

GINEVRA CERRINA FERONI

INTELLIGENZA ARTIFICIALE E SISTEMI DI SCORING SOCIALE. TRA DISTOPIA E REALTÀ

SOMMARIO: 1. Una distopia che è già realtà. — 2. La Repubblica Popolare Cinese: il *Social Credit System*. — 2.1. Pianificazione ed implementazione. — 2.2. Fattori di valutazione del comportamento dei cittadini. — 2.3. Effetti del *Social Credit System*. — 3. I Paesi Bassi: il *Toeslagenaaf faire*. — 3.1. I fatti. — 3.2. Aspetti giuridici rilevanti. — 4. L'esperienza italiana. Il *rating* reputazionale nella giurisprudenza del Garante per la protezione dei dati personali. — 4.1. Il caso *Mevaluate*. — 4.2. L'ambito scolastico. — 5. Riflessioni conclusive. Dall'algoritmica all'algoritmica.

1. UNA DISTOPIA CHE È GIÀ REALTÀ.

Immaginiamo se ogni nostra azione, interazione, movimento, venisse ridotta ad una valutazione su una scala a cinque punti. Un punteggio più alto apre le porte a favolose opportunità e vantaggi, mentre un punteggio basso può isolarci dal resto della società. Così, per esempio, in ragione di come abbiamo risposto al cameriere del bar dove abbiamo ordinato un caffè, potrebbe esserci interdetto salire su un mezzo pubblico o, a seconda delle valutazioni del nostro precedente vicinato, preclusa la possibilità di trasferirci in un altro quartiere popolato da soggetti il cui punteggio è superiore ad una certa soglia. Oggetto di una delle più famose puntate della serie fantascientifica *Black Mirror*, questo, per quanto semplificato, è come funziona un sistema di punteggio o credito sociale, ovvero di *social scoring* o *rating* reputazionale. Ma quella che ancora suona come la trama distopica di una narrazione di fantascienza è una realtà che già ci circonda ed interessa direttamente la vita dei cittadini di alcune regioni del mondo.

Uno dei sistemi di credito sociale più noti ed evidenti è quello introdotto dal Governo cinese, sebbene la Cina non sia l'unico ordinamento che sta adottando soluzioni di monitoraggio sociale. A gennaio 2022, il governo del Regno Unito ha rivelato che avrebbe utilizzato un *software* di riconoscimento facciale in tempo reale per le strade di Londra per trovare sospetti ricercati dalla polizia e sistemi simili sono già utilizzati dalle forze dell'ordine e da agenzie di *intelligence* di tutto il mondo. Oltre alle autorità pubbliche, altri

soggetti, come compagnie assicurative, hanno annunciato o già implementato sistemi che autorizzerebbero i propri operatori da raccogliere efficacemente informazioni sul comportamento delle persone da utilizzare in scelte aziendali. Ad esempio, le compagnie specializzate in assicurazioni sulla vita, a New York, sono autorizzate ad assumere decisioni sui propri clienti utilizzando le informazioni trovate nei *social network*. In fondo, il settore delle assicurazioni automobilistiche utilizza già delle *black box* installate sulle vetture come parte del contratto, in grado di calcolare il *malus* sulla base dello stile di guida che percepiscono.

Esistono programmi che consentono ai titolari di attività di ristorazione o intrattenimento di creare liste di persone non autorizzate ad entrare a causa di precedenti comportamenti scorretti. E tutti i servizi di *sharing economy*, come Airbnb, Uber o servizi di consegna a domicilio, dispongono di una sorta di sistema di punteggio che valuta coloro che partecipano al servizio sia come lavoratori che come consumatori. Con l'emergenza sanitaria da Covid-19, la questione del monitoraggio dei cittadini si è imposta in maniera ancora più pervasiva, accentrando su di sé le radicali divergenze culturali, prima che politiche e giuridiche, circa la possibilità che autorità di regolazione di tutto il mondo potessero o meno ipotizzare soluzioni per rintracciare i cittadini infetti o non vaccinati e, sulla base dei loro dati di salute, consentire o proibire loro accessi, opportunità e persino l'esercizio ed il godimento di diritti e libertà garantite costituzionalmente.

Per affrontare il tema vale la pena soffermarsi, seppur a grandi linee, sull'esperienza ad oggi più evoluta di sistema di credito sociale, ovvero il caso della Repubblica Popolare Cinese. Passeremo poi a richiamare una vicenda di credito sociale che ha interessato i Paesi Bassi ed è stata oggetto di un ampio dibattito pubblico, per poi analizzare il caso italiano, anche alla luce della più significativa giurisprudenza del Garante della protezione dei dati personali.

2. LA REPUBBLICA POPOLARE CINESE: IL *SOCIAL CREDIT SYSTEM*.

Il Sistema di Credito Sociale (SCS) è un'iniziativa implementata dalla Repubblica Popolare Cinese fin dai primi anni Duemila con il fine di sviluppare un sistema nazionale per classificare la reputazione dei propri cittadini. Attraverso il SCS si assegna ad ogni cittadino un punteggio rappresentante il suo "credito sociale", sulla base di informazioni possedute dal Governo, come quelle riguardanti la sua condizione economica e sociale, le sue preferenze e le sue abitudini¹, operando così sostanzialmente come un meccanismo di sorveglianza di massa affidato a tecnologie

¹ Tra i lavori più recenti che hanno provato a definire sistematicamente il tema,

X. DAI, *Toward A Reputation State: A Comprehensive View of China's Social Credit*

di AI per l'analisi dei *big data*. Oltre ai singoli cittadini, come si vedrà, il SCS è stato esteso anche alle attività di tutte le imprese che operano nel mercato cinese ².

2.1. Pianificazione ed implementazione.

Il Sistema di Credito Sociale si inserisce all'interno dell'approccio di c.d. "*top-level design*" cinese ed è coordinato dal "Gruppo Dirigente Centrale per l'Approfondimento di Riforme Comprensive". Secondo la "Pianificazione della Creazione del Sistema di Credito Sociale (2014-2020)" pubblicata dal Consiglio di Stato, il Sistema di Credito Sociale si concentra su quattro aree: "onestà negli affari di governo", "integrità commerciale", "integrità sociale" e "credibilità giudiziaria".

L'obiettivo dell'iniziativa è duplice: da una parte vi è il piano individuale che attiene alla sfera etica delle azioni del singolo. In questa dimensione il SCS è volto a individuare e punire le devianze e a incentivare la sincerità, l'onestà e l'integrità dei singoli cittadini cinesi ³. Dall'altra parte — e forse questo è l'aspetto più trascurato del fenomeno — viene espressamente dichiarato l'intento di creare condizioni prodromiche al perfezionamento e alla regolamentazione dell'economia di un mercato socialista ⁴. Il rafforzamento della capacità dirigistica dello Stato e la sua abilità di indirizzare il comportamento e le azioni dei cittadini sono infatti finalizzati non solo ad esigenze di controllo sociale, ma anche al bisogno di uniformazione del mercato cinese. Da quest'ultimo punto di vista il progetto di SCS persegue due fondamentali obiettivi economici: in piccola scala, eliminare patologie del mercato, come truffe e contraffazioni; in larga scala, risolvere il problema della affidabilità finanziaria *all'interno e da parte* del mercato cinese e, in questo modo, contribuire a realizzare il

System Project, in O. Everling (a cura di), *Social Credit Rating*, Springer, Wiesbaden, 2020, 139-165; S. HOFMANN, *Social Credit: Technology-enhanced authoritarian control with global consequences*, International Cyber Policy Center, Policy Brief Report n. 6/2018, online su https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2018-06/Social%20credit_1.pdf?O3X2xnkGONvJFjK4Z57Xlf06lget_MID=; F. Sapia, *The Meanings and Extensions of 'Social Credit System'*, *China Social Credit System*, online su <https://china-social-credit.com/2017/07/meaningextension-social-credit-system/>; S. AHMED, *Consumer Protection in the Chinese Social Score System*, in *Digital Credit Observatory*, 2017, online su www.digitalcreditobservatory.org; G. KO-

STKA, *China's social credit systems and public opinion: Explaining high levels of approval*, in *New Media and Society*, vol. 21, n. 7, 2019, 1565-1593.

² M. PERSSON, M. VLASKAMP, F. OBEMA, *China rates its own citizens — including online behaviour*, su *de Volkskrant*, 25 aprile 2015.

³ Non a caso R. CREEMERS, *China's Social Credit System: An Evolving Practice of Control*, in www.papers.ssrn.com, 2, nota come in mandarino la parola "credito" (*xinyong*) sia semanticamente identica al termine "integrità".

⁴ J. CHIN W G. WONG, *China's New Tool for Social Control: A Credit Rating for Everything*, in *Wall Street Journal*, 28 novembre 2016.

connubio tra armonia sociale e armonico sviluppo economico del Paese ⁵.

Il Sistema di Credito sociale, per ora, risulta confinato alla Cina continentale, e non anche ad Hong Kong e Macao. Tuttavia i piani per l'implementazione del sistema non parlano di una distinzione tra imprese cinesi e imprese straniere che operano nel mercato cinese, implicando la possibilità che anche le imprese straniere che operano in Cina siano sottoposte alla classificazione all'interno del sistema.

Il sistema è già stato coinvolto in parecchie controversie, in particolare per quanto riguarda il modo in cui sarà applicato alle persone ed alle imprese, e tuttavia né le cause vinte di fronte ai tribunali né gli errori compiuti dall'algoritmo e riconosciuti dalle stesse autorità come tali hanno portato ad una modifica del processo di funzionamento dello stesso ⁶.

Per quanto riguarda i cittadini, sono già riscontrabili esempi di punizioni causate dalla violazione dei protocolli sociali: il sistema ha già negato a nove milioni di persone con "punteggi bassi" il diritto di comprare voli domestici. Il sistema risulta utilizzato anche per escludere la possibilità a genitori di iscrivere i loro figli in determinate scuole, proibire a coloro che hanno un punteggio basso di affittare camere d'albergo, di usare carte di credito, e per inserire in una *black list* determinati individui con l'obiettivo di precludergli l'ottenimento di lavoro ⁷. Il sistema è applicato in funzione anche di valutare le abitudini di navigazione su internet della gente (troppo tempo passato giocando ai video giochi riduce il punteggio), le abitudini riguardanti gli acquisti e una varietà di azioni personali e assolutamente innocue che non hanno alcun impatto sulla comunità ⁸.

La "Notifica del Consiglio di Stato riguardante l'emanazione della Pianificazione della Creazione del Sistema di Credito Sociale (2014-2020)" è stata emanata dal Consiglio di Stato cinese il 14

⁵ Così D. MAC SETHING, M. SIEMS, *The Chinese Social Credit System: a Model for other Countries*, in *Modern Law Review*, vol. 82, n. 6, 2019, pp. 1034-1071, spec. 1048 ss.; afferma poi R. CREEMERS, *China's Social Credit System: An Evolving Practice of Control*, cit., p. 4, « *The context in which the concept of social credit evolved is perhaps best characterized as one of repeated efforts at building an effective, powerful and prosperous state. The term was mentioned first in the Nineties, in the context of the market-economic reforms that had started under the leadership of Deng Xiaoping* ». Sulla strumentalizzazione del concetto di moralità e di capacità normativa dell'autorità morale in Cina, v. P.M. THORNTON, *Disciplining the state: virtue, violence,*

and state-making in modern China, Cambridge (MA), Harvard University Press, 2007.

⁶ M. v. BLONBERG, *The Social Credit System in China's Rule of Law*, 100 ss. conclude che le rettifiche che seguono alla vittoria di controversie davanti ai tribunali sono pensate per dispiegare solo effetti *inter partes*, ma non hanno alcuna influenza nel correggere le storture del sistema.

⁷ Tra i molti che hanno riportato gli esempi, D. MAC SETHING, M. SIEMS, *The Chinese Social Credit System: a Model for other Countries*, cit., 1036 ss.

⁸ A. FENWICK ELLIOTT, *China is banning people with bad 'social credit' from using planes and trains*, in *The Telegraph*, 19 marzo 2018.

giugno 2014. Nel 2015 è stata concessa la licenza ad otto aziende, affinché cominciassero a sviluppare dei prototipi di sistemi di credito. Le imprese selezionate sono Ant Financial dell'Alibaba Group, lo sviluppatore di *software* Tencent e altre sei. I sistemi di credito creati da queste imprese utilizzavano banche-dati come il Sesame Credit all'inizio della fase di sperimentazione⁹. Apparentemente, nessuna nuova licenza è stata concessa dal 2017¹⁰.

Fino a ora non esiste nessun sistema di credito sociale comprensivo ed esteso a tutta la nazione, ma sono stati implementati molteplici progetti sperimentali che stanno testando il sistema su scala locale e in alcuni settori dell'economia. Un programma di questo tipo è stato reso operativo a Shanghai per mezzo dell'app Honest Shanghai, che utilizza un *software* di riconoscimento facciale per cercare all'interno di archivi del governo e classificare gli utenti in questo modo. I punteggi possono basarsi anche su informazioni estrapolate dalle abitudini di ricerca *online* dei cittadini cinesi.

L'infrastruttura del Sistema di Credito Sociale è stata completata nel 2020, ma le restrizioni sui cittadini e le imprese con "punteggi bassi" hanno iniziato ad avere effetto a partire dal 1° maggio 2018.

Per quanto riguarda specificamente le imprese, invece, il Sistema di Credito Sociale è stato ideato come meccanismo di regolazione del mercato. L'obiettivo è quello di stabilire una struttura di regolazione auto-imposta, alimentata da *big data*, all'interno della quale le imprese si controllano tra di loro. L'idea di base è che con un sistema di credito funzionale attivo, le aziende si impegneranno a conformarsi alle politiche e alle disposizioni del Governo per evitare che il loro punteggio si abbassi. Per come è stato progettato, le imprese con punteggi alti avranno dei vantaggi come migliori condizioni sui prestiti, tasse più basse e più opportunità di investimento. Le imprese con punteggi bassi riceveranno condizioni svantaggiose per nuovi prestiti, tasse più alte, restrizioni sugli investimenti e opportunità minori di partecipare a progetti finanziati dal settore pubblico. I piani del

⁹ V. M. BLOMBERG, *The Social Credit System in China's Rule of Law*, cit., 93 riporta il funzionamento dell'esperimento: « *The smartphone application calculates a score from 300 to 950 for every user. Categories of information include credit history, fulfilment capacity, and personal characteristics, which are applied by traditional credit rating services in other countries as well. A novel category Sesame Credit applies is interpersonal relationships. Being friends with a low scoring peer on Sesame Credit will lower one's own score* ». Interessante

notare come anche le differenti preferenze di acquisto concorrono a formare il credito: la scelta di offerte promozionali particolarmente basse sulla piattaforma *shopping* Alibaba riduce il credito Sesame, mentre l'acquisto di strumenti per cucina, sport o lavori fai-da-te, evidentemente giudicate attività utili alla salute psico-fisica del cittadino cinese, vengono premiate con extra credito.

¹⁰ *Planning Outline for the Construction of a Social Credit System (2014-2020)*, su *China Copyright and Media*, 14 giugno 2014.

Governo prevedono anche un monitoraggio in tempo reale delle attività delle imprese. In quel caso, infrazioni da parte di un'impresa potrebbero concretizzarsi in un abbassamento istantaneo nel punteggio.

2.2. *Fattori di valutazione del comportamento dei cittadini.* _____

Il Sistema di credito sociale cinese si basa su tre variabili:

1. Un primo fattore, non dissimile invero da quello preso in considerazione nei sistemi di credito sociale “privati” (es. Schufa in Germania, o FICO negli USA) è quello legato alla capacità di ogni cittadino di ottemperare regolarmente alle proprie obbligazioni finanziarie. Una persona che paga le proprie bollette puntualmente può essere considerata come affidabile e solvibile e, per questo motivo, riceverà un punteggio più alto.

2. Un secondo fattore riguarda la capacità di ogni individuo di adempiere ai propri doveri così come definiti nei contratti stipulati. Ad esempio, un individuo con un cospicuo livello di risparmi riceverà un punteggio più alto rispetto ad una persona che fatica ad arrivare alla fine del mese.

3. Un terzo fattore, quello che presenta maggiori criticità, riguarda il comportamento sociale estensivamente considerato: gli orientamenti d'acquisto su portali *online* come Alibaba, i *trend* degli acquisti *offline* pagati con Alipay, le interazioni sociali su WeChat, la cronologia *online* di Baidu, le aree di frequentazione (individuate tramite geolocalizzazione del cellulare), cioè in genere quegli elementi che possono dare un'idea onnicomprensiva della personalità di ogni cittadino. Ad esempio, un utente che acquista pannolini è con molta probabilità un genitore e, pertanto, anche un cittadino più responsabile. Al contrario, un cittadino che spende ingenti somme per giochi *online* è considerato più pigro e meno produttivo per la società.

Ben si capisce come tali risultati possono ricavarsi solo da un sistema di valutazione algoritmica il cui codice sorgente è altamente influenzato da valori sociali di interesse collettivo ¹¹.

2.3. *Effetti del Social Credit System.*

Il sistema cinese non si limita ad analizzare il comportamento del singolo individuo ¹². È infatti previsto che il punteggio di ogni

¹¹ Da qui si capisce l'equiordinato valore che la civiltà cinese attribuisce alle norme sociali e alla legge. Anzi, come è stato scritto proprio con riferimento al SCS, « *Laws must often fade into the background (...) keep adapting to evolving social practices* » (Z. Zuo, *Governance by Algorithm: China's Social Credit System*, University of Cambridge Working Paper, 16 giugno 2020, online su www.finance.group.cam.ac.uk).

¹² J. CHIN E L. LIN, *Stato di sorveglianza. La via cinese verso una nuova era del controllo sociale*, Bollati Boringhieri, Torino, 2022, 219 ss.

individuo sia influenzato anche dal comportamento delle persone a lui più vicine. Ad esempio, un soggetto che infrange una o più leggi del Partito Comunista Cinese influenzerà negativamente anche partner e persone allo stesso collegate ¹³.

Gli effetti del *Social Credit System* possono riguardare una serie di attività.

a) Divieti di volo: attualmente, il divieto di volo per persone considerate non affidabili è già pratica comune in Cina ¹⁴.

b) Esclusione da scuole private: se il punteggio dei genitori dovesse essere al di sotto di una certa soglia, i figli di questi verrebbero esclusi dalle migliori scuole nella regione.

c) Rallentamento della connessione internet: i cittadini considerati inaffidabili potrebbero vedere la propria connessione internet rallentata o addirittura venire esclusi completamente dall'accesso a specifici siti internet.

d) Esclusione da lavori ad alto prestigio: il punteggio del SCS cinese potrebbe essere in futuro parte del *curriculum vitae* di una persona e fornire quindi informazioni essenziali a potenziali datori di lavoro. In tal modo a persone con un punteggio basso, sarebbe quasi sicuramente precluso il raggiungere posizioni lavorative più ambite e qualificate.

e) Registrazione su una *blacklist* pubblica: un prototipo di *blacklist* esiste attualmente in Cina e tutti coloro i quali vengono registrati su tale lista potrebbero aspettarsi una o più delle penalità menzionate in precedenza.

Un altro passaggio cruciale nel processo di costruzione del SCS è rappresentato anche dal parallelo sistema premiale che viene prospettato ai cittadini volto a incentivare questi ultimi a comportarsi secondo le regole per potere godere di facilitazioni in riferimento allo stesso tipo di attività che agli altri vengono impedito. Il "circolo virtuoso" che ne consegue si sostanzia soprattutto nel vantaggio che lo Stato cinese non è tenuto a fornire alcun intervento diretto né di controllo né correttivo su questo tipo di attività ¹⁵. Tra le misure premiali, si possono ricordare l'accesso facilitato a finanziamenti (nel progetto pilota di *Sesame Credit*, coloro che raggiungono un punteggio di 600, possono richiedere un prestito fino ad un ammontare di 5.000 Yuan) e ad affitti o noleggi (tra cui l'esenzione dall'obbligo di deposito cauzionale), agevolazioni degli spostamenti (a seconda del punteggio, vengono richiesti meno documenti per giustificare la richiesta di

¹³ *China's Chilling 'Social Credit' Blacklist*, su *Human Rights Watch*, 12 dicembre 2017.

¹⁴ *China Releases Investigative Journalist After Almost Year in Jail*, su *Newsweek*, 3 agosto 2014. È il caso di chi ha

scritto, ad esempio, post critici nei riguardi di dirigenti del partito.

¹⁵ M. V. BLOMBERG, *The Social Credit System in China's Rule of Law*, in *Mapping China Journal*, n. 2, 2018, 78, 162, spec. 93 ss.

visti di viaggio) e l'accrescimento di *status* sociale che segue al suo riconoscimento ufficiale (il proprio punteggio può venir usato come *status symbol* su piattaforme *social* ed app di incontri)¹⁶.

3. I PAESI BASSI: IL *TOESLAGENAFFAIRE*.

Sebbene il caso cinese di profilazione su base reputazionale rimanga certamente il più estremo, nondimeno esso costituisce solo una parte di un fenomeno più ampio. Non si può non notare, infatti, che, favorito dal crescente potere dell'algoritmo, l'utilizzo del *social scoring* sul piano pubblicitario abbia trovato delle prime applicazioni su scala globale e, persino, in Europa. Il caso più eclatante, che nel gennaio 2021 portò alle dimissioni dell'allora Primo Ministro Mark Rutte, riguarda il c.d. *Toeslagenaffaire*, in ragione del quale fu portato alla luce che l'amministrazione fiscale olandese e, in particolare, la sua unità Toeslagen (il dipartimento del *welfare*) aveva illegittimamente richiesto il rimborso retroattivo degli assegni di sussidio percepiti da circa 35.000 genitori. È stato proprio l'utilizzo dell'automazione illimitata di algoritmi di *machine-learning* ad aver giocato un ruolo significativo nello scandalo che, per giunta, ha provocato alla finanza pubblica olandese un danno di circa 500 milioni di euro, rivelandosi così l'antieconomicità dell'intera operazione¹⁷.

3.1. *I fatti.*

Verso la fine del 2011 la stampa olandese rivelò una serie di casi di frode alla previdenza sociale aventi ad oggetto versamenti effettuati dall'amministrazione fiscale in favore di persone che risiedevano in altri Stati membri dell'UE. Questa serie di casi, battezzata dalla stampa *Bulgarenfraude* (frode bulgara)¹⁸, ha rappresentato l'incentivo per riformare il sistema di *welfare*, al fine di aumentare l'efficienza del processo di individuazione delle frodi e di recupero degli emolumenti non dovuti. Tale efficientamento è stato perseguito attraverso diversi mezzi tra cui l'implementazione di algoritmi di *machine-learning* per automatizzare l'identificazione di errori colposi o di truffe nell'elaborazione dei moduli di previdenza sociale trasmessi dai genitori. In pratica

¹⁶ Come riporta S. JOHNSON, *How China's 'social credit score' will punish and reward citizens*, su *Big Think*, 26 aprile 2018, il sito Baihe, ad esempio, permette già ora ai propri utenti di pubblicare il proprio punteggio.

¹⁷ Secondo i calcoli dell'Antoriteit Persoonsgegevens, *Werkwijze Belastingdienst in strijd met de wet en discriminerend*, 17 luglio 2020, gli eventi avrebbero provocato

perdite di incassi netti di circa 10 milioni di euro per il *Belastindienst*, mentre il Governo olandese è dovuto correre ai ripari creando un fondo di 500 milioni per risarcire le vittime.

¹⁸ L'appellativo era legato al fatto che alcune di queste frodi riguardavano famiglie di nazionalità bulgara: *Uitspraak Bulgarenfraude: hoe zat het ook alweer?*, su *rtl nieuws*, 19 maggio 2015.

però le stesse sanzioni, ovvero la sospensione degli assegni e la richiesta di restituzione di quanto indebitamente percepito, sono state irrogate non solo per reati di falso ideologico dolosi, ma anche per errori materiali e semplici sviste formali, come ad esempio l'erronea compilazione di una casella o l'omissione di una firma. Ciò è dovuto al fatto che l'art. 26 della legge sui regimi di reddito (AWIR) non contemplava l'obbligo per l'amministrazione di applicare il principio di proporzionalità¹⁹. L'impiego di questo algoritmo di AI per il rilevamento del rischio "a tolleranza zero" ha obbligato, dunque, la restituzione integrale del debito senza possibilità di una sua dilazione temporale e spesso con l'aggiunta del pagamento di una mora, anche ad individui a cui si potevano imputare solo errori materiali nella compilazione dei moduli di richiesta del contributo previdenziale.

Il Dipartimento del *welfare*, inoltre, aveva utilizzato un algoritmo di *machine-learning* anche nel processo di valutazione del rischio per la selezione dei beneficiari degli assegni per l'assistenza all'infanzia da sottoporre a specifiche verifiche. Come è tipico degli algoritmi di *machine-learning*, il sistema ricavava l'esistenza di fattori di rischio sulla base dell'analisi dei dati storici, ovvero di precedenti casi noti di frode. Si tratta, come è noto da tempo, di un metodo che non è scevro da effetti distorsivi come il produrre pregiudizi e discriminazioni, che a tal fine richiede un duplice protocollo di garanzie: innanzitutto i dati elaborati sia durante che dopo il processo di apprendimento automatico devono rappresentare accuratamente la popolazione *target*, ovvero il destinatario della misura di *welfare*; in secondo luogo i fattori di rischio (i cosiddetti "pesi" nella terminologia del *machine-learning*) devono essere il più possibile liberi da qualsiasi tipo influenza che riguardi la condotta genericamente considerata tenuta dalle minoranze o da individui di *status* socio-economico inferiore. Non a caso, quando si dice che gli algoritmi (e anche gli algoritmi predittivi) si basano su decisioni già prese, ciò significa che non è l'algoritmo a contenere il *bias*, poiché ad essere discriminatoria è la sequenza storica delle decisioni che sta a monte. Per queste ragioni, l'aggiramento di tale distorsione richiede vigilanza, monitoraggio costante, verifiche a campione e anche un certo grado di scetticismo nei confronti delle conclusioni dell'algoritmo²⁰.

Nel *Toeslagenaffaire* l'Autorità olandese per la protezione dei dati (AP) e il Servizio nazionale di audit (ADR) hanno dimostrato

¹⁹ V. anche Venice Commission CoE, *Opinion no. 1030/2021*, 5. La legittimità di tale legge era stata confortata in più occasioni dalla Corte amministrativa suprema olandese che aveva ritenuto lecita la man-

cata applicazione del principio di proporzionalità alla gravità del fatto.

²⁰ Parlementaire ondervragingscommissie Kinderopvangtoeslag, *Tweede Kamer der Staten-Generaal* 11 luglio 2020.

come l'atteggiamento tenuto dall'amministrazione fiscale fosse in realtà di segno completamente opposto. L'unità *Toeslagen* non avrebbe vigilato con prudenza né avrebbe verificato se l'algoritmo fosse privo di *bias*, inducendo indirettamente l'algoritmo a restituire risultati distorti. A riprova di ciò vi è il cosiddetto “*twin test*”²¹ effettuato dalla AP e da cui è emerso che, a parità di altre condizioni, l'algoritmo segnalava un rischio di frode più elevato da parte degli individui che non erano di nazionalità olandese rispetto a cittadini olandesi che presentavano caratteristiche analoghe²².

Sommandosi a poco a poco, questi arbitrari *targeting* manuali di stranieri nel sistema di *machine-learning* — che apprende e aggiorna costantemente i fattori di rischio sulla base dei dati storici inseriti dai funzionari del Fisco — hanno indotto una vera e propria distorsione nell'algoritmo, che ha concluso euristica-mente che i beneficiari del *welfare* di origine e cittadinanza straniera fossero più inclini alla frode. Ancora più problematico è il fatto che i beneficiari del *welfare*, ritenuti potenziali truffatori, avessero subito l'interruzione dell'indennità senza nemmeno alcuna valutazione *ex post* da parte di operatori umani.

3.2. Aspetti giuridici rilevanti.

Il primo insegnamento da trarre dal *Toeslagenaffaire* è che non si può automatizzare un processo che ha delle tare originarie, poiché l'automatizzazione ne amplificherà solo la portata. Algoritmi di apprendimento automatico sono stati utilizzati nel *Toeslagenaffaire* per ridurre i gravi arretrati nel processo di recupero dell'assegno per l'assistenza all'infanzia. Tuttavia, poiché il processo di recupero era in primo luogo gravemente viziato — segnatamente a causa della mancanza dell'applicazione del principio di proporzionalità in caso di errori materiali da parte dei beneficiari del *welfare* —, gli algoritmi hanno determinato un risultato completamente opposto a quanto previsto: hanno cioè ulteriormente aumentato l'arretrato presso il *Belastingdienst* e il

²¹ Un “*twin test*” è una procedura semplice in cui si creano due profili fittizi identici, tranne che per una caratteristica, in questo caso la caratteristica era “Olandese/non Olandese”.

²² Così Autoriteit Persoonsgegevens, *Belastingdienst belooft ambtenaren niet te straffen om toeslagenaffaire*, 20 maggio 2019, secondo cui, in sostanza, l'algoritmo discriminava i residenti stranieri. Nel suo rapporto, l'AP descrive un caso particolarmente inquietante, in cui, in base alle segnalazioni riguardanti uno specifico istituto di assistenza all'infanzia, in cui circa 100 ge-

nitrici di origine ghanese erano sospettati di frode, l'unità *Toeslagen* ha arbitrariamente deciso di indagare tutti i 6.047 genitori di origini ghanesi residenti nei Paesi Bassi. In tali condizioni, dove le presunte decisioni “obiettive” perché “guidate dai dati”, sono in realtà arbitrarie, parziali e discriminatorie, è difficile negare che la macchina abbia adottato una decisione discriminatoria come forse avrebbero fatto anche i funzionari umani del fisco. Peraltro il caso di genitori ghanesi non è l'unico in cui gli stranieri sono stati oggetto di una decisione arbitraria da parte della dirigenza del *Belastingdienst*.

numero di ricorsi promossi dai beneficiari del *welfare* contro l'unità *Toeslagen* ²³.

La seconda conclusione è che gli Stati dovrebbero utilizzare il *machine-learning* e l'automazione predisponendo *ex ante* le condizioni giuridiche e tecnologiche affinché questi algoritmi possano essere utilizzati in modo lecito ed etico. Ben prima del *Toeslagenaffaire*, era stato più volte segnalato come l'apprendimento automatico fosse portato a creare seri rischi di pregiudizi e discriminazioni, come è emerso più volte nel corso degli ultimi anni, ad esempio, con riferimento al diritto a un processo equo, a una buona amministrazione e ad altri principi ricavabili da norme costituzionali o interposte ²⁴. A partire da ciò, l'adozione di algoritmi di *machine-learning*, in assenza di tutele rafforzate e della protezione di questi diritti, è inevitabilmente foriero di effetti perversi.

Il che porta alla terza lezione ricavabile dal caso olandese: il necessario controllo da parte dell'operatore umano. L'adozione sistematica di strumenti di *machine-learning* coincide all'incirca con la crisi finanziaria globale del 2008-2011, in cui le misure di austerità hanno portato alla riduzione della forza lavoro delle amministrazioni fiscali nell'intera Unione europea. Ciò dovrebbe essere di per sé motivo di preoccupazione, poiché è probabile che una riduzione del personale incida sul numero dei funzionari addetti ai reclami *ex post* dei contribuenti, e quindi sulla loro capacità di riesame. Così facendo non si stanno creando le condizioni per un'IA antropocentrica, ma piuttosto un sistema decisionale incentrato sull'intelligenza artificiale incontrollata, del tutto antitetico rispetto agli obiettivi che si intende promuovere con i progetti di *governance* algoritmica di stampo comunitario.

La vicenda riportata è, dunque, di grande interesse in quanto è facile immaginare che per i presupposti che l'hanno determinata non sia un fenomeno isolato, ma che possa al contrario ripetersi in altri Stati membri dell'UE.

4. L'ESPERIENZA ITALIANA. IL RATING REPUTAZIONALE NELLA GIURISPRUDENZA DEL GARANTE PER LA PROTEZIONE DEI DATI PERSONALI.

La proliferazione di applicazioni o di banche dati private che

²³ S. RANCHORDAS, *Bureaucracy and Vulnerability in the (Digital) Administrative State*, in *I-CONnect* 9 Settembre 2020.

²⁴ Cfr. C. CASTRO, *What's Wrong with Machine Bias*, in *Ergo*, vol. 5, n. 15, 2019-2020; J. Agwin et al., *Machine Bias*, in *ProPublica*, 23 maggio 2016, online su <http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Per la dottrina italiana, E. STRADELLA, *Ste-*

reotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale, in *Consulta Online*, 20 marzo 2020, online su www.giurcost.org, in part. 9 ss.; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo* — Speciale 2019, 101-130; A. GATTI, *L'algoritmo tra volontà e rappresentazione*, in *DPCE online*, n. 3, 2020, 3457-3461.

raccogliono ed archiviano indiscriminatamente dati personali al fine di assumere decisioni anche con effetti e conseguenze negative sulla vita privata, sociale ed economica degli individui ha raggiunto anche il nostro Paese. Tali progetti aggregano informazioni provenienti da diverse fonti, non sempre di carattere oggettivo — e.g. anche *post* di profili *social* — ed estendono il modello di valutazione, ormai consolidato alle strutture ricettive o di ristorazione, direttamente anche alla valutazione delle persone.

Si tratta di sistemi che puntano, in sostanza, a rendere misurabile la reputazione dei soggetti censiti, sollevando più di una preoccupazione in tema di protezione dei dati personali, nonostante le finalità dichiarate siano quasi sempre giustificate dal perseguimento di rilevanti obiettivi quali, ad esempio, la lotta alla corruzione, alle false identità, alle frodi, ecc. Ma l'idea stessa di affidare la "recensione" di una persona ad un algoritmo rischia di aprire davvero una deriva assai pericolosa. Sicché s'impone sin da subito una riflessione che vada oltre l'applicazione tecnicistica delle norme del GDPR (ed in particolare dall'art. 22, in materia di decisioni automatizzate comprendenti anche la profilazione e successive in tema di sicurezza), per soffermarsi piuttosto sui principi di carattere generale da questo sanciti.

La reputazione è un elemento complesso di un'intera personalità, corrispondente a più aspetti che interessano molteplici dati personali anche significativi e sensibili, pertanto come tale è protetta dal GDPR e dal Codice privacy ed è garantita da un'Autorità indipendente. Il pregiudizio alla reputazione è espressamente annoverato tra i possibili danni conseguenti ad una violazione di dati personali dell'interessato, ai sensi del Considerando 85 del GDPR. Un caso emblematico può verificarsi proprio nell'ipotesi di pubblicazione di dati personali trattati per finalità di profilazione.

È evidente che la reputazione degli individui che si vorrebbe quantificare per attribuire punteggi di affidabilità è strettamente legata alla persona considerata nella sua proiezione sociale e risulta intimamente connessa con la dignità, elemento cardine della disciplina in materia di protezione dei dati personali, che guarda proprio all'impatto di trattamenti sconsiderati, specie se automatizzati e di profilazione, sui diritti e le libertà degli individui, prestando attenzione a che l'interessato possa sempre esercitare il proprio controllo sui suoi dati trattati da altri.

Molti dubbi riguardano l'effettiva qualità dei dati raccolti — destinati appunto ad essere elaborati da specifici algoritmi — o la genuinità delle eventuali recensioni inserite da terzi che possono facilmente prestarsi ad usi distorti, diffamatori o lesivi. Giudizi che rischiano di compromettere gravemente il diritto all'integrità dell'identità personale e i cui effetti pregiudizievoli vengono irrimediabilmente amplificati dal libero accesso degli utenti, attra-

verso i motori di ricerca. E cristallizzati nell'ipermnesia della rete per un tempo indefinito.

Impongono valutazioni ulteriori sia le concrete modalità di gestione di tali sistemi di *rating*, sia la discrezionalità dei criteri di misurazione individuati da parte delle società che intendono offrire sul mercato tali servizi. Non è un mistero che in passato i gestori, in alcuni casi, hanno addirittura consapevolmente alterato il mercato di riferimento inserendo false recensioni²⁵. Si tratta di progetti e pratiche che, a fronte delle diverse criticità sollevate, hanno un notevole impatto sulla sfera personale degli utenti, sul loro diritto all'autodeterminazione informativa. Dimensione questa la cui centralità è stata ribadita dalla giurisprudenza della Corte di giustizia dell'Unione europea, che dal 2014 ha riconosciuto l'esercizio di un nuovo diritto all'oblio digitale²⁶.

Un essere umano non è una ragione sociale, un asset commerciale, un servizio e non può e non deve diventarlo²⁷. La Corte di Lussemburgo ha stabilito che, in caso di conflitto, il mero interesse economico soccombe, quando ricorrono determinate condizioni, rispetto alla categoria di diritti fondamentali che sono espressione dei principi di dignità contenuti nella Carta dei diritti²⁸.

4.1. Il caso Mevaluate.

Con il provvedimento n. 488/2016, il Garante Privacy ha giudicato come non lecita la piattaforma dell'Associazione "Mevaluate Onlus" che raccoglieva numerose categorie di dati personali (attinenti alle seguenti sfere: penale, fiscale, civile, lavoro ed impegno civile, studi e formazione) per il calcolo di un *rating* reputazionale del richiedente (persona fisica o giuridica), al fine di « rendere maggiormente efficienti, trasparenti e sicuri i rapporti socio-economici ». Tali informazioni erano acquisite sulla base del consenso degli interessati e sono fornite da fonti diverse dagli interessati, eventualmente integrate da « articoli stampa, radio/

²⁵ *Recensioni false sul web e trasparenza nell'informazione*, su dirittodellinformazione.it, 22 febbraio 2022.

²⁶ CGUE sent. del 13 maggio 2014, *Google Spain SL e Google Inc. contro Agencia Española de Protección de Datos (AEPD) e Mario Costeja González* (ECLI:EU:C:2014:317).

²⁷ Il Presidente Antonello Soro già nel 2015 sosteneva: « Sulla base di tali presupposti, diventa allora difficile implementare banche dati o servizi finalizzati alla profilazione di una persona anche sul piano morale e relazionale per valutarne — eventual-

mente — l'asserita affidabilità. L'ordinamento europeo tutela la privacy, riconosciuta come diritto fondamentale dell'uomo e vieta di "recensire" una qualunque persona in assenza di un interesse pubblico che certo non può essere invocato quando si entra nel mondo delle relazioni personali e sentimentali di chiunque disponga di un account delle reti sociali ». Antonello Soro, *La reputazione online tra principi di dignità e diritto all'oblio*, Huffington Post, 6 ottobre 2015.

²⁸ CGUE sent. del 17 marzo 2005, *Karl Robert Kranemann c. Land Nordrhein-Westfalen* (ECLI:EU:C:2005:187).

TV » e altri documenti a supporto prodotti facoltativamente dagli interessati.

A titolo di premessa il Garante ha osservato che il *rating* elaborato dal sistema « potrebbe ripercuotersi pesantemente sulla vita (anche privata) degli individui censiti, influenzandone scelte e prospettive e condizionando la loro stessa ammissione a (o esclusione da) specifiche prestazioni, servizi o benefici »; e che occorre affrontare la questione con estrema cautela, dal momento che « la reputazione che [la piattaforma] vorrebbe misurare, [...] strettamente correlata alla considerazione delle persone e alla loro stessa proiezione, sociale, risulta intimamente connessa con la loro dignità, elemento cardine della disciplina di protezione dei dati personali ».

Con riferimento alla liceità del trattamento il Garante ha sottolineato che il trattamento non si basava su un'ideale fonte di regolazione, dal momento che l'ordinamento riconosce e regola esclusivamente il "*rating* d'impresa"²⁹ e non già quello relativo a persone fisiche. Inoltre, il consenso degli interessati, per essere conforme a legge, avrebbe dovuto essere manifestato liberamente, presupposto che non ricorre nel caso di autonoma elaborazione di documenti liberamente utilizzabili (e.g. le sentenze) da parte di un sistema automatizzato, sia con riferimento ai profili relativi agli utenti che hanno espressamente acconsentito al trattamento, sia con riferimento ai "profili contro terzi" relativi a soggetti non iscritti alla piattaforma; senza contare che il trattamento si basava su un'informativa inadeguata.

Il trattamento in esame appariva poi in violazione dei principi di minimizzazione dei dati, poiché si basava su una raccolta massiva di dati e documenti, e di proporzionalità del trattamento, in quanto la rilevanza e la pertinenza dei dati e documenti raccolti appariva dubbia e, comunque, indimostrata ed il periodo di conservazione dei dati non era adeguatamente giustificato. Infine, il trattamento, coinvolgendo un numero potenzialmente molto elevato di soggetti, avrebbe avuto attendibili significative ripercussioni per i diritti degli interessati in caso di violazione delle misure di sicurezza e queste ultime non erano dotate di una sufficiente affidabilità.

L'Associazione Mevaluate ha impugnato il provvedimento del Garante davanti al Tribunale di Roma che, con sentenza n. 5715/2018, ha stabilito che la mancanza di una disciplina normativa sul *rating* reputazionale non ostasse allo sviluppo di infrastrutture capaci di attribuirlo. Di avviso opposto rispetto al Garante, che aveva ritenuto che il complesso sistema di raccolta e

²⁹ Gestito peraltro dall'Autorità Nazionale Anti-Corruzione sulla base dell'art.

83, comma 10 del d.lgs n. 50 del 18 aprile 2016.

di trattamento dei dati personali in oggetto sarebbe stato in grado di incidere sia sulla rappresentazione economica e sociale di un'ampia categoria di soggetti, sia sulla vita privata degli individui censiti, il giudice ordinario rilevava come siano ormai largamente diffusi gli organismi privati di valutazione e di certificazione, riconosciuti anche a fini di attestazione di qualità e/o di conformità a norme tecniche.

Il Tribunale, stabilendo che « non può negarsi all'autonomia privata la facoltà di organizzare sistemi di accreditamento di soggetti, fornendo servizi in senso lato 'valutativi', in vista del loro ingresso nel mercato, per la conclusione di contratti e per la gestione di rapporti economici », ha considerato così il trattamento conforme al requisito di liceità di trattamento, partendo dall'assunto che « le attività di caricamento delle informazioni e di validazione e certificazione dei documenti sono soggette al consenso dell'interessato e alla volontarietà della sua azione ». Sulla base di tali argomentazioni, pertanto, ha annullato il provvedimento del Garante, fermo il divieto di trattare i dati personali di soggetti non iscritti alla piattaforma, ancorché tratti da documenti liberamente conoscibili.

Con riferimento, invece, ai profili contro terzi e alle altre questioni sollevate dal Garante, il Tribunale ha confermato l'illiceità del trattamento posto in essere da Mevaluate.

Sulla questione della validità del consenso prestato dall'interessato si è pronunciata, per ultima, la Corte di Cassazione, la quale, nella sentenza 14381/2021, ha accolto il ricorso del Garante e ha espresso il seguente principio di diritto: « in tema di trattamento di dati personali, il consenso è validamente prestato solo se espresso liberamente e specificamente in riferimento a un trattamento chiaramente individuato; ne segue che nel caso di una piattaforma web (con annesso archivio informatico) preordinata all'elaborazione di profili reputazionali di singole persone fisiche o giuridiche, incentrata su un sistema di calcolo con alla base un algoritmo finalizzato a stabilire i punteggi di affidabilità, il requisito di consapevolezza non può considerarsi soddisfatto ove lo schema esecutivo dell'algoritmo e gli elementi di cui si compone restino ignoti o non conoscibili da parte degli interessati » ³⁰.

Di fatto, la Corte ha sottolineato che ai fini della liceità del trattamento basato sul consenso, l'art. 23 del d.lgs n. 196/2003 (i.e. il Codice privacy pre-GDPR, vigente all'epoca dei fatti come normativa di riferimento) presupponeva non solo il consenso *tout court*, ma anche che il consenso fosse validamente prestato ³¹. Nello specifico, difatti, l'art. 23 disponeva che « (a) il trattamento

³⁰ Provvedimento del Garante per la protezione dei dati personali n. 488 del 24 novembre 2016 [doc. web n. 5796783].

³¹ v. anche Corte di Cassazione, sentt. n. 17278/2018 e n. 16358/2018.

di dati personali da parte di privati o di enti pubblici economici è ammesso solo con il consenso espresso dell'interessato; (b) il consenso può riguardare l'intero trattamento ovvero una o più operazioni dello stesso; (c) il consenso è validamente prestato solo se è espresso liberamente e specificamente in riferimento ad un trattamento "chiaramente individuato", se è documentato per iscritto, e se sono state rese all'interessato le informazioni di cui all'art. 13; (d) il consenso è manifestato in forma scritta quando il trattamento riguarda dati sensibili ».

Nel caso di specie, quindi, il trattamento era (ed è) funzionale alla determinazione del profilo reputazionale dei soggetti e, di conseguenza, la valutazione di liceità di un simile trattamento, basata sul consenso, non poteva essere prospettata dal Tribunale senza una previa considerazione degli elementi suscettibili di incidere sulla serietà della manifestazione, compresi gli elementi implicati e considerati nell'algoritmo afferente, il funzionamento del quale è essenziale al calcolo del *rating*.

In particolare la Cassazione ha evidenziato che la scarsa trasparenza dell'algoritmo impiegato allo specifico fine non è stata ritenuta decisiva dall'impugnata sentenza, la quale ha semplicemente affermato la scarsa incisività dei dubbi relativi al sistema automatizzato di calcolo per la definizione del *rating* reputazionale, sul rilievo che la validità della formula riguarderebbe « il momento valutativo del procedimento », a fronte del quale spetterebbe invece al mercato « stabilire l'efficacia e la bontà del risultato ovvero del servizio prestato dalla piattaforma ». Tale impostazione non è, quindi, condivisa dalla Suprema Corte la quale ha sostenuto che il presupposto della liceità del trattamento è proprio costituito dalla validità del consenso che si assume prestato al momento dell'adesione. E non può logicamente affermarsi che l'adesione a una piattaforma da parte dei consociati comprenda anche l'accettazione di un sistema automatizzato, che si avvale di un algoritmo, per la valutazione oggettiva di dati personali, laddove non siano resi conoscibili lo schema esecutivo in cui l'algoritmo si esprime e gli elementi all'uopo considerati.

Come ben noto le osservazioni della Suprema Corte, pertinenti e precise, sono state ampiamente confermate dal GDPR dove l'art. 4, par. 11, stabilisce che il consenso dell'interessato, in modo inequivocabile, deve essere innanzitutto libero poiché se l'interessato non compie una scelta reale e si sente obbligato a prestare il proprio consenso anche per evitare conseguenze negative nel caso rifiutasse, allora il consenso non potrà essere considerato come valido. Se, ad esempio, il consenso è inserito in una parte non negoziabile di termini e condizioni si presume che lo stesso non sia stato fornito liberamente. Inoltre il consenso deve essere specifico, poiché se utilizzato per giustificare molteplici trattamenti deve essere dato liberamente per ciascuno. Gli interessati devono essere in grado di scegliere per quali scopi accon-

sentono il trattamento. Infine, il consenso deve essere informato: senza informazioni accessibili, gli interessati non possono prendere decisioni informate e quindi, chiarisce il WP29 (nelle proprie linee guida del 2017), « il controllo dell'utente diventerebbe illusorio e il consenso non valido per l'elaborazione ».

Con riferimento poi, alle forme di profilazione automatica, va evidenziato che in ambito di decisioni basate unicamente su un trattamento automatizzato, come previsto dal già menzionato art. 22, il Regolamento introduce la necessità di fornire all'interessato maggiori informazioni sulle modalità di creazione ed utilizzo di questi processi. A riguardo, difatti, il GDPR ravvisa possibili rischi significativi per i diritti e le libertà degli individui riconnessi alla tendenziale opacità dei processi e meccanismi automatizzati, che porta spesso l'individuo, oggetto di profilazione, a non esserne a conoscenza ed alla creazione, da parte del titolare, di dati nuovi, aggiuntivi rispetto agli originali che potrebbero ridurre l'interessato ad una categoria a cui non si riconosce condizionando così le sue scelte e, in alcuni casi, portando anche a forme di discriminazione.

Pertanto, il GDPR, per correggere questo dislivello informativo tra titolare e interessato ed evitare pregiudizi alla sfera giuridica di quest'ultimo, individua una serie di requisiti su cui concentrarsi per rendere questi trattamenti automatizzati conformi alla normativa: specifiche prescrizioni in tema di trasparenza e correttezza; maggiori obblighi di *accountability*; basi giuridiche specifiche per la legittimazione del trattamento; garanzie per gli individui in tema di diritto di opposizione alla profilazione e, in particolare, alla profilazione per finalità di marketing; esecuzione di una valutazione d'impatto sulla protezione dei dati laddove non siano soddisfatte certe condizioni.

In particolare, l'art. 13, par. 2, lett. f) e l'art. 15, par. 1, lett. h) stabiliscono il diritto dell'interessato di conoscere l'esistenza del processo decisionale automatizzato e, in particolare, di ottenere informazioni significative sulla logica utilizzata (i criteri assunti per raggiungere la decisione, senza che con ciò si debba necessariamente fornire una spiegazione complessa degli algoritmi utilizzati) e sulle conseguenze previste di tale trattamento (attraverso esempi bisognerà fornire informazioni su come il processo automatizzato potrebbe influenzare in futuro la persona interessata).

Tenuto conto dei rischi rilevanti sui diritti e libertà dell'interessato per queste tipologie di trattamento, il Regolamento, da un lato obbliga il titolare ad attuare misure appropriate e rafforzate di tutela (importante sarà anche prevedere modalità che verifichino con regolarità la correttezza dei processi per limitare errori di classificazione o valutazione con impatto negativo sui soggetti profilati), dall'altro lato, riconosce il potere all'interessato di ottenere l'intervento umano da parte del titolare, di esprimere la

propria opinione e di contestare la decisione, nei casi in cui tale decisione sia prevista per contratto o consentita dall'interessato (art. 22, par. 3). Tra l'altro, un processo decisionale automatizzato che coinvolga categorie particolari da dati, di cui all'art. 9, par. 1, è consentito solo in presenza del consenso esplicito dell'interessato o per motivi di interesse pubblico rilevante sulla base del diritto dell'Unione o degli Stati membri.

Vale la pena ricordare, infatti, quanto stabilito dall'art. 22, par. 1 GDPR: « l'interessato ha diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona ».

Per il Gruppo dei Garanti europei ³², pertanto, l'art. 22, par. 1, istituisce un divieto per quel processo decisionale individuale completamente automatizzato, compresa la profilazione, che abbia un effetto legale o analogo sull'interessato. Per « decisione basata unicamente sul trattamento automatizzato » si deve intendere una decisione presa senza il coinvolgimento di un essere umano che possa influenzare ed eventualmente cambiare il risultato attraverso la sua autorità o competenza.

Perché sia riconosciuto il diritto dell'interessato a non essere sottoposto ad una decisione basata unicamente sul trattamento automatizzato è necessario che tale decisione « produca effetti giuridici o incida in modo analogo significativamente sulla sua persona ». Il riferimento agli “effetti giuridici” concerne, ovviamente, l'impatto che una decisione automatizzata può produrre sulla sfera giuridica dell'individuo (e.g. penalizzando il diritto di associazione, di voto, di libertà negoziale, di libera circolazione eccetera). Oltre a questo, la norma apre anche alle circostanze che “in modo analogo” possono potenzialmente e significativamente influenzare i comportamenti e le scelte degli individui interessati.

Il Considerando 71 del GDPR cita, come esempi di decisioni automatizzate che possono incidere sui diritti e le libertà degli individui in maniera rilevante, il rifiuto automatico di una domanda di credito online o pratiche di assunzione elettronica senza interventi umani.

L'art. 22 al par. 2 prevede tre eccezioni al divieto generale di un processo decisionale completamente automatizzato che determini effetti nella sfera giuridica dell'individuo: quando la decisione è necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; quando la decisione è

³² EDPB, *Guidelines on Automated individual decision-making and Profiling for*

the purposes of Regulation 2016/679, 22 agosto 2018.

autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento; quando la decisione si basa sul consenso esplicito dell'interessato.

In riferimento al primo punto, i Garanti europei chiariscono che la necessità di utilizzare decisioni automatizzate per l'esecuzione o conclusione di un contratto deve essere interpretata in modo restrittivo: ciò significa che il titolare deve essere in grado di dimostrare che la profilazione è necessaria (principio di necessità) e non sono disponibili mezzi alternativi meno invasivi (principio di proporzionalità), sulla scorta di quanto già stabilito dalla giurisprudenza CEDU, con particolare riferimento all'art. 8 della Convenzione³³. In riferimento al secondo punto, la legislazione degli Stati membri può, in casi specifici, autorizzare il ricorso ad un processo di decisione automatizzata per il monitoraggio e la prevenzione delle frodi e dell'evasione fiscale o per garantire la sicurezza e l'affidabilità di un servizio fornito dal titolare. Infine, per quanto riguarda la terza deroga, il Regolamento richiede il consenso esplicito dell'interessato, ossia confermato da una dichiarazione espressa e non desunto da comportamento concludente.

Sul fronte amministrativo, non bisogna infine dimenticare che anche il Consiglio di Stato³⁴ nel valutare la legittimità dell'adozione di un algoritmo per lo svolgimento di un'attività amministrativa ha rilevato che bisogna tenere in debita considerazione nell'esame e nell'utilizzo degli strumenti informatici tre principi fondamentali. In primo luogo, il principio di conoscibilità, per cui ognuno ha diritto a conoscere l'esistenza di processi decisionali automatizzati che lo riguardino ed in questo caso a ricevere informazioni significative sulla logica utilizzata. Il secondo principio è definibile come il principio di non esclusività della decisione algoritmica. Nel caso in cui una decisione automatizzata "produca effetti giuridici che riguardano o che incidano significativamente su una persona", questa ha diritto a che tale decisione non sia basata unicamente su tale processo automatizzato. Il terzo principio è quello della non discriminazione algoritmica, secondo cui è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, mettendo in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori.

³³ Cfr. *S & Marpel v. UK*, ricorsi nn. 30562 e 30566, 4 dicembre 2008, §§ 35 e 119.

³⁴ Consiglio di Stato, sentenza n. 8472/2019.

4.2. *L'ambito scolastico.*

Utile ad arricchire il quadro è la questione del *rating* reputazionale in ambito scolastico.

Il Garante per la protezione dei dati personali ha inviato una richiesta di informazioni all'associazione Crop News (dove "Crop" è un acronimo per Cronache reputazionali oggettive personalizzate, che opera all'interno del gruppo Mevaluate). Notizie di stampa rivelavano che l'associazione, nell'ambito del Progetto "Virtute 4 Students", sperimentasse nei confronti degli studenti il *rating* reputazionale elaborato sulla base di algoritmi dalla stessa Piattaforma Mevaluate e che al progetto avesse aderito un istituto di istruzione superiore.

Il Garante, considerata la delicatezza del progetto che si rivolge a soggetti particolarmente vulnerabili (studenti e minori), ha chiesto chiarimenti all'associazione, in particolare il funzionamento della piattaforma e della connessa banca dati al fine di consentire all'Autorità di valutare l'impatto dell'uso degli algoritmi e gli effetti che essi possono determinare sugli studenti, nonché le misure eventualmente adottate a loro tutela ³⁵.

L'Associazione ha dichiarato che il fine dell'iniziativa fosse quello di stimolare gli studenti a costruirsi una reputazione reale che coincidesse con la reputazione virtuale attraverso il proprio *rating* reputazionale digitalizzato per promuovere un modello economico comportamentale trasparente. Virtute calcola le singole reputazioni pubblicate dal periodico *online* Crop News, promuovendo quindi comportamenti legali fondati sul timore di comparire sulla testata. La riduzione di illeciti e inadempimenti contrattuali garantirebbe sia gli operatori economici, che consumatori e utenti, ma in ragione del "ricatto" morale di un *rating* reputazionale automatizzato, che peraltro attribuisce compensi variabili in relazione al contributo fornito da ciascuno per qualificare la reputazione documentata e tracciabile propria e delle controparti nei rapporti obbligatori.

Sulla base delle informazioni fornite e di ulteriori indagini, l'Autorità ha ritenuto il trattamento illecito, in violazione delle garanzie obbligatorie dell'art. 22 del GDPR e, soprattutto, contrario ai principi di liceità (poiché effettuato in assenza di una base giuridica sufficiente), correttezza (poiché in grado d'ingenerare discriminazioni e pregiudizi irragionevoli e sproporzionati), trasparenza (in quanto effettuato in assenza di un'informativa adeguata), minimizzazione (trattando quantità di dati indeterminate ed indeterminabili *a priori*), limitazione (poiché produce

³⁵ "Scuola: *rating* reputazionale sotto la lente del Garante privacy. Avviata istruttoria su una piattaforma rivolta agli stu-

denti", comunicato del Garante per la protezione dei dati personali 3 maggio 2022.

nuovi dati da potersi trattare per il perseguimento di finalità diverse e non necessariamente lecite), e limitazione della conservazione (affidando i dati ad una testata *online*). Crop News è stata quindi ammonita, mentre si è aperta una nuova istruttoria nei confronti di Mevaluate, collettore di tutti i dati raccolti.

5. RIFLESSIONI CONCLUSIVE. DALL'ALGOCRAZIA ALL'ALGORETICA.

Le esperienze e gli esempi mostrati nel corso del contributo ci permettono una panoramica sulle pratiche realizzazioni, per lo più sperimentali, che interessano ad oggi la profilazione reputazionale. Tuttavia non c'è bisogno di dire che, per le sue implicazioni, il tema necessita di essere affrontato ad un livello più alto, cioè a livello dei suoi fondamenti teorici e filosofici che ci portano a fare i conti con il fenomeno della c.d. algocrazia, ovvero della crescente dipendenza da algoritmi e processi decisionali automatizzati in vari aspetti della società. Questa tendenza è stata guidata dai progressi dell'intelligenza artificiale e del *machine learning*, che hanno reso possibile automatizzare attività che in precedenza erano ritenute dominio esclusivo del giudizio e delle competenze umane. Tuttavia, poiché l'uso degli algoritmi è diventato più diffuso, molte sono le preoccupazioni sorte circa le implicazioni etiche di questi sistemi e il loro impatto sugli individui e sulla società nel suo insieme.

In risposta a questi interrogativi, si è assistito ad uno spostamento dell'attenzione sull'algoetica, ovvero lo studio delle implicazioni etiche degli algoritmi e dei processi decisionali automatizzati. L'algoetica cerca di affrontare questioni come il potenziale per gli algoritmi di perpetuare e amplificare le disuguaglianze e i pregiudizi esistenti, nonché la mancanza di trasparenza e responsabilità in questi sistemi.

Per affrontare questi e altri profili etici, si registra una spinta per lo sviluppo di linee guida, appunto etiche, e di meccanismi per la supervisione e la responsabilizzazione nell'uso degli algoritmi. Ciò include gli sforzi per garantire che i dati utilizzati per addestrare gli algoritmi siano diversi e rappresentativi e per progettare algoritmi che risultino trasparenti e spiegabili nei loro processi decisionali.

Nel complesso, il passaggio dall'algocrazia all'algoetica riflette un crescente riconoscimento della necessità di considerare le implicazioni etiche degli algoritmi e dei sistemi decisionali automatizzati. Sebbene queste tecnologie abbiano il potenziale per migliorare l'efficienza e l'obiettività nel processo decisionale, è importante garantire che vengano utilizzate in modo da promuovere l'equità e la trasparenza. Ciò richiederà un'attenzione costante alle implicazioni etiche e sociali dell'algocrazia e un impe-

gno ad affrontare eventuali conseguenze negative che potrebbero sorgere ³⁶.

Machine-learning, AI e *Big Data* stanno indubbiamente cambiando il modo di leggere e gestire la realtà. Queste tecnologie non solo aiutano l'uomo incrementandone le potenzialità, ma in sempre maggiori situazioni danno luogo a sistemi, bot o robot, completamente autonomi, che quindi lo sostituiscono. In particolare, le intelligenze artificiali sono artefatti tecnologici differenti dagli artefatti prodotti fino ad oggi. Tutti gli strumenti che abbiamo congegnato hanno consentito all'uomo di svolgere alcuni compiti potenziandone le capacità fisiche: dalle clave primitive fino alle grandi macchine industriali, tutti questi strumenti hanno funto a svolgere meglio, più velocemente ed in maniera più efficace, dei compiti precisi. Le AI, tanto nei bot quanto nei robot, superano il concetto conosciuto fino ad adesso. Tutti i meccanismi automatici costruiti durante la rivoluzione industriale sono stati progettati in funzione di quello che era il loro scopo e, pertanto, assolvevano esclusivamente a quella specifica mansione.

Oggi le AI non vengono progettate così: non sono *software* programmati, ma sistemi addestrati ³⁷, superando il modello classico "*if this then that*" in cui un ingegnere prevedeva prima tutte le possibili occorrenze. Le AI rispondono in maniera autonoma a un problema che gli viene posto. Questi artefatti sono una nuova specie nelle macchine. Delle *machine sapiens* ³⁸ che condividono il mondo con l'*homo sapiens*, ragionando però diversamente.

I *data scientist* sottolineano come il problema sia legato alla qualità ed alla quantità dei dati di cui dispongono le macchine per assumere le loro decisioni: quando avremo un *database* perfetto su cui far girare i sistemi di AI — così dicono — la macchina farà scelte "perfette".

Ma è davvero così? Già in passato abbiamo avuto questa impressione. Laplace sosteneva che qualora noi avessimo conosciuto la posizione in uno stesso istante di tutte le particelle che contengono l'universo saremmo stati in grado di predire tutto il futuro e di conoscere tutto il passato dell'universo. Ma anche se fossimo in grado di creare una mappa che è l'esatta copia della realtà, includendo al suo interno tutto, questa mappa sarebbe comunque inutile, poiché risulterebbe complessa come la realtà e, quindi, troppo complessa per assumere decisioni più efficaci di quelle che già prendiamo.

³⁶ I primi quattro capoversi del par. 5. *Conclusione. Dall'algocrazia all'algoretica*, sono stati generati da un'intelligenza artificiale: ChatGPT di OpenAI.

³⁷ Sul funzionamento e le implicazioni delle decisioni algoritmiche nei vari campi della vita sociale, v. A. ADINOLFI, A. SIMON-

CINI (a cura di), *Protezione dei dati personali e nuove tecnologie. Ricerca interdisciplinare sulle tecniche di profilazione e le loro conseguenze*, Napoli, ESI, 2022.

³⁸ P. BENANTI, *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Marietti, Bologna, 2018, 38 ss.

Ci troveremmo cioè di fronte al noto paradosso raccontato da Jorge Luis Borges in un frammento *Del rigore della scienza*, l'ultimo di *Storia universale dell'infamia* pubblicato per la prima volta nel 1935. Come sua abitudine, l'Autore argentino attribuisce la citazione a un libro che in realtà non esiste: « ... In quell'Impero, l'arte della cartografia giunse a una tal perfezione che la mappa di una sola Provincia occupava tutta una città, e la mappa dell'impero tutta una Provincia. Col tempo, queste mappe smisurate non bastarono più. I collegi dei cartografi fecero una mappa dell'Impero che aveva l'immensità dell'Impero e coincideva perfettamente con esso (Suárez Miranda, *Viajes de varones prudentes*, libro IV, cap. XIV, Lérida, 1658) »³⁹.

I dati sono una mappa della realtà, rappresentano una riduzione della realtà e per questo sono utili per prendere decisioni. Inoltre le AI, oltre che sui *database*, lavorano sui sensori che a loro volta non sono in grado di leggere tutta la realtà: ne apprendono solo una parte trasformandola in dati. Questo è il punto chiave della questione. Siccome le intelligenze artificiali fondano loro decisioni sui dati, e poiché questi non sono una copia perfetta della realtà, non è pensabile aprioristicamente che la macchina dotata di intelligenza artificiale possa fare una scelta priva di errori. La macchina *sapiens* sarà sempre e costitutivamente fallibile.

Dal momento che le intelligenze artificiali possono sbagliare, occorre comprendere come gestire tali errori, assegnando alle AI un'etica. Occorre cioè trovare un sistema etico condiviso perché l'utilizzo di questi sistemi non produca ingiustizie, non danneggi le persone e non crei dei forti disequilibri globali.

L'esistenza di macchine *sapiens* chiede di mettere in piedi un nuovo linguaggio universale che sappia tradurre queste direttrici etiche in direttive eseguibili dalla macchina e poiché la dimensione dell'era digitale è regolata dagli algoritmi, ecco che si impone l'algocrazia⁴⁰, il dominio dell'algoritmo, a fronte del quale è urgente sviluppare il linguaggio comune dell'algoretica.

Nella misura in cui vogliamo affidare competenze umane, di comprensione, di giudizio e di autonomia di azione a dei sistemi *software* di AI dobbiamo capire il valore, in termini di conoscenza e capacità di azione, di questi sistemi che pretendono di essere intelligenti e cognitivi, ed è per questo che il problema è innanzitutto filosofico ed epistemologico. Le AI "funzionano" secondo schemi che connettono dati; ma che tipo di conoscenza è questa? Che valore ha? Come va trattata e considerata?

³⁹ J.L. BORGES, *Del rigore della scienza*, in *Storia Universale dell'infamia*, Adelphi, Milano, 1997, 70.

⁴⁰ Su tutti cfr. P. BENANTI, *Oracoli. Tra algoretica e algocrazia*, Luca Sossella, Roma, 2018.

È innanzitutto necessario chiarire in che senso si parla di valore. Infatti gli algoritmi lavorano su valori di natura numerica, mentre l'etica pondera il valore morale. Servirebbe dunque stabilire un linguaggio che sappia tradurre il valore morale in un qualcosa di computabile per la macchina. La percezione del valore etico è una capacità puramente umana e la capacità di lavorare dei valori numerici è invece l'abilità della macchina. Ma nella relazione tra uomo e macchina il vero conoscitore e portatore di valore è la parte umana. La trattazione sulla dignità umana ed i diritti umani non può che insegnare come sia l'uomo da proteggere nella relazione tra uomo e macchina.

Questa evidenza ci fornisce l'imperativo etico fondamentale per la macchina *sapiens*: dubita di se stessa. Ciò significa dover mettere in grado la macchina di nutrire un certo senso di incertezza. Tutte le volte che la macchina non sa se sta tutelando con certezza il valore umano deve richiedere l'azione dell'uomo. Questa direttiva fondamentale si ottiene introducendo dei paradigmi statistici all'interno delle AI. Dev'essere questa capacità di incertezza il cuore del decidere della macchina, poiché così facendo la macchina stessa pone l'umano al centro.

Abstract

The article analyses several experiences of so-called social scoring, starting from the well-known Chinese example, and some European cases, stressing the point that the antidote to artificial intelligence classification of human beings resides in the exclusively human ethical dimension.