Scuola Internazionale Superiore
di Studi Avanzati

Il dibattito pubblico sulla governance degli algoritmi

Laboratorio Interdisciplinare per le Scienze Naturali e Umanistiche

Master di primo livello in Comunicazione della Scienza "Franco Prattico"

Candidata

Chiara Sabelli

Relatrice

Prof.ssa Mariachiara Tallacchini

Anno Accademico 2016-2017

Tesi non ancora discussa. Versione del 7 febbraio 2018.
Data della discussione: 28 febbraio 2018.

Chiara Sabelli. ***Il dibattito pubblico sulla governance degli algoritmi.***
Tesi di Master di primo livello. Scuola Internazionale Superiore di Studi Avanzati

EMAIL: chiara.sabelli@gmail.com

Indice

Introduzione

Viviamo in un'epoca dominata dai dati, per lo più dati digitali. La nostra società è stata definita *data society*, la nostra economia *data economy*, l'espressione *big data* è diventata un mantra che promette rivoluzioni, profitti e allo stesso tempo minaccia l'organizzazione del lavoro così come la conosciamo oggi.

Ma ciò che rende preziosi i dati è la nostra capacità di analizzarli, di trovargli un significato, di ridurre grandi quantità di micro-informazioni in conoscenze di più alto livello che siano utili per prendere decisioni, siano esse riguardanti le campagne di marketing, le politiche di salute pubblica, le diagnosi sanitarie o l'amministrazione della giustizia.

Dal 2005 a oggi la percentuale di famiglie europee con accesso al web è passata dal 40% all'80%, come si legge nel rapporto ITU [2017]. Nel 2017 abbiamo prodotto ogni giorno 2,5 miliardi di GB di dati digitali[1] e il ritmo è destinato ad aumentare con la diffusione di dispositivi e sensori a basso costo. Allo stesso tempo siamo stati in grado di progettare *algoritmi*, procedure informatiche, in grado di processare queste grandi quantità di dati. La vera rivoluzione è avvenuta con i sistemi di *machine learning*, algoritmi non completamente pre-programmati dallo sviluppatore, che basano parte del loro comportamento sui dati stessi. Questa classe di algoritmi *apprende* dai dati.

Gradualmente gli algoritmi sono stati impiegati come strumenti di assistenza alla decisione in diversi campi: prima di tutto la finanza (per prevedere l'andamento dei titoli sul mercato e stabilire le strategie di investimento), poi le assicurazioni e l'accesso al credito (per quantificare il rischio di incidente stradale o la capacità di ripagare un debito), i motori di ricerca (che rappresentano ormai la lente attraverso la quale guardiamo il mondo), il marketing (per creare profili sempre più dettagliati dei consumatori e inviargli le offerte più adatte), la gestione delle risorse umane, la giustizia, la polizia, l'intelligence, infine i social network (l'algoritmo di *news feed* di Facebook apprende i nostri "gusti" per proporci i contenuti per noi più rilevanti e coinvolgenti),

infine la sanità (per automatizzare la diagnosi e definire i piani terapeutici migliori).

L'utilizzo di sistemi automatizzati in questi contesti impone una riflessione etica. Le prime domande da porsi riguardano le possibili violazioni del diritto alla privacy, inteso non solo come diritto a mantenere riservati i dati personali, ma anche come diritto all'autodeterminazione, del diritto alla presunzione di innocenza, del diritto ad avere le stesse opportunità indipendentemente dal sesso, dalla provenienza geografica e dalle condizioni sociali di partenza. Questi rischi si acuiscono quando i sistemi in questione utilizzano algoritmi di machine learning che, apprendendo dai dati, tendono a riprodurre e perpetrare i comportamenti discriminatori già presenti nella nostra società.

Vigilare su questa tendenza è particolarmente importante. I dati, e di conseguenza gli algoritmi che li analizzano o che da essi apprendono, sono spesso guardati come strumenti oggettivi, privi cioè dei *bias* che viziano i giudizi degli esseri umani. La loro adozione è stata spesso giustificata in questi termini, ma, come vedremo più avanti, si tratta di una convinzione fallace.

Questa consapevolezza sta gradualmente emergendo. Negli Stati Uniti, dove il sistema di welfare è in mano a società private e dove la *digital revolution* è nata, queste vulnerabilità sono emerse prima che in Europa. Da anni attivisti per i diritti, giornalisti e accademici (Pariser [2012], O'Neil [2016], Pasquale [2015]) si battono perché gli algoritmi vengano sottoposti a uno scrutinio accurato, siano trasparenti e non diluiscano eccessivamente la responsabilità fino a renderla irrintracciabile. In Europa, seppure con un po' di ritardo, sta avvenendo lo stesso, con una maggiore attenzione all'educazione, al coinvolgimento e all'aumento di consapevolezza dei cittadini.

L'oggetto di questa tesi è approfondire le strategie che i governi e i vari attori della società civile in USA e Europa hanno messo in campo per fronteggiare questa situazione. Intendiamo concentrarci su due aspetti:

- gli interventi dei Governi, in termini di strumenti legislativi o di *policy*, volti a stabilire delle prime regole di *governance* degli algoritmi;
- l'attività di ricerca, che i data scientist portano avanti sul tema degli algoritmi, sia riguardo al problema della privacy che a quello dell'equità;
- le iniziative di promozione della *cittadinanza digitale*, ovvero del coinvolgimento del pubblico e dei diversi settori industriali in una riflessione etica.

Nel Capitolo 1 passeremo in rassegna i campi di applicazione degli algoritmi, cercando di metterne in evidenza gli aspetti più controversi dal punto di

vista dei diritti e delle libertà. Nel Capitolo 2 analizzeremo il contenuto di due gruppi di documenti istituzionali, uno relativo agli Stati Uniti, in particolare i rapporti redatti dal Big Data Working Group istituito all'interno dell'Executive Office del Presidente Obama EOP [2016], e l'altro riguardante l'Unione Europea, nello specifico il Regolamento UE 2016/679 in materia di tutela della privacy, chiamato Regolamento Generale per la Protezione dei dati (GDPR)[a], che entrerà in vigore il 25 maggio del 2018. Nel Capitolo 3 affronteremo il problema dal punto di vista matematico e informatico, rendendo conto del dibattito nato all'interno della comunità dei data scientist sui metodi per progettare algoritmi che siano equi, non violino cioè diritti personali e collettivi. Infine nel Capitolo 4 ci occuperemo delle questioni relative alla *cittadinanza digitale*: qual è il livello di consapevolezza dei cittadini sull'impiego di sistemi informatici per assistenza alla decisione? Quale il livello di consapevolezza dei diversi attori sociali e imprese private relativamente a questo tema? Per provare a rispondere a queste domande analizzeremo due progetti di *public engagement* promossi rispettivamente in Francia e negli Stati Uniti. Si tratta del dibattito Éthique Numérique, promosso in Francia dalla Commission Nationale Informatique & Libertés (CNIL) da gennaio a ottobre del 2017, e del Civic online debate "Governing The Rise Of Artificial Intelligence", promosso dalla The Future Society alla Harvard Kennedy School nell'ambito della AI Initiative. L'analisi di questi progetti, in particolare dei contenuti e della scelta degli interlocutori coinvolti, ci aiuterà a comprendere come i diversi Paesi stanno affrontando il problema, e come lo inquadrano all'interno del rapporto tra tecnologia e diritti[b].

[a] "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC", Official Journal of the European Union L 119:1-88, May 2016. Consultabile al: http://eur-lex.europa.eu/eli/reg/2016/679/oj. Data ultimo accesso: 20 dicembre 2017.

[b] Nel testo sono stati distinti tre tipi di riferimenti: le fonti normative sono state riportate come note a piè di pagina, le fonti giornalistiche come note nella sezione **Note** a pagg.74-78, le pubblicazioni scientifiche e i documenti istituzionali come voci bibliografie nella sezione **Bibliografia** a pagg.79-84.

Capitolo 1

Algoritmi e scelte sociali: le incognite delle "automated decisions" tra numeri e valori

In questo capitolo cercheremo di dare concretezza al problema oggetto di studio. Per farlo passeremo in rassegna diversi casi giunti all'attenzione di accademici appartenenti a diversi settori (dai giuristi ai *data scientist*) e successivamente agli organi di stampa internazionali.

Prima di iniziare questa rassegna però, vogliamo chiarire due punti. In primo luogo collocare la nostra riflessione nel più ampio dibattito sull'etica e la *governance* dell'intelligenza artificiale. In secondo luogo ripercorrere brevemente la storia dell'intelligenza artificiale ricercando i precursori degli attuali sistemi di *machine learning*.

1.1 Quale intelligenza artificiale?

Il dibattito etico e politico sull'intelligenza artificiale si è concentrato negli ultimi anni principalmente sulla robotica. Si è affermata infatti la convinzione che robot sempre più intelligenti ed economici sostituiranno gli esseri umani nei lavori manuali, cambiando il volto del mondo del lavoro in modo radicale. Questa convinzione è accompagnata dalla paura che i governi, le organizzazioni e le industrie non siano pronti ad affrontare questa rivoluzione e che lasceranno vittime le componenti più deboli della società, il cui sostentamento è basato proprio su lavori manuali.

La sensazione è che questa rivoluzione, pur apparendo sempre più verosimile, non sia imminente. Per capirlo consideriamo il settore della logistica, che per fatturato e domanda di posti di lavoro sta gradualmente sostituendo l'industria manifatturiera, a causa della crescita del mercato degli acquisti online, che stanno sostituendo di fatto la grande distribuzione fino nel settore alimentare. Ebbene un recente rapporto della divisione Ricerca e Sviluppo del gruppo Deutsche Post DHL mostra che solo il 5% degli stabilimenti dell'azienda presenta un buon livello di automazione[2], il 15% è solo meccanizzato, il restante 80% affida agli esseri umani i compiti determinanti. Il motivo è che, al contrario di quanto avviene nel settore manifatturiero, le capacità cognitive richieste a un robot che deve prelevare dagli scaffali oggetti di dimensioni, pesi e forme molto diversi e depositarli in una scatola, sono molto più sofisticate di quelle necessarie per ripetere lo stesso gesto alla catena di montaggio.

Esistono, chiaramente, degli importanti problemi etici che riguardano sistemi robotici come le auto a guida auotonoma o i droni militari, che richiedono un aggiornamento profondo del quadro concettuale di riferimento sul tema, formalizzato per la prima volta da Asimov nelle sue tre leggi della robotica[3], ed è importante che questa riflessione vada avanti e si aggiorni.

Tuttavia l'intelligenza artificiale su cui intendiamo concentrarci in questa tesi è di natura diversa. In una singola espressione potremmo dire che ci interessano i pericoli che provengono dal solo *software*, e non dall'*hardware*.

1.2 Breve storia dell'intelligenza artificiale

L'articolo che fonda il campo dell'intelligenza artificiale viene pubblicato nel 1950 da Alan Turing sulla rivista MIND Turing [1950]. Il titolo è eloquente "Computing Machinery and Intelligence" e va dritto al punto ponendosi la domanda: "Can machines think?" Turing chiarisce subito che rispondere a questa domanda è un compito assurdo se si considerano i significati con cui vengono usate comunemente le parole *machine* e *think*. È in questo contesto che Turing introduce il famoso *imitation game*. Come primo passo Turing considera il seguente gioco. Ci sono tre partecipanti: un uomo A e una donna B che siedono in una stanza e un intervistatore C (uomo o donna, è indifferente), posizionato in un'altra stanza. Lo scopo del gioco per l'intervistatore è indovinare il sesso di A ponendogli delle domande tramite una telescrivente (in modo che il timbro di voce non lo influenzi) e per A quello di indurre C in errore. Ora Turing costruisce un secondo gioco: sostituisce A con una

macchina e si chiede: "la probabilità che C indovini il sesso di A è uguale nei due scenari?". Questa domanda, dice Turing, rimpiazza la domanda iniziale "Can machines think?" La versione più nota dell'imitation game, a cui ci si riferisce di solito come Test di Turing, ignora l'identità sessuale dei due partecipanti e viene riformulato così: A e B sono rispettivamente una macchina e un essere umano e l'intervistatore comunica con loro tramite una telescrivente. L'intervistatore è in grado di distinguere la macchina dall'essere umano ponendogli delle domande? Anche se altri prima di Turing avevano formulato l'idea che si potessero usare i computer come strumenti sperimentali per studiare la natura del pensiero umano, l'articolo pubblicato su Mind nel 1950 consolida l'idea che si possa programmare una macchina perché esibisca un comportamento intelligente Buchanan [2005].

Nei dieci successivi a questo articolo, l'evoluzione dell'intelligenza artificiale (IA) è stata condizionata notevolmente dai limiti di velocità e di capacità di memoria degli apparecchi dell'epoca. Si deve arrivare al 1963 per la pubblicazione della prima raccolta di programmi di IA funzionanti, intitolata "Computers and Thoughts". La raccolta contiene il programma per il gioco della dama scritto da Arthur Samuel, che vedremo più avanti, che per la prima volta è in grado di apprendere dall'esperienza. L'apprendimento viene di certo considerato come un'attività caratterizzante per un sistema intelligente, come afferma Marvin Minsky, un altro dei protagonisti di questa storia, in un articolo di rassegna del 1961 Minsky [1961]. Tra i contributi di questi primi dieci anni vale la pena citare il Logic Theorist di Newell e Simon, in grado di inventare le dimostrazioni a teoremi logici, e il Pandemonium System di Oliver Selfridge, in grado di riconoscere visivamente le lettere dell'alfabeto, che evolvette poi nei sistemi di classificazione o *pattern recognition*. Durante gli anni '60 l'interesse si sposta da sistemi basati su una rigorosa formalizzazione dei meccanismi del ragionamento verso sistemi basati su ampi corpi di conoscenza, i cosiddetti *knowledge-based systems* o sistemi esperti.

1.2.1 I sistemi esperti

I sistemi esperti nella definizione di Jackson [1999], sono "sistemi computerizzati in grado di emulare l'attività di decisione di un essere umano esperto in un certo argomento".

L'inizio dei sistemi esperti viene ricondotto al programma Stanford Heuristic Programming Project, guidato da Edward Feigenbaum al dipartimento di Computer Science di Stanford negli anni '70. I primi problemi che il gruppo di Stanford cercò di affrontare riguardavano la diagnosi di malattie infetti-

ve o il riconoscimento di molecole complesse. L'idea alla base dei sistemi esperti era quella di autmotizzare il ragionamento sul corpo di conoscenze acquisite su un certo problema. I sistemi esperti si componevano di due parti: la base di conoscenza e l'algoritmo di inferenza che, impiegando una logica di ragionamento sulla base di conoscenza, era deputato a prendere decisioni. L'evoluzione dei sistemi esperti, che ha visto un'esplosione negli anni '80, è stata fortemente segnata dal progresso tecnologico e in particolare dalla diffusione dei Personal Computer che, in breve tempo, hanno offerto performance migliori e più economiche dei costosi *mainframe*. Questa rivoluzione tecnologica favorì la diffusione dei sistemi esperti all'interno delle imprese. Per rendere conto dell'importanza di questi sistemi nella storia dell'intelligenza artificiale è utile leggere la risposta che Allen Newell, vincitore del Turing Award insieme a Herbert Simon[4], diede agli editor della rivista Journal of Artificial Intelligence che nel 1985 gli chiedevano qual era stata la novità più importante per il settore dell'intelligenza artificiale nell'ultimo decennio Bobrow and Hayes [1985]:

> There is no doubt, as far as I am concerned, that the development of expert systems is the major advance in the field during the past decade...The emergence of expert systems has transformed the enterprise of AI, not only because it has been the main driver of the current wave of commercialization of AI, but because it has set the major scientific problems for AI for the next few years...

Tra i sistemi esperti quelli che ebbero la maggiore diffusione e che segnarano in qualche modo il passaggio a una fase successiva sono le reti bayesiane Pearl [1985, 1988]. Queste portarono a un cambiamento radicale nell'approccio all'IA, introducendo il concetto di "agente intelligente" o "agente razionale" ben consolidato nel campo dell'economia e della teoria della decisione. Questo concetto spinse a considerare i sistemi di intelligenza artificiale come sistemi in grado di risolvere un compito circoscritto interagendo con l'ambiente esterno e massimizzando le probabilità di successo. L'introduzione del ragionamento probabilistico in condizioni di incertezza fu, infatti, uno degli elementi chiave che traghettò l'IA da una fase discendente, quella della seconda metà degli anni '80, verso una nuovo periodo di successi caratterizzato dalla centralità dei dati.

1.2.2 Machine learning

Il primo utilizzo documentato del termine machine learning risale al 1959 ed è attribuito ad Arthur Samuel [1959], un dipendente della IBM che pubblicò

un lavoro riguardante un programma informatico in grado di giocare a dama. Come vedremo in seguito, i giochi hanno rappresentato e continuano a rappresentare una palestra per gli algoritmi di machine learning. È interessente leggere l'introduzione all'articolo:

> Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8-10 hours of machine playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations.

Con queste parole Arthur Samuel solleva molti dei problemi su cui gli esperti di etica dell'intelligenza artificiale si interrogano ancora oggi. Da una parte la possibilità di costruire un'intelligenza *super umana*, con i timori di perderne il controllo. Dall'altra la consapevolezza che l'intelligenza in senso umano è la capacità di saper risolvere compiti molto diversi tra loro. L'algoritmo progettato da Samuel era un *search tree*, un albero decisionale che, data la posizione delle pedine sulla scacchiera, passava in rassegna le mosse possibili e per ognuna calcolava le possibili configurazioni future del gioco per valutare quali erano vincenti. Per la precisione Samuel aveva messo a punto una tecnica per esplorare solo le configurazioni interessanti, poiché la capacità di calcolo che aveva a disposizione era ridotta. Oggi quella stessa tecnica è utilizzata in algoritmi di machine learning (ML) più evoluti.

Un momento decisivo nell'evoluzione dei sistemi di ML è il 10 febbraio 1996, quando il programma DeepBlue sviluppato da IBM sconfigge il campione di scacchi Garry Kasparov[5].

Altri momenti salienti sono:

- febbraio 2011: IBM Watson vince al quiz Jeopardy;
- febbraio 2016: Alpha Go batte il campione di Go Lee Sedol a Go;
- ottobre 2017: Alpha Go Zero batte i programmi più esperti al gioco del Go impiegando solamente le regole del gioco e allenandosi contro se stesso, senza utilizzare "vere" partite giocate da esseri umani esperti;

- dicembre 2017: Alpha Zero impara a giocare in 24 ore al gioco del Go, scacchi e shogi (un gioco da tavola giapponese simile agli scacchi) sempre ricevendo come unico input le regole del gioco.

1.3 Il dibattito etico sull'intelligenza artificiale

La riflessione etica sull'intelligenza artificiale è ben sintetizzata da Bodstrom and Yudkowski [2014]. I due autori identificano tre grandi categorie di intelligenza artificiale, ciascuna delle quali pone diversi quesiti etici. Due di queste tre categorie sembrano echeggiare le parole con cui Arthur Samuel introduceva il suo sistema nel 1959. Sono la *superintelligence* e la *artificial general intelligence*. Ma la categoria che sembra porre i problemi più imminenti è la prima, quella del machine learning e più in generale dei *domain-specific AI algorithms*, ovvero gli algoritmi sviluppati per risolvere uno specifico compito. Per spiegare quali questioni etiche questa categoria solleva, i due autori utilizzano un esempio.

Immaginiamo una banca che impiega un algoritmo per valutare le richieste di mutuo. Analizzando la storia creditizia dei candidati, l'algoritmo approva o rifiuta la richiesta. Immaginiamo poi che uno dei candidati rifiutati faccia causa alla banca, accusandola di discriminazione razziale. La banca risponde che questo è impossibile: nel progettare l'algoritmo è stato esplicitamente vietato l'utilizzo di informazioni relative alla provenienza dei candidati. Nonostante questa accortezza, un'analisi statistica mostra che la percentuale di approvazione delle richieste depositate da risparmiatori di origine africana tende a diminuire nel tempo e a essere sempre minore rispetto a quella dei bianchi. Anche facendo l'esperimento di sottoporre all'algoritmo dieci richieste indistinguibili per un essere umano, si vede che il sistema rifiuta i neri e accetta i bianchi. Cosa sta succedendo? Trovare una risposta a questa domanda potrebbe non essere semplice. Se l'algoritmo di machine learning è una rete neurale profonda o un algoritmo genetico, potrebbe essere impossibile capire perché e persino come il sistema giudica le richieste sulla base della provenienza geografica. Se, invece, l'algoritmo è una rete Bayesiana o un albero decisionale è molto più semplice analizzarne il funzionamento. Questo vuol dire che un revisore riuscirebbe a capire, per esempio, che il sistema usa le informazioni relative all'indirizzo di residenza che, nelle società più segregate, sono una buona approssimazione della *razza*. Oltre alla necessità di trasparenza, nel senso di possibilità di essere valutati, gli algoritmi che

agiscono in contesti sociali dovrebbero essere *prevedibili* e con *responsabilità note*. Quando un sistema del genere sbaglia, chi è responsabile per l'errore? La società che ha sviluppato e venduto il software o l'istituzione che lo ha utilizzato?

Bodstrom e Yudkowski concludono la loro riflessione sugli algoritmi progettati per sostituire gli esseri umani nell'esercizio di funzioni sociali con queste parole:

> Responsibility, transparency, auditability, incorruptibility, predictability, and a tendency to not make innocent victims scream with helpless frustration: all criteria that apply to humans performing social functions; all criteria that must be considered in an algorithm intended to replace human judgment of social functions; all criteria that may not appear in a journal of machine learning considering how an algorithm scales up to more computers. This list of criteria is by no means exhaustive, but it serves as a small sample of what an increasingly computerized society should be thinking about.

Nelle sezioni che seguono vedremo alcuni esempi di sistemi di intelligenza artificiale che già sostituiscono gli esseri umani in compiti con una forte valenza sociale.

1.4 Gestione del personale

1.4.1 Selezione

Un esperimento condotto nel 2003 da Bertrand and Mullainathan [2003], esperte di economia del lavoro, mostrò che negli Stati Uniti i processi di selezione del personale soffrono ancora di una forte discriminazione razziale. Sottoponendo ai reclutatori *curriculum vitae* fittizi con nomi tipicamente bianchi e tipicamente afro-americani, questi ultimi venivano chiamati a colloquio molto meno dei primi a parità di competenze.

Se gli algoritmi di selezione del personale impiegati oggi apprendono cos'è un buon candidato dai dati del passato, non potranno essere molto più neutrali degli esseri umani[6]. Negli Stati Uniti nel 2016 il 72% dei curriculum sono stati rifiutati senza che alcun essere umano li abbia mai letti. A prendere la decisione sono stati degli algoritmi, senza alcuna supervisione umana.

Un altro strumento largamente impiegato nei processi di selezione del personale è il test della personalità. Il più diffuso, in un certo senso lo standard del settore, è basato sul *five factor model*. Il modello sintetizza la personalità in cinque fattori Goldberg [1993]: estroversione-introversione, gradevolezza-sgradevolezza, coscienziosità-negligenza, nevroticismo-stabilità emotiva, apertura mentale-chiusura mentale. L'idea è che sia possibile valutare questi cinque fattori sottoponendo l'individuo a un questionario. Il motivo è che si osservano delle associazioni frequenti tra certe espressioni verbali e certi tratti della personalità. Lo stesso Goldberg afferma che il modello a cinque fattori si è mostrato di grande utilità per la selezione del personale, poiché i tratti della personalità sono sistematicamente correlati con diversi criteri di performance lavorativa.

Un modello così semplificato è facilmente automatizzabile, e questo è quello che è successo in un numero crescente di aziende, soprattutto quelle che devono passare in rassegna un grande numero di candidature spesso per compiti di medio e basso livello. Ma i candidati ne sono consapevoli? Se ricevono un rifiuto, sapranno mai perché? A questo proposito è interessante la storia di Kyle Behm[7]. Kyle aveva sofferto di disturbo bipolare della personalità e per questo aveva sospeso per un anno e mezzo gli studi universitari. Quando si sentì sufficientemente in salute per tornare a studiare, Kyle si mise anche alla ricerca di un lavoro part-time e inviò la sua candidatura a una grossa catena di supermercati. Si trattava di una posizione a salario minimo, niente di ambizioso. Ma la sua candidatura venne rifiutata e Kyle non venne mai richiamato. Un amico, che lavorava per quella catena, chiese spiegazioni e scoprì che Kyle non aveva superato il test della personalità, un test molto simile a quello cui era stato sottoposto al momento della diagnosi di disturbo bipolare. La storia non finisce qui. Kyle venne rifiutato a molti altri test, finché non ne parlò con il padre avvocato che approfondì la questione e scoprì che era una pratica comune quella di utilizzare i questionari per la personalità da parte delle grandi compagnie. Scoprì inoltre che raramente i candidati rifiutati facevano causa alla società. Il padre di Kyle intraprese una class action contro sette grandi aziende, tuttora pendente. Il punto è stabilire se il questionario, offerto dalla società di gestione delle risorse umane Kronos, sia o non sia un test medico. Se lo fosse sarebbe vietato usarlo nelle procedure di assunzione, come prescritto dall'American with Disabilities Act del 1990.

1.4.2 Misura della performance

La misura delle performance del personale è un altro ambito in cui gli algoritmi si stanno gradualmente diffondendo. L'esempio che trattiamo riguarda un programma di valutazione degli insegnanti pubblici negli Stati Uniti[8].

Alla fine del mese di maggio del 2011 Sarah Wysocki stava concludendo il secondo anno come insegnante in una scuola media di Washington D.C.. I suoi superiori, gli studenti e i loro genitori erano entusiasti del suo lavoro. Eppure, due mesi più tardi, Sarah viene licenziata[9]. La decisione era stata presa in base ai risultati del programma di valutazione IMPACT[10], un sistema introdotto dal responsabile dell'istruzione della città nel tentativo di riformare il sistema scolastico su mandato del nuovo sindaco. IMPACT, tutt'oggi attivo, assegna a ciascun insegnante un punteggio. Coloro che si posizionano nel 2% più basso della classifica vengono licenziati. Il punteggio è calcolato per metà in base all'osservazione di esperti durante le lezioni in classe e per l'altra metà dal risultato del Value Added Model messo a punto dalla società di consulenza Mathematica Policy Research.

Il Value Added Model ha l'obiettivo ambizioso di stabilire quanta parte del progresso scolastico annuale di ciascuno studente sia da attribuire all'insegnante. I progressi o i peggioramenti degli studenti sono il risultato di numerosi fattori: la condizione familiare, le possibilità economiche, la loro interazione con il resto della classe, il background culturale della famiglia. Per tenere conto di tutti questi elementi, il modello utilizza solo due informazioni: il diritto dello studente all'esenzione dal pagamento del pasto in mensa e le sue eventuali "learning disabilities" (quello che in Italia chiamiamo "bisogni educativi speciali"). Aggregando i dati relativi a tutte le classi che un insegnante ha avuto in un anno scolastico, il modello formula una previsione su qual è il progresso atteso per ogni singolo studente e lo confronta con quello effettivamente misurato nei test di fine anno[11].

Se un insegnante ha lavorato con una sola classe, il suo operato sarà valutato sulla base della performance di 15-20 studenti, un numero insufficiente ad ottenere risultati solidi dal punto di vista statistico. Anche ammesso di avere a disposizione un campione più numeroso, la procedura ha dei punti deboli. Per prima cosa il punteggio dipende molto dal voto ottenuto nei test di profitto dell'anno precedente: se i voti sono truccati, ad esempio, il punteggio ne risentirà. Il secondo limite, forse quello più grave, è l'estrema semplificazione con cui misura l'impatto delle vicende personali e familiari sull'andamento scolastico dei ragazzi.

Sarah Wysocki voleva sapere come era stato calcolato il suo punteggio e quello degli altri 200 insegnanti licenziati a Washington D.C. quell'anno. Molto presto scopre che è sostanzialmente impossibile capire il funzionamento dell'algoritmo[12]: come utilizza i dati, come soppesa i diversi fattori. Tuttavia ha un sospetto. All'inizio dell'anno aveva notato che tutti gli studenti della sua classe avevano ottenuto voti sopra la media del distretto ai test dell'anno precedente, soprattutto nella lettura. Ma già durante i primi mesi Sarah si era resa conto che la loro abilità nella comprensione dei testi scritti era tutt'altro che eccellente. Un'inchiesta del giornale USA TODAY[13] rivelò, qualche mese più tardi, che almeno 70 scuole del distretto avevano truccato i test. Questo è il feedback negativo generato dall'algoritmo. Gli insegnanti modificano i loro comportamenti per conformarsi a ciò che questo premia. Purtroppo Wysocki non riuscì ad accumulare prove sufficienti per far riconsiderare il suo punteggio. E qui il paradosso: spesso agli esseri umani sono richieste prove molto più stringenti di quanto non si chieda a un algoritmo per provare la sua affidabilità.

Le storia di Sarah Wisoky e Kyle Behm fanno parte della tendenza più generale ad applicare metodi analitici per prevedere le performance dei lavoratori: si tratta della cosiddetta *people's analytics*[14].

1.5 Polizia predittiva

L'idea che i dati e la statistica possano aiutare le attività di sorveglianza e investigazione delle forze dell'ordine non è nuova. Già nel 1994 la polizia di New York aveva messo a punto CompStat Shapiro [2017], un programma informatico che analizzava le distribuzioni temporali e geografiche dei reati commessi in città ed emetteva dei bollettini periodici con i risultati. Gli agenti avevano infatti notato che a New York i crimini si concentravano in pochi isolati, addirittura pochi incroci, i cosiddetti *crime hotspot*.

Circa 20 anni più tardi, alla fine degli anni 2000, alla University of California Los Angeles l'antropologo Jeffrey Brantingham e il matematico Andrea Bertozzi adattarono modelli formulati per descrivere l'attività sismica alla previsione dei crimini, dal furto d'auto, alla rapina, fino all'omicidio[15].

Sfruttando solo tre variabili, data, luogo e tipo di crimine, i due svilupparono il software PredPol, un sistema in grado di indicare le dieci o venti zone della città dove è più probabile che venga commesso il prossimo crimine. Concentrando i pattugliamenti in queste zone la probabilità di sventare i

reati aumenta. Nessuno conosce i dettagli dell'algoritmo, poiché è coperto da segreto commerciale, che vale diversi milioni di dollari. Nel 2012 infatti Brantingham e Bertozzi hanno fondato una startup, grazie al finanziamento di 3,7 milioni di dollari ricevuto da fondi di *venture capital*. PredPol oggi è utilizzato da oltre 50 corpi di polizia negli Stati Uniti[16].

Proprio mentre Brantingham e Bertozzi progettavano PredPol, Mario Venturi, un assistente capo della Polizia di Stato presso la questura di Milano, sviluppava il software KeyCrime. Testato nel 2007 e utilizzato regolarmente dalla Polizia di Milano a partire dal 2008, KeyCrime viene impiegato per controllare le rapine a danno di esercizi commerciali e banche. L'algoritmo si basa sulla ricostruzione delle serie criminali per prevedere il prossimo evento della serie. Per ogni rapina denunciata, gli agenti registrano nel sistema alcune centinaia di informazioni, la maggior parte riguardano il comportamento dell'autore, o degli autori, del crimine. Si tratta delle informazioni contenute nei verbali di denuncia delle vittime, ma anche dell'analisi delle immagini registrate dalle telecamere a circuito chiuso effettuata dal personale di polizia. Il nuovo evento criminoso viene poi confrontato con gli altri presenti nel database e l'algoritmo di KeyCrime cerca i più *simili* e propone all'agente una serie (un insieme di eventi che il sistema ritiene essere stati commessi dalla stessa mano criminale). Sulla base di questa serie il software produce delle previsioni sul prossimo colpo, che vengono poi comunicate alle pattuglie attraverso le 'Note ricerche', qualcosa del tipo: "tra domani e lunedì verrà commessa una rapina ai danni di uno dei supermercati di questa zona, in questa fascia oraria".

In una intervista telefonica Mario Venturi ha dichiarato: "ho voluto sviluppare KeyCrime per lasciare ai colleghi che continueranno a lavorare dopo di me la mia esperienza sul comportamento criminale, maturata durante la mia carriera. Il successo di KeyCrime è dovuto al fatto che il 70% delle rapine commesse non sono isolate, ma seriali", prosegue Venturi e aggiunge "ogni volta che, grazie alle previsioni di KeyCrime, l'autore di una rapina viene arrestato perché colto sul fatto, verrà indagato anche per tutte le rapine precedenti di cui è sospettato e, se giudicato colpevole, verrà condannato a una detenzione più lunga"[17].

Ma KeyCrime funziona? A condurre uno studio accurato dell'efficacia di questo sistema è stato Giovanni Mastrobuoni, economista del Collegio Carlo Alberto di Torino, che ha analizzato i dati relativi a due anni di rapine, dal gennaio 2008 al dicembre 2009 Mastrobuoni [2017]. L'analisi di Mastrobuoni mostra che, dalla quarta rapina in poi, l'impiego di KeyCrime permette di risolvere il 9% in più di casi (arresto in flagranza di reato o immediatamente

dopo). Ottenere questa misura non è affatto banale. L'efficacia di un sistema del genere non può essere valutata guardando all'andamento temporale del numero di rapine in una città, perché questo può essere dovuto ad altri fattori ed è impossibile isolare il contributo dell'algoritmo KeyCrime. Non si può nemmeno confrontare lo stesso periodo in due città diverse, una in cui viene usato il software e l'altra no, perché nessuna città è paragonabile all'altra dal punto di vista dell'attività criminale. Per poter misurare l'impatto di uno strumento simile è necessario avere a disposizione due campioni confrontabili di rapine, relativi cioè allo stesso territorio e allo stesso periodo, di cui uno sia stato trattato con i mezzi tradizionali e l'altro con il software. Questo è stato possibile per Mastrobuoni perché a Milano, come nelle altre grandi aree metropolitane italiane, Polizia e Carabinieri si alternano nel sorvegliare la città, ruotando ogni sei ore su tre zone. I Carabinieri non usano KeyCrime, mentre la Polizia si. Dunque le rapine su cui i Carabinieri hanno indagato e quelle su cui invece ha indagato la Polizia sono due campioni statisticamente confrontabili.

Una valutazione di efficacia affidabile, come quella eseguita su KeyCrime, manca del tutto per molti degli altri sistemi che abbiamo nominato finora, ad esempio PredPol. "Spesso questo accade per una mancanza di trasparenza, sia nel caso che il software sia sviluppato autonomamente dagli agenti perché vogliono difendere gli investimenti fatti, sia quando è un prodotto commerciale", commenta Giovanni Mastrobuoni e aggiunge "chi esegue la valutazione non deve aver alcun conflitto di interessi, deve essere un'entità terza e indipendente, prima di tutto per garantire ai cittadini che l'investimento fatto è stato utile"[18].

L'unica analisi indipendente condotta su PredPol riguarda il rischio di discriminazione razziale. Da anni attivisti, giornalisti e accademici negli Stati Uniti lamentano il rischio che l'impiego di algoritmi predittivi, soprattutto di machine learning, da parte della polizia possa danneggiare le minoranze. Nel 2016 gli statistici dello Human Rights Data Analysis Group (HRDAG) hanno analizzato le previsioni di PredPol sulla città di Oakland in California, relative al consumo illegale di droga Lum and Isaac [2016]. Nonostante i dati del dipartimento Health and Human Services mostrino che l'uso illecito di sostanze è diffuso uniformemente in tutta la città, le zone a cui PredPol assegna un rischio maggiore sono quelle abitate principalmente da cittadini afro-americani e con basso reddito. Il motivo di questa discrepanza, secondo HRDAG, è che il sistema riceve come input i dati storici sulle denunce e gli arresti e prevede, di conseguenza, il comportamento delle vittime e della polizia piuttosto che quello dei criminali. In altre parole apprenderebbe i pregiudizi radicati nella società americana e li rinforzerebbe con le sue previsioni.

A partire dal 2014 esperimenti simili sono nati anche in Europa. Ad esempio le polizie di Zurigo[19], Aargau[20], Berlino[21] e Monaco[22] hanno testato con successo il software Pre Crime Observation System, detto PRECOBS, e sono decise ad adottarlo stabilmente. Il software non si occupa di tutti i tipi di crimine, ma si concentra sulle rapine. La sua efficacia è stata tuttavia messa in dubbio dal giornale Zeitung[23]. PRECOBS è stato oggetto di critica in Germania, soprattutto per il timore che incentivi le forze di sicurezza alla raccolta indiscriminata di dati, anche personali, violando la privacy e il diritto alla presunzione di innocenza. Alla fine di febbraio del 2017 Dieter Schürmann, al ministero dell'interno del länder Nordreno-Vestfalia, ha dichiarato di voler integrare PRECOBS con i dati riguardanti le infrastrutture, il reddito, il consumo d'acqua, l'uso delle reti di telecomunicazioni, il consumo di energia, l'utilizzo del trasporto pubblico e perfino le ricerche di mercato. Tutte informazioni che l'Internet of Things promette di rendere disponibili con facilità e a basso costo. Il pericolo è che venga autorizzata una " pesca a strascico" di dati che potrebbe spingere ad attingere anche dai *social network*, incoraggiata da studi come quello realizzato dal Predictive Technology Laboratory della University of Virginia Gerber [2014].

Risale alla fine del 2014 anche il primo sistema di polizia predittiva testato dalla polizia di Londra, la Metropolitan Police Force, che si è concentrata sui crimini commessi dai membri delle gang. Il software, sviluppato dalla società di consulenza multinazionale Accenture[24], stima il rischio che un membro già noto di una gang commetta nel futuro un crimine.

In Francia, precisamente nella regione dell'Oise poco a nord di Parigi, da giugno 2016 la *gendarmerie* sta testando l'algoritmo PredVol [25], specializzato nel prevedere i furti di automobile, una vera e propria piaga per quella zona.

L'esperienza e l'evoluzione delle tecniche di machine learning, hanno fatto nascere, anche in questo settore, nuovi strumenti. Un esempio è HunchLab Shapiro [2017], che oltre a indicare le zone ad alto rischio, specifica che tipo di crimine è probabile che venga commesso e, prendendo in considerazione le posizioni degli agenti sul territorio in quel momento, suggerisce la tattica migliore per portare a casa un risultato con il minor danno possibile.

Di natura diversa, ma sempre di intelligenza artificiale si tratta, è VALCRI (Visual analytics for sense making in criminal intelligence analysis)[26], un progetto sviluppato dal gruppo di Computer Science della Middlesex University a Londra in collaborazione con la West Midlands Police Force (Gran Bretagna) e la Federale Gerechtelijke Politie Antwerpen Noordersingel, la polizia federale di Anversa (Belgio). Il finanziamento ricevuto dalla Commissione Europea nel 2014 ammonta a oltre 16,5 milioni di euro[27]. Attualmente la

West Midlands Police lo sta testando su 6,5 milioni di casi raccolti in tre anni, anonimizzando i dati[28]. Si tratta di un sistema di supporto al lavoro dell'investigatore che: analizza semanticamente i verbali secondo un modello investigativo, estrae così delle caratteristiche e calcola le somiglianze a partire dalla distanza euclidea pesata tra i vettori di caratteristiche. Tutto il processo è condiviso con l'investigatore attraverso un'interfaccia visuale molto sofisticata. Sostanzialmente l'idea è che si possa automatizzare il ragionamento dell'investigatore, insegnando al computer tutta la conoscenza accumulata nel tempo dal corpo di polizia a cui appartiene.

All'Imperial College di Londra, infine, è appena partito ICONIC (Inference, COmputation and Numerics for Insights into Cities)[29], un progetto finanziato con quasi 3 milioni di sterline dal Engineering and Physical Sciences Research Council[30].

Una buona sintesi sulle implicazioni etiche e giuridiche dell'utilizzo di questi sistemi è contenuta in un articolo dei ricercatori del progetto VALCRI Schlehahn et al. [2015].

1.6 Amministrazione della giustizia

Anche nell'ambito della giustizia sono impiegati sistemi automatizzati di assistenza alla decisioine, progettati per lo più per aiutare il giudice a stabilire la "giusta" sentenza. Per capire di cosa si tratta andiamo, ancora una volta, negli Stati Uniti. Molti tribunali utilizzano sistemi automatici per stabilire il rischio di recidiva di un soggetto arrestato in attesa di giudizio o dell'imputato di un processo. Questi sistemi si basano su modelli che, prendendo come input una serie di dati sull'accusato, calcolano un *risk recidivism score*, ovvero la probabilità che lui o lei commetta un nuovo reato in futuro.

Il più antico e diffuso di questi sistemi è il Level of Service Inventory (LSI) e la sua versione aggiornata LSI-Revised. Sviluppato nel 1995 dalla società canadese Multi-Health-Systems, si basa su un questionario che viene compilato dal trasgressore appena arrestato. Le domande riguardano eventuali arresti o condanne passati, ma anche i membri della famiglia e gli amici frequentati, il contesto sociale, economico e culturale, l'eventuale dipendenza da droga o alcol. Pur non riferendosi direttamente al gruppo etnico di appartenenza, al reddito o al livello di istruzione, il questionario in qualche modo ricostruisce queste informazioni. Basta immaginare un giovane cresciuto in un quartiere borghese: molto probabilmente non ha mai avuto rapporti con la giusti-

zia prima. Un afro-americano, al contrario, sarà stato molto probabilmente sorpreso a fumare marijuana o a bere alcolici prima dell'età legale.

Per calcolare il rischio di recidiva il sistema individua quali informazioni sono maggiormente correlate con la ricaduta nel reato e il modello gli assegna un peso maggiore nel calcolare il risk recidivism score. Nel migliore dei casi, questo punteggio è utilizzato per inserire i condannati nei programmi di recupero attivi nelle prigioni, nel caso peggiore aiuta il giudice a stabilire la lunghezza della pena: più alto il punteggio, più tempo in galera. Il Risk Recidivism Model rischia così di diventare una profezia che si autoadempie. Scontando una pena più lunga la probabilità di trovare un lavoro, una volta liberi, è molto più bassa e di conseguenza aumenta la probabilità di commettere nuovamente un crimine. Una volta arrestato, in attesa di un nuovo processo, il detenuto avrà già una condanna alle spalle e il suo risk recidivism score schizzerà alle stelle. Uno studio condotto nel 2013 ha stabilito che il software tende ad assegnare punteggi più alti ai membri delle minoranze Oliver [2014].

Partendo dal LSI-Revised Tim Brennan, professore di statistica alla University of Colorado, e fondatore insieme a Dave Wells, direttore di un centro correzionale, della società di consulenza Northpointe, ha sviluppato l'algoritmo COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). COMPAS si basa su un questionario di 137 domande che amplia e in parte corregge quello del LSI-Revised. In particolare aggiunge domande di natura psicologica, chiedendo ad esempio di esprimere accordo o disaccordo con affermazioni come "A hungry person has a right to steal" oppure "If people make me angry or lose my temper, I can be dangerous."

A maggio del 2016 ProPublica ha condotto un'inchiesta approfondita sull'algoritmo COMPAS[31] concludendo che penalizza gli afro-americani. Il team di Pro Publica ha ottenuto i risk recidivism score assegnati da COMPAS a oltre 7000 persone arrestate nella contea di Broward in Florida tra il 2013 e il 2014, e ha poi controllato quanti di questi sono stati accusati di nuovi crimini nei due anni successivi (la stessa procedura di validazione utilizzata dagli sviluppatori dell'algoritmo)[32]. Ebbene: le previsioni di recidiva dell'algoritmo sbagliano in maniera diversa per bianchi e neri. In particolare la percentuale di arrestati che pur avendo ricevuto un risk recidivism score elevato non hanno commesso reati nei dua anni seguenti (falsi positivi) sono il 23% tra i bianchi e il 44.9% tra i neri. Al contrario coloro che avendo ricevuto un punteggio basso hanno comesso nuovi reati (falsi negativi) sono il 47.7% tra i bianchi e il 28% tra i neri. In altre parole l'inchiesta ha svelato che COMPAS sovrastima il rischio di recidiva per i neri e lo sottostima per i bianchi.

La Northpointe si è difesa dalle accuse di ProPublica dicendo che l'algoritmo garantisce la stessa accuratezza (percentuale di arrestati valutati ad alto rischio di recidiva che hanno commesso nuovi reati nei due anni successivi) per le due popolazioni, bianchi e neri. Il punto è che quando l'algoritmo sbaglia la sua previsione lo fa in direzioni diverse. L'inchiesta di ProPublica ha destato l'interesse di quattro diversi gruppi di ricercatori negli Stati Uniti, che hanno cercato di capire se è possibile "correggere" l'algoritmo affinché rispetti entrambi i vincoli: uguale livello di accuratezza e stessa percentuale di falsi positivi (o negativi) per le due popolazioni. Tutti e quattro i gruppi hanno concluso che non è possibile progettare un simile algoritmo[33]. Il motivo è che bianchi e neri sono rappresentati in proporzioni diverse nel campione di dati sugli arresti nella contea di Broward (ci sono più arresti tra i neri che tra i bianchi). L'unica soluzione, sembrano suggerire i ricercatori, sarebbe quella di usare due strumenti diversi per le due popolazioni. Va in questa direzione un contributo recente Zafar et al. [2017] presentato alla "31st Conference on Neural Information Processing Systems" in cui i ricercatori hanno introdotto una nozione di equità basata sulle preferenze dei diversi gruppi sociali presenti nel campione.

L'algoritmo COMPAS è stato infine oggetto di analisi del gruppo di data scientist del Darmouth College che ha confrontato le sue performance con quelle di un gruppo di volontari, reclutati sulla piattaforma di crowdsourcing Amazon Mechanical Turk, a cui sono stati sottoposti dei profili sintetici degli arrestati. Lo studio ha rivelato che l'accuratezza di COMPAS è sostanzalmente uguale a quella del gruppo di volontari. Inoltre i giudizi dei volontari non hanno penalizzato gli afroamericani più di COMPAS. I risultati dello studio sono stati pubblicati a gennaio 2018 su Science Advances Dressel and Farid [2018].

1.7 Intelligence

A maggio del 2015 la testata online The Intercept ha pubblicato dei documenti riservati della National Security Agency sul programma Skynet, in possesso di Edward Snowden[34].

Obiettivo di questo programma è individuare potenziali terroristi di Al Qaida analizzando i dati relativi al traffico mobile di 55 milioni di cittadini pakistani che possiedono un cellulare. In una prima fase, chiamata *training*, l'algoritmo di machine learning "impara" quali sono le abitudini dei terroristi nell'utilizzo del cellulare: frequente spegnimento del dispositivo, numerosi cambi di

SIM, spostamenti durante la notte e nel weekend su percorsi ben precisi. Si effettua il training dell'algoritmo su un campione di 10000 soggetti, tra cui 6 dei 7 terroristi le cui identità sono note ai servizi di intelligence americani. Nella fase successiva si valuta la capacità dell'algoritmo di individuare il settimo terrorista. Superato questo test, il sistema viene applicato all'intera popolazione, assegnando così a ciascun individuo un punteggio: il punteggio è tanto più alto quanto più il suo modo di utilizzare il cellulare è simile a quello dei terroristi.

Si tratta di una tecnica statistica, e come tale è soggetta a errore. I data scientist della NSA dichiarano che l'algoritmo produce circa il 50% di falsi negativi (potenziali terroristi etichettati come innocenti) e lo 0.18% di falsi positivi (cittadini innocenti etichettati come terroristi). Su una popolazione di 55 milioni di persone questo vuol dire 99000 innocenti scambiati per membri di Al Qaida.

Utilizzare questo algoritmo non sembra molto diverso da lanciare una moneta. "La base di apprendimento è incredibilmente limitata, soprattutto su una fenomenologia così variegata, per apprendere un comportamento deviante e applicarlo a un'intera popolazione", commenta Fosca Giannotti, direttrice del laboratorio Knowledge Discovery and Data Mining (KDD) del CNR di Pisa. La scelta di includere sostanzialmente tutti coloro che in Pakistan possiedono un cellulare, ha delle pericolose implicazioni etiche. Dino Pedreschi, co-direttore del KDD, afferma: "applicare un algoritmo simile a un'intera popolazione significa rinunciare alla presunzione di innocenza. Se analizzo i dati di mobilità e comunicazione di tutti i cittadini indiscriminatamente, sto assumendo che tutti i cittadini sono potenzialmente terroristi. Equivale a vendere la libertà di comunicazione e di spostamento in cambio della sicurezza". "Sarebbe diverso se queste analisi fossero usate una volta identificati dei sospetti per supportare l'attività investigativa", conclude la Giannotti[35].

Queste riflessioni appaiono ancora più importanti se si considera il fatto che almeno una parte degli oltre 400 attacchi con droni compiuti in Pakistan dal 2004 a oggi, sono stati pianificati sulla base di questi algoritmi[36].

La sorveglianza dei cittadini può avere poi scopi diversi dalla sicurezza. La Banca Centrale Cinese ha recentemente dato completo accesso all'attività online dei cinesi a compagnie private come TenCent o AliBaba, colossi delle comunicazioni che gestiscono la rete di social network cinesi, o Ping An Insurance, una società di assicurazioni. L'obiettivo è sviluppare modelli di *credit rating* basati su indicatori non tradizionali, così da aumentare il numero di privati e piccole imprese che hanno accesso al credito, pur non avendo una storia finanziaria alle spalle. Il programma si chiama Social Credit System[37].

Non si conoscono dettagli sui modelli impiegati, ma si sa che diventerà operativo a partire dal 2020. Le informazioni rilevanti saranno i prodotti acquistati online: se acquisti una lavastoviglie o degli accessori per neonato il tuo punteggio aumenta; se invece compri videogame diminuisce. Sarà importante anche la rete sociale, secondo il motto "good people are friends with good people". Esiste già oggi un sito che mostra il proprio punteggio, per ora basato solo su parametri fiscali, ed elenca le circa 90000 persone da cui è consigliato allontanarsi per aumentarlo. Ciò che desta ulteriore preoccupazione è che il *social credit score* influenzerà anche la qualità dei servizi sanitari a cui i cinesi avranno accesso e le scuole che frequenteranno i loro figli.

Se siamo ormai consapevoli che utilizzando dispositivi mobili lasciamo delle tracce, siamo meno abituati a pensare che anche i nostri movimenti, i nostri gesti e le interazioni con le persone in uno spazio pubblico potrebbero diventare numeri, ed essere analizzati. Come ha raccontato la giornalista Sharon Weinberger [2010] su Nature, già da anni negli aeroporti americani centinaia di agenti speciali sono incaricati di osservare i passeggeri che si muovono tra i controlli di sicurezza per individuare comportamenti anomali: gesti delle mani, posizione del corpo, tono della voce e perfino micro-espressioni facciali. Qualsiasi cosa sia utile a tradire l'intenzione di compiere atti pericolosi.

Alla base di questo programma di sicurezza, chiamato SPOT da Screening of Passengers by Observation Techniques, c'è la teoria psicologica di Paul Ekman, oggi professore emerito di psicologia all'università di San Francisco. I recenti progressi delle tecniche di visione artificiale, che permettono di trasformare in scene tridimensionali le immagini bidimensionali registrate dalle telecamere di sorveglianza, potrebbero supportare queste attività di controllo. Il movimento di un braccio diventa una traiettoria con velocità e curvatura da confrontare con le traiettorie considerate normali. Se devia troppo il sistema potrebbe allertare l'agente che analizzerà più attentamente il comportamento della persona, per esempio facendo uno zoom sul viso e osservando le micro espressioni facciali.

"La tecnologia non consente ancora di sostituire l'attività di questi agenti con sistemi di *computer vision* accoppiati ad algoritmi di analisi statistica. Per esempio l'analisi delle micro-espressioni facciali richiede immagini ad altissima definizione dei volti di ciascun passeggero, con un'illuminazione costante", dichiara Alessandro Vinciarelli[38], professore presso il Dipartimento di Informatica dell'università di Glasgow e coordinatore del network Social Signal Processing promosso dalla Commissione Europea. E prosegue: "l'affidabilità di questi sistemi di controllo è da valutare con grande cautela. Prima di tutto la teoria di Ekman non è condivisa dalla comunità scientifica. An-

che accogliendo completamente l'idea che un certo *pattern* di contrazione dei muscoli facciali corrisponde a un determinato stato emotivo dell'indivisuo, ci sono moltissimi fattori che condizionano le espressioni delle persone in un aeroporto: il jet lag, la stanchezza, la paura di volare". Il programma SPOT va avanti però, e presto si evolverà in una versione più tecnologica che controllerà anche alcuni dati biometrici dei passeggeri.

Capitolo 2

Politiche degli algoritmi: un approccio comparativo

In questo capitolo analizziamo gli strumenti legislativi e di *policy* che sono stati sviluppati per governare gli algoritmi, in particolare quelli dedicati all'analisi dei big data. Nel primo paragrafo studieremo gli Stati Uniti, concentrandoci sul lavoro del Big Data Working Group istituito all'interno dell'Executive Office of the President durante la Presidenza Obama, mentre il secondo paragrafo sarà dedicato all'Europa. Come vedremo il problema della governance degli algoritmi si intreccia con il diritto alla tutela della privacy, inteso sia come diritto alla riservatezza dei dati personali che come diritto all'autodeterminazione. Per questo motivo in entrambi i contesti, americano ed europeo, la nostra riflessione partirà da una rapida rassegna della storia del diritto alla privacy.

2.1 Stati Uniti

Durante la Presidenza di Barack Obama è stata dedicata grande attenzione alle potenzialità e ai rischi derivanti dal progresso delle tecnologie digitali, in particolare quelle connesse ai big data.

È in quegli anni, quindi, che gli Stati Uniti hanno formulato le riflessioni più mature sulla possibilità che l'impiego degli algoritmi possa violare la privacy dei cittadini e alimentare decisioni discriminatorie. Per questo motivo ci concentreremo sui documenti istituzionali relativi a quel periodo.

Prenderemo in considerazione i rapporti redatti da due organi: il Big Data Working Group interno all'Executive Office of the President, lo staff esecutivo più vicino al Presidente, e la Federal Trade Commission (FTC), un'agenzia del governo federale degli Stati Uniti indipendente dal Presidente. Il primo gruppo di esperti si è concentrato principalmente sull'utilizzo dei dati e degli algoritmi predittivi da parte del Governo, mentre la FTC ha studiato in dettaglio il settore privato.

2.1.1 Le fonti

Nell'*exit memo* dell'Office for Science and Technology Policy (OSTP) dell'amministrazione Obama una sezione è dedicata al tema "Understanding the implication of AI, machine learning and big data".

Il mandato dell'OSTP è:

> provide the President and the President's senior staff with accurate, relevant, and timely advice on the scientific and technological aspects of all issues before them; ensure the policies and programs developed across the Executive Branch are informed by sound science; and ensure that Federal investments in science and technology (S&T) are making the greatest possible contribution to economic prosperity, public health, environmental quality, and national security.

In risposta alla richiesta di Barack Obama di immaginare il futuro degli Stati Uniti, avanzata durante la White House Frontiers Conference dell'ottobre del 2016, l'OSTP ha identificato venti frontiere su cui gli Stati Uniti possono diventare guida del processo di innovazione. L'innovazione su scala nazionale, sempre secondo OSTP, riguarderà lo sviluppo e gestione delle potenzialità dell'intelligenza artificiale, in particolare sulla scienza dei dati, il machine learning, l'automazione e la robotica.

Nel formulare la sua posizione su questo argomento l'OSTP fa riferimento a tre documenti di nostro interesse:

- "Big Data: seizing opportunities, preserving values" EOP [2014];
- "Big Data: Seizing Opportunities and Preserving Values: Interim Progress Report" EOP [2015];
- "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" EOP [2016].

I tre rapporti sono il risultato di una consultazione condotta dal gruppo di lavoro sui big data, iniziata a gennaio del 2014 e conclusasi nel 2016 sotto la guida di John Podesta, consigliere di Obama. Per la parte tecnologica il gruppo si è riferito al rapporto "Big Data and Privacy: A Technological Perspective", redatto dal Council of Advisors on Science and Technology CAST [2014].

La Federal Trade Commission, istituita nel 1913 col Federal Trade Commission Act del Presidente Woodrow Wilson, si occupa di protezione dei consumatori. In un primo documento FTC [2014] ha considerato i problemi connessi alla raccolta e all'analisi dei dati, concentrandosi sul ruolo dei *data brokers*. In un secondo documento FTC [2016] ha invece studiato i rischi di violazione etica e legale connessi all'utilizzo dei risultati delle analisi dei dati dei consumatori.

2.1.2 Non si tratta di privacy si tratta di uguaglianza

Il documento finale EOP [2016] pone al centro la necessità di sviluppare il quadro normativo sulle tecnologie associate ai big data basandosi sul prinicipio di *equal opportunity by design*. Questa è la conclusione a cui giunge il gruppo di lavoro dopo due anni di analisi, in cui si è confrontato con accademici, attivisti per i diritti, rappresentanti del settore privato e cittadini. Ed è una conclusione affatto banale se si guarda invece al punto di partenza EOP [2014], tutto incentrato sulla necessità di aggiornare le leggi a tutela della privacy. Si passa dunque dall'idea che il diritto minacciato dall'analisi algoritmica dei dati sia quello alla privacy, anche inteso nel suo significato più ampio di diritto all'autonomia, all'idea che il diritto da proteggere sia quello dell'accesso alle stesse opportunità indipendentemente dal sesso, dalla provenienza geografica, dalla condizione socioeconomica di partenza. Questo è il passaggio concettuale più siginificativo che emerge da questi rapporti.

Dopo aver ripercorso, nella sezione 2.1.3, l'evoluzione delle leggi a tutela della privacy, nelle sezioni 2.1.4 e 2.1.5 diamo conto dei casi di studio e delle indagini realizzate rispettivamente nel settore pubblico e in quello privato, che hanno portato a questo cambio di paradigma.

2.1.3 Evoluzione delle leggi sulla privacy

Il diritto alla privacy negli Stati Uniti si è evoluto in modo strettamente correlato alla tecnologia. Ne è la prova il fatto che l'articolo di Samuel Warren

e Louis Brandeis "The Right to Privacy" Warren and Brandeis [1890] fu catalizzato dalla diffusione delle prime macchine fotografiche portatili:

> Recent inventions and business methods call attention to the next step which must be taken for the protection of the person, and for securing to the individual what Judge Cooley calls the right "to be let alone"[4]. Instantaneous photographs and news paper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops".

"The Right to privacy" è considerato il primo testo di analisi giuridica che introduce e difende il diritto alla privacy negli Stati Uniti e lo formula come il *diritto di essere lasciati soli* dallo Stato e dagli altri cittadini.

Un ulteriore avanzamento tecnologico favorisce l'evoluzione di questo diritto. Nel 1928 in "Olmstead v. United States", la Corte stabilì che intercettare le conversazioni telefoniche di un cittadino era legittimo da parte dello Stato perché il dispositivo per l'intercettazione era posizionato fuori dalla casa, e dunque dalla proprietà privata, di Olmstead. La sentenza del caso Olmstead rimase di riferimento fino al 1967 quando la corte stabilì che in "Katz v. United States" l'FBI aveva violato la *reasonable expectation to privacy* piazzando un dispositivo per le intercettazioni nelle immediate vicinanze, anche se all'esterno, di una cabina telefonica.

Con l'evoluzione dell'informatica e dei sistemi computerizzati di gestione dei dati il problema della privacy assunse una nuova dimensione. Nel 1973 lo U.S. Department of Health, Education, and Welfare pubblicò il rapporto "Records, Computers, and the Rights of Citizens" DHEW [1973]. È interessante leggere la prefazione del rapporto che, per molti versi, ricalca il dibattito sugli algoritmi che stiamo affrontando:

> Computers linked together through high-speed telecommunications networks are destined to become the principal medium for making, storing, and using records about people. Innovations now being discussed throughout government and private industry recognize that the computer-based record keeping system, if properly used, can be a powerful management tool.

e più avanti

> Nonetheless, it is important to be aware, as we embrace this new technology, that the computer, like the automobile, the skyscra-

> per, and the jet airplane, may have some consequences for American society that we would prefer not to have thrust upon us without warning. Not the least of these is the danger that some recordkeeping applications of computers will appear in retrospect to have been oversimplified solutions to complex problems, and that their victims will be some of our most disadvantaged citizens.

Il rapporto contiene dei principi volti a salvaguardare i cittadini nelle pratiche di raccolta e analisi dei dati personali, i cosiddetti "Fair Information Practice Principles" (FIPP's):

- i cittadini hanno il diritto di sapere quali informazioni un'organizzazione ha raccolto su di loro e come intende usarle prima che questo avvenga;
- i cittadini hanno il diritto di obiettare alcuni degli utilizzi di questi dati e di correggere i dati stessi se sono sbagliati o non aggiornati;
- l'organizzazione che raccoglie le informazioni deve assicurare che queste sono affidabili e deve custodirle in modo sicuro;
- per assicurarsi che le aziende seguano questi principi, devono essere predisposte delle misure efficaci (sia di autoregolamentazione da parte di chi raccoglie i dati, sia sanzioni stabilite dal governo).

I FIPP's hanno costituito la base del "Privacy Act" del 1974[a], che regola la raccolta, il mantenimento, l'uso e la diffusione di informazioni sugli individui presenti nei sistemi di archiviazione delle agenzie federali.

Da quel momento in poi l'approccio normativo è stato quello di riferirsi al "Bill of rights", in particolare al Quarto Emendamento alla Costituzione degli Stati Uniti, per i valori generali riguardanti il diritto alla privacy, e di regolare le istanze specifiche settore per settore: credito, salute, educazione, intelligence. Sono state di conseguenza promulgate diverse leggi:

- "Fair Credit Reporting Act", 1970
- "Family Education Rights and Privacy Act", 1974
- "Cable Communications Policy Act", 1984
- "Electronic Communications Privacy Act", 1986

[a] U.S. Code, Title 5, §552a, "Records Computers and the Rights of Citizens", July 1974. Consultabile al: https://www.gpo.gov/fdsys/pkg/USCODE-2012-title5/pdf/USCODE-2012-title5-partI-chap5-subchapII-sec552a.pdf. Data ultimo accesso: 15 dicembre 2017.

- "Computer Fraud and Abuse Act", 1986
- "Heatlh Insurance Portability and Accountability Act", 1996
- "Children's Online Privacy Protection Act", 1998
- "Video Privacy Protection Act", 1998
- "Genetic Information Nondiscrimination Act", 2008

Particolarmente rilevanti in questo elenco sono il "Fair Credit Reporting Act"[b] e l'"Health Insurance Portability and Accountability Act"[c].

Il "Fair Credit Reporting Act" (FCRA) viene promulgato nel 1970 per promuovere l'accuratezza, l'equità e la protezione della privacy riguardo le informazioni raccolte dalle *consumer reporting agencies* con lo scopo di valutare l'idoneità dei cittadini a ricevere prestiti e acquistare polizze assicurative, a essere assunti da una certa azienda, a diventare affittuari di una casa. La legge dà ai cittadini, intesi come *consumers*, il diritto di accedere alle informazioni raccolte dalle agenzie sul loro conto e di correggerle se sbagliate o non aggiornate. La legge impone alle agenzie il dovere di garantire l'accuratezza e la completezza dei dati e prescrive che il consumatore che si vede negato un credito o un'assicurazione, venga avvisato dall'agenzia responsabile del suo fascicolo. Sull'FCRA torneremo nella sezione 2.1.5, quando analizzeremo le pratiche in materia di protezione e analisi dei dati attive nel settore privato.

L'"Health Insurance Portability and Accountability Act" viene approvato dal Congresso nel 1996, durante la prima presidenza di Bill Clinton, con l'obiettivo di garantire che i cittadini che perdono o cambiano lavoro non debbano rinunciare all'assicurazione sanitaria. Di nostro interesse è il Title II, che regola la pubblicazione di informazioni sulla salute personale dei cittadini da parte di specifiche organizzazioni, chiamate *covered entities*[d] e stabilisce gli standard necessari affinché i cittadini possano capire e avere un controllo su

[b] U.S. Code, Title 15, §1681 et seq., "Fair Credit Reporting Act", October 1970. Consultabile al: https://www.gpo.gov/fdsys/pkg/STATUTE-84/pdf/STATUTE-84-Pg1114-2.pdf. Data ultimo accesso: 15 dicembre 2017.

[c] 104th Congress, Public Law 191, "Health Insurance Portability and Accountability Act". Consultatible al: https://www.gpo.gov/fdsys/pkg/STATUTE-110/pdf/STATUTE-110-Pg1936.pdf. Data ultimo accesso: 17 dicembre 2017.

[d] Nel "Summary of the HIPAA Privacy Rule, U.S. Department of Health and Human Services" vengono definite le "covered entities and business associates" come: "health care clearinghouses, employer sponsored health plans, health insurers, and medical service providers that engage in certain transactions". Consultabile al: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html, https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html. Data ultimo accesso: 17 dicembre 2017.

come le loro informazioni sanitarie sono trattate e per che scopi sono utilizzate. Il Title II stabilisce il principio del *minimum necessary*: le *covered entities* devono sempre divulgare il minimo quantitativo necessario di informazioni sanitarie personali.

L'ultimo atto di questo processo di evoluzione è rappresentato dalla proposta di legge "Consumer Privacy Bill of Rights", che si basa sui Fair Information Practice Principles, ma è direttamente indirizzato alla protezione dei dati relativi ai consumatori, volgendo dunque l'attenzione al settore privato dove, come vedremo fra poco, si concentrano le questioni etiche più controverse. Il "Consumer Privacy Bill of Rights" è stato pubblicato nel febbraio 2012 inserito in un quadro più ampio che sollecitava l'avvio da parte del governo di processi di consultazione con i principali attori dei diversi settori privati affinché i principi generali del "Consumer Privacy Bill of Rights" potessero essere declinati in maniera più specifica.

Il punto più interessante del "Consumer Privacy Bill of Rights" riguarda il rispetto del contesto: i consumatori hanno il diritto di assumere che le società raccolgano, utilizzino e divulghino i loro dati personali in maniera coerente con il contesto in cui il consumatore ha comunicato le informazioni personali. Se da una parte il "Consumer Privacy Bill of Rights" cerca di estendere i diritti di trasparenza e responsabilità che il "Fair Credit Reporting Act" prevede per una certa categoria di compagnie private, ovvero le *consumer reporting agencies*, dall'altra rende il diritto alla privacy del consumatore, cioè del cittadino nei confronti di un'azienda con cui scambia informazioni allo scopo di acquistare o usufruire gratuitamente di beni o servizi, un diritto fondamentale. Lascia però all'autoregolamentazione il compito di declinare in maniera più dettagliata questo principio, prevedendo la possibilità, ma non stabilendo l'obbligo, che il processo verso l'autoregolamentazione porti alla formulazione di nuovi strumenti legislativi.

Con questo approccio l'amministrazione Obama sperava di rinforzare alcuni diritti, senza però bloccare o inibire l'innovazione che deriva dall'utilizzo dei dati sia nel settore privato che in quello pubblico.

Il "Consumer Privacy Bill of Rights" ha avuto poca fortuna e il Congresso non lo ha approvato prima della fine della Presidenza di Barack Obama. Nel frattempo però il confronto *multi-stakeholder* è proseguito in alcuni settori (telecomunicazioni, riconoscimento facciale, social network) registrando tuttavia l'abbandono dei tavoli di negoziazione da parte dei rappresentanti della società civile, scoraggiati dalle resistenze dell'industria nel sottoscrivere impegni che potessero in qualche modo limitare il loro potenziale di innova-

zione, e dunque di profitto. Lo stato di cose è ben riassunto dalla giornalista Natasha Singer in un articolo pubblicato a febbraio del 2016[39] in cui si legge:

> The clash continues between consumer advocates who warn that unfettered commercial data collection could chill the daily routines of Americans — where they go, what they say, how they shop — and industry advocates who warn that any restrictions on data collection could chill innovation.

2.1.4 Il settore pubblico

La gestione dei dati nel settore pubblico negli Stati Uniti ha sempre avuto come obiettivo quello di bilanciare il potere tra lo Stato e cittadini.

Particolarmente controverso è l'uso secondario dei dati, utilizzo non direttamente collegato allo scopo per cui i dati sono stati raccolti. Un esempio risale alla seconda guerra mondiale, quando i dati raccolti per il censimento dei cittadini vennero utilizzati per ottenere gli indirizzi di residenza degli americani di origine giapponese allo scopo di arrestarli e inviarli in campi di concentramento.

Allo stesso tempo, proprio gli utilizzi secondari dei dati dei cittadini hanno aperto possibilità di inclusione prima di oggi inaspettate. Consideriamo ad esempio il "Patient Protection and Affordable Care Act"[e], chiamato più brevemente "Affordable Care Act" o Obamacare, e abbreviato con l'acronimo ACA. Le disposizioni contenute nell'ACA sono entrate in vigore a tutti gli effetti dal 2016, dimezzando il numero di cittadini americani senza alcuna assicurazione sanitaria. Questo risultato è stato raggiunto da una parte ampliando il numero di persone idonee a ricevere assistenza sanitaria nel programma Medicaid, dall'altra riformando il mercato delle assicurazioni sanitarie individuali (quelle cioè non sponsorizzate dal datore di lavoro). In particolare gli assicuratori sono tenuti ad accettare tutti i candidati e applicare gli stessi prezzi, indipendentemente dalle loro condizioini preesistenti (di salute e non) e dal loro sesso. Questa disposizione ha posto un freno all'utilizzo dei dati sulle prescrizioni di farmaci per valutare l'accesso a una polizza sanitaria, come accadeva in passato. È famoso il caso dei coniugi Walter e Paula Shelton della Louisiana che nel 2008 si videro negare l'assicurazione

[e] 111th Congress, Public Law 148, "Patient Protection and Affordable Care Act", March 2010. Consultabile al: https://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf. Data ultimo accesso: 17 dicembre 2017.

sanitaria perché Paula aveva assunto dei farmaci antidepressivi in passato e per brevi periodi[40].

L'"Affordable Care Act" introduce però delle altre vulnerabilità. Favorisce infatti il ricorso a una medicina sempre più *personalizzata*, *predittiva* e *di precisione*. Per farlo, promuove la partecipazione dei cittadini a programmi di salute e benessere: chi dimostra di partecipare a uno di questi programmi ha diritto a uno sconto sull'assicurazione. Per dimostrare di essere iscritto a una di queste iniziative è necessaria la divulgazione di dati che riguardano non più le condizioni sanitarie ma gli stili di vita. Perché il paradigma della medicina personalizzata si realizzi è poi necessario combinare le informazioni sanitarie con dati riguardanti le proprie condizioni economiche e i propri dati genetici. È proprio la combinazione di dati provenienti da diversi ambiti che crea la nuova dimensione del problema: l'approccio settoriale tenuto finora sembra insufficiente, a meno di richiedere a questi programmi di rispettare contemporaneamente le leggi di tutela della privacy pensate separatamente per ciascuno dei settori di pertinenza dei dati.

Gli altri settori pubblici che richiedono un'attenzione particolare verso il trattamento dei dati e la loro analisi sono quello dell'istruzione (proliferano ormai i corsi online, da cui è possibile da una parte mettere a punto nuove strategie educative, dall'altra raccogliere informazioni su molti aspetti sensibili della vita dei bambini e dei ragazzi), la sicurezza e la difesa contro il terrorismo (la nascita, dopo l'11 settembre 2001, del Department of Home Security dalla fusione di 22 agenzie federali ha richiesto un'armonizzazione delle pratiche e dei database che è tutt'altro che conclusa), polizia e giustizia (anche in questo settore i problemi etici e legali emergono dalla combinazioni di dati provenienti da fonti molto diverse tra loro, si veda la sezione 1.6).

2.1.5 Il settore privato

Il settore privato ha tutto da guadagnare dall'analisi di dati sui suoi potenziali clienti, aggregando fonti molto diverse fra loro: commerciali (studi di settore), database governativi, dati pubblicamente disponibili (social media, blog, internet).

Un'azienda può infatti modellare le sue offerte commerciali sulla base dall'analisi di queste informazioni, costruendo dei profili molto dettagliati. Il settore della pubblicità e del marketing è stato in effetti rivoluzionato dalla disponibilità di dati e dalla capacità di analizzarli. Il punto debole di questo sistema sta però nel fatto che il consumatore non entra quasi mai in contatto

con l'entità che analizza i dati e vende le informazioni ottenute. Spesso l'utente interagirà con il provider dei servizi internet (piattaforme di blogging, social network, siti di informazione), che poi venderà i dati a una pletora di compagnie intermedie che si occupano di *consumer analytics*. Questi intermediari sono i data broker, ed è su questi soggetti che si concentra l'analisi della Federal Trade Commission.

Il rapporto FTC [2014] rende conto di un'indagine condotta presso nove società di *data brokering*. La FTC distingue tre categorie:

- i data broker che analizzano i dati sui consumatori per stabilire la loro idoneità a ricevere un prestito, stipulare un'assicurazione e ottenere una casa in affitto, le cosiddette *consumer reporting agencies* sottoposte all'FCRA;
- i data broker che vendono prodotti dell'analisi dei dati per scopi di marketing;
- i data broker che raccolgono e analizzano i dati degli utenti per verificarne l'identità o rilevare tentativi di frode.

Il caso delle *consumer reporting agencies* è paradgmatico e vale la pena analizzarlo nel dettaglio. I pionieri di questo settore sono stati i *credit bureau* degli anni '60, che giudicavano l'accettabilità della richiesta di prestito da parte di un cittadino sulla base di caratteristiche del tutto scorrelate dall'ambito finanziario, come il grado di cura con cui si presenteva il suo cortile. L'FCRA regolò queste pratiche, richiedendo prima di tutto che venissero usate informazioni rilevanti e pertinenti e poi che i dossier relativi alla richiesta fossero consultabili dal richiedente. Tuttavia l'FCRA non bastò a rendere trasparente il sistema di valutazione di merito creditizio. In poco tempo infatti nacquero una serie di servizi di monitoraggio, messi in piedi dagli stessi credit bureau, che aiutavano i cittadini a mantenere la buona condotta per ottenere il prestito desiderato ma che allo stesso tempo offuscavano i siti governativi nati per consultare le pratiche. Parallelamente venivano sviluppati i primi modelli matematici per valutare la probabilità di un debitore di andare in bancarotta e non poter ripagare il prestito, il cosiddetto *credit score*. Il primo di questi modelli venne introdotto nel 1989 dalla società Fair, Isaac, and Company e chiamato *FICO score*. Gradualmente il *credit score* ha abbandonato il suo contesto originale, ed è stato utilizzato per valutare l'affidabilità dei cittadini in ambiti molto diversi da quello finanziario.

In altre parole il settore del credito ha inaugurato l'idea che si possa costruire una *scored society*, l'idea cioè che si possano costruire degli algoritmi che restituiscano come risultato una misura della reputazione dei consumatori.

Il problema di questo approccio è che mentre il settore del credito e delle assicurazioni è regolato dall'FCRA e dal " Equal Credit Opportunity Act"[f], il mondo dei *consumer score* non è sottoposto ad alcuna regola e questo può generare problemi.

Nel suo rapporto la FTC ne evidenzia tre. Il primo riguarda la trasparenza: se i dati per le attività di analisi a scopo marketing vengono raccolti da una società che non è la stessa a condurre l'analisi, esiste un modo efficace per garantire al consumatore l'accesso e il controllo delle informazioni personali in possesso del data broker e la possibilità di scegliere a quali attività essere sottoposti e quali rifiutare? L'esperienza dei consensi informati e della notifica sul sito web che raccoglie i dati si è dimostrata insufficiente, perché spesso poco chiara. Il secondo problema riguarda invece le conseguenze di un'iper-personalizzazione delle offerte commerciali che può arrivare a ridurre in maniera ingiusta le possibilità di scelta offerte ai consumatori. Si veda ad esempio il caso del *differential pricing* affrontato nel rapporto Council of Economic Adviders [2015] o quello dei *for-profit college* che puntano ai consumatori più deboli per offrirgli un'educazione universitaria più costosa e spesso meno valida di quella statale[41]. Il problema nasce quando profili ideati per scopi di marketing vengono utilizzati per stabilire, di fatto, l'accesso all'istruzione, all'assicurazione sanitaria o automobilistica. Ad esempio le società di data brokering che vendono profili costruiti per personalizzare la pubblicità online a soggetti come i *for-profit college* non sono sottoposti ad alcuna legge, non essendo registrate come *consumer reporting agencies*.

Questi problemi sono molto bene evidenziati nel libro "The black box society. The secret algorithms that control money and information" Pasquale [2015] scritto da Frank Pasquale, giurista alla University of Maryland, di cui riportiamo un passaggio significativo:

> America's patchwork of weak privacy laws are no match for the threats posed by this runaway data, which is used secretly to rank, rate, and evaluate persons, often to their detriment and often unfairly. Without a society-wide commitment to fair data practices, digital discrimination will only intensify.

e più avanti

> Even with that commitment, we can't forget that access to data is just the first and smallest step toward fairness in a world of

[f] U.S. Code, Title 15, §1691 et seq., "Equal Credit Opportunity Act", October 1974. Consultabile al: https://www.law.cornell.edu/uscode/text/15/1691?qt-us_code_temp_noupdates=1#qt-us_code_temp_noupdates. Data ultimo accesso: 17 dicembre 2017.

> pervasive digital scoring, where many of our daily activities are processed as "signals" for rewards or penalties, benefits or burdens. Critical decisions are made not on the basis of the data per se, but on the basis of data analyzed algorithmically: that is, in calculations coded in computer software. Failing clear understanding of the algorithms involved—and the right to challenge unfair ones— disclosure of underlying data will do little to secure reputational justice.

2.1.6 L'indirizzo politico e normativo

L'analisi del settore privato, in particolare delle pratiche dei *data broker*, la volontà ferma di non voler inibire il potenziale innovativo derivante dall'analisi di grandi quantità di dati di diversa natura e l'abitudine all'autoregolamentazione, hanno modellato le raccomandazioni di policy formulate dal Big Data Working Group di Obama, contenute nel documento finale EOP [2016]. Queste si concentrano molto più sul tema delle uguali opportunità che sulla protezione della privacy, intesa come diritto alla riservatezza dei dati personali:

- Sostenere la ricerca volta a trovare soluzioni tecnologiche al problema della discriminazione che può essere causata dal ricorso a sistemi automatizzati di assistenza alla decisione. La ricerca in questo settore deve essere caratterizzata da una forte interdisciplinarità. Al riguardo, si veda il Capitolo 3, in cui diamo una breve panoramica dell'evoluzione della ricerca in questo campo negli ultimi decenni.
- Incoraggiare gli operatori di mercato a progettare "algoritmi migliori", più trasparenti e di cui sia più facile comprendere il funzionamento e identificare le responsabilità, proseguendo il dibattito avviato dalla Federal Trade Commission.
- Promuovere la ricerca sull'*auditing* degli algoritmi per monitorare il loro intero ciclo di vita: dal campione di dati utilizzati come input, al funzionamento dell'algoritmo vero e proprio, fino agli output prodotti.
- Allargare la partecipazione: favorire la scelta di percorsi di istruzione superiore relativi alla scienza dei dati e all'informatica, aumentare la *data literacy* dei cittadini americani. A questo proposito si vedano i programmi Computer Science for All e TechHire.

- Assicurarsi che il dibattito cominciato con il lavoro del Big Data Working Group informi le regole che verranno stabilite in futuro sull'utilizzo dei dati.

2.2 Le istituzioni europee: UE e Consiglio d'Europa

Fra poco più di tre mesi, il 25 maggio del 2018, entrereà in vigore il Regolamento Generale sulla Protezione dei Dati (GDPR)[g]. Il Regolamento sostituisce la Direttiva sulla Protezione dei Dati Personali del 1995 (Direttiva 95/46)[h], con l'obiettivo di aggiornarne i contenuti rispetto alle tecnologie digitali che nel frattempo hanno posto nuove domande riguardo al diritto alla privacy dei cittadini europei. Inoltre il Regolamento cerca di introdurre un'ulteriore armonizzazione territoriale che non riguardi solo i dati che attraversano i confini degli Stati membri dell'Unione, ma regolino anche la raccolta e il trattamento dei dati effettuati da società private non incorporate sul territorio europeo.

Due aspetti differenziano l'approccio normativo europeo in materia di privacy da quello statunitense. Da una parte il fatto che il diritto alla privacy è riconosciuto come un diritto fondamentale dal 1950, come stabilito dalla "Convenzione Europea dei Diritti dell'Uomo"[i]. Dall'altra il nuovo Regolamento GDPR si concentra sul concetto di *privacy by design* piuttosto che di *equal opportunity by design*, lasciando le disposizioni in materia di decisioni automatizzate pressocché invariate rispetto alla Direttiva 95/46. Quest'ultima sì stabiliva il diritto a non essere sottoposti a decisioni basate esclusivamente su procedure automatizzate, ma introduceva deroghe sufficientemente ampie da rendere la norma di fatto di limitata applicabilità. A questo proposito un elemento di novità nel GDPR riguarda la trasparenza: i cittadini europei hanno il diritto di conoscere "la logica" utilizzata per il trattamento dei dati personali (automatizzati o meno). Se da una parte si tratta di una posizione

[g] Si veda la nota a.

[h] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data", Official Journal of the European Union L 281:31-50, November 1995. Consultabile al: http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:31995L0046. Data ultimo accesso: 19 gennaio 2018.

[i] "Convenzione Europea dei Diritti dell'Uomo", novembre 1950, Corte Europea dei Diritti Umani. Disponibile al: http://www.echr.coe.int/Documents/Convention_ITA.pdf. Data ultimo accesso: 20 dicembre 2017.

forte – il Regolamento ha validità di legge in tutti i Paesi dell'Unione senza necessità che i singoli Stati implementino strumenti legislativi nazionali – dall'altra sembra non raccogliere la complessità intrinseca di certe pratiche di analisi dati, come alcuni algoritmi di machine learning, la cui logica può rimanere oscura agli stessi responsabili del trattamento se non addirittura agli sviluppatori del codice.

2.2.1 Le fonti

Considereremo due gruppi di documenti:

- testi di legge
  - la Direttiva sulla Protezione dei Dati Personali del 1995 (Direttiva 95/46)[j]
  - Regolamento Generale sulla Protezione dei Dati del 2016 (GDPR)[k]
- testi di *soft-law*
  - la Raccomandazione (2010)13 del Consiglio d'Europa COE [2010]
  - le dichiarazioni dell'Article 29 Working Party (un organo di consulenza formato dai rappresentati delle autorità per la protezione dei dati personali di ciascun Paese membro dell'Unione, istituito dall'Articolo 29 della Direttiva 95/46) Article29 WP [2014], Article29 WP [2017]
  - la Risoluzione del Parlamento europeo European Parliament [2017]

Come nel caso degli Stati Uniti, la nostra analisi comincia con una rassegna dell'evoluzione storica del diritto alla privacy nell'Unione Europea, per poi approfondire lo stato attuale.

[j] Si veda la nota h.
[k] Si veda la nota a.

2.2.2 Evoluzione delle leggi sulla privacy e la protezione dei dati in Europa

Il diritto alla privacy in Europa è considerato un diritto umano fondamentale. Nella "Convenzione Europea dei Diritti dell'Uomo"[1], firmata a Roma il 4 novembre 1950, l'Articolo 8 recita:

> Diritto al rispetto della vita privata e familiare:
>
> 1. Ogni persona ha diritto al rispetto della propria vita privata e familiare, del proprio domicilio e della propria corrispondenza.
>
> 2. Non può esservi ingerenza di una autorità pubblica nell'esercizio di tale diritto a meno che tale ingerenza sia prevista dalla legge e costituisca una misura che, in una società democratica, è necessaria alla sicurezza nazionale, alla pubblica sicurezza, al benessere economico del paese, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, o alla protezione dei diritti e delle libertà altrui.

Un buon punto di partenza per ripercorrere l'evoluzione delle leggi sulla privacy in Europa è la Seconda Guerra Mondiale[42]. Durante il conflitto i cittadini di diversi Paesi europei vennero sottoposti a forti violazioni della privacy da parte dei governi, desiderosi di sfruttare tutte le informazioni possibili per proteggere la sicurezza nazionale. Tuttavia le intrusioni nella vita privata dei cittadini spesso andarono oltre le necessità imposte dallo stato di guerra. Questo successe in particolare in Germania, per identificare i membri di gruppi sociali (per etnia o opinione politica) oggetto di discriminazione da parte del regime nazista. Nel dopoguerra la Corte Costituzionale della Repubblica Federale di Germania, la Germania Ovest, stabilì i confini in cui poi si mosse tutta la legislazione europea in materia di privacy.

Nel 1982 il Governo Federale aveva promulgato il "Volkszählungsgesetz" (in inglese "Population Census Act"), una legge che stabiliva la necessità di effettuare un censimento approfondito della popolazione. Il censimento prevedeva oltre 150 domande riguardanti: nome, indirizzo, numero di telefono, sesso, data di nascita, stato civile, affiliazione religiosa, fonte di sostentamento, tipo di impiego, datore di lavoro, istruzione e mezzi di trasporto utilizzati. Prevedeva inoltre che queste informazioni fossero trasmesse ai governi locali

[1] Si veda nota i

per aggiornare i registri dei residenti. La cittadinanza reagì con forza contro questa legge, memore dell'esperienza del nazismo e a causa dell'emergente consapevolezza che l'analisi di grandi quantità di dati personali rappresentava uno strumento di controllo molto potente.

Questo fu il contesto che portò la vicenda all'attenzione della Corte Costituzionale nel caso "Volkszählungsurteil" o "Population Census Case". La Corte formulò una serie di meccanismi di tutela alla base del diritto alla privacy in Europa. In particolare introdusse il diritto all'*informational self-determination*, facendolo discendere direttamente dal diritto costituzionale alla dignità, alla libertà degli individui e allo sviluppo autonomo della personalità. L'idea è che ogni cittadino, per poter sviluppare in maniera libera la propria personalità, deve divulgare alla società le informazioni che ritiene opportune: se non sa che tipo di analisi è condotta sui suoi dati personali, potrebbe decidere di non partecipare a iniziative collettive, fondamentali per il sano svolgimento della vita democratica.

Inoltre venne stabilita:

- la necessità di specificare il contesto e gli scopi per cui venivano raccolti i dati;
- il dovere di garantire l'accuratezza delle informazioni e l'accesso ai cittadini per controlli e rettifiche;
- un limite di tempo oltre il quale le informazioni devono essere distrutte;
- il principio del *minimum necessary*, che abbiamo già trovato enunciato nell'"Health Insurance Portability and Accountability Act"[m];
- l'indipendenza delle autorità a garanzia della privacy.

La sentenza della Corte Costituzionale tedesca già recepiva i "Fair Information Practice Principles" statunitensi, enunciati nel 1973, che avevano anche ispirato l'Organizzazione per la cooperazione e lo sviluppo economico (OCSE) nel formulare le sue "Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data"[n]. Le linee guida dell'OCSE, oltre a introdurre le prime riflessioni su come regolare il flusso dei dati personali

[m] Si veda la nota c.

[n] "Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data of 23 September 1980 [C(80)58/FINAL].", 23 settembre 1980. Disponibile al: http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm. Data ultimo accesso: 29 gennaio 2018.

attraverso i confini nazionali, hanno influenzato le leggi nazionali sulla privacy dei Paesi membri e in generale hanno stabilito l'indirizzo normativo dei decenni successivi.

Un altro passaggio rilevante è la pubblicazione, il 28 gennaio 1981 da parte del Consiglio d'Europa, della "Convention for the protection of individuals to automatic processing of personal data", nota anche come "Convention no.108"[o], che nel 1985 otterrà un numero sufficiente di ratifiche per poter entrare in vigore. Il trattato si concentra sulla protezione dei dati personali, relativamente alla possibilità che questi siano processati in modo automatico. La diffusione dell'informatica infatti porta il problema della raccolta e analisi dei dati personali su una nuova scala: si possono archiviare quantità di dati maggiori e su questi si può effettuare una grande varietà di transazioni molto velocemente. La situazione dei Paesi membri del Consiglio d'Europa, un'organizzazione fondata nel 1949 dal Trattato di Londra con il mandato di "promuovere la democrazia, i diritti umani, l'identità culturale europea e la ricerca di soluzioni ai problemi sociali in Europa" e che è tuttora distinta dalle istituzioni dell'Unione Europea, era molto variegata e frammentata relativamente alla protezione dei dati. L'obiettivo della Convenzione era dunque anche quello di favorire un'armonizzazione delle legislazioni nazionali per assicurare che il diritto fondamentale della privacy fosse assicurato ai cittadini europei indipendentemente dalla loro residenza.

L'entrata in vigore del Trattato di Lisbona, il 13 dicembre 2007, ha reso vincolante la "Carta Europea dei Diritti Fondamentali"[p] che, agli Articoli 7 e 8, chiarisce la distizione tra diritto alla privacy e diritto alla protezione dei dati personali:

> Articolo 7
> **Rispetto della vita privata e della vita familiare**
>
> Ogni persona ha diritto al rispetto della propria vita privata e familiare, del proprio domicilio e delle proprie comunicazioni.
>
> Articolo 8
> **Protezione dei dati di carattere personale**

[o] "Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Convention No. 108", Council of Europe, January 1981. Disponibile al: https://rm.coe.int/1680078b37. Data ultimo accesso: 20 dicembre 2017.

[p] "Carta dei diritti fondamentali dell'Unione Europea, 2012/C 326/02", Gazzetta Ufficiale dell'Unione Europea C 326/391, Dicembre 2000. Consultabile al: http://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN. Data ultimo accesso: 20 dicembre 2017.

1. Ogni persona ha diritto alla protezione dei dati di carattere personale che la riguardano.

2. Tali dati devono essere trattati secondo il principio di lealtà, per finalità determinate e in base al consenso della persona interessata o a un altro fondamento legittimo previsto dalla legge. Ogni persona ha il diritto di accedere ai dati raccolti che la riguardano e di ottenerne la rettifica.

3. Il rispetto di tali regole è soggetto al controllo di un'autorità indipendente.

Questa rapida rassegna storica sembra indicare che la formulazione dei FIPP's segna un punto di svolta per la legislazione europea sulla privacy. Da quel momento in poi l'attività normativa e di policy si concentra maggiormente sulla tutela della privacy intesa come protezione dei dati personali, lasciando in secondo piano l'originario significato introdotto da Warren e Brandeis Warren and Brandeis [1890] di diritto all'autoderminazione. Non mancano tuttavia le eccezioni a questo approccio. Particolarmente rilevante in questo senso è il caso "Guerra and others v. Italy"[q], in cui la Corte Europea dei Diritti Umani ha stabilito all'unanimità che c'era stata una violazione dell'Articolo 8 della "Convezione Europea dei Diritti dell'Uomo"[r] a causa dell'assenza d'informazione della popolazione di Manfredonia in Puglia sui rischi corsi e sui provvedimenti da adottare in caso di incidente in una industria chimica nelle vicinanze, la Enichem.

2.2.3 Direttiva 95/46

Con la nascita dell'Unione Europea, sancita dal trattato di Maastricht del 1993, si rende necessario fornire un quadro normativo di riferimento per i Paesi membri dell'Unione relativamente alla privacy e alla protezione dei dati personali. A questo scopo il Parlamento europeo e il Consiglio dell'Unione Europea adottano nel 1995 la "Direttiva 95/46/CE del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati"[s], chiamata brevemente

[q] "Case of Guerra and Others v. Italy (116/1996/735/932)", European Court of Human Rights, February 1998. Disponibile al: http://hudoc.echr.coe.int/fre?i=001-58135. Data ultimo accesso: 20 dicembre 2017. Si veda anche "Case of Guerra and others v. Italy", *The International Journal of Human Rights* **2**(2):93-97, 1998. Disponibile al: https://doi.org/10.1080/13642989808406737. Data ultimo accesso: 10 gennaio 2018.

[r] Si veda nota i.

[s] Si veda la nota h

"Data Protection Directive" o Direttiva 95/46, tuttora valida. In Italia la Direttiva ha portato alla formulazione della legge a "Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali"[t], entrata in vigore a maggio del 1997, poi sostituita dal "Codice in materia di protezione dei dati personali" del 2003[u].

Nella Direttiva 95/46 vengono incorporati i principi stabiliti nella "Convention no.108" del 1981, che a loro volta recepivano i "Fair Information Practice Principles" statunitensi. Viene introdotto il *data subject* e si affermano i principi del "lawful processing" o "trattamento lecito" (il trattamento è connsiderato lecito se la persona ha acconsentito, se è necessario ai fini dell'esecuzione di un accordo di cui la persona è parte in causa, se è necessario per motivi legali o di esercizio di funzioni pubbliche, ...) e della "data quality" (i principi di quailtà dei dati richiedono che il loro trattamento sia lecito e leale; inoltre è vietato il trattamento di dati personali che rivelino l'origine razziale o etnica, le opinioni pubbliche, le convinzioni religiose o filosofiche, l'appartenenza sindacale, nonché il trattamento di dati riguardanti la salute e la vita sessuale). Inoltre il *data subject* ha il diritto di ottenere informazioni sul trattamento dei suoi dati, di accedere ai suoi dati e di opporsi al trattamento se ha buone ragioni per farlo. Importante anche il concetto di *purpose limitation*, simile al concetto di *contesto* che abbiamo visto nella proposta di legge di Obama "Consumer Privacy Bill of Rights". Nella sintesi della Direttiva 95/46[v] si legge infatti:

> [...] i dati personali devono essere trattati lealmente e lecitamente e rilevati per finalità determinate, esplicite e legittime. Essi devono inoltre essere adeguati, pertinenti, non eccedenti, esatti e, se necessario, aggiornati. I dati personali devono essere conservati

[t] Legge 31 dicembre 1996, n.675, "Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali", Gazzetta Ufficiale Serie Generale n.5 del 08-01-1997 - Suppl. Ordinario n. 3. Consultabile al: http://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=1997-01-08&atto.codiceRedazionale=097G0004&elenco30giorni=fals. Data ultimo accesso: 19 gennaio 2018.

[u] Decreto Legislativo 30 giugno 2003, n. 196, "Codice in materia di protezione dei dati personali", Gazzetta Ufficiale Serie Generale n.174 del 29-07-2003 - Suppl. Ordinario n. 123. Consultatible al: http://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2003-07-29&atto.codiceRedazionale=003G0218. Data ultimo accesso: 20 dicembre 2017.

[v] La sintesi della Direttiva 95/46 è consultabile al: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:l14012&from=IT&isLegissum=true. Data ultimo accesso: 21 dicembre 2017.

non oltre il tempo necessario ed esclusivamente per le finalità per le quali sono stati rilevati.

Particolarmente rilevante per la nostra discussione è l'Articolo 15 della Direttiva:

> **Articolo 15**
> Decisioni individuali automatizzate
>
> 1. Gli Stati membri riconoscono a qualsiasi persona il diritto di non essere sottoposta ad una decisione che produca effetti giuridici o abbia effetti significativi nei suoi confronti fondata esclusivamente su un trattamento automatizzato di dati destinati a valutare taluni aspetti della sua personalità, quali il rendimento professionale, il credito, l'affidabilità, il comportamento, ecc.
>
> 2. Gli Stati membri dispongono, salve le altre disposizioni della presente direttiva, che una persona può essere sottoposta a una decisione di cui al paragrafo 1, qualora una tale decisione:
>
>   (a) sia presa nel contesto della conclusione o dell'esecuzione di un contratto, a condizione che la domanda relativa alla conclusione o all'esecuzione del contratto, presentata dalla persona interessata sia stata accolta, oppure che misure adeguate, fra le quali la possibilità di far valere il proprio punto di vista garantiscano la salvaguardia del suo interesse legittimo, oppure
>
>   (b) sia autorizzata da una legge che precisi i provvedimenti atti a salvaguardare un interesse legittimo della persona interessata.

Come afferma Bygrave [2017], dell'Università di Oslo, il primo limite dell'Articolo 15 è l'appicabilità: numerose condizioni devono verificarsi affinché una decisione sia regolata da questo articolo (deve essere "presa una decisione", questa deve avere un impatto "significativo", deve essere basata *esclusivamente* su una procedura automatizzata e deve avere come obiettivo la valutazione di caratteristiche personali). Il secondo limite è rappresentato dalle deroghe espresse nel paragrafo 2, che sono piuttosto ampie e si prestano a interpretazioni molto diverse fra loro.

2.2.4 Regolamento Generale sulla Protezione dei Dati Personali

La sempre maggiore diffusione e importanza della digitalizzazione della vita pubblica e privata dei cittadini europei ha reso necessario un aggiornamento della Direttiva 95/46. Il cambiamento più importante introdotto dal nuovo Regolamento[w], riguarda i soggetti che devono rispettare la privacy e la protezione dei dati dei cittadini europei. Mentre la Direttiva 95/46 non chiariva se la raccolta e analisi dei dati personali dei cittadini europei da parte di società non incorporate sul suolo europeo e che archiviano e analizzano i dati in centri collocati fuori dall'Unione Europea fosse soggetta alle stesse regole e principi validi per i Governi e le aziende europee, il GDPR stabilisce che[x]:

> the biggest change to the regulatory landscape of data privacy comes with the extended jurisdiction [...] it will apply to the processing of personal data by controllers and processors in the EU, regardless of whether the processing takes place in the EU or not. The GDPR will also apply to the processing of personal data of data subjects in the EU by a controller or processor not established in the EU, where the activities relate to: offering goods or services to EU citizens (irrespective of whether payment is required) and the monitoring of behaviour that takes place within the EU.

L'altro elemento di novità è rappresentato dal concetto di *privacy by design*, ovvero la necessità di progettare i sistemi per l'archiviazione e l'analisi dei dati in modo che questi già rispettino i principi chiave stabiliti dal GDPR senza che sia necessario prevedere degli interventi a posteriori.

Ai fini della nostra discussione è però particolarmente rilevante l'Articolo 22 del Regolamento, che conserva, seppure con piccole modifiche, l'Articolo 15 della Direttiva 95/46:

> Articolo 22
> **Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione**
>
> 1. L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo

[w] Si veda la nota a.

[x] "GDPR Key changes", consultabile al: https://www.eugdpr.org/key-changes.html. Data ultimo accesso: 21 dicembre 2017.

riguardano o che incida in modo analogo significativamente sulla sua persona.

2. Il paragrafo 1 non si applica nel caso in cui la decisione:
   (a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento;
   (b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato;
   (c) si basi sul consenso esplicito dell'interessato.
3. Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione.
4. Le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato.

L'Articolo 22 del GDPR conserva, in gran parte, le condizioni di applicabilità dell'Articolo 15 della Direttiva 95/46 tranne per due aspetti. Da una parte amplia le deroghe alle prescrizioni del paragrafo 1 escludendo le decisioni per cui è stato ottenuto il consenso, dall'altra impone al responsabile delle procedure automatizzate l'obbligo di garantire al *data subject* il diritto a chiedere un intervento manuale. Non è chiaro se questo obbligo implichi il diritto del cittadino a ottenere una spiegazione *ex post* sull'operato dell'algoritmo. È certo che nei casi in cui la procedura automatizzata è particolarmente complessa, l'esercizio di questo diritto può avere delle limitazioni pratiche.

2.2.5 Oltre il GDPR

Abbiamo visto come il Regolamento Generale sulla Protezione dei Dati Personali affronta in maniera solo marginale il problema degli algoritmi sociali,

così come lo abbiamo descritto nel Capitolo 1. In particolare lo affronta solo in nome del diritto alla privacy, non considerando mai diritti di natura collettiva, come il diritto a uguali opportunità e, inoltre, non tratta i casi in cui i dati sono anonimizzati o quelli in cui un cittadino, pur non avendo condiviso alcuna informazione personale, è comunque esposto al rischio di discriminazione semplicemente perché appartiene a un certo gruppo sociale.

Tuttavia altri soggetti istituzionali europei si sono occupati della questione in maniera che potremmo definire più compiuta, in particolare intuendo che le attività legate al *profiling* dei consumatori potessero minacciare il diritto alla non discriminazione, più che quello alla privacy e alla protezione dei dati personali.

Particolarmente siginificativa a questo proposito è la Raccomandazione del Consiglio d'Europa COE [2010] dedicata proprio al *profiling*, in cui gli algoritmi vengono definiti come strumenti in grado di rivelare correlazioni e tendenze altrimenti nascoste e per questo in grado di dedurre informazioni personali che i soggetti non hanno volontariamente divulgato. La Raccomandazione individua il rischio di violare il diritto alla non discriminazione anche nel *targeted marketing*:

> Considering, however, that profiling an individual may result in unjustifiably depriving her or him from accessing certain goods or services and thereby violate the principle of non-discrimination.

Stabilisce inoltre che:

> every person should know the logic involved in profiling.

A questo proposito sono di particolare rilievo le Guidelines pubblicate dell'Article 29 Working Party a ottobre del 2017 Article29 WP [2017], che hanno affrontato, tra le altre cose, l'applicabilità dell'Articolo 22 alle attività di profiling, sottolineando che l'Articolo 14 del GDPR assicura, al paragrafo 2 lettera (g), il diritto del cittadino di conoscere la logica utilizzata per il trattamento dei dati e le conseguenze previste:

> (g) l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato.

Resta però difficile da valutare la attuabilità in pratica di questo diritto, soprattutto in relazione ad algoritmi di machine learning la cui logica può arrivare a essere sconosciuta al programmatore stesso.

Basandosi sulle analisi dello European Data Protection Supervisor EDPS [2015], EDPS [2016] e sulla dichiarazione Article29 WP [2014] dell'Article 29 Working Party, il Parlamento Europeo ha adottato la Risuoluzione European Parliament [2017] che ha confermato la necessità di governare le conseguenze del massiccio utilizzo dei big data e delle tecniche di analisi connesse, nel rispetto del diritto alla privacy, alla non discriminazione e alla libertà di espressione, aggiungendo inoltre la necessità di vigliare sul rischio che venga limitato l'accesso ad un ambiente informativo pluralistico. Il riferimento qui è al fatto che la iper-personalizzazione introdotta dagli algoritmi che regolano i social network e i motori di ricerca impedisca l'accesso a tutte le informazioni necessarie per l'esercizio degli altri diritti stabiliti dalla Carta Europea dei Diritti Fondamentali.

Capitolo 3

Traduzione algoritmica dei problemi giuridici

Parallelamente al dibattito riguardante gli strumenti di *policy* e normativi, gli accademici di diversi settori si sono interrogati sui problemi che l'analisi statistica dei dati implica per il diritto alla privacy e alle pari opportunità. In questo capitolo rendiamo conto in particolare del lavoro fatto dalla comunità dei *data scientist* per progettare algoritmi che rispettino la privacy e siano *fair*, ovvero equi. Particolarmente significativa in questo senso è l'attività del gruppo di lavoro Fairness Accountability and Transparency of Machine Learning (FATML)[43].

3.1 Algoritmi che proteggono la privacy

Il primo approccio alla progettazione di algoritmi in grado di proteggere la privacy, ovvero di consegnare agli analisti di dati un campione che sottoposto a qualsiasis tipo di analisi non violasse la privacy dei partecipanti al database, è stato quello della de-identificazione o anonimizzazione dei campioni di dati.

Questo approccio ha mostrato molto presto i suoi limiti. È piuttosto famoso il caso del Netflix Prize. Quando un utente accede al suo account sulla piattaforma di streaming Netflix, il sistema gli suggerisce alcuni film da poter vedere. Per determinare tali raccomandazioni l'algoritmo ha confrontato la cronologia delle valutazioni dell'utente con quelle di altri clienti Netflix al fine di determinare a quale utente assomiglia di più. Una volta che il software ha identificato il cliente più simile, il suo *nearest neighbour*, cerca i film o le

serie TV a cui lui o lei ha dato il punteggio più alto e lo raccomanda. (Questi algoritmi sono anche chiamati *collaborative rating algorithms*). La precisione dell'algoritmo equivale alla sua capacità di predire la prossima valutazione. Migliore è la previsione del software sulla valutazione che l'utente assegnerà a diversi film, maggiore sarà la soddisfazione del cliente e quindi la sua volontà di acquistare più prodotti. Per migliorare l'accuratezza dell'algoritmo di raccomandazione, Netflix ha lanciato nel 2006 una competizione tra sviluppatori e scienziati di dati chiamata "Netflix prize". Il team in grado ottenere risultati migliori nella previsione del rating rispetto al sistema di raccomandazione Netflix, avrebbe ricevuto un milione di dollari. Il *training data set*, il campione di dati su cui l'algoritmo si allena, includeva 100'480'507 valutazioni che 480'189 utenti avevano espresso su 17'770 film. Le singole voci del campione avevano la forma ⟨user, film, data, rating⟩. La valutazione era un numero intero da 1 (valutazione più bassa) a 5 (valutazione massima). Una volta progettato l'algoritmo, i concorrenti lo hanno addestrato sul training set e successivamente lo hanno testato su un campione contenente 1'408'342 voci del tipo ⟨utente, film, data,?⟩. Per proteggere la privacy dei clienti, gli utenti sono stati etichettati con numeri interi, così come i film. Una preoccupazione per la privacy degli utenti è stata sollevata qualche tempo dopo, quando due scienziati della University of Texas di Austin hanno pubblicato un articolo Narayanan and Shmatikov [2008] che mostrava che erano in grado di identificare gli utenti nel campione di dati del Netflix Prize usando l'Internet Movie Database, anche se l'identità dell'utente era stata cancellata. In seguito a questa pubblicazione Netflix ha deciso di interrompere la competizione alla sua terza edizione, che era già in preparazione. Questo episodio ha messo in luce i limiti posti dalle procedure di anonimizzazione: sono abbastanza forti da resistere alla reidentificazione, data la grande quantità di dati (ridondanti) che le persone producono e rendono disponibili su internet?

Un altro ambito in cui sono emersi i limiti della anonimizzazione dei database come misura per tutelare la privacy dei cittadini è quello dei Genome Wide Association Studies (GWAS). La condivisione di campioni di dati anonimizzati relativi al sequenziamento del DNA è diventata una pratica comune in genomica, anche se nel 2008 David Craig e collaboratori hanno dimostrato che, nonostante il GWAS avesse rilasciato solo statistiche riassuntive sui partecipanti, le informazioni contenute nel campione di DNA di un individuo potevano determinare se lui/lei aveva partecipato allo studio Homer et al. [2008]. Una preoccupazione maggiore è stata sollevata da Gymrek e collaboratori nel 2013 riguardo alla possibilità di re-identificare i partecipanti a un GWAS, utilizzando le informazioni pubblicamente disponibili su internet attraverso il cognome Gymrek et al. [2013]. I cognomi sono ereditati dal

padre nella maggior parte delle società umane e quindi si può trovare una correlazione tra cognomi e aplotipi del cromosoma Y (la porzione non ricombinante del cromosoma Y che è passata quasi invariata da padre in figlio). Sulla base di questo fatto, sono nati numerosi progetti di genealogia genetica, con lo scopo di ricollegare rami di famiglie collegate per via patrilineare. Attualmente ci sono almeno otto database, contenenti centinaia di migliaia di coppie aplotipo-cognome. Questi database sono disponibili pubblicamente su Internet e alcuni sono gratuiti. Usando queste informazioni i ricercatori sono riusciti ad abbinare il 12% dei genomi maschili dello studio con i loro proprietari.

L'approccio che ha superato quello della de-identificazione è quello della *differential privacy* introdotta da Cynthia Dwork, di cui sono riassunti i tratti salienti in Dwork [2008]. L'idea è che gli analisti di dati non abbiano mai accesso ai dati grezzi, che possono contenere molte informazioni sensibili, ma solo a una versione "depurata" dei dati. Il depuratore del campione di partenza è esso stesso un algoritmo, che deve assicurare che le conclusioni che l'analista può ottenere lavorando sul campione di dati depurato, siano indipendenti dalla partecipazione di un singolo individuo. In altre parole se cancelliamo i dati relativi all'individuo A, oppure aggiungiamo i dati relativi all'individuo B o scambiamo i dati di A con quelli di B, l'analista deve raggiungere le stesse conclusioni.

3.2 "It's not privacy and it's not fair"

La definizione di *differential privacy* assicura dunque che l'analista non impari nulla di specifico sui soggetti i cui dati sono contenuti nel database, ma l'analista di certo imparerà qualcosa di nuovo sulla popolazione di cui il database è un campione rappresentativo. Questo, del resto, è il suo obiettivo. In quest'ottica ogni individuo che può essere considerato membro di quella popolazione, potrebbe essere toccato dalle conclusioni dell'analista, sia che abbia partecipato alla raccolta dei dati, sia che non lo abbia fatto.

Questo concetto è ben espresso da Dwork in un seminario del Novembre 2016 presso l'Institute of Advanced Studies di Princeton in occasione del "Differential Privacy Symposium: Four Facets of Differential Privacy"[44]. Dwork considera il caso di uno studio clinico che trova un'associazione tra abitudine al fumo e probabilità di sviluppare il cancro ai polmoni. Un fumatore che non ha partecipato allo studio sarà comunque influenzato da questa conclu-

sione. Ad esempio potrebbe vedere il premio della sua assicurazione sanitaria aumentare.

Gli interessi di ricerca di Cynthia Dwork si spostano dunque dal problema della privacy a quello della *fairness* Dwork and Mulligan [2013], secondo l'idea di Latanya Sweeney, "Computer science got us into this mess. Can computer science get us out of it?" Il punto di vista di Dwork è ben raccontato in un'intervista con Kevin Hartnett pubblicata su Quanta Magazine[45].

A ben vedere la ricerca delle soluzioni al problema degli algoritmi, come minaccia alla privacy e all'uguaglianza di opportunità, ha portato alla coproduzione tra scienza e *policy* e tra scienza e diritto. In effetti il lavoro di Dwork e dei suoi colleghi, ma anche quello di alcuni giuristi, come Frank Pasquale, ha informato i documenti di policy del Big Data Working Group istituito da Barack Obama.

3.3 Algoritmi equi, ma per chi?

Ma è possibile progettare algoritmi che siano equi, indipendentemente dai dati su cui verranno applicati e dalle decisioni che aiuteranno a prendere?

La risposta tende a essere negativa, ma la ricerca su questi temi è ancora gli inizi. In un articolo del 2011 intitolato "Fairness through awarness" Dwork e colleghi affrontano il problema nel caso degli algoritmi di classificazione Dwork et al. [2011], mostrando come sia possibile tradurre matematicamente la richiesta di equità in termini di un vincolo geometrico a un problema di ottimizzazione. Tuttavia questo richiede di essere consapevoli dei modi in cui l'algoritmo può essere discriminatorio. Se il classificatore agisce su una popolazione formata da un gruppo maggioritario e una serie di gruppi minoritari, questi saranno penalizzati per un semplice fatto statistico: la qualità delle conclusioni che possiamo trarre su un campione è proporzionale alla numerosità del campione. Il gruppo minoritario, di conseguenza, sarà trattato secondo le leggi dedotte da quello maggioritario. Per evitare che questo accada sarà necessario far emergere la caratteristica tipica del gruppo minoritario, che tuttavia viene spesso considerata una informazione sensibile e omessa dal database.

Un caso interessante in cui la scienza dei dati è stata impiegata per cercare di curare un algoritmo notoriamente discriminatorio è quello di COMPAS, che abbiamo trattato nella sezione 1.6.

Capitolo 4

Il coinvolgimento dei cittadini

In questo capitolo analizzeremo due dibattiti pubblici avviati in Francia e negli Stati Uniti sulle implicazioni etiche degli algoritmi e sulle strategie per affrontarle. A seconda dei casi si tratterà di un dibattito tra esperti e cittadini, esperti e decisori politici, esperti e aziende private che operano in diversi settori (tecnologia, finanza, assicurazioni, ecc.). L'identità dei promotori di questi dibattiti e la scelta degli interlocutori è di per sé informativa: i governi o le università che hanno organizzato i confronti chi hanno deciso di coinvolgere?

Abbiamo scelto di analizzare:

- il dibattito francese "Éthique Numérique", promosso dal Commission Nationale Informatique & Libertés (CNIL) su mandato della "Loi pour une République Numérique", svoltosi da gennaio a ottobre del 2017[46];
- il dibattito online e su scala internazionale organizzato dalla AI Initiative[47], un progetto dedicato allo sviluppo dell'Intelligenza Artificiale promosso da The Future Society alla Harvard Kennedy School, iniziato a settembre 2017 e che si concluderà a marzo del 2018[48].

Nella sezione 4.1 introduciamo le caratteristiche principali dei processi partecipativi e le dinamiche che ne hanno favorito la diffusione, seguendo l'impostazione di Nanz and Fritsche [2014]. Nelle sezioni 4.2 e 4.3 passeremo ad analizzare i due dibattiti nel dettaglio.

4.1 Metodo di analisi

I processi partecipativi, o deliberativi, prevedono il coinvolgimento dei cittadini in questioni controverse. Rispondono alla richiesta della popolazione di essere maggiormente coinvolta nelle questioni che la riguardano da vicino. Se da una parte la domanda di partecipazione può essere interpretata come un sintomo della crisi delle democrazie rappresentative, è probabilmente più corretto considerare i processi partecipativi come uno strumento di aiuto allo svolgimento della vita di una democrazia rappresentativa, più che una minaccia alla sua sopravvivenza.

La scienza e la tecnologia, e soprattutto il loro impatto sulla società e sulle scelte politiche, hanno rappresentato un ambito particolarmente favorevole al coinvolgimento dei cittadini nel processo decisionale. L'Office of Technology Assessment (OTA), istituito dal "Technology Assessment Act 1972: Sec. 1" nel 1972, rappresenta una delle esperienze più precoci in questo senso. Nel 1978 l'OTA si rivolse a 5000 cittadini per stabilire le questioni tecnologiche più urgenti da affrontare. Tuttavia le volontà dei cittadini vennero prese in scarsa considerazione dall'OTA a causa di successivi conflitti che si svilupparono all'interno del gruppo dirigente. L'analisi di Redelfs and Stanke [1988] mostra infatti che la partecipazione dei cittadini era prevista a livello organizzativo solo con funzione consultiva (nel Technology Assessment Advisory Council sedevano 10 cittadini, scelti di volta in volta tra i gruppi sociali portatori di interesse riguardo una certa tecnologia) e non a livello decisionale.

In contesti locali, i processi partecipativi sono nati invece per risolvere controversie ambientali o stabilire il bilancio di un comune o di una regione. Il primo esempio in questi senso è il bilancio partecipativo di Porto Alegre, portato a termine per la prima volta nel 1989 su proposta del Partito dei Lavoratori Brasiliano Wainwright [2016].

Le questioni con un importante contenuto tecnico o scientifico ma che hanno implicazioni valoriali per la società, dunque si prestano particolarmente a questa forma di confronto e deliberazione con la cittadinanza.

Affinché un processo di coinvolgimento della popolazione possa essere considerato un processo partecipativo, è necessario che venga creata una situazione dialogica e paritaria che permetta il raggiungimento di una conclusione mediata che tenga conto della complessità della situazione.

Come è facile immaginare però esiste una grande varietà di processi partecipativi. Per una prima panoramica delle enormi differenze che esistono tra i

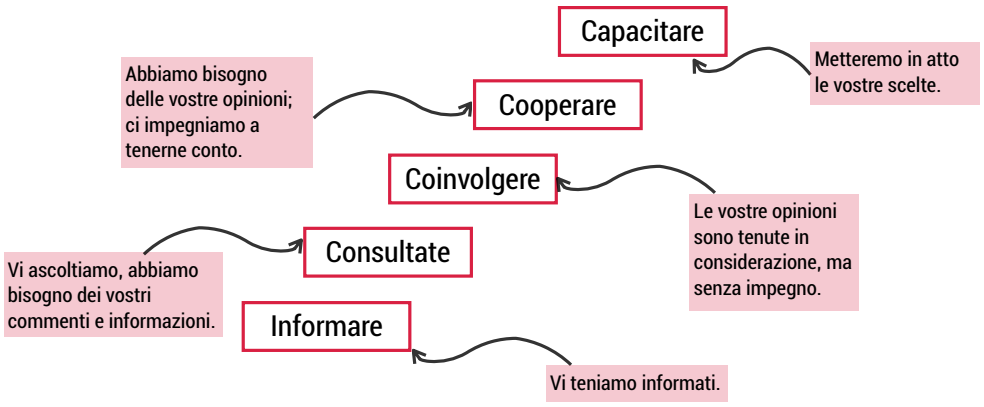

Figura 4.1: A ladder of citizen participation. Immagine rielaborata a partire da Nanz and Fritsche [2014], pag.41.

diversi tipi di processi è utile considerare il modello sviluppato in Arnstein [1969], in cui viene proposta una scala di intensità del coinvolgimento che tiene conto degli obiettivi che gli organizzatori del processo si sono prefissati. Questa scala è rappresentata dallo schema in Figura 4.1. Il modello di valutazione proposto da Arnstein ha il pregio di essere semplice e diretto, ma ha il limite di non considerare il contributo dei cittadini.

Per questo motivo, soprattutto allo scopo di confrontare diversi processi partecipativi, è utile considerare lo schema proposto in Fung [2006]. Si basa su quattro caratteristiche fondamentali.

1. **Durata e numero dei partecipanti**: puntuale/continuativo, sono presenti sempre le stesse persone o alcune escono di scena?
2. **Reclutamento e selezione dei partecipanti**. Permette di stabilire la rappresentatività sociale del processo partecipativo. L'autoselezione, il campionamento casuale e il campionamento mirato, sono tre possibili modalità di scelta dei partecipanti.
    - l'autoselezione è spesso considerata la modalità che conduce alla minore rappresentatività sociale, poiché è stato osservato che le persone che spontaneamente decidono di prendere parte a iniziative del genere sono spesso in possesso di un buon livello di istruzione (superiore rispetto alla media della popolazione).
    - Il campionamento casuale cura in parte i limiti dell'autoselezione, ma non è detto che i cittadini sorteggiati siano poi effettivamente disposti a partecipare.

- Il campionamento mirato seleziona invece rappresentanti di particolari gruppi sociali che, secondo gli organizzatori, sono portatori di interessi rilevanti per la questione oggetto di discussione.

3. **Principali forme di comunicazione**: chi parla e chi ascolta? Come si diffondono le informazioni di background necessarie alla discussione? Se i cittadini coinvolti svolgono solo l'attività di ascolto o osservazione, senza avere occasione di esprimere il loro punto di vista, non si tratta di processi partecipativi. Perché si possa parlare di partecipazione deve essere favorita una delle seguenti modalità di comunicazione.

    - Articolazione di interessi: i cittadini espongono le loro posizioni ed eventualmente le modificano attraverso gruppi di discussione di piccola o media dimensione. Non è necessario arrivare a una posizione condivisa.

    - Negoziazione: con l'obiettivo di raggiungere una posizione di compromesso. Questa modalità viene usata per moderare il confronto tra interessi conflittuali.

    - Scambio di argomentazioni e deliberazioni: il successo dipende da: espressione, reciprocità, responsività. I passi che attraversa un processo di questo tipo sono: confronto, apprendimento, posizione condivisa.

4. **Funzioni del processo partecipativo**.

    - Utilità personale e affinamento delle conoscenze individuali (un esempio sono i forum dei cittadini).

    - Influenza sull'opinione pubblica e la società: questa influenza può essere esercitata grazie al fatto che i media locali o nazionali raccontano lo svolgimento e le conclusioni del processo partecipativo. L'influenza sull'opinione pubblica può tradursi da un lato in un aumento della pressione sui decisori politici, dall'altro, nel caso di situazioni particolarmente conflittuali, nella facilitazione di una posizione mediata tra quelle più lontane. Alcuni esempi sono le Open Space Technology e le Raccomandazioni dei cittadini.

    - Consultazione e presa di posizione dei cittadini: i cittadini formulano delle raccomandazioni che i decisori si dichiarano pronti a considerare durante l'iter decisionale. È ideale se i decisori si impegnano prima dell'inizio del processo a dare un feedback su come le sue risultanze verranno prese in considerazione nella fase

deliberativa e su quali raccomandazioni verranno accolte o a spiegare la decisione di non tenerne conto affatto. Esempi di processi partecipativi con questa funzione sono le Conferenze di Consenso, i Citizen's Panel e i Town Meeting del XXI secolo.

- Co-decisione/co-governance tra l'amministrazione pubblica e la politica da un lato e la società civile dall'altro. Esempi in questo senso sono i Town Meeting del XXI secolo e i Bilanci Partecipativi.

È importante sottolineare che, al di là degli scopi deliberativi che, come abbiamo visto, non rappresentano una conclusione necessaria, i processi partecipativi possono diventare luoghi di apprendimento di *competenze democratiche* per la popolazione.

Le quattro caratteristiche fondamentali che abbiamo appena elencato costituiscono lo schema attraverso il quale analizzeremo i due dibattiti sui rischi etici connessi all'utilizzo degli algoritmi come sistemi di assistenza alla decisione. Prima di procedere all'analisi però diamo dei brevi cenni riguardo i processi partecipativi organizzati online.

4.1.1 Processi partecipativi online

La crescente diffusione dell'accesso a internet ha permesso di sviluppare dei processi partecipativi completamente digitali. I precursori di queste esperienze possono essere identificati nei progetti di *e-Government*, in cui una serie di funzioni e servizi offerti dalle Amministrazioni Pubbliche ai propri cittadini vengono resi accessibili online. In questi progetti i cittadini vengono considerati però come 'clienti'. Si parla invece di *e-Participation* quando i cittadini vengono chiamati a esprimere il loro parere digitalmente (l'esempio più semplice è l'invito a contribuire con un parere nella forma di testo scritto) su una questione che li riguarda, dandogli così la possibilità di influenzare più o meno direttamente l'iter decisionale.

L'utilizzo di internet per coinvolgere i cittadini in un dibattito pubblico presenta opportunità e limiti. Da una parte abbatte i vincoli spazio temporali che possono rendere difficile, e a volte impedire, la partecipazione delle persone. Inoltre la mediazione digitale aiuta anche i più timidi o coloro che si sentono meno sicuri ed esperti della materia a esprimere ugualmente un punto di vista. Dall'altro, la necessità di competenze tecnologiche avanzate e di una buona padronanza della lingua (il contributo dei partecipanti dovrà quasi sempre passare attraverso un testo scritto) esclude delle fasce di popo-

lazione (anziani, immigrati, cittadini con basso livello di istruzione) anche se allo stesso tempo permette di raggiungere più efficacemente i giovani.

Gli obiettivi di un processo partecipativo online sono molto simili a quelli dei processi partecipativi in presenza e variano dall'arricchimento personale e la presa di coscienza fino all'influenza dell'iter decisionale o politico. I forum di discussione sono una versione molto semplice di processo partecipativo online. Esistono poi le petizioni online, che hanno scopo consultivo, fino a piattaforme più evolute che permettono video-chat con esperti, politici e decisori o tra i partecipanti stessi.

4.2 Il dibattito francese: Éthique Numérique

La "Loi pour une République Numérique", promulgata il 7 ottobre 2016, ha dato mandato alla Commission nationale de l'informatique et des libertés (CNIL) di guidare la riflessione sulle implicazioni etiche degli algoritmi. Il dibattito si è svolto durante 10 mesi, dal 23 gennaio al 14 ottobre 2017. Il resoconto del dibattito è stato realizzato dalla CNIL e pubblicato in internet il 15 dicembre 2017 CNIL [2017].

Il punto di partenza del dibattito è stato la volontà di riportare la discussione sulle questioni etiche connesse allo sviluppo e alla diffusione dall'intelligenza artificiale su un piano meno apocalittico e più concreto. La discussione pubblica su questi temi tende infatti a polarizzarsi su due visioni estreme del futuro: da una parte l'idea che i robot svolgeranno i compiti più gravosi e noiosi al posto degli esseri umani e gli algoritmi assicureranno la completa oggettività e neutralità, dall'altra quella che vede i robot stravolgere il mercato del lavoro (scompariranno i lavori di bassa manodopera e con basso contenuto creativo) e gli algoritmi prendere il sopravvento, imponendo discriminazione e ingiustizia senza possibilità di appello.

Quando la "Loi pour une République Numérique" ha dato mandato alla CNIL di avviare il dibattito su questi temi, la CNIL si è resa subito conto di non possedere tutte le competenze necessarie e per questo motivo ha deciso di coinvolgere dei partner, appartenenti a diversi settori, nell'organizzazione e realizzazione di tutto il processo.

Il dibattito si è svolto attraverso 40 incontri animati da 65 partner sotto la coordinazione della CNIL, cui hanno preso parte quasi 3000 professionisti del mondo dell'istruzione, della finanza, delle assicurazioni, della sanità della

gestione del personale, e una concertazione cittadina, svoltasi a Montpellier il 14 ottobre 2017, a cui hanno preso parte circa 40 persone.

Gli obiettivi del dibattito erano principalmente due:

- alimentare e informare il dibattito pubblico (informa l'opinione pubblica di qual è la posta in gioco quando si parla di algoritmi);
- formulare principi e raccomandazioni che, auspicabilmente, informino il dibattito politico e che contribuiscano a stabilire un modello francese di governance degli algoritmi.

4.2.1 Il dibattito tra professionisti

In occasione dell'evento di lancio dell'intera iniziativa, il 23 gennaio 2017, la CNIL ha incoraggiato tutti i soggetti interessati (istituzioni pubbliche, società civile, imprese) a organizzare un incontro o una manifestazione sull'argomento, di cui poi redigere un resoconto da condividere con la CNIL. La scelta di questa strategia era tesa a lasciar emergere i punti ritenuti critici da parte degli attori coinvolti, piuttosto che stabilire la direzione del dibattito dall'alto e a priori.

L'appello della CNIL è stato raccolto da 60 partner di diversa natura: organizzazioni del mondo dell'istruzione, associazioni di assicuratori, il Ministero della Cultura, associazioni dei gestori delle risorse umane, sindacati. Questi hanno organizzato 45 eventi in tutta la Francia, di cui 27 a Parigi, 14 in provincia (Caen, Lille, Bordeaux, Toulouse, Ax-Les-Termes, Marsiglia, Lione) e 4 negli Stati Uniti (in collaborazione con la "AI initiative" promossa da The Future of Society alla Harvard Kennedy School). Gli eventi si sono svolti da marzo a ottobre 2017 e vi hanno preso parte circa 3000 persone. Il ruolo della CNIL è stato di coordinamento degli incontri.

Il primo effetto di queste manifestazioni è stato di portare il tema all'attenzione pubblica, permettendogli così di uscire dai circoli di esperti e raggiungere i cittadini comuni.

La prima parte dei dibattiti si è concentrata sull'identificazione di sei questioni etiche riguardanti l'utilizzo degli algoritmi come strumenti di assistenza alla decisione:

1. La fiducia verso la neutralità e oggettività degli algoritmi rischia di spingere gli esseri umani a scaricare su di essi la responsabilità delle scelte più delicate e controverse.

2. Gli algoritmi rischiano di agire sulla base di pregiudizi, generare discriminazione ed esclusione sociale. Questo può avvenire per volontà o nella consapevolezza di coloro che li programmano e mettono in opera, ma anche a loro insaputa.

3. Le attività di profilazione volte a modellare le offerte commerciali rischiano di intaccare delle logiche collettive essenziali per la vita della nostra società (pluralismo democratico e culturale, distribuzione del rischio).

4. Da una parte la raccolta e l'analisi di grandi quantità di dati provenienti da diverse fonti ha grandi potenzialità. Dall'altra la legge promuove una logica di minimizzazione. Le promesse dell'intelligenza artificiale giustificano una revisione della legge?

5. È necessario promuovere un atteggiamento critico verso gli algoritmi e i sistemi automatizzati e autonomi in generale.

6. Si può arrivare a parlare degli algoritmi come soggetti dotati di una propria etica?

Nella seconda parte del dibattito sono state formulate delle possibili risposte alle sei questioni etiche individuate come fondamentali, nella forma di due principi fondanti di un'intelligenza artificiale al servizio degli esseri umani:

1. **Principio di lealtà**, lealtà non solo verso i singoli individui, ma verso la comunità.

2. **Principio di vigilanza e riflessività**. Allo scopo di contrastare l'inconoscibilità e l'opacità degli algoritmi, in particolare di quelli di machine learning, e arginare la diluizione delle responsabilità, occorre predisporre delle procedure di *auditing* periodiche e concrete da parte di tutti i soggetti coinvolti nelle attività degli algoritmi.

Insieme sono state sviluppate due considerazioni sull'approccio normativo europeo, quello contenuto nel GDPR:

1. ripensare l'obbligo di intervento umano nelle decisioni;

2. organizzare il processo di comprensione degli algoritmi e di identificazione delle responsabilità.

Queste riflessioni sono confluite nelle **sei raccomandazioni finali** rivolte tanto ai decisori politici quanto ai diversi membri della società civile:

1. educare all'etica tutti gli anelli della catena algoritmica (sviluppatori, aziende, cittadini);

2. rendere gli algoritmi comprensibili agli utenti e favorire una mediazione sul loro funzionamento;
3. progettare gli algoritmi in modo che siano al servizio delle libertà fondamentali;
4. costruire una piattaforma nazionale di *auditing* degli algoritmi;
5. incoraggiare la ricerca su un'intelligenza artificiale etica, lanciando un progetto di interesse nazionale;
6. rafforzare gli uffici che si occupano di etica nelle aziende.

4.2.2 La concertazione cittadina

L'evento conclusivo del dibattito è stato una concertazione cittadina svoltasi a Montpellier il 14 ottobre 2017, che ha visto 37 partecipanti selezionati per autocandidatura in risposta all'appello degli organizzatori diffuso su diverse piattaforme (stampa, social media, consigli di quartiere[a], fiera delle associazioni di Montpellier). L'evento è stato organizzato e realizzato da cinque facilitatori professionisti dell'associazione Lisode, specializzata in processi partecipativi.

La giornata si è svolta in quattro sessioni.

I. Una sessione interattiva, con video, giochi e laboratori, in cui scoprire cosa sono gli algoritmi, in quali ambiti vengono usati e quali effetti potrebbero avere per la società.

II. Una sessione plenaria per:

   - presentare più compiutamente la tematica oggetto di discussione e fornire alcuni elementi di base riguardanti le leggi e i regolamenti rilevanti;
   - presentare quattro casi di studio;
   - costituire quattro gruppi di lavoro.

III. Lavoro in sotto-gruppi, ciascuno su un caso di studio (salute, lavoro, motori di ricerca e profilazione, istruzione) per:

[a] strutture istituite nel 2002 nei comuni con più di 80 mila abitanti. Hanno il compito di essere il riferimento per i cittadini di un quartiere e di favorire la loro partecipazione democratica.

- identificare le opportunità offerte dall'utilizzo degli algoritmi in questo settore;
- identificare i rischi etici che l'utilizzo degli algoritmi in questo settore comporta;
- formulare delle raccomandazioni per uno sviluppo etico degli algoritmi in questo settore.

IV. Riunione dei sotto-gruppi e discussione collettiva dei quattro lavori:
   - prima con la tecnica di conversazione informale del world café;
   - poi chiedendo di esprimere a ciascun partecipante un grado di consenso per ciascuna delle raccomandazioni formulate in precedenza per i quattro casi di studio.

I risultati sono raccolti in dettaglio in CNIL [2017b] e possono essere riassunti nei seguenti punti:

- perdita di responsabilità e di competenza per impiegati e medici, che ricorreranno sempre più sistematicamente all'aiuto di un sistema automatizzato;
- gli algoritmi gestiscono l'incertezza e le eccezioni in modo peggiore rispetto alle persone, inoltre mancano di umanità (numerosi i riferimenti all'algoritmo che gestisce l'ammissione alle università francesi);
- mancanza di trasparenza;
- in relazione ai motori di ricerca e ai social media viene sottolineata una progressiva e pericolosa diluizione della responsabilità;
- il rischio di discriminazione e perpetuazione dei pregiudizi è percepito come particolarmente preoccupante nella gestione delle risorse umane, meno nelle attività di profilazione online.

4.3 Civic online debate: Governing The Rise Of Artificial Intelligence

Il "Civic online debate: Governing The Rise Of Artificial Intelligence " si svolge su un arco di tempo di 7 mesi sulla piattaforma digitale Assembl, sviluppata dalla società di Open Innovation francese Bluenove. Al 28 dicembre 2017 i partecipanti registrati erano 401.

Il dibattito è organizzato da The Future Society della Harvard Kennedy School in collaborazione con l'associazione Institute of Electrical and Electronics Engineers (IEEE) e la Japanese Society for Artificial Intelligence.

Il dibattito intende coinvolgere cittadini, professionisti, esperti e ricercatori nel settore dell'intelligenza artificiale, della robotica, delle politiche pubbliche, delle relazioni internazionali e dell'economia. L'obiettivo è formulare delle proposte e delle indicazioni di policy su una strategia di governance internazionale dell'intelligenza artificiale che sensibilizzino l'opinione pubblica e gli esperti di vari settori, spingendoli così a organizzare altri dibattiti pubblici.

La struttura del dibattito è quella tipica della piattaforma Assembl. Si articola in quattro fasi.

I. *Discovery*: la fase in cui il tema viene presentato rapidamente nei suoi aspetti fondamentali (7 settembre - 16 novembre 2017). Questa fase ha suddiviso il dibattito in quattro temi: "The AI Revolution", "AI for the common good", "AI Impact on the workforce", "Possible futures by 2045".

II. *Ideation/Divergence*: in cui i partecipanti partecipano a discussioni più dettagliate in forum suddivisi per temi (17 novembre 2017 - 31 gennaio 2018). Ciascun partecipante può esprimere la propria posizione rispetto a una questione specifica sollevata da altri utenti o dai moderatori e reagire alle opinioni altrui. I temi identificati per questa fase sono sei: "Reinvent Men and Machine Relationship", "AI safety and security", "Governance Frameworks", "Adapting the workforce for the age of AI", "Drive AI for Public Good", "Imaginaries of AI".

III. *Exploration*: in cui l'attenzione dei partecipanti viene convogliata su una selezione di idee emerse nella seconda fase (1 febbraio - 15 marzo 2018).

IV. *Convergenze*: le proposte migliori sono selezionate e organizzate in un documento che ne descriva le strategie di implementazione (15 - 31 marzo 2018).

Ad animare il dibattito ci sono quattro figure.

- Il coordinatore del dibattito: si occupa di organizzare il processo, pubblicizzarlo e coinvolgere il maggior numero di persone interessate, stabilire i contenuti iniziali per avviare il dibattito, identificare il formato che più si presta a riassumere i risultati del processo.

- Il *knowledge manager*: che segue lo sviluppo delle consultazioni in tutte le sue fasi eseguendo un'attività di *fact-checking* sulle informazioni condivise dai partecipanti, eventualmente proponendo dei contenuti utili allo svolgimento del dibattito.
- L'*harvester*: che seleziona le idee e i punti salienti emersi durante la fase di discussione libera e li riassume per proporli nel modo più efficace possibile durante la fase di esplorazione.
- Il *revealer*: cerca di identificare una cultura comune emergente dal dibattito.

4.4 Confronto

Nella Tabella 4.4 riassumiamo e confrontiamo le caratteristiche fondamentali dei due dibattiti che abbiamo considerato, secondo lo schema descritto nella sezione 4.1. Si tratta di due processi di portata molto diversa e con scopi

	Francia	**USA**
numero dei partecipanti	> 3000	> 400
durata	10 mesi	7 mesi
metodi di reclutamento	autoselezione	autoselezione
principali forme di comunicazione	articolazione di interessi negoziazione	articolazione di interessi
funzione del processo	consultazione	affinamento conoscenze personali influenzare opinione pubblica

Tabella 4.1: Tabella di confronto delle caratteristiche principali dei due dibattiti considerati.

molto diversi tra loro. Il dibattito francese nasce per volontà del legislatore, che ne dà mandato in un testo di legge e dunque si impegna a tener conto dei risultati che emergeranno dal dibattito. Il dibattito statunitense invece abbraccia uno spettro più ampio di tematiche riguardanti l'intelligenza artificiale con l'obiettivo di sollecitare dibattiti in aree più specifiche. Per avere un panorama più ampio dell'approccio statunitense al coinvolgimento dei cittadini vale la pena sottolineare che sia il Big Data Working Group che la Federal Trade Commission hanno previsto durante le loro attività l'apertura

della consultazione ai cittadini, oltre che ad aziende, associazioni e organizzazioni di varia natura. E alcuni cittadini hanno effettivamente espresso dei pareri. Non è chiaro però in quale considerazione siano stati tenuti nella stesura del documento finale, visto che dovevano essere confrontati e sintetizzati con le posizioni di società private del calibro di IBM.

Ciò che infine è interessante sottolineare è come la consultazione francese si sia trasformata anche in un momento di riflessione sugli strumenti normativi già in vigore, o in procinto di entrare in vigore, come il Regolamento Generale per la Protezione dei Dati. Il coinvolgimento con esperti e semplici cittadini ha permesso di evidenziare i limiti di applicabilità di alcuni diritti stabiliti dal Regolamento.

Conclusioni

Alcuni hanno definito la società contemporanea *data society*, la nostra economia *data economy*. Tre elementi di cambiamento hanno contribuito a rendere i *dati* così centrali. In primo luogo il fatto che i nostri comportamenti producono dati in grandissima quantità: nel 2017 abbiamo generato 2,5 miliardi di GB al giorno. In secondo luogo abbiamo a disposizione macchine computazionali sempre più potenti: il computer più potente nel 1945 era in grado di eseguire 500 operazioni al secondo, mentre il supercomputer più potente oggi ne esegue 100 mila miliardi ogni secondo. Terzo sappiamo progettare algoritmi sempre più sofisticati, in particolare capaci di analizzare e "apprendere" dai dati, offrendo delle soluzioni (delle previsioni) a quei problemi per cui esiste scarsa o nulla conoscenza del fenomeno osservato.

Così i dati hanno acquisito la posizione di primo piano che occupano oggi nell'industria, nella finanza, nella politica e nell'informazione. Gli algoritmi sanno processare i dati che produciamo ed estrarne delle informazioni utili ed è per questo che sono diventati strumenti di assistenza alla decisione, nei settori più diversi: sicurezza, giustizia, gestione e selezione del personale, accesso all'istruzione, accesso alle cure sanitarie, accesso al credito, assicurazioni. Ed è proprio negli ambiti in cui si prendono le decisioni più controverse che gli algoritmi sono apparsi come una soluzione ai problemi di oggettività e imparzialità che caratterizzano le scelte di qualsiasi essere umano. L'idea che le decisioni fossero basate sui risultati di procedure matematiche condotte sui dati ha convinto molti che l'automatizzazione fosse una soluzione. Chiaramente l'altra caratteristica dell'automatizzazione è l'efficienza, i tagli alle spese. E questa caratteristica ha reso ulteriormente allettante questa strategia.

Ma gli algoritmi sono tutt'altro che imparziali, tutt'altro che oggettivi. Del resto sono progettati da esseri umani e, soprattutto, apprendono dai dati storici sui nostri comportamenti. In quei dati sono scritti i nostri valori, le dinamiche sociali, incluse quelle ingiuste e discriminatorie, che gli algoritmi

apprendono e replicano su una scala e a una velocità irraggiungibile per gli esseri umani. Per essere chiari facciamo un esempio. Immaginiamo il direttore dell'ufficio del personale di una grande azienda che vuole rendere più efficiente le procedure di selezione dei nuovi impiegati. Ha a disposizione i dati storici sulle assunzioni passate e mette a lavoro un gruppo di informatici con il mandato di trovare quali caratteristiche degli impiegati in passato sono correlate con i percorsi 'di successo'. Per andare avanti gli informatici hanno però bisogno di definire un impiegato di successo e il direttore gli risponde che per lui 'successo' vuol dire che l'impiegato è rimasto in azienda almeno cinque anni dopo l'assunzione ed è stato promosso almeno due volte. Gli informatici svilupperanno un codice in grado di analizzare i CV degli impiegati al momento dell'assunzione per capire quali dettagli sono più frequenti negli impiegati di 'successo', per come li ha definiti il direttore. Ebbene supponiamo che trovino che c'è una caratteristica, su tutte, che presenta la correlazione maggiore: l'impiegato deve essere maschio. Bene: la conclusione è che gli impiegati maschi sono investimenti più sicuri per l'azienda. L'algoritmo ignora i motivi per i quali il sesso maschile garantisca il successo, ma noi supponiamo di indagare tra i dipendenti della società e scoprire che è frequente che le donne vengano marginalizzate dopo la maternità o addirittura licenziate. È chiaro, dunque, che non c'è nessun rapporto di causa-effetto tra sesso e 'successo', ma la correlazione trovata dall'algoritmo fotografa un fenomeno le cui cause sono da condannare. Eppure, basando le procedure di selezione su questo algoritmo, le donne non arriveranno neanche al colloquio e pregiudizio e discriminazione saranno perpetrate, sotto l'aura di oggettività della matematica e dei dati.

Dunque se da una parte il binomio algoritmi e big data promette di essere un vettore per il rilancio dell'economia e per la conquista di una sempre migliore qualità della vita, dall'altra occorre vigilare perché non ci siano vittime senza possibilità di appello in questo processo. La particolarità di questo problema risiede nel fatto che spesso gli algoritmi sono scatole nere, di cui possiamo sapere poco o niente. In alcuni casi perché protette dal segreto commerciale, in altri perché il loro sempre crescente grado di complessità li rende di difficile comprensione per coloro su cui vengono prese le decisioni.

Il problema è arrivato ormai all'attenzione pubblica in diversi contesti, con l'obiettivo non solo di denunciare l'attuale stato di cose, ma anche di trovare delle strategie per la *governance* degli algoritmi, che tengano conto della dimensione sociale e giuridica. A iniziare la discussione sono stati tre gruppi sociali: giuristi, data scientist, attivisti per i diritti. Per questo motivo, la tesi ha analizzato il dibattito sulla governance degli algoritmi da tre diversi punti di vista: quello legislativo e di policy, quello matematico e informatico

e infine quello compiutamente pubblico, ovvero della società civile. In questa analisi abbiamo confrontato il contesto statunitense con quello europeo, rintracciando differenze e tratti comuni sia nel metodo che nel contenuto.

Dal **punto di vista giuridico** e di policy il punto di partenza, comune sia al contesto americano che europeo, è la legislazione sulla privacy. Fortemente influenzato dagli avanzamenti tecnologici, il diritto alla privacy ha una duplice interpretazione: da una parte il diritto all'autodeterminazione e dall'altra il diritto alla protezione dei dati personali. All'inizio il diritto alla protezione dei dati personali era funzionale al diritto all'autonomia nello stabilire la direzione della propria vita personale e familiare: i cittadini devono essere liberi di condividere con i Governi solo le informazioni che ritengono opportune al fine di definire la propria personalità nel contesto della vita democratica. Con il passare del tempo e in particolare con l'informatizzazione dei processi di acquisizione e analisi dei dati e con la crescita del settore privato legato alla gestione e analisi dei dati, questi due diritti si sono in qualche senso disaccoppiati e la protezione dei dati personali è diventato un diritto da tutelare indipendentemente dal diritto all'autodeterminazione. Questa cesura è stata particolarmente rilevante nella legislazione comunitaria europea, prima tramite la Direttiva 95/46 sulla privacy e poi con il Regolamento Generale sulla Protezione dei Dati, che entrerà in vigore il 25 maggio del 2018. Negli Stati Uniti, al contrario, si è scelta la strada dell'autoregolamentazione, soprattutto nel settore privato.

Se il punto di partenza è comune, il punto di arrivo dei due dibattiti, americano ed europeo, è molto diverso. Nei rapporti pubblicati nel 2016 dal Big Data Working Group dell'Executive Office di Obama, viene infatti sottolineato come i diritti minacciati dall'automatizzazione incontrollata di alcuni processi di decisione riguardino i criteri di giustizia distributiva, l'uguaglianza e l'accesso equo e non la vita privata o il trattamento dei dati personali. In particolare si solleva il problema del diritto a uguali opportunità indipendentemente dal sesso, la provenienza geografica e le condizioni socioeconomiche di partenza. La conclusione del rapporto stabilisce che il principio guida deve essere quello dell'equità. Le raccomandazioni finali includono: la necessità di finanziare e favorire la ricerca sulla progettazione di algoritmi che siano intrinsecamente equi (*equity-by-design*), la necessità di progettare procedure di *auditing* dell'intero processo di automatizzazione della decisione (dal campione di dati, all'algoritmo che li analizza, fino ai modi in cui vengono utilizzati i risultati dell'algoritmo), la necessità di rendere i cittadini sempre più consapevoli delle implicazioni tecniche e giuridiche degli strumenti e dei servizi digitali di cui si servono ogni giorno.

Nel contesto europeo l'approccio è rimasto molto più fermo sulla tutela della privacy come diritto alla riservatezza dei propri dati personali, ribadendo l'importanza di progettare procedure di archiviazione e analisi dei dati che rispettino il principio di *privacy-by-design*. Probabilmente lo sforzo normativo impiegato nella stesura del Regolamento Generale sulla Protezione dei Dati ha in qualche senso totalizzato l'attenzione dei legislatori. All'interno del Regolamento il problema degli algoritmi è affrontato nell'Articolo 22, che riguarda la decisione automatizzata. L'articolo stabilisce che nessuna decisione significativa per la vita di un cittadino dovrebbe essere presa senza l'intervento di un essere umano, ma da una parte non è chiaro quali decisioni possano essere classificate come 'significative' e dall'altro sembra poco plausibile richiedere l'intervento umano in un numero di situazioni che si preannuncia molto elevato. L'Articolo 22 stabilisce inoltre che il cittadino che ritiene ingiusta la decisione dell'algoritmo ha diritto a una revisione della decisione e a conoscere la logica del software che lo ha 'giudicato'. Questi diritti sembrano difficili da esercitare. Da una parte richiedono che il cittadino sia consapevole che a determinare una certa circostanza sia stato un algoritmo, dall'altra presuppone che la logica dell'algoritmo sia di facile spiegazione e rintracciabilità.

L'Article 29 Working Party, un organo consultivo che raggruppa le autorità garanti della privacy degli Stati membri dell'Unione, ha sollevato proprio questi problemi, aggiungendo un punto importante. Anche le attività di profilazione condotte a scopi di marketing possono violare il diritto alla privacy, inteso come diritto all'autodeterminazione. L'iper-personalizzazione dei risultati di un motore di ricerca online come Google può infatti nascondere alcuni elementi informativi importanti per prendere decisioni in fatto di salute, istruzione, ecc.. Lo stesso vale per i social network, che offrendoci una dieta informativa che sia il più *engaging* (coinvolgente) possibile, ci chiudono in bolle che rinforzano certe nostre convinzioni esponendoci raramente a posizioni diverse dalla nostra. Sulla base dell'analisi dell'Article 29 WP e dello European Data Protection Supervisor, il Parlamento Europeo ha adottato una Risoluzione che conferma la necessità di governare le conseguenze del massiccio utilizzo dei big data e delle tecniche di analisi connesse, nel rispetto del diritto alla privacy, alla non discriminazione e alla libertà di espressione, aggiungendo inoltre l'importanza di vigliare sul rischio che venga limitato l'accesso a un ambiente informativo pluralistico.

Il **secondo punto di vista** da cui abbiamo riflettuto sul problema della governance degli algoritmi è quello matematico e della scienza dei dati più in generale. La sensibilità sempre crescente della comunità verso questo tema è testimoniata dalla nascita di gruppi di lavoro e cicli di conferenze, come

il "Fairness Accountability and Transparency of Machine Learning" giunto alla sua quarta edizione. L'equità degli algoritmi come argomento di ricerca è naturalmente evoluto da quello della crittografia e dei *privacy preserving algorithms*. A partire dagli anni '90 Cynthia Dwork, oggi Gordon McKay Professor of Computer Science presso la Harvard Paulson School of Engineering and Applied Sciences, e i suoi collaboratori hanno introdotto e sviluppato l'idea della *differential privacy*. La definizione di differential privacy risponde alla domanda: cosa vuol dire che un data base rispetta la *privacy* dei soggetti che vi hanno contribuito? La risposta di Dwork e colleghi è: il data base rispetta la privacy se i risultati dell'analisi condotta su quel campione non cambiano se sostituisco il contributo di un individuo con quello di un altro individuo o se cancello o aggiungo i dati relativi a un singolo individuo. L'idea è dunque che sia sufficiente intervenire sulla base dati per tutelare i diritti dei cittadini. Tuttavia l'analisi di un data base comporta, necessariamente, l'acquisizione di nuove informazioni sulla popolazione di cui il campione è rappresentativo. In questo senso è possibile che ogni cittadino appartenente a quella popolazione possa subire le conseguenze di questa analisi, pur non avendo ceduto alcuna informazione personale. Un esempio semplice è uno studio clinico di associazione tra fumo e rischio di sviluppare un tumore ai polmoni. Se lo studio clinico, condotto diciamo su 1000 casi, conclude che l'abitudine al fumo è correlata con un rischio aumentato di ammalarsi di tumore ai polmoni, è possibile che tutti i fumatori vedano aumentare il premio della loro assicurazione sanitaria, anche se non erano affatto coinvolti nello studio clinico.

Sono queste riflessioni che hanno modificato la domanda di ricerca di alcuni gruppi di data scientist: come possiamo progettare algoritmi equi? Dwork e collaboratori hanno affrontato per primi il problema, inquadrandolo da un punto di vista del tutto generale. Questo approccio si è finora scontrato sul piano dei fatti. Un buon esempio e l'algoritmo COMPAS. COMPAS stima il rischio di recidiva di un imputato di un crimine partendo da un questionario di 137 domande di diversa natura: sul contesto familiare e sociale ("If you lived with both parents and they later separated, how old were you at the time?", "How many of your friends have ever been arrested?", "How often have you moved in the last 12 months?"), comportamentali e psicologiche ("A hungry person has a right to steal. [Say yes or no]", "If people make me angry or lose my temper, I can be dangerous. [Say yes or no]"), o addirittura rivolte all'agente che ha in custodia il presunto criminale ("Based on the screener's observations, is this person suspected or admitted gang member?"). COMPAS assegna all'imputato un punteggio da 1 a 10, a seconda che il rischio che lui o lei commetta un nuovo crimine nel prossimo futuro sia

basso (da 1 a 3) o medio/alto (da 4 a 10). A partire dagli anni '80 molti tribunali americani utilizzano algoritmi simili per aiutare i giudici a stabilire se l'arrestato deve aspettare il processo in carcere o può essere rilasciato, oppure a decidere la lunghezza della sentenza di condanna al termine di un processo. Un'inchiesta del giornale ProPubica, pubblicata nel 2016, ha dimostrato che l'algoritmo penalizza i neri rispetto ai bianchi. In particolare ha dimostrato che la percentuale di falsi positivi (individui che hanno ricevuto un punteggio alto dall'algoritmo, ma che non hanno commesso nuovi crimini nei due anni successivi) è più alta tra i neri (44.9%) che tra i bianchi (23.5%). La società di consulenza che ha sviluppato e commercializzato il software ha risposto alle critiche dicendo che l'algoritmo ha la stessa accuratezza (percentuale di persone che, avendo ricevuto un punteggio alto, effettivamente commettono un nuovo crimine nei due anni successivi) nelle due popolazioni. Si tratta, in effetti, di due nozioni diverse di equità: quella che prevede un uguale numero di falsi positivi nelle due popolazioni e quella che prevede un uguale livello di accuratezza nelle due popolazioni, la cosiddetta *predictive parity*. Quattro diversi gruppi di ricercatori negli Stati Uniti hanno concluso che non è possibile progettare un algoritmo che rispetti contemporaneamente i due principi di equità. Il motivo è che le due popolazioni sono rappresentate in proporzioni diverse nel campione di dati: i neri vengono arrestati più dei bianchi. La soluzione sarebbe quindi di usare strumenti diversi per le due popolazioni, ma questo richiederebbe di rendere visibile all'algoritmo l'informazione sul colore della pelle dell'arrestato, un'informazione che viene considerata sensibile e quindi omessa. Questa conclusione appare contraddittoria, ma rende esplicito un concetto semplice: l'unica forma di equità possibile è quella che passa dalla consapevolezza delle diversità e delle disuguaglianze che caratterizzano le nostre società. Cynthia Dwork lo ha sintetizzato bene con l'espressione *fairness through awareness*.

Il **terzo punto di vista** attraverso il quale abbiamo studiato il problema della governance degli algoritmi è quello dei cittadini. Lo abbiamo fatto considerando due processi partecipativi: il dibattito "Éthique numérique" organizzato dalla "Commission Nationale Informatique & Libertés" (CNIL) in Francia e il "Civic online debate: Governing The Rise Of Artificial Intelligence" della Harvard Kennedy School negli Stati Uniti. Nel confrontare questi due dibattiti ci siamo concentrate sulle dinamiche che li hanno prodotti e sulle finalità per cui sono stati organizzati.

Il dibattito francese è avvenuto su mandato della "Loi pour une République numérique", la legge promulgata a ottobre del 2016 in vista del Regolamento Generale sulla Protezione dei Dati con l'obiettivo di guidare la transizione digitale mettendo al centro i cittadini e i loro diritti. L'idea è quindi che

il legislatore tenga in considerazione le conclusioni del processo nella sua attività futura. Diversamente il "Civic online debate" statunitense ha come obiettivo quello di sensibilizzare l'opinione pubblica, ma più precisamente gli esperti dei vari settori per cui i dati sono diventai vitali, al problema delle implicazioni etiche dell'intelligenza artificiale. Il risultato finale del processo dovrebbe essere quello di favorire dibattiti specifici nei diversi ambiti.

"Éthique numérique" si è svolto su un arco di dieci mesi e ha coinvolto oltre 3000 esperti di diversi settori in cui i dati e gli algoritmi sono e saranno sempre più importati (salute, assicurazioni, istruzione, pubblica amministrazione) e ha infine coinvolto circa 40 cittadini in un dibattito conclusivo. Dal dibattito sono emersi due tipi di conclusioni. Da una parte i rischi connessi all'automatizzazione di alcuni processi decisionali: la fiducia cieca nell'oggettività dei numeri e delle procedure informatiche, la perpetuazione di pregiudizi e ingiustizie contenuti nei dati su cui gli algoritmi si allenano, la diluizione delle responsabilità, la restrizione dell'accesso a un panorama informativo pluralistico. Dall'altra il dibattito ha formulato delle raccomandazioni, di cui elenchiamo le più rilevanti: costruire una piattaforma nazionale di *auditing* degli algoritmi, rendere gli algoritmi comprensibili agli utenti e contemporaneamente favorire una mediazione sulla comprensione del loro funzionamento, educare all'etica tutti gli anelli della catena algoritmica.

Il dibattito statunitense ha una natura completamente diversa. Prima di tutto per le finalità, come abbiamo già detto, poi per la natura degli organizzatori, principalmente accademici, e infine per le modalità con cui viene portato avanti. Si tratta infatti di un processo partecipativo online sviluppato su un arco di sette mesi e articolato in quattro fasi ("discovery", "ideation/divergence", "exploration", "convergence") che si concluderà a marzo del 2018.

Dalla nostra analisi emerge una caratteristica fondamentale del problema della governance degli algoritmi: è necessario il coinvolgimento di diversi gruppi sociali per riuscire a progettare una strategia efficace. Il legislatore e il politico hanno bisogno di comunicare con i cittadini, per capire l'applicabilità, nella pratica, di alcune procedure che hanno previsto in teoria. Un esempio su tutti è l'Articolo 22 del GDPR: ci sarà bisogno di un confronto con i cittadini e con le imprese per capire quanto sia possibile 'condividere' il contenuto degli algoritmi viste le competenze a disposizione dei cittadini e le procedure in atto nelle aziende. Allo stesso tempo legislatori e politici hanno bisogno di comunicare con i tecnici e gli scienziati, che devono comprendere quali principi etici e quali diritti devono essere incorporati negli algoritmi e tradurli in linguaggio matematico. Se in questa attività di 'traduzione' si scoprisse

poi che certi principi non sono matematicamente compatibili con altri, allora sarebbe fondamentale per gli scienziati tornare a confrontarsi con i legislatori e i politici per capire quali compromessi sono possibili.

L'idea che emerge dalla nostra analisi è che sarà necessario attraversare una fase laboratoriale, ovvero sperimentale, dal punto di vista normativo, politico e tecnologico. Sarà necessario, a nostro avviso, considerare problemi circoscritti, trovare soluzioni temporanee da sperimentare sul campo per testarne l'efficacia tecnica e l'accettabilità sociale, per poi tornare a riflettere ed eventualmente consolidare le esperienze positive e imparare da quelle negative. Particolare attenzione dovrebbe essere dedicata al coinvolgimento dei cittadini nell'attività del legislatore. Sull'impronta del dibattito francese, si dovrebbe considerare la comunicazione tra legislatore e cittadini come un momento fondamentale nel processo di formulazione delle leggi, ripensando e integrando le forme di consultazione pubblica fin qui attuate.

Note

Introduzione

[1]**2,5 miliardi di GB al giorno**: IBM Marketing Cloud, "10 Key Marketing Trends 2017", 2017. https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN. Data ultimo accesso: 9 dicembre 2017.

Capitolo 1

[2]**mostra che solo il 5% degli stabilimenti dell'azienda sono automatizzati**: Tom Bonkenburg, "Robotics in Logistics. A DPDHL perspective on implications and use cases for the logistics industry", *DHL Customer Solutions & Innovation*, May 2016. http://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/dhl_trendreport_robotics.pdf. Data ultimo accesso: 8 dicembre 2017.

[3]**le tre leggi della robotica di Asimov hanno bisogno di un aggiornamento**: Mark Robert Anderson, "After 75 years, Isaac Asimov's Three Laws of Robotics need updating", *The Conversation UK*, 17 March 2017. https://theconversation.com/after-75-years-isaac-asimovs-three-laws-of-robotics-need-updating-74501. Data ultimo accesso: 8 dicembre 2017.

[4]**Allen Newell**: nel 1955-56 Allen Newell programmò insieme a Herbert Simon Logic Theorist, il primo sistema informatico che cercava di imitare l'attività di *problem solving* degli esseri umani. Viene considerato il primo esempio di intelligenza artificiale.

[5]**DeepBlue sviluppato da IBM sconfisse il campione di scacchi Garry Kasparov**: Alex Q. Arbuckle, "1996-1997. The Kasparov-Deep Blue chess matches. An unprecedented battle of wits between man and machine", *Mashable*. http://mashable.com/2016/02/10/kasparov-deep-blue/#uKAFHgWuXEqc. Data ultimo accesso: 8 dicembre 2017.

[6]**gli hiring algotihms difficilmente saranno più neutrali degli esseri umani**: Gideon Mann, Cathy O'Neil, "Hiring Algorithms Are Not Neutral", *Harvard Business Review*, December 09, 2016. https://hbr.org/2016/12/hiring-algorithms-are-not-neutral. Data ultimo accesso: 8 dicembre 2017.

[7]**la storia di Kyle Behm**: Cathy O'Neil, " How algorithms rule our working lives', *The Guardian*, September 01, 2016. https://www.theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives. Data ultimo accesso: 8 dicembre 2017.

[8]Il paragrafo è tratto da: Chiara Sabelli, "Algoritmi ingiusti", *Scienza in rete*, 17 luglio 2017. http://www.scienzainrete.it/articolo/algoritmi-ingiusti/chiara-sabelli/2017-07-17. Data ultimo accesso: 28 gennaio 2018.

[9]**Sarah Wisoki viene licenziata a luglio 2011**: Bill Turque, " Creative, motivated and fired", *The Washington Post*, March 6, 2012. https://www.washingtonpost.com/local

/education/creative--motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html?utm_term=.a005a9d00f5c. Data ultimo accesso: 8 dicembre 2017.

[10]**programma di valutazione IMPACT**: sito web del District of Columbia Public Schools, " IMPACT: An Overview". https://dcps.dc.gov/page/impact-overview. Data ultimo accesso: 8 dicembre 2017.

[11]**il Value Added Model implementato da Mathematica Policy Research per il programma IMPACT**: Eric Isenberg, Heinrich Hock, " Measuring School and Teacher Value Added for IMPACT and TEAM in DC Public Schools", *Mathematica Policy Research*, August 20, 2010. https://www.mathematica-mpr.com/our-publications-and-findings/publications/measuring-school-and-teacher-value-added-for-impact-and-team-in-dc-public-schools. Data utimo accesso: 8 dicembre 2017.

[12]**Un'altra insegnante di Washington D.C., Sarah Bax, aveva tentato invano di ottenere dettagli sul suo punteggio**: Gfbrandenburg's Blog, " DCPS Administrators Won't or Can't Give a DCPS Teacher the IMPACT Value-Added Algorithm", February 27, 2011. https://gfbrandenburg.wordpress.com/2011/02/27/. Data ultimo accesso: 8 dicembre 2017.

[13]**un'inchiesta del giornale USA TODAY**: Greg Toppo, " Memo warns of rampant cheating in D.C. public schools", *USA TODAY*, April 11, 2013. https://www.usatoday.com/story/news/nation/2013/04/11/memo-washington-dc-schools-cheating/2074473/. Data ultimo accesso: 8 dicembre 2017.

[14]**applicare metodi analitici e statistici per prevedere le performance dei lavoratori**: Don Peck, " They're Watching You at Work. What happens when Big Data meets human resources? The emerging practice of "people analytics" is already transforming how employers hire, fire, and promote.", *The Atlantic*, December 2013. https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/.Data ultimo accesso: 8 dicembre 2017.

[15]**adattarono i modelli di previsione dei terremoti**: Ellen Huet, "Server And Protect: Predictive Policing Firm PredPol Promises To Map Crime Before It Happens", *Forbes*, March 2, 2015. https://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing/#2deff0ae4f9b. Data ultimo accesso: 7 dicembre 2017.

[16]**approfondimento su PredPol**: Nate Berg, "Predicting Crime, LAPD-style", *The Guardian*, 24 June 2014. https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report. Data ultimo accesso: 7 dicembre 2017.

[17]**aumenta il periodo di detenzione media**: Mario Venturi, intervista telefonica 24 novembre 2017, 5 dicembre 2017.

[18]**la trasparenza dei sistemi di polizia predittiva è di fondamentale importanza**: intervista a Giovanni Mastrobuoni, 1 dicembre 2017.

[19]**Zurigo**: "Spezial-Software: Polizei verhindert Einbrüche, bevor sie geschehen", *Aargauer Zeitung*, 1 Dezember 2015. https://www.aargauerzeitung.ch/aargau/kanton-aargau/spezial-software-polizei-verhindert-einbrueche-bevor-sie-geschehen-128979133. Data ultimo accesso: 7 dicembre 2017.

[20]**Aargau**: Werner Schüepp, "Achtung, bei Ihnen droht ein Einbruch", *Tages Anzeiger*, 4 September 2015. https://www.tagesanzeiger.ch/zuerich/stadt/Achtung-bei-Ihnen-droht-ein-Einbruch/story/29427848. Data ultimo accesso: 7 dicembre 2017.

[21]**Berlino**: Sarah Griffiths, "Predicting crimes BEFORE they happen: Berlin police adopt Minority Report-style software that seeks out criminal behaviour", *Daily Mail*, 2 December 2014. http://www.dailymail.co.uk/sciencetech/article-2857814/Predict-crimes-happen-Berlin-police-adopt-Minority-Report-style-software-seeks-criminal-behaviour.html. Data ultimo accesso: 7 dicembre 2017.

[22]**Monaco**: "Herrmann berichtet über Erfahrungen des Precobs-Tests in München und Mittelfranken", *Pressemitteilungen Bayerisches Staatsministerium des Innern*, 24 Juni 2015. https://www.stmi.bayern.de/med/pressemitteilungen/pressearchiv/2015/204/index.php. Data ultimo accesso: 7 dicembre 2017.

[23]**dubbi sull'efficacia di PRECOBS**: Kai Biermann, "Noch hat niemand bewiesen, dass Data Mining der Polizei hilft", *Zeitung*, 29 März 2015. http://www.zeit.de/digital/datenschutz/2015-03/predictive-policing-software-polizei-precobs. Data ultimo accesso: 7 dicembre 2017.

[24]**Accenture**: "London Metropolitan Police Service and Accenture Police Solutions complete analytics pilot program to fight gang crime", 2015. https://www.accenture.com/t20160415T055944__w__/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_8/Accenture-London-Metropolitan-Police-Service-And-Acn-Police-Solutions-v2.pdf. Data ultimo accesso: 4 dicembre 2017.

[25]**PredVol**: Mickaël Sizine, "Quand les gendarmes prédisent les vols de voitures", *Le-Parisien*, 30 janvier 2017. http://www.leparisien.fr/espace-premium/oise-60/quand-les-gendarmes-predisent-les-vols-de-voitures-30-01-2017-6635961.php. Data ultimo accesso: 6 dicembre 2017.

[26]**VALCRI**: sito web del progetto VALCRI, 2017. http://valcri.org/. Data ultimo accesso: 4 dicembre 2017.

[27]**ammonta a 16,5 milioni di euro**: Community Research and Development Information Service della Commissione Europea, Project Details, 2014. http://cordis.europa.eu/project/rcn/188614_en.html. Data ultimo accesso: 4 dicembre 2017.

[28]**la West Midlands Police sta testando VALCRI su 6,5 milioni e mezzo di dati raccolti in tre anni**: Timothy Revell, "AI detective analyses police data to learn how to crack cases", *New Scientist*, May 10, 2017. https://www.newscientist.com/article/mg23431254-000-ai-detective-analyses-police-data-to-learn-how-to-crack-cases/. Data ultimo accesso: 4 dicembre 2017.

[29]**progetto ICONIC all'Imperial College London**: Hayley Dunning, "Predictive policing research gets a boost from £3m grant", *Imperial College News*, 21 March 2017. http://www3.imperial.ac.uk/newsandeventspggrp/imperialcollege/newssummary/news_21-3-2017-13-40-12. Data ultimo accesso: 4 dicembre 2017.

[30]**finanziamento progetto ICONIC**: Engineering and Physical Sciences Research Council, Details of grant, 2017. http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/P020720/1. Data ultimo accesso: 7 dicembre 2017.

[31]**inchiesta di ProPublica sull'algoritmo COMPAS**: Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks", *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Data ultimo accesso: 7 dicembre 2017.

[32]Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "How we analyzed the COMPAS recidivism algorithm", *ProPublica*, May 23, 2016. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm/. Data ultimo accesso: 7 dicembre 2017.

[33]Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say", ProPublica, December 30, 2016. https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say. Data ultimo accesso: 7 dicembre 2017.

[34]**The Intercept ha pubblicato dei documenti riservati della National Security Agency sul programma Skynet**:

Cora Currier, Glenn Greenwald, Andrew Fishman, "U.S. Government designated prominent Al Jazeera journalist as 'member of Al Qaeda' ", *The Intercept*, May 8, 2015. https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/. Data ultimo accesso: 11 dicembre 2017.

"SKYNET: Applying Advanced Cloud-based Behavior Analytics", *The Intercept*, May 8, 2015. https://theintercept.com/document/2015/05/08/skynet-applying-advanced-cloud-based-behavior-analytics/. Data ultimo accesso: 11 dicembre 2017.

"SKYNET: Courier Detection via Machine Learning", *The Intercept*, May 8, 2015. https://theintercept.com/document/2015/05/08/skynet-courier/. Data ultimo accesso: 11 dicembre 2017.

[35] **commenti di Fosca Giannotti, direttrice del laboratorio Knowledge Discovery and Data Mining (KDD) del CNR di Pisa, e di Dino Pedreschi, co-direttore del KDD**: intervista telefonica del 23 luglio 2016.

[36]**almeno una parte degli oltre 400 attacchi con droni compiuti in Pakistan dal 2004 a oggi, sono stati pianificati sulla base di questi algoritmi**: Christian Grothoff and J.M. Porup, "The NSA's SKYNET program may be killing thousands of innocent people", *Ars Technica*, February 16, 2016. https://arstechnica.com/information-technology/2016/02/the-nsas-skynet-program-may-be-killing-thousands-of-innocent-people/. Data ultimo accesso: 11 dicembre 2017.

[37]**Il programma si chiama Social Credit System**: Shai Oster, "China Tries Its Hand at Pre-Crime", *Bloomberg Businessweek*, March 3, 2016. https://www.bloomberg.com/news/articles/2016-03-03/china-tries-its-hand-at-pre-crime. Data ultimo accesso: 11 dicembre 2017..

[38]**dichiara Alessandro Vinciarelli**: intervista telefonia del 22 luglio 2016.

Capitolo 2

[39]**stato della discussione sul "Consumer Privacy Bill of Rights"**: Natasha Singer, "Why a Push for Online Privacy Is Bogged Down in Washington", *The New York Times*, February 28, 2016. https://www.nytimes.com/2016/02/29/technology/obamas-effort-on-consumer-privacy-falls-short-critics-say.html?_r=0. Data ultimo accesso: 17 dicembre 2017.

[40]**il caso di Walter e Paula Shelton**: Chad Terhune, "They know what's in your medicine cabinet", *Bloomberg Business Week*, July 22, 2008. https://www.bloomberg.com/news/articles/2008-07-22/they-know-whats-in-your-medicine-cabinet. Data ultimo accesso: 17 dicembre 2017.

[41]**for-profit college**: Patricia Cohen, "For-Profit Colleges Accused of Fraud Still Receive U.S. Funds", *The New York Times*, October 12, 2015. https://www.nytimes.com/2015/10/13/business/for-profit-colleges-accused-of-fraud-still-receive-us-funds.html. Data ultimo accesso: 17 dicembre 2017.

[42]**la Seconda Guerra Mondiale e le leggi sulla privacy in Europa**: Thomas Show, "Privacy Law and History: WWII-", *International Association of Privacy Professionals*, March 1, 2013. https://iapp.org/news/a/2013-03-01-privacy-law-and-history-wwii-forward/. Data ultimo accesso: 20 dicembre 2017.

[43]**Fairness Accountability and Transparency of Machine Learning**: sito web del gruppo https://www.fatml.org/.

[44]**Differential Privacy Symposium: Four Facets of Differential Privacy, Institute of Advanced Studies**: Cynthia Dwork, "The Definition of Differential Privacy", November 12, 2016. https://www.youtube.com/watch?v=lg-VhHlztqo. Data ultimo accesso: 17 dicembre 2017.

[45]**intervista di Kevin Hartnett a Cynthia Dwork**: Kevin Hartnett, "How to Force Our Machines to Play Fair", *Quanta Magazine*, November 26, 2016. https://www.quantamagazine.org/making-algorithms-fair-an-interview-with-cynthia-dwork-20161123. Data ultimo accesso: 17 dicembre 2017.

Capitolo 3

[46]**dibattito pubblico Éthique Numérique**: https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de. Data ultimo accesso: 22 dicembre 2017.

[47]**civic online debate "Governing The Rise Of Artificial Intelligence"**: http://ai-initiative.org/ai-consultation/. Data ultimo accesso: 22 dicembre 2017.

[48]siamo consapevoli dei limiti di un'analisi condotta su un dibattito ancora in corso, ma ci soffermeremo sugli aspetti organizzativi e sui dati di partecipazione rilevati fino a questo momento

Bibliografia

Sherry R. Arnstein. A ladder of citizen participation. *Journal of the American Institute of Planners*, 35(4):216–224, 1969. doi:10.1080/01944366908977225. URL https://doi.org/10.1080/01944366908977225.

Article29 WP (Article 29 Data Protection Working Party). Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU (14/EN, WP221), September 2014. URL http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf. Data ultimo accesso: 20 dicembre 2017.

Article29 WP. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (17/EN, WP251), October 2017. URL http://ec.europa.eu/newsroom/just/document.cfm?doc_id=47963. Data ultimo accesso: 20 dicembre 2017.

Marianne Bertrand and Sendhil Mullainathan. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. Working Paper 9873, National Bureau of Economic Research, July 2003. URL http://www.nber.org/papers/w9873.

D.G. Bobrow and P.J. Hayes. Artificial intelligence-where are we? *Artificial Intelligence*, 25:385, 1985. In responce to question 2 in a questionnaire.

Nick Bodstrom and Eliezer Yudkowski. *The Ethics of Artificial Intelligence*. In *The Cambridge Handbook of Artificial Intelligence, 2014, Cambridge University Press, UK*, pages 316–334. Cambridge University Press, UK, 2014. ISBN 978-0521691918. doi:10.1007/11681878_14.

Bruce G. Buchanan. A (very) brief history of artificial intelligence. *AI Magazine*, 26(4):53–60, 2005.

Lee A. Bygrave. EU data protection law falls short as desirable model for algorithmic regulation. Center for Analysis of Risk and Regulation, King's College London, DISCUSSION PAPER, 85:31–33, Semptember 2017. URL https://www.kcl.ac.uk/law/research/centres/telos/assets/DP85-Algorithmic-Regulation-Sep-2017.pdf.

CAST (Council of Advisors on Science and Technology of President). Big data and privacy: a technological perspective, May 2014. URL https://obamawhitehouse.archives.gov/sites/default/files/m

icrosites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf. Data dell'ultimo accesso: 17 dicembre 2017.

CNIL (Commission national de l'informatique et des libertés). Comment permettre à l'homme de garder la main?, December 2017. URL https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf. Data ultimo accesso: 24 dicembre 2017.

CNIL. Concertation citoyenne sur les enjeux ethiques lies a la place des algorithmes dans notre vie quotidienne: Synthese de la journee, December 2017. URL https://www.cnil.fr/sites/default/files/atoms/files/cr_concertation_citoyenne_algorithmes.pdf. Data ultimo accesso: 24 dicembre 2017.

COE (Council of Europe). Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling, November 2010. URL https://rm.coe.int/16807096c3. Data ultimo accesso: 20 dicembre 2017.

Council of Economic Adviders. Big Data and Differential Pricing, February 2015. URL https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf. Data dell'ultimo accesso: 17 dicembre 2017.

DHEW (U.S. Department of Health, Education & Welfare). Records computers and the rights of citizens, report of the secretary's advisor committee on automated personal data systems, July 1973. URL https://www.justice.gov/opcl/docs/rec-com-rights.pdf. Data dell'ultimo accesso: 15 dicembre 2017.

J Dressel and Hany Farid. The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, (1), January 2018. doi:10.1126/sciadv.aao5580.

Cynthia Dwork. *Differential Privacy: A Survey of Results*. In *Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings*, pages 1–19. Springer, Berlin, Heidelberg, 2008. ISBN 978-3-540-79228-4. doi:10.1007/978-3-540-79228-4_1.

Cynthia Dwork and Deirdre K. Mulligan. It's not privacy and it's not fair. *Stanford Law Review Online*, 66:35–40, September 2013. URL https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/DworkMullliganSLR.pdf.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. Fairness through awareness. *CoRR*, arXiv:1104.3913 [cs.CC], 2011. URL http://arxiv.org/abs/1104.3913.

EDPS (European Data Protection Supervisor). Opinion 7/2015. Meeting the challenges of big data. A call for transparency, user control, data protection by design and accountability, November 2015. URL https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf. Data ultimo accesso: 20 dicembre 2017.

EDPS. Opinion 8/2016. EDPS Opinion on coherent enforcement of fundamental rights in the age of big data, September 2016. URL https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf. Data ultimo accesso: 20 dicembre 2017.

EOP (Executive Office of the President). Big Data: seizing opportunities, preserving values, May 2014. URL https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. Data dell'ultimo accesso: 13 dicembre 2017.

EOP. Big Data: Seizing Opportunities and Preserving Values: Interim Progress Report, February 2015. URL https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf. Data dell'ultimo accesso: 13 dicembre 2017.

EOP. Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, May 2016. URL https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf. Data dell'ultimo accesso: 25 ottobre 2017.

European Parliament. Resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2017/2225(INI)), March 2017. URL http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//EN. Data ultimo accesso: 20 dicembre 2017.

FTC (Federal Trade Commission). Data Brokers. A call for transparency and accountability, May 2014. URL https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf. Data dell'ultimo accesso: 13 dicembre 2017.

FTC. Big data. A Tool for Inclusion or Exclusion? Understanding the issues, January 2016. URL https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf. Data dell'ultimo accesso: 13 dicembre 2017.

Archon Fung. Varieties of participation in complex governance. *Public Administration Review*, 66:66–75, 2006. ISSN 1540-6210. doi:10.1111/j.1540-6210.2006.00667.x. URL http://dx.doi.org/10.1111/j.1540-6210.2006.00667.x.

Matthew S. Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61(Supplement C):115 – 125, 2014. ISSN 0167-9236. doi:10.1016/j.dss.2014.02.003.

Lewis R. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34, 1993. URL http://projects.ori.org/lrg/PDFs_papers/Goldberg.Am.Psych.1993.pdf.

Melissa Gymrek, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich. Identifying personal genomes by surname inference. *Science*, 339(6117):321–324, January 2013. ISSN 0036-8075. doi:10.1126/science.1229566.

Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V. Pearson, Dietrich A. Stephan, Stanley F. Nelson, and David W. Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLOS Genetics*, 4(8):1–9, 08 2008. doi:10.1371/journal.pgen.1000167. URL https://doi.org/10.1371/journal.pgen.1000167.

Peter Jackson. *Introduction to Expert Systems, Third Editrion*. Addison-Wesley, International Computer Science Series, USA, 1999. ISBN 0201876868, 978-0201876864.

Kristian Lum and William Isaac. To predict and serve? *Significance*, 13(5): 14–19, 2016. ISSN 1740-9713. doi:10.1111/j.1740-9713.2016.00960.x.

Giovanni Mastrobuoni. Crime is terribly revealing: Information technology and police productivity, February 2017. URL https://ssrn.com/abstract=2989914. Working paper, disponibile su SSRN.

M. Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49 (1):8–30, Jan 1961. ISSN 0096-8390. doi:10.1109/JRPROC.1961.287775.

Patrizia Nanz and Miriam Fritsche. *La partecipazione dei cittadini: un manuale. Metodi partecipativi: protagonisti, opportunità e limiti*. Regione Emilia Romagna, Assemblea Legislativa, 2014. URL http://partecipazione.regione.emilia-romagna.it/tecnico-di-garanzia/documentazione/la-partecipazione-dei-cittadini-un-manuale/documenti/la-partecipazione-dei-cittadini-un-manuale/at_download/file/partecipazione_totale_web.pdf. Data ultimo accesso: 22 dicembre 2017.

Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP '08, pages 111–125, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-0-7695-3168-7. doi:10.1109/SP.2008.33.

M. E. Olver, K. C. Stockdale, and J. S. Wormith. Thirty years of research on the level of service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, 26(1):156–176, March 2014. doi:10.1037/a0035080.

Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, New York, NY, USA, 2016. ISBN 0553418815, 9780553418811.

Eli Pariser. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Group, USA, 2012. ISBN 0143121235, 978-0143121237.

Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, Cambridge, MA, USA, 2015. ISBN 0674368274, 9780674368279.

Judea Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. In *In Proceedings of the 7th Conference of the Cognitive science society, University of California, Irvine, CA*, pages 329–334, April 1985. URL http://ftp.cs.ucla.edu/tech-report/198_-reports/850017.pdf.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Series in Representation and Reasoning, San Francisco, California, USA, 1988. ISBN 1558604790, 978-1558604797.

Manfred Redelfs and Michael Stanke. Citizen participation in technology assessment: Practice at the congressional office of technology assessment. *Impact Assessment*, (1):55–70, 1988. doi:10.1080/07349165.1988.9725622.

A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, July 1959. ISSN 0018-8646. doi:10.1147/rd.33.0210.

E. Schlehahn, P. Aichroth, S. Mann, R. Schreiner, U. Lang, I. D. H. Shepherd, and B. L. W. Wong. Benefits and pitfalls of predictive policing. In *2015 European Intelligence and Security Informatics Conference*, pages 145–148, September 2015. doi:10.1109/EISIC.2015.29.

Aaron Shapiro. Reform predictive policing. *Nature*, 541:458–460, January 2017. URL https://www.nature.com/polopoly_fs/1.21338!/menu/main/topColumns/topLeftColumn/pdf/541458a.pdf.

Alan Mathison Turing. I.—computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950. doi:10.1093/mind/LIX.236.433.

International Telecommunication Union. *Measuring the Information Society Report, Volume 1*. ITU, Place des Nations, CH-1211 Geneva, Switzerland, 2017. ISBN 978-92-61-24511-5.

Hilary Wainwright. Making a people's budget in porto alegre. *NACLA Report on the Americas*, (5), 2016. doi:10.1080/10714839.2003.11724548.

Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5):193–220, 1890. ISSN 0017811X. URL http://www.jstor.org/stable/1321160.

Sharon Weinberger. Airport security: Intent to deceive? *Nature*, 465:412–415, May 2010. doi:10.1038/465412a.

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, Krishna P. Gummadi, and Adrian Weller. From parity to preference-based notions of fairness in classification. arXiv:1707.00010 [stat.ML], 2017. URL https://arxiv.org/abs/1707.00010. To appear in Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017).