntualità nei pagamenti risultanti da estratti conto delle carte di credito, sul presupposto che chi paga i propri debiti alla scadenza sia più incline a rispettare le regole, arrivare in orario sul luogo di lavoro, svolgere i compiti assegnati. Chi viene escluso sulla base di queste informazioni, tuttavia, potrebbe essere un ottimo lavoratore, le cui difficoltà economiche sono state generate da accadimenti gravi e imprevisti, come una separazione o la malattia di un familiare, che sono inevitabilmente destinate ad aggravarsi se resta fuori dal mercato del lavoro e non trova un impiego proprio a causa di dati ininfluenti e privi di valore rispetto alle sue reali capacità. Nel modello predittivo usato dalla società ZestFinance, già ricordato, che propone un tasso di interesse sui prestiti sensibilmente più basso di quelle che offre il mercato, rilevano variabili apparentemente ultronee come il tempo e la cura impiegati per compilare la domanda di accesso al finanziamento, la presenza di errori ortografici, se ci si è dedicati a leggere le condizioni di contratto. Sottintesa a queste osservazioni è che la fedeltà alle regole, anche grammaticali, ovvero l'interesse per il contenuto del contratto cui ci si vincola sia sintomatico di una maggiore propensione a rispettare gli impegni in termini più generali. Sennonché la padronanza della lingua scritta e la capacità di comprensione delle condizioni contrattuali finisce inevitabilmente per svantaggiare persone con una bassa alfabetizzazione o soggetti immigrati di recente, rispetto alle quali i dati considerati non sono per nulla significativi.

Similmente, quando il training dell'algoritmo si basa su decisioni pregresse, etichettate come corrette, ma che includono in realtà pregiudizi o criteri discriminatori, benché non intenzionali, questi si riprodurranno con precisione matematica nell'elaborazione computerizzata. Istruttivo è il caso della Medical School del St George Hospital di Londra, che aveva introdotto un sistema automatico di valutazione delle domande per le posizioni di specializzazione messe a disposizione annualmente. Per "insegnare" al sistema il modo di operare, vi immette i giudizi delle commissioni di selezione degli anni precedenti. Poiché la valutazione

avveniva tenendo conto anche di elementi come la scarsa conoscenza della lingua inglese, dimostrata ad esempio da errori ortografici o sintattici nella presentazione della domanda, il computer impara a istituire una correlazione diretta tra il luogo di nascita, il cognome o la provenienza da sobborghi prevalentemente popolati da immigrati – fattori condivisi da una buona porzione delle domande rigettate – e una peggiore qualità dei candidati.[64]

Ancora, un recente lavoro pubblicato su Science dimostra come un algoritmo in uso presso il sistema sanitario statunitense discrimina gli afroamericani poiché usa la spesa sostenuta per le cure come un *proxy* per i bisogni effettivi di terapia; riflette, in questo modo, le disuguaglianze nella distribuzione della ricchezza nella popolazione americana.[65]

La costruzione dell'algoritmo deve permettere decisioni granulari: ad esempio, un'assicurazione che considera più rischiosa la guida in ore notturne e pratica prezzi più elevati a coloro che hanno tale abitudine, non scevera chi frequenta discoteche e locali dove si servono alcolici da chi effettua turni di notte e magari guida per recarsi al lavoro o tornare a casa. Chi appartiene a quest'ultimo gruppo deve di conseguenza sopportare costi più elevati per la classificazione grossolana che lo associa ad una certa propensione al rischio, quando in realtà non presenta un'attitudine agli incidenti superiore alla media. Un sistema equo deve dunque consentire la scomposizione del gruppo in sottoclassi che tengano conto di dati aggiuntivi.[66]

Infine, le decisioni algoritmiche devono essere verificabili, ossia spiegate e sottoposte a controprova. A meno che l'astrusità dei risultati sia patente, in mancanza di un controllo esperto non esiste la possibilità pratica di correggere gli errori predittivi. I *bias* che affliggevano il software COMPAS usato in diversi Stati americani per determinare la pericolosità sociale e quindi il rischio di recidiva sono stati disvelati dall'associazione ProPublica, che ha messo a confronto le predizioni dell'algoritmo con

---

[64] S. BAROCAS, A.D. SELBST, *Big Data's Disparate Impact*, 104 *California L. Rev.* 2016, 682.
[65] Z. OBERMEYER *et al.*, *Dissecting racial bias in an algorithm used to manage the health of populations*, in *Science*, 2019, vol. 336, 447 ss.
[66] A. MANTELERO, *Personal data for decisional purposes*, cit., 247.

l'effettivo ritorno a delinquere degli arrestati.[67] Nel caso appena descritto del St George Medical Hospital, soltanto dopo che la Commissione per l'uguaglianza razziale del Governo britannico condanna la scuola di specializzazione per discriminazione razziale nelle sue *policies* di ammissione l'algoritmo viene corretto, rendendolo cieco rispetto ai dati anagrafici e facendo rilevare solo le competenze di tipo medico acquisite tramite la formazione.

Come evidenziano queste esperienze, un simile controllo impegna notevoli risorse e non può pertanto attuarsi grazie alle contestazione individuale, per le ragioni già viste, e neppure per iniziativa di gruppi organizzati, data la difficoltà di aggregare gli interessi lesi intorno a un denominatore comune riconoscibile *ex ante*.[68]

Occorre pertanto, per un verso, definire i modi di corretta costruzione di un modello algoritmico attraverso regole di tipo sostanziale e strumenti di certificazione della qualità del sistema; per un altro verso, istituzionalizzare il monitoraggio delle risposte che sono state prodotte per poterne vagliare la correttezza e l'imparzialità.

Modelli di riferimento cui attingere per l'introduzione di questa architettura regolatoria sono presenti nello stesso GDPR e, in parte, essi possono già attualmente svolgere una funzione di disciplina nel senso indicato. Lo strumento del *Data Protection Impact Assessment* (art. 35, comma 1) dovrà essere adottato in quanto obbligatorio per i trattamenti che presentano un "rischio elevato" per i diritti e le libertà delle persone fisiche, categoria nella quale verosimilmente ricadono quelli che adoperano meccanismi di *machine learning*.[69] La profilazione, anche quando non condotta esclusivamente con mezzi automatizzati,[70] è del

---

[67] Si è dimostrato che il software, messo a punto da una compagnia privata, prediceva una più alta propensione alla criminalità per le persone di colore, producendo circa il doppio di falsi positivi che per i bianchi: J. Angwin *et al.*, ProPublica, *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

[68] A. MANTELERO, *Personal data for decisional purposes*, cit., 251, sottolinea come gli interessi colpiti dalle decisioni automatizzate sono di varia natura e la classificazione in gruppi avviene spesso all'insaputa dei singoli.

[69] L. EDWARDS, M. VEALE, *op. cit.*, 78.

[70] ART. 29 DATA PROTECTION WORKING PARTY, *op. cit.*, 29.

resto espressamente richiamata tra le ipotesi esemplificative in cui il DPIA è richiesto (art. 35, comma 3, lett. a). L'istituzione di sistemi di certificazione (art. 42) risponde anch'essa a una logica di controllo *ex ante* della qualità del sistema che processa i dati e può estendersi alle forme automatizzate di trattamento.

Questi strumenti presentano una declinazione specifica alla *data protection*, ma oltre a istituire valide forme di controllo rispetto al profilo della qualità dei dati nella loro portata attuale,[71] possono servire come schemi di disciplina cui guardare con l'obiettivo della *governance* algoritmica.

Oltre che di strumenti appropriati di regolazione, quest'ultima necessita di una struttura apicale che sia investita delle funzioni di monitoraggio e di controllo della *compliance* degli operatori. A questo riguardo si prospetta un'alternativa di fondo: l'approccio settoriale[72] ovvero quello generale, che affiderebbe a un unico soggetto, eventualmente già esistente, il ruolo di autorità di riferimento.[73]

Il primo, che è compatibile con una regolazione sostanziale omogenea contenente le linee di fondo, ma a sua volta suscettibile di differenziarsi sulla base degli ambiti di applicazione e dei rischi maggiori o minori in essi presenti,[74] sembra preferibile. Poiché i contesti in cui si impiegano sistemi automatizzati di decisione sono molteplici e diversificati, è poco realistico costruire competenze, a livello apicale e a livello operativo (ad esempio, degli enti di certificazione), che siano capaci di coprire tale

---

[71] E si consideri l'orientamento del DPIA, non ristretto alla valutazione sulla tutela dei dati, ma rivolto più in generale alla protezione dei diritti e delle libertà fondamentali. V. anche *considerando* 75.

[72] THE ROYAL SOCIETY, *Machine learning: the power and promise of computers that learn by example*, 2017, 98 ss.

[73] È la proposta di A. MANTELERO, *Personal data for decisional purposes*, cit., 251 ss., che osserva come tutte le forme di *big data analytics* abbiano in comune operazioni di trattamento dei dati. Questo coinvolge necessariamente le competenze dell'Autorità garante per i dati personali, che dovrebbe di conseguenza non limitare la propria attività al profilo della *data protection*, ma essere investita anche della valutazione dell'impatto etico e sociale dei processi algoritmici.

[74] Questo approccio alla regolazione, basato sul livello di rischio dei sistemi algoritmici, apprezzato in base al settore in cui operano e al modo di funzionamento, emerge dal White Paper *On Artificial Intelligence* appena emanato dalla Commissione europea (cfr. *retro*, nt. 30).

complessità. Viceversa, l'uso di sistemi esperti in medicina o di algoritmi di calcolo nel campo dei servizi finanziari, bancari o assicurativi potrebbe essere soggetto alla vigilanza degli organismi esistenti. A sua volta, la raccolta massiva di dati nel commercio *on line*, nel mercato dei *social network* o di altri servizi digitali e la pubblicità personalizzata si propongono come fenomeni che interessano gli apparati di tutela dei consumatori e le misure contro le pratiche commerciali sleali.