

LQ *The Lab's Quarterly*

2018 / a. XX / n. 4 (ottobre-dicembre)



DIRETTORE

Andrea Borghini

COMITATO SCIENTIFICO

Albertini Françoise (Corte), Massimo Ampola (Pisa), Gabriele Balbi (Lugano), Matteo Bortolini (Padova), Massimo Cerulo (Perugia), Marco Chiuppesi (Pisa), Franco Crespi (Perugia), Sabina Curti (Perugia), Gabriele De Angelis (Lisboa), Paolo De Nardis (Roma), Teresa Grande (Cosenza), Elena Gremigni (Pisa), Roberta Iannone (Roma), Anna Giulia Ingellis (València), Mariano Longo (Lecce), Domenico Maddaloni (Salerno), Stefan Müller-Doohm (Oldenburg), Gabriella Paolucci (Firenze), Massimo Pendenza (Salerno), Walter Privitera (Milano), Cirus Rinaldi (Palermo), Antonio Viedma Rojas (Madrid), Vincenzo Romania (Padova), Angelo Romeo (Perugia), Giovanni Travaglini (Kent).

COMITATO DI REDAZIONE

Luca Corchia (segretario), Roberta Bracciale, Massimo Cerulo, Cesar Crisosto, Elena Gremigni, Antonio Martella, Gerardo Pastore

CONTATTI

thelabs@sp.unipi.it

I saggi della rivista sono sottoposti a un processo di double blind peer-review.

La rivista adotta i criteri del processo di referaggio approvati dal Coordinamento delle Riviste di Sociologia (CRIS): cris.unipg.it

I componenti del Comitato scientifico sono revisori permanenti della rivista.

Le informazioni per i collaboratori sono disponibili sul sito della rivista:

<https://thelabs.sp.unipi.it>

ISSN 1724-451X



Quest'opera è distribuita con Licenza

Creative Commons Attribuzione 4.0 Internazionale

“The Lab’s Quarterly” è una rivista di Scienze Sociali fondata nel 1999 e riconosciuta come rivista scientifica dall’ANVUR per l’Area 14 delle Scienze politiche e Sociali. L’obiettivo della rivista è quello di contribuire al dibattito sociologico nazionale ed internazionale, analizzando i mutamenti della società contemporanea, a partire da un’idea di sociologia aperta, pubblica e democratica. In tal senso, la rivista intende favorire il dialogo con i molteplici campi disciplinari riconducibili alle scienze sociali, promuovendo proposte e special issues, provenienti anche da giovani studiosi, che riguardino riflessioni epistemologiche sullo statuto conoscitivo delle scienze sociali, sulle metodologie di ricerca sociale più avanzate e incoraggiando la pubblicazione di ricerche teoriche sulle trasformazioni sociali contemporanee.

2018 / a. XX / n. 4 (ottobre-dicembre)

Gli algoritmi come costruzione sociale

A cura di

Antonio Martella, Enrico Campo e Luca Ciccarese

Enrico Campo, Antonio Martella, Luca Ciccarese	<i>Gli algoritmi come costruzione sociale. Neutralità, potere e opacità</i>	7
SAGGI		
Massimo Airoidi, Daniele Gambetta	<i>Sul mito della neutralità algoritmica</i>	25
Chiara Visentin	<i>Il potere razionale degli algoritmi tra burocrazia e nuovi idealtipi</i>	47
Mattia Galeotti	<i>Discriminazione e algoritmi. Incontri e scontri tra diverse idee di fairness</i>	73
Biagio Aragona, Cristiano Felaco	<i>La costruzione socio-tecnica degli algoritmi. Una ricerca nelle infrastrutture di dati</i>	97
Aniello Lampo, Michele Mancarella, Angelo Piga	<i>La (non) neutralità della scienza e degli algoritmi. Il caso del machine learning tra fisica fondamentale e società</i>	117
Luca Serafini	<i>Oltre le bolle dei filtri e le tribù online. Come creare comunità "estetiche" informate attraverso gli algoritmi</i>	147
Costantino Carugno, Tommaso Radicioni	<i>Echo chambers e polarizzazione. Uno sguardo critico sulla diffusione dell'informazione nei social network</i>	173

LIBRI IN DISCUSSIONE

Irene Psaroudakis	Mario Tirino, Antonio Tramontana, <i>I riflessi di «Black Mirror». Glossario su immaginari, culture e media della società digitale</i> , Roma, Rogas Edizioni, 2018, 280 pp.	203
Junio Aglioti Colombini	Daniele Gambetta, <i>Datacrazia. Politica, cultura algoritmica e conflitti al tempo dei big data</i> , Roma, D Editore, 2018, 360 pp.	209
Paola Imperatore	Safiya Umoja Noble, <i>Algorithms of Oppression: How Search Engines Reinforce Racism</i> , New York, New York University Press, 2018, 265 pp.	215
Davide Beraldo	Cathy O'Neil, <i>Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy</i> , New York, Broadway Books, 2016, 272 pp.	223
Letizia Chiappini	John Cheney-Lippold, <i>We Are Data: Algorithms and The Making of Our Digital Selves</i> , New York, New York University Press, 2017, 320 pp.	229



GLI ALGORITMI COME COSTRUZIONE SOCIALE

Neutralità, potere e opacità

di Enrico Campo, Antonio Martella, Luca Ciccarese¹

- È stato un duro compito - disse pacato Pensiero Profondo.
- Quarantadue ? - urlò Loonquawl
- È tutto quello che hai da dirci dopo sette milioni e mezzo di anni di lavoro?
- Ho controllato con grande minuziosità - disse il computer
- e questa è la risposta veramente definitiva. Credo, se devo essere franco, che il problema stia nel fatto che voi non avete mai realmente saputo quale fosse la domanda.

1. INTRODUZIONE

La rivoluzione delle *Information and Communication Technologies* (ICTs) ha rappresentato uno dei principali fattori che, grazie alla digitalizzazione dei contenuti, ha portato a interpretare la società contemporanea nei termini di una società caratterizzata dall'abbondanza di flussi informativi, capitali, immagini, simboli, etc. e in cui la "conoscenza" è diventata un'essenziale fonte di produttività e potere (Castells, 2002). I processi di distribuzione, elaborazione e selezione delle grandi quantità di dati rendono necessaria l'adozione di strumenti tecnologici in grado di effettuare questi compiti in tempi sempre più brevi e senza l'intervento umano: gli algoritmi.

Grazie alla pervasività delle tecnologie digitali, gli algoritmi sono diventati quindi una presenza costante nelle nostre vite in grado di produrre ed elaborare un'enorme quantità di dati a partire dalle nostre

¹ ENRICO CAMPO è ricercatore postdoc presso l'Università della Corsica e affiliato al laboratorio CNRS, LISA - UMR 6240. Si occupa del rapporto tra processi sociali e processi cognitivi e nella tesi di dottorato, intitolata *Essere qui, essere altrove. L'attenzione e la sua crisi*, ha indagato in particolare l'attenzione. Email: enricocampo@gmail.com

ANTONIO MARTELLA è dottorando in Scienze Politiche presso l'Università di Pisa e membro del MediaLaB Unipi. Si occupa di comunicazione politica, in particolare di leader, populismo e social media. Discuterà la sua tesi di dottorato, intitolata *Il populismo come stile comunicativo: i leader politici italiani nei social media*, a pochi giorni dalla pubblicazione di questo numero. Email: antonio.martella@sp.unipi.it

LUCA CICCARESE è PhD in Sociologia presso il Dipartimento di Scienze Politiche dell'Università di Pisa. Si occupa di metodologia della ricerca sociale, con particolare attenzione alla Grounded Theory. Nella sua tesi di dottorato, dal titolo *L'effetto NEET come miraggio identitario. Politiche del sé e cornici istituzionali. Una Grounded Theory*, ha indagato i processi psicosociali che sottendono a condizioni di marginalità formativa e professionale. Email: ciccarese.luca@gmail.com

attività quotidiane. Queste ultime, a loro volta, vengono scandite e influenzate dalle operazioni che gli algoritmi effettuano sui dati. Si pensi ad esempio, molto semplicemente, ai risultati di una ricerca su Google, o alle offerte di un supermercato (spesso elaborate attraverso algoritmi di *market basket analysis*), oppure alla concessione di un prestito, determinata dalle sempre più pervasive tecniche di *credit scoring* (Hvistendahl, 2017), fino ad arrivare all'uso degli algoritmi di *machine learning* per le decisioni che concernono la recidiva, anche sulla base dati demografici e relativi alla storia giuridica dell'imputato (Livni, 2017). La loro ubiquità ha pertanto effetti potenzialmente rivoluzionari tanto nell'ambito della vita quotidiana, quanto nella ricerca sociale.

Non c'è da stupirsi, dunque, se gli algoritmi siano diventati un argomento centrale tanto nel dibattito pubblico (Gilbert, 2018; Gillespie, 2014) che nella comunità accademica. Ne è una prova la recente pubblicazione di numeri di riviste scientifiche interamente dedicati alla tematica (Amoore, 2019; Beer, 2017; Boccia Artieri, Marinelli, 2018; Ziewitz, 2016) che, da punti di vista e con obiettivi differenti, hanno scandagliato i molteplici aspetti di quello che tecnicamente può essere definito come «un procedimento di calcolo esplicito e descrivibile con un numero finito di regole che conduce al risultato dopo un numero finito di operazioni, cioè di applicazioni delle regole» (Treccani, 2019). La semplicità di questa formula è però solo apparente (Barocas, Hood, Ziewitz, 2013; Seaver, 2017) e contrasta in maniera netta rispetto all'estrema complessità dell'impatto sociale degli algoritmi.

Indagare le modalità attraverso cui questi operano non è per nulla agevole. Un primo approccio per iniziare a discutere di algoritmi potrebbe consistere nel chiedersi quali siano le operazioni che compiono: selezionano le informazioni rilevanti, scartano quello che è considerato irrilevante, strutturano le priorità, aiutano nei processi di ricerca e in quelli decisionali attraverso complessi sistemi di selezione e raccomandazione. Si tratta di un'azione sempre più necessaria anche a causa della continua espansione delle reti, e del web, all'interno della *network society* (van Dijk, 2012), che implica l'esplosione del volume dei contenuti presenti in esse e impone la necessità di soggetti intermediari – come i motori di ricerca, i sistemi di raccomandazione, etc. – strumenti ormai imprescindibili per orientarsi nella sovrabbondanza informativa (Boccia Artieri, 2014). Spesso, queste tecnologie sembrano essere addirittura in anticipo rispetto agli utenti: ne conoscono i bisogni e, frequentemente, prendono decisioni per conto degli esseri umani. Si pensi ai sistemi di *anticipatory shipping* di Amazon, che

prevedono i probabili acquisti dei clienti² o i servizi antispam di Google che adottano tecniche di deep learning per riconoscere i messaggi da scartare³. A questi esempi è possibile aggiungere anche l'impatto sulla vita quotidiana degli algoritmi che regolano il funzionamento delle piattaforme di social network, sia sul piano dei contenuti che su quello delle relazioni sociali. Difatti, le *timeline* degli utenti sono spesso costruite sulla base di criteri non del tutto comprensibili. Da questo punto di vista risulta esemplare il caso di Facebook: la logica di funzionamento della sua *timeline*, paragonabile ai sistemi di raccomandazione di Netflix o Spotify (Hargreaves *et al.*, 2019), influenza, attraverso la selezione dei contenuti da mostrare, i contatti all'interno della rete sociale (Rader, 2017) e di conseguenza l'immagine di sé (Willson, 2017). Eppure, è necessario porre in primo piano due elementi che stridono rispetto all'enorme potere sociale che gli algoritmi sembrano avere. Da un lato essi sono intrinsecamente opachi: per diversi motivi – che affronteremo brevemente di seguito – è molto difficile sapere come e perché un determinato algoritmo (o meglio una famiglia di algoritmi) abbia prodotto proprio quello specifico risultato (Pedreschi *et al.*, 2018). In alcuni casi è praticamente impossibile, anche per i soggetti che lo hanno effettivamente scritto e progettato. Inoltre, venendo al secondo aspetto, gli algoritmi sfuggono alla percezione diretta⁴: è possibile osservarne il risultato finale, il prodotto del loro operare, che molto spesso non viene presentato come l'esito di un processo di selezione tra diverse possibilità, ma come un semplice dato di fatto. Quando nel 2014 venne mostrato che un'alterazione dell'algoritmo che determina la costruzione della *timeline* di Facebook può avere un'influenza sulle emozioni degli utenti (Kramer, Guillory, Hancock, 2014), furono posti a tema sia il rischio di manipolazione collettiva che l'esistenza dell'algoritmo in sé (Willson, 2017).

² <https://www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/>

³ <https://forbes.it/2019/02/07/google-spam-tensorflow-ai/>

⁴ Le loro operazioni si realizzano a un livello che sembra sfuggire alla coscienza desta e per interpretare il quale alcuni autori hanno ripreso il concetto di “inconscio tecnologico”, di Nigel Thrift. Il geografo britannico aveva usato questa espressione, in *Knowing Capitalism*, in riferimento ai sistemi di *track and change* che garantiscono la possibilità del corretto posizionamento degli oggetti attraverso un apparato socio-tecnico che sfrutta sia conoscenze formalizzate di logistica e di indirizzamento che sofisticate tecnologie di tracciamento (dai codici a barre agli strumenti di identificazione a radiofrequenza fino ai GPS). Tutto questo sistema si muove, sostiene Thrift, in maniera automatica, quasi come un sistema vivente autonomo. È probabilmente questo aspetto che attira maggiormente l'attenzione dei ricercatori che si interessano di algoritmi. Si veda anche Citton sul tema (2016).

È importante tenere ben presente questa distinzione tra il potere sociale che gli algoritmi hanno in quanto “codice” – ovvero come programmi che regolano il funzionamento di un’ampia varietà di meccanismi e pratiche sociali attraverso un’azione sui dati – e quello che hanno invece all’interno di un discorso sociale che li sostiene e legittima (o che viceversa li scredita). In questo senso è possibile dire che gli algoritmi come oggetto di ricerca si biforcano immediatamente (Beer, 2017): da un lato come concetto inquadrato in uno specifico ambito discorsivo e dall’altro come codice implementato ed eseguito dalle macchine. Anche se analiticamente separati, i due piani sono strettamente interrelati: il loro potere in quanto nozione sociale diffusa dipende anche da quello che fanno (e da come lo fanno). Pertanto, l’indagine dei due aspetti deve essere vista comunque come un’impresa di ricerca unitaria: è necessario comprendere sia il codice che il ruolo sociale che svolge la nozione di “algoritmo”, la sua funzione politica. Studiare il potere sociale degli algoritmi o il potere sociale che opera attraverso di essi significa «to understand the power of algorithms as code whilst also attempting to understand how notions of the algorithm move out into the world, how they are framed by the discourse and what they are said to be able to achieve» (Ivi, 10). Chiaramente, per portare a termine un tale compito bisogna affrontare molte difficoltà, in primo luogo di ordine epistemologico e metodologico, poiché i livelli ontologici implicati nello studio degli algoritmi sono molto diversi: i linguaggi, i codici, la formalizzazione, la regolamentazione, le infrastrutture, i dati, le piattaforme, gli attori sociali coinvolti a vario titolo (gli utenti, i programmatori, i committenti, le grandi aziende del digitale, ecc.). Per aprire la black box degli algoritmi (Seaver, 2017) sarà dunque necessario chiamare a raccolta saperi che hanno matrici disciplinari tra loro molto diverse. Pur consapevoli della difficoltà che una tale scelta comporta, anche semplicemente rispetto alla possibilità di tradurre linguaggi specialistici tanto diversi, questo numero speciale ha voluto raccogliere la sfida, come la diversa formazione degli autori dimostra. Il numero infatti raccoglie e mette a confronto le competenze di informatici, matematici, fisici, sociologi, filosofi e ricercatori con una formazione interdisciplinare che hanno collaborato al fine di costruire un terreno di dialogo comune⁵. Proprio al fine di mettere in risalto

⁵ Ad esempio, nella costruzione del numero monografico sono emerse le differenze tra convenzioni formali e informali dei rispettivi ambiti di ricerca, a partire dagli editor di testo utilizzati fino ad arrivare alle pratiche di citazione consolidate o al diverso utilizzo dei riferimenti bibliografici. In questo senso, si ritiene importante sottolineare lo sforzo degli autori e dei revisori non solo per il contributo alla riuscita del numero ma anche per la

questa matrice comune, nel prosieguo dell'introduzione si metteranno in evidenza gli elementi tematici che legano i diversi contributi.

2. NEUTRALITÀ E POTERE DEGLI ALGORITMI

Tutti i lavori partono da una premessa fondamentale: per studiare gli algoritmi e il loro potere è necessario considerarli come il risultato dell'incontro di complessi processi sociali e il cui esito non è mai scontato. Gli algoritmi pertanto producono i loro effetti nel tessuto sociale e nelle vite degli individui poiché sono il prodotto di un contesto sociale animato da interessi molteplici – spesso in contrasto tra loro – e condizionato anche da limiti tecnici. Si tratta di un aspetto messo in luce dal lavoro di Biagio Aragona e Cristiano Felaco, i quali analizzano le fasi di progettazione degli algoritmi intervistando i soggetti coinvolti in diversi centri di ricerca ed evidenziando come lo stesso processo di ricerca sia frutto di contrattazione tra interessi e necessità differenti. Gli algoritmi non sono quindi una soluzione univoca e meramente tecnica a un problema ma implicano sempre delle scelte a diverse fasi della loro progettazione e implementazione: hanno anche una dimensione discrezionale e dunque *politica*. Non a caso, uno degli argomenti centrali di questo numero riguarda la supposta neutralità degli algoritmi, ed è in contrasto con la retorica sociale che li vorrebbe come strumenti oggettivi ed efficienti che non sono condizionati dalle distorsioni e dai limiti tipici dell'azione umana (Sandvig, 2015). Il saggio di Massimo Airoldi e Daniele Gambetta ha proprio l'obiettivo di decostruire il mito della “neutralità algoritmica”, considerata come la nozione centrale del discorso egemonico sugli algoritmi (Natale, Ballatore, 2017). Al fine di mettere in discussione la presunta oggettività delle decisioni prese dagli algoritmi, gli autori si concentrano sui casi in cui questi hanno prodotto degli errori, non hanno adempiuto allo scopo per cui erano stati progettati, oppure la loro implementazione ha portato a esiti non previsti⁶. È questo il caso, ad esempio, del sistema di raccomandazione Amazon che ha riscosso un certo interesse pubblico grazie a un'inchiesta giornalistica che mostrò un fatto alquanto curioso: cercando uno specifico agente chimico all'interno della piattaforma, il sistema di raccomandazione consigliava l'acquisto di altri oggetti “*Frequently*

disponibilità nel cercare un terreno comune di dialogo. A loro va il nostro più sincero ringraziamento.

⁶ Da questo punto di vista, il loro contributo può essere considerato in dialogo con un recente e fecondo interesse negli studi sui media per la categoria del fallimento. A tal proposito si veda Magudda e Balbi (2018).

bought together” che di fatto erano i componenti necessari a costruire una bomba (Tsang, 2017).

La dimensione politica degli algoritmi emerge anche a livello della loro formalizzazione matematica, un tema raramente indagato dagli scienziati sociali (Burrell, 2016; Mazzotti, 2015), ma presente da tempo nell'agenda delle scienze “dure” (Friedler, Scheidegger, Venkatasubramanian, 2016; Pedreschi *et al.*, 2018). Le procedure di formalizzazione in questo caso vengono analizzate in quanto parte del processo sociale complessivo e quindi anche rispetto alle loro ricadute politiche. L'articolo di Mattia Galeotti segue appunto questa impostazione per studiare una questione di grande attualità, la *fairness* degli algoritmi, ovvero la loro capacità di ridurre e rimuovere le discriminazioni rispetto a determinati attributi sensibili (pensiamo ad esempio ai casi in cui l'algoritmo decide di non concedere un prestito sulla base dell'etnia). Il saggio mostra chiaramente come determinate idee e visioni del mondo strutturino necessariamente le procedure di formalizzazione: il modo di intendere la discriminazione incide cioè nella codificazione matematica della *fairness*. Chiaramente, le modellizzazioni possono anche essere incompatibili tra loro e si basano su degli assunti che provengono da ipotesi di fondo molto diverse: se infatti pensiamo all'assenza di discriminazione come una situazione di uguaglianza delle probabilità o delle opportunità avremo una formalizzazione di un algoritmo “*fair*” molto diversa da quella che avremmo se invece pensassimo alla *fairness* come una condizione di parità statistica tra gruppi. Inoltre, Galeotti sottolinea come l'idea di ottenere un algoritmo privo di “*bias*” (distorsioni) sia discutibile sul piano epistemologico perché è fondata sull'assunto implicito che l'algoritmo debba semplicemente avvicinarsi il più possibile a un contesto di selezione “giusto”, che esiste “là fuori”, indipendentemente dall'algoritmo stesso. La possibilità di risolvere discriminazioni e disparità di trattamento presenti nella società attraverso una programmazione il più possibile *unbiased* si rivela quindi insoddisfacente: in primo luogo perché gli assiomi iniziali di un modello condizionano necessariamente i passaggi successivi e, per di più, la realtà su cui gli algoritmi operano, e i dati che questi utilizzano, non esistono indipendentemente dall'algoritmo, che contribuisce in parte a creare e a mettere in forma quella stessa realtà. Da questo punto di vista, le conseguenze sociali di queste criticità non risolte (e probabilmente irrisolvibili se affrontate solo da una prospettiva tecnica) sono quelle note come “effetto San Matteo” (Merton, 1968; Mingo, Bracciale, 2016), che consistono cioè nella riproduzione e amplificazione delle disparità sociali, in questo caso nascoste dal velo di imperscrutabilità imposto dal-

l'automatizzazione dalle scelte di selezione, classificazione, identificazione delle relazioni operate da algoritmi sempre più complessi.

Gli aspetti della matematizzazione e quello della presunta neutralità pongono quindi come ineludibile la necessità di comprendere le modalità di esercizio del potere, che gli algoritmi abilitano o favoriscono. Una delle questioni essenziali riguarda la possibilità che i processi informatici prendano decisioni a proposito di ambiti socialmente sensibili in forma del tutto automatizzata, quindi senza l'intervento umano e senza che sia possibile ricostruire i passaggi intermedi che hanno portato a una determinata scelta. La "governance algoritmica" è anche una "*governance by numbers*" (Katz, 2017), sostanziata da un potere che sembra esercitarsi in maniera impersonale, astratta e oggettiva. Un potere guidato solo da criteri di efficienza tecnica. Si pensi ad esempio ai tentativi di attualizzare le teorie di matrice lombrosiana, ma inserite in un contesto in cui il riconoscimento "fisiognomico" del criminale è delegato alla macchina (Wu, Zhang, 2016), riportate da Airoldi e Gambetta come caso esemplare di ingenuità in una costruzione del dato che – seppur elaborato da "algoritmi neutrali" – restituiva risultati discriminatori.

Impersonalità, oggettività, razionalità, discrezionalità politica resa tecnica: gli algoritmi richiamano quasi immediatamente i meccanismi di funzionamento e i criteri di legittimazione tipici delle burocrazie occidentali. È dunque possibile chiedersi se il potere che passa attraverso gli algoritmi possa essere considerato come un'estensione e intensificazione delle tradizionali forme amministrativo-burocratiche. L'articolo di Chiara Visentin affronta proprio questo genere di interrogativi e ricostruisce un dibattito che negli ultimi anni ha mostrato una certa vitalità. Le analogie tra l'idealtipo della burocrazia e degli algoritmi sono in effetti molteplici. Ad esempio, entrambi si basano su regole prestabilite e sono espressione massima della razionalità formale e della calcolabilità. Non per questo però è scontato poter considerare i secondi come estensioni e intensificazioni della burocrazia. Permangono infatti differenze sostanziali (Aneesh, 2009), anche rispetto ai principi centrali che presidono al funzionamento dei due sistemi:

L'algcrazia – scrive Visentin – non va confusa con una forma di "iper-burocrazia" [...]. I sistemi algocratici, infatti, a differenza di quelli burocratici, strutturano il campo delle azioni possibili senza bisogno che gli agenti interiorizzino il rispetto per regole e leggi, né vi siano indotti dalla cognizione di punizioni: la loro azione è controllata dando forma all'ambiente in cui si svolge, e facendo in modo che siano presenti solo alternative programmate.

Rimane cioè una differenza sostanziale nelle modalità di esercizio del potere che si mostra chiaramente quando si tratta di opporvisi: chi è soggetto al potere burocratico può, almeno in linea di principio, comprenderne il funzionamento ed eventualmente contestarlo. Gli algoritmi invece sono fondamentalmente imperscrutabili per gli attori sociali e in alcuni casi sono coperti da segretezza.

In effetti, il potere algoritmico non si configura come un riferimento statico e monolitico per l'attore, poiché predispone e modella un ambiente e un sistema di riferimenti entro il quale il soggetto si orienta – o reagisce – individualmente. In questa sua attitudine mimetica e al contempo vincolante, l'algoritmo – come rilevato da Dominique Cardon (2016) – riproduce in parte ciò che Thaler e Sunstein (2008) identificano come “paternalismo libertario”. In particolare, il riferimento è al ruolo dei *nudge*, ovvero delle “spinte gentili”, piccoli incentivi e “pungoli” che – anche se privi di coercitività – possono influenzare il comportamento degli attori sociali. L'ambiente algoritmico, di fatto, si modella costantemente attorno al soggetto e alle tracce della sua biografia digitale; è mimetico e presenta certamente diverse analogie col “*nudging*” che, come pratica di regolazione comportamentale, non riguarda la creazione e l'interiorizzazione di valori e norme, ma si fonda sulla costruzione di particolari architetture di scelta e di comportamento che circoscrivono le possibilità di azione degli attori sociali. Il *nudging*, tuttavia, come pratica di *policy* consolidatasi negli ultimi decenni, è ancora uno strumento di regolazione sociale di natura statica e legato a una scansione temporale di ampio respiro che è (ancora) gestibile dall'uomo. Le architetture di scelta e comportamento, poi, possono cambiare, ma sulla base degli effetti che vengono rilevati nel lungo periodo. Diversamente, il meccanismo che gli algoritmi consolidano può essere definito *hypernudging*. Yeung infatti argomenta:

Big Data driven nudging is therefore nimble, unobtrusive and highly potent, providing the data subject with a highly personalised choice environment – hence I refer to these techniques as “hypernudge”. Hypernudging relies on highlighting algorithmically determined correlations between data items within data sets that would not otherwise be observable through human cognition alone (or even with standard computing support [Shaw, 2014]) thereby conferring ‘salience’ on the highlighted data patterns, operating through the technique of ‘priming’, dynamically configuring the user’s informational choice context in ways intentionally designed to influence her decisions (2017, 122).

L'*hypernudging*, dunque, presenta caratteri di dinamismo estremo, rapidità di calcolo e personalizzazione continua; ciò significa che

l'algoritmo può operare in tempo reale – su scale temporali non-umane – e attraverso modelli non osservabili, sulle informazioni e sulle tracce digitali che i soggetti producono costantemente con le loro scelte operate in ambito digitale. In questo, la sua azione è ricorsiva, estremamente dinamica, necessariamente mutevole e flessibile, poiché esso agisce nell'immediatezza di una ricorsività tra soggetti e macchina.

In pratica allora, se il potere burocratico può essere sfidato e contestato proprio in relazione alla sua configurazione intellegibile, il potere algoritmico presenta caratteristiche *altre* che rendono più complesso il “conflitto”, rendendo ardua per i soggetti la sfida costante alla malleabilità degli output. Queste considerazioni introducono al secondo nucleo tematico dei saggi presenti in questo numero: il problema della “scatola nera”.

3. BLACK BOX ED ECHO CHAMBERS

La rilevanza pubblica degli algoritmi è quindi strettamente legata al loro essere delle vere e proprie “scatole nere” il cui funzionamento è difficile o impossibile da decifrare (Pasquale, 2015). Le ragioni di questa difficoltà di interpretazione possono essere le più diverse – ad esempio non si ha accesso ai dati, controllati dalle grandi aziende del digitale, su cui la macchina si esercita – ma seguendo Burrell (2016) possono essere ricondotte prevalentemente a tre categorie tra loro strettamente legate: a) gli algoritmi e i dati sono coperti da segretezza ; b) spesso è difficile comprendere il linguaggio con cui l'algoritmo è stato scritto; c) le specifiche caratteristiche del *machine learning* e soprattutto del *deep learning* effettuano scelte che non necessariamente sono comprensibili agli esseri umani. Complessivamente, le tre forme di opacità, che spesso operano congiuntamente, rendono estremamente complessa l'interpretabilità degli algoritmi e risulta quindi pressoché impossibile capirne fino in fondo gli effetti e gli eventuali errori o distorsioni (Pasquale, 2015; Pedreschi *et al.*, 2018). Su questo punto sembrano convergere tutti i saggi inclusi in questa raccolta. In particolare, risulta molto interessante che sia il saggio di Aragona e Felaco sia quello di Aniello Lampo, Michele Mancarella e Angelo Piga, seppur da ambiti culturali e metodologici completamente differenti, si interrogino sulla validità epistemologica dell'adozione degli algoritmi nei processi di produzione della conoscenza.

Su un piano più generale, è difficile sottostimare tale questione se si pensa al fatto che gli algoritmi sono determinanti tanto nella selezione delle informazioni rilevanti che nella costruzione delle reti sociali. Ad

esempio, sono ben note le distorsioni incorporate dall'algoritmo di Google nella restituzione dei risultati di ricerca, dovuti anche a finalità commerciali – riportati nel saggio *Algorithms of Oppression: How Search Engines Reinforce Racism* (2018) di Safiya Umoja Noble (recensito in questo numero) – ai quali si aggiungono gli effetti della costruzione algoritmica delle *timeline* dei social network site. Da un lato, infatti, diversi studi evidenziano il ruolo crescente dei social media come fonte di informazioni sia per gli “utenti comuni” (Newman, Fletcher, Kalogeropoulos, Levy, Nielsen, 2017) che per gli stessi giornalisti e media tradizionali (von Nordheim, Boczek, Koppers, 2018). Dall'altro, la costruzione algoritmica delle timeline incide “invisibilmente” sulla fruizione dei contenuti e sulla gestione della propria rete sociale di contatti (Rader, 2017), nonostante nell'esperienza d'uso i social network sembrino fondarsi sul mito dello “*us now*”, per riprendere l'espressione di Nick Couldry. Un mito particolarmente seduttivo secondo cui la comunità di Facebook (e di altri social network) formano una sorta di collettività naturale: «on this story, media institutions, at least in their normal form, drop out altogether from the picture of “what is happening”. This myth offers a story focussed entirely on what “we” do when, as humans like to, we keep in touch with each other» (2015b, 620). Questa narrazione fa da supporto all'illusione che le relazioni sui social network siano immediate (2015a) quando invece sono mediate in primo luogo dagli algoritmi della piattaforma, la cui logica può portare gli utenti a essere esposti maggiormente a contenuti coerenti con la propria visione del mondo. Tali meccanismi sono infatti centrali nella creazione delle “camere dell'eco” e delle “bolle dei filtri” (Pariser, 2012), affrontata in due saggi di questo numero speciale. Nell'ipotesi di Luca Serafini, la formazione di comunità estetico-ludiche, tendenzialmente chiuse e con relazioni fortemente omofile, implicherebbe una sorta di ritorno delle grandi narrazioni moderne, dei *grand récit* come sistemi narrativi onnicomprensivi e coerenti, che il pensiero postmoderno ha fortemente criticato. All'interno di queste comunità estetiche, le informazioni dissonanti rispetto al proprio sistema di credenze vengono ignorate o tenute all'esterno e quelle coerenti, anche se inverosimili, si diffondono più facilmente e anzi fondano la loro autorità proprio sulla diffusione virale. Al fine di limitare l'eccessiva chiusura delle comunità estetiche e la conseguente polarizzazione del dibattito pubblico, Serafini propone di esplorare tutte quelle soluzioni, anche tecniche, che tendono a favorire la “serendipità” online e promuovere forme di partecipazione che non si limitino a premiare solo la visibilità dei contenuti. L'obiettivo è cioè di

favorire un'architettura della rete che incoraggi la «predisposizione nei confronti delle esperienze non programmate, un atteggiamento aperto all'ibridazione, ad incontri non voluti, non cercati, ma che contribuiscono ad arricchire l'esperienza che abbiamo del mondo». D'altro canto, è comunque necessario avere una certa cautela prima di generalizzare questi processi a tutti i social network e a tutte le forme di comunità online: ogni piattaforma possiede le sue specifiche *affordances*, pratiche d'uso e logiche algoritmiche. Per cui è necessario studiare empiricamente la strutturazione delle relazioni sociali online per evitare di incorrere in generalizzazioni indebite. Il contributo di Costantino Carugno e Tommaso Radicioni si muove proprio in questa direzione per ricostruire le reti di relazione e discussione, su Facebook e Twitter, che si sono formate nel dibattito che in Italia ha ruotato attorno alla delicata questione dell'obbligo vaccinale. Il caso di studio scelto è particolarmente utile a indagare la dinamica relazionale online proprio perché la discussione pubblica si è fortemente polarizzata tra favorevoli e contrari al decreto legge "Lorenzin" (Scalari, 2018) in merito all'aumento del numero di vaccinazioni obbligatorie⁷. In questo contesto, i due autori individuano la presenza di *echo-chambers*, sia evidenziando la prevalenza di parole d'ordine specifiche tra comunità di utenti sia ricostruendo la morfologia della rete delle relazioni discorsive (tecnicamente possibile solo su Twitter) mostrando l'esistenza di cluster ben distinti e non comunicanti intorno a fonti di informazione schierate. Il tema della polarizzazione e della formazione delle *echo-chambers* è argomento di dibattito tra gli studiosi dei social media (Barberá, 2015b, 2015a; Conover *et al.*, 2010; Conover, Gonçalves, Flammini, Menczer, 2012) con risultati anche contrastanti. Ciò che si evidenzia è che questi processi emergono come mix tra le azioni umane e delle macchine, e nei social network portano a rafforzare le proprie opinioni all'interno di una bolla informativa creata sulla base di preferenze, legami e abitudini che gli algoritmi interpretano come *pattern* dell'utente⁸. Le attività degli algoritmi sui dati, al fine di rintracciarne *pattern* e relazioni, introducono al terzo nucleo evidenziato dai saggi presenti in questa raccolta: il rapporto tra azione algoritmica e riproduzione del contesto sociale, ovvero la riproposizione e l'amplificazione delle discriminazioni presenti nella società e nei "dati" che dovrebbero in qualche modo rappresentarla.

⁷ <http://www.salute.gov.it/portale/vaccinazioni/dettaglioContenutiVaccinazioni.jsp?lingua=italiano&id=4824&area=vaccinazioni&menu=vuoto>

⁸ https://medium.com/@_vecna/algoritmo-di-facebook-ha-deciso-che-i-fascistoidi-vedono-solo-foto-88a9352f38df

4. ALGORITMI E RIPRODUZIONE DEL CONTESTO SOCIALE

Un ulteriore aspetto problematico che emerge nel dibattito sugli algoritmi e sui loro effetti riguarda il rischio di riprodurre ed esasperare processi sociali e visioni del mondo egemoniche nel tessuto sociale complessivo. Pensiamo ancora una volta ai casi di discriminazione che, come riportato nella letteratura scientifica (Cheney-Lippold, 2011; Pedreschi *et al.*, 2018), nei saggi divulgativi (Cheney-Lippold, 2017; O'Neil, 2016; Umoja Noble, 2018) e nel dibattito pubblico più generale (Antwan, 2016; Cain Miller, 2015; Devlin, 2016; Lee, 2018), sono sempre più frequenti e negli ambiti più disparati (come la concessione di un prestito, la selezione per un lavoro, la profilazione su base etnica o di genere per obiettivi pubblicitari, fino alla scelta, presa dalla macchina, di escludere i quartieri più poveri dai servizi di Amazon Prime). Rispetto al dibattito presente in letteratura scientifica, il problema della riproduzione di alcuni processi sociali e della retroazione generata dagli algoritmi sembra articolarsi su almeno tre punti principali: 1) la costruzione del dato, 2) le difficoltà nella rimozione o riduzione degli aspetti discriminatori presenti nei dati, 3) l'interpretazione dei risultati.

In relazione alla costruzione del dato, il saggio di Galeotti, cui abbiamo già in precedenza accennato, mette in luce come nella costruzione di una base di dati, su cui gli algoritmi esercitano la propria azione di selezione e categorizzazione, vi sia l'intervento umano tanto nella scelta degli indicatori che nella loro costruzione per la rappresentazione della realtà da analizzare. Come chiariscono Friedler e colleghi:

In order to define a feature space, we must answer questions about what features should be included and how (and when) they should be measured. This description illuminates our first important distinction from a common set-up of such a problem: the feature space itself is a representation of a chosen set of possibly hidden or unmeasurable constructs (2016, 2).

In altre parole, il problema della costruzione stessa del dato non è esente dalla presenza di quelli che Pedreschi e colleghi definiscono «human biases and prejudices, as well as collection artifacts» (2018, 1) che si riflettono sia nella scelta degli indicatori che nei metodi di misurazione e che vengono riprodotti dalle procedure automatizzate di analisi algoritmica. Questo aspetto viene sottolineato anche dal saggio di Lampo, Mancarella, Piga, da un punto di vista epistemologico, soprattutto rispetto ai risultati delle analisi effettuate dagli algoritmi: «i dati sperimentali candidati alla verifica logica di una teoria sono sempre contaminati dalle assunzioni insite in un paradigma, e i dati stessi sono

sempre theory-laden (carichi di teoria)». Allo stesso modo, il contributo di Aragona e Felaco sottolinea come anche la produzione di un algoritmo, insieme a quella del dato, sia influenzata dalle «scelte soggettive di chi è direttamente coinvolto in questo processo che, a loro volta, saranno condizionate dalle competenze in materia, dall'appartenenza ad uno specifico settore scientifico, quindi, dalla propria formazione e da quell'insieme di valori, credenze ed esperienze».

Un ulteriore livello di complessità si raggiunge quando sono gli stessi algoritmi, come le reti neurali, a selezionare le *feature*, ovvero le caratteristiche e gli indicatori discriminanti all'interno di un database che, se da un lato semplificano e velocizzano i processi di ricerca e analisi, dall'altro non risultano immediatamente comprensibili. Come afferma Galeotti, che in questo senso sviluppa il lavoro di Friedler *et al.* (2016), i numerosi tentativi di riduzione o eliminazione dei *bias* presenti nei dati si scontrano con la correlazione statistica tra attributi. Ad esempio, la rimozione del dato sensibile "etnia" può essere insufficiente se il luogo di residenza corrisponde a quartieri o zone connotate etnicamente, come nel già citato caso dei servizi di Amazon Prime (Antwan, 2016). D'altro canto, le pratiche di "aggiustamento" degli output forniti dagli algoritmi, riassumibili sotto il nome di *fairness algoritmica*, non sembrano in grado di affrontare adeguatamente il problema della discriminazione, soprattutto nei casi in cui le popolazioni presentano una stessa caratteristica in proporzioni differenti. È questo il caso di Compas (*Correctional Offender Management Profiling for Alternative Sanctions*), un algoritmo considerato in grado di prevedere la recidività degli individui e riportato da diversi contributi presenti in questo numero speciale (Airoidi e Gambetta; Galeotti; Mancarella, Piga e Lampo), in cui è emersa l'impossibilità matematica di conciliare le probabilità individuali e di gruppo quando a monte vi è disparità in determinati attributi (Courtland, 2018). Anche in questo caso molti autori dei saggi di questa raccolta, pur partendo da punti di vista e campi di applicazione differenti, evidenziano il problema etico posto da questo tipo di tecnologie: quando l'algoritmo decide su ambiti in cui c'è in gioco la vita degli individui, ciò che può essere statisticamente accettabile in un determinato dominio, ad esempio quello della ricerca e dell'analisi statistica, non lo è necessariamente (o in assoluto) in altri campi.

L'ultima criticità riguarda l'interpretazione degli esiti delle procedure algoritmiche. Come sostengono Parisi (2019) e Amore (2019) il *machine learning* è l'opposto della programmazione, nel senso che l'aspetto più problematico non è tanto l'interpretazione degli output ma piuttosto la scelta di un algoritmo adeguato – ovvero in grado di

produrre un determinato output – che, come già evidenziato, è il più delle volte frutto di un processo di contrattazione continua tra soggetti con obiettivi e conoscenze differenti e, in questo senso, non neutrale. Inoltre, come sostengono Bolin e Andersson Schwarz, i risultati delle analisi, anche se ottenuti attraverso tecniche di Big Data, sono riportate spesso a parametri sociali “tradizionali”, introducendo quella che gli autori definiscono *heuristics of the algorithm*. In altri termini, l'interpretazione di output prodotti da processi a volte imperscrutabili stimola la riconduzione dei risultati a categorie preesistenti che non riflettono la logica algoritmica applicata ai big data:

In the context of Big Data, to solve the ‘problem’ with varied understandings of how the algorithms function, those who work with predictive, algorithmically based targeting have a need to ‘translate back’ the correlational data into normal-type, normal-curve terminology in order to make the information intelligible for the buyer, since most buyers are used to thinking in terms of age, gender, income, education, etc. (2015, 8).

La consapevolezza di questo limite ha spinto Beer (2017) a sostenere con forza la necessità di una maggiore comprensione dei processi endogeni ed esogeni alle analisi automatizzate. Su questo punto convergono anche Mancarella, Piga e Lampo, i quali sottolineano che il problema dell'interpretazione, che si pone in domini controllati come quello della fisica fondamentale, a maggior ragione, risulta imprescindibile per contesti in cui i sistemi risultano meno controllabili e le variabili in gioco non sono sperimentali.

La ricchezza di approcci e di prospettive disciplinari contenuta in questo numero serve a raccogliere la sfida che Burrell esplicita a conclusione del suo saggio: per affrontare adeguatamente l'opacità degli algoritmi è necessario impegnarsi in una «partnerships between legal scholars, social scientists, domain experts, along with computer scientists [that] may chip away at these challenging questions of fairness in classification in light of the barrier of opacity» (2016, 10). In questo senso, si ritiene importante sottolineare il tentativo e lo sforzo di trovare un terreno di dialogo comune a proposito del rapporto tra algoritmi, processi sociali e ricerca scientifica, tra ricercatori e ricercatrici con esperienze e conoscenze molto differenti in un periodo storico in cui anche l'accademia è fortemente spinta in direzione della specializzazione e della settorializzazione. L'auspicio dei curatori è che questo numero possa risultare una piattaforma di dialogo, ovvero una prima tappa di una discussione comune sul tema che abbia l'ambizione di creare una comunità di ricerca sempre più interdisciplinare.

RIFERIMENTI BIBLIOGRAFICI

- AMOOORE, L. (2019). Introduction: Thinking with Algorithms: Cognition and Computation in the Work of N. Katherine Hayles. *Theory, Culture & Society*, 36(2), 3-16.
- ANEESH, A. (2009). Global Labor: Algocratic Modes of Organization. *Sociological Theory*, 27(4), 347-370.
- ANTWAN, J. (2016). Amazon Prime and the Economics of Race. *Huffington Post*, 5 ottobre.
- BARBERÁ, P. (2015a). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 23(1), 76-91.
- BARBERÁ, P. (2015b). How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. *Job Market Paper, New York University*, 46.
- BAROCAS, S., HOOD, S., ZIEWITZ, M. (2013). Governing Algorithms: A Provocation Piece. *SSRN Electronic Journal*.
- BEER, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1-13.
- BOCCIA ARTIERI, G. (2014). La rete dopo l'overload informativo. La realtà dell'algoritmo da macchia cieca a bene comune. *ParadoXa*, 2.
- BOCCIA ARTIERI, G., MARINELLI, A. (2018). Introduzione: piattaforme, algoritmi, formati. Come sta evolvendo l'informazione online. *Problemi dell'informazione*, 3, 349-368.
- BOLIN, G., ANDERSSON SCHWARZ, J. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society*, 2(2), 1-12.
- BURRELL, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12.
- CAIN MILLER, C. (2015). When Algorithms Discriminate. *New York Times*. July 9.
- CARDON, D. (2016). *Cosa sognano gli algoritmi? Le nostre vite al tempo dei big data*. Milano: Mondadori.
- CASTELLS, M. (2002). *La nascita della società in rete*. Milano: Università Bocconi.
- CHENEY-LIPPOLD, J. (2011). A New Algorithmic Identity. *Theory, Culture & Society*, 28(6), 164-181.
- CHENEY-LIPPOLD, J. (2017). *We Are Data. Algorithms and The Making of Our Digital Selves*. New York (NY): New York University Press.
- CITTON, Y. (2016). Notre inconscient numérique. Comment les infrastructures du web transforment notre esprit. *La Revue du Crieur*,
-

- (4), 144-158.
- CONOVER, M. D., GONÇALVES, B., FLAMMINI, A., MENCZER, F. (2012). Partisan asymmetries in online political activity. *EPJ Data Science*, 1(1), 6.
- CONOVER, M. D., RATKIEWICZ, J., FRANCISCO, M., GONC, B., FLAMMINI, A., MENCZER, F. (2010). Political Polarization on Twitter. *Icwsm*, 133, 89-96.
- COULDRY, N. (2015a). Illusions of immediacy: rediscovering Hall's early work on media. *Media, Culture & Society*, 37(4), 637-644.
- COULDRY, N. (2015b). The myth of 'us': digital networks, political change and the production of collectivity. *Information, Communication & Society*, 18(6), 608-626.
- COURTLAND, R. (2018). Bias detectives: the researchers striving to make algorithms fair. *Nature*, 558(7710), 357-360.
- DEVLIN, H. (2016). Discrimination by algorithm: scientists devise test to detect AI bias. *The Guardian*. December 19.
- FRIEDLER, S. A., SCHEIDEGGER, C., VENKATASUBRAMANIAN, S. (2016). On the (im)possibility of fairness. *arXiv.org*, 1609.07236.
- GILBERT, A. S. (2018). Algorithmic culture and the colonization of life-worlds. *Thesis Eleven*, 146(1), 87-96.
- GILLESPIE, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, K. A. Foot (eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-194). Cambridge (MA): MIT Press.
- HARGREAVES, E., AGOSTI, C., MENASCHÉ, D., NEGLIA, G., REIFFERS-MASSON, A., ALTMAN, E. (2019). Fairness in online social network timelines: Measurements, models and mechanism design. *Performance Evaluation*, 129, 15-39.
- HVISTENDAHL, M. (2017). Inside China's vast new experiment in social ranking. *Wired*, December 19.
- KATZ, Y. (2017). Manufacturing an Artificial Intelligence Revolution. *SSRN Electronic Journal*, 1-21.
- KRAMER, A. D. I., GUILLORY, J. E., HANCOCK, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
- LEE, D. (2018). Facebook accused of job ad gender discrimination. *BBC*, September 19.
- LIVNI, E. (2017). In the US, some criminal court judges now use algorithms to guide decisions on bail. *Quartz*. February.
- MAGUDDA, P., BALBI, G. (2018). *Fallimenti digitali. Un'archeologia dei*
-

- «nuovi» media. Milano: Unicopli.
- MAZZOTTI, M. (2015). Per una sociologia degli algoritmi. *Rassegna Italiana di Sociologia*, 56(3-4), 465-477.
- MERTON, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, 159(3810), 56-63.
- MINGO, I., BRACCIALE, R. (2016). The Matthew Effect in the Italian Digital Context: The Progressive Marginalisation of the “Poor”. *Social Indicators Research*, 135(2), 629-659.
- NATALE, S., BALLATORE, A. (2017). Imagining the thinking machine. *Convergence: The International Journal of Research into New Media Technologies*, 135485651771516.
- NEWMAN, N., FLETCHER, R., KALOGEROPOULOS, A., Levy, D., Nielsen, R. K. (2017). *Reuters Institute Digital News Report 2017*.
- O’NEIL, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York (NY): Broadway Books.
- PARISER, E. (2012). *Il filtro. Quello che internet ci nasconde*. Milano: il Saggiatore.
- PARISI, L. (2019). Critical Computation: Digital Automata and General Artificial Thinking. *Theory, Culture & Society*, 36(2), 89-121.
- PASQUALE, F. (2015). *The Black Box Society*. Cambridge, MA: Harvard University Press.
- PEDRESCHI, D., GIANNOTTI, F., GUIDOTTI, R., MONREALE, A., PAPPALARDO, L., RUGGIERI, S., TURINI, F. (2018). Open the Black Box. Data-Driven Explanation of Black Box Decision Systems. *ArXiv Preprint*, 1(1), 1-15.
- RADER, E. (2017). Examining user surprise as a symptom of algorithmic filtering. *Journal of Human Computer Studies*, 98, 72-88.
- SANDVIG, C. (2015). Seeing the sort: The aesthetic and industrial defense of “the algorithm”. *Journal of the New Media Caucus*, 11(1), 35-51.
- SCALARI, A. (2018). Sui vaccini diamoci tutti una calmata. *Valigia Blu*, 10 agosto.
- SEEVER, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 205395171773810.
- THALER, R. H., SUNSTEIN, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth and Happiness*. New Haven (CT): Yale University Press.
- TRECCANI. (2019). Algoritmo. In *Treccani*. Retrieved from <http://www.treccani.it/vocabolario/algoritmo/>
-

- TSANG, A. (2017). Amazon 'Reviewing' Its Website After It Suggested Bomb-Making Items. *The New York Times*. September 27.
- UMOJA NOBLE, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York (NY): New York University Press.
- VAN DIJK, J. (2012). *The Network Society* (3rd ed.). SAGE Publications Ltd.
- VON NORDHEIM, G., BOCZEK, K., KOPPERS, L. (2018). Sourcing the Sources. *Digital Journalism*, 6(7), 807-828.
- WILLSON, M. (2017). Algorithms (and the) everyday. *Information, Communication & Society*, 20(1), 137-150.
- Wu, X., Zhang, X. (2016). Automated Inference on Criminality using Face Images. *ArXiv Preprint*, 1-11.
- YEUNG, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118-136.
- ZIEWITZ, M. (2016). Governing Algorithms. *Science, Technology, & Human Values*, 41(1), 3-16.
-

SUL MITO DELLA NEUTRALITÀ ALGORITMICA

di *Massimo Airoidi, Daniele Gambetta**

Abstract

From the recommendation of cultural content to the identification of potential criminals, a growing number of activities are ordinarily delegated to algorithms and AI systems. These are narrated as neutral technologies which make complex processes more efficient and lead to objective results. However, a wide literature argues that algorithms are social products that reflect the particular interests, cultural assumptions and biases of individuals and organizations. The present contribution aims to deconstruct in a Foucaultian way the algorithmic neutrality myth, illustrating its genesis, discursive facets and weaknesses, also drawing from a series of empirical cases. In the conclusion, we propose a counternarrative of the algorithm focused on explainability and collective sovereignty.

Keywords

Algorithms; Neutrality; Machine Learning; Foucault; STS

* MASSIMO AIROLDI è Assistant Professor, Emlyon Business School.

Email: airoidi@em-lyon.com

DANIELE GAMBETTA è ricercatore indipendente.

Email: daniele.gambetta7@gmail.com

1. INTRODUZIONE

Una discussione filosofica sull'uomo come capolavoro del creato a causa delle sue capacità di deduzione logica, oggi ha perso ogni significato, dal momento che le stesse capacità sono dimostrate da un buon cervello elettronico. Non è che perda di valore e di senso un discorso sull'uomo: solo che l'uomo non andrà più visto come animale sillogizzante, ma come animale capace di costruire macchine sillogizzanti e di porsi nuovi problemi (inediti) circa il loro uso (Eco, 2017, 368)

There's a lot that we don't know about these services to which we hand over so much of our lives. Despite their claims of objectivity and neutrality, they are constantly making value-laden, controversial decisions. They help create the world they claim to merely 'show' us (Pasquale, 2015, 10)

Con il diffondersi delle tecnologie di cattura e analisi di dati in tempo reale e la parallela ascesa di social media e piattaforme digitali, le nostre vite sono divenute la materia prima informazionale di processi algoritmici votati a un ampio ventaglio di scopi (cfr. Amoore, Piotukh, 2016). Anche per questo motivo, da semplice ricetta computazionale relegata a manuali di ingegneria, analisi finanziarie e conferenze di computer science, l'algoritmo è diventato qualcosa di più: una parola chiave della contemporaneità. Degli algoritmi di Google e Facebook hanno iniziato a parlare politici, operatori culturali, giornalisti, persino comuni cittadini – preoccupati perché i post delle rispettive vacanze non appaiono mai nel *newsfeed*. I casi di delega algoritmica oggetto di dibattito pubblico negli ultimi anni sono stati vari. Per esempio, si è parlato di Compas, un modello di calcolo di rischio di recidività del crimine soggetto a *bias* razziali (cfr. Pedreschi *et al.*, 2018); ha fatto ugualmente scalpore il caso di Tay, il *twitter-bot* di Microsoft ritirato perché diventato antisemita a causa dell'azione congiunta di vari utenti, così come il *software* di Google News che scriveva notizie con un forte pregiudizio sessista e le raccomandazioni automatiche di ingredienti "esplosivi" da parte di Amazon¹.

Con un po' di ritardo (come di consueto), anche le scienze umane e sociali hanno visto un proliferare di articoli accomunati da un aggettivo prima sconosciuto: *algorithmic identity* (Cheney-Lippold, 2011), *algorithmic culture* (Striphas, 2015), *algorithmic life* (Amoore, Piotukh, 2016), *algorithmic decision* (Zarsky, 2016), *algorithmic memory* (Esposito, 2017), *algorithmic states of exception* (McQuillan, 2015), *algorithmic brands* (Carah, 2017), solo per citarne alcuni. Ciò che questi e

¹ Si veda www.nytimes.com/2017/09/20/technology/uk-amazon-bomb.html.

molti altri contributi hanno provato a teorizzare – o, più raramente, a indagare empiricamente – è come l'esternalizzazione di scelte umane alla macchina stia trasformando società, individui e relazioni sociali.

Di recente, numerosi autori hanno inoltre contribuito alla problematizzazione del presunto carattere neutrale di quelli che Gillespie (2014) ha definito «algoritmi di rilevanza pubblica», utilizzati da piattaforme e servizi per filtrare contenuti e personalizzare le esperienze degli utenti. Dalla definizione e operativizzazione dei modelli statistici alla scelta delle fonti di dati, passando per l'implementazione e continua modifica delle funzionalità, i sistemi algoritmici sono essenzialmente prodotti storici e sociali che riflettono le scelte – nonché gli assunti culturali – di esseri umani e organizzazioni (Seaver, 2017; Striphas, 2015; Gillespie, 2014, 2016; Rieder, 2017). Ciononostante, il mito della neutralità algoritmica continua a circolare nell'immaginario contemporaneo, componente centrale del più ampio discorso pubblico intorno a tecnologie digitali e intelligenza artificiale (Natale e Ballatore, 2018).

Questo articolo, rifacendosi al filone di ricerca internazionale tracciato dagli autori sopra citati, si pone il seguente obiettivo: decostruire il discorso intorno alla neutralità della macchina e al suo supposto agire in modo scevro da quei pregiudizi ed errori che inficiano le decisioni umane. Questo mito è legato a doppio filo con quello del calcolo matematico e della sua presunta oggettività, e ha radici storiche ben più antiche della nascita di Google. Il presente contributo punta a: descriverne, in breve, la genesi (Paragrafo 3); illustrarne le fragilità (e pericolosità) attraverso una serie di esempi empirici (Paragrafo 4); rispolverare la critica epistemologica alla neutralità della scienza (Paragrafo 5); e, infine, sulla base di essa, proporre nelle conclusioni una «contronarrazione» (Pasquale, 2016) dell'algoritmo con al centro spiegabilità e sovranità condivisa (Pedreschi *et al.*, 2018; Rojo, 2018). Aniché soffermarci sulle proprietà tecniche e applicazioni di specifici algoritmi (cfr. Rieder, 2017), intendiamo seguire la linea teorica suggerita da Beer (2017) e Seaver (2017): l'algoritmo sarà trattato come un oggetto discorsivo, nuova parola d'ordine del mito moderno della neutralità di scienza e tecnologia. Anche e soprattutto per questo motivo, lo sguardo teorico adottato in queste pagine è ispirato al metodo genealogico di Michel Foucault e, in particolare, all'idea di una necessaria sovrapposizione tra la produzione discorsiva di verità sul mondo e i meccanismi del potere (cfr. Foucault, 2011).

2. L'ALGORITMO COME DISCORSO

Nell'introduzione alla special issue *The Social Power of Algorithms* recentemente apparsa nella rivista *Information, Communication & Society*, il sociologo David Beer (2017) sottolinea la necessità di prestare attenzione non soltanto a come gli algoritmi utilizzati da piattaforme, aziende e governi modellano la vita sociale in tempo reale, ma anche e soprattutto all'algoritmo in quanto discorso, nel senso foucaultiano del termine.

Per Foucault, la produzione sociale di discorsi e il loro carattere di verità sono legati a doppio filo con l'esercizio del potere: «non c'è potere che si eserciti senza una serie di intenti e di obiettivi» (2011, 84). Gerarchizzando, normando, indirizzando, censurando, i regimi discorsivi orchestrano le pratiche sociali, come nel caso della sessualità e della sua evoluzione storica. Qualcosa di simile, secondo Beer, accade anche nel caso degli algoritmi e, più in generale, della costellazione di narrazioni che gravitano intorno a piattaforme, intelligenza artificiale e big data (cfr. anche Beer, 2018; Natale e Ballatore, 2018; Seaver, 2017).

Da un lato, gli algoritmi che popolano le piattaforme digitali, classificando e ordinando contenuti in base al calcolo predittivo della loro rilevanza o similarità, riproducono un certo tipo di discorso sul sociale, oggettificato negli *output* dell'elaborazione statistica e, a monte, negli assunti culturali cristallizzati nel codice. Con le parole dell'autore, «there is a truth to which the algorithm might adhere or a truth that its actions might produce» (Beer, 2017, 9). È il caso, ad esempio, delle *meaningful interactions* privilegiate dalla più recente formulazione dell'algoritmo *EdgeRank* di Facebook²: l'operativizzazione di *meaningful* specificata e riprodotta dal modello statistico (Gillespie, 2016), così come i contenuti che vengono algoritmicamente etichettati come significativi e dunque inclusi nel *newsfeed* dell'utente, hanno entrambi a che fare con la costruzione e riproduzione discorsiva di una verità.

Dall'altro lato, gli algoritmi sono a loro volta oggetti discorsivi, parole chiave (Seaver, 2017) che risuonano con quello che – sempre Beer (2018) – chiama «immaginario dei dati», veicolando «narrazioni» e «contronarrazioni» (Pasquale, 2016), proto-teorie sul sociale, così come «intenti» e «obiettivi» che – Foucaultianamente – si legano ai «vari meccanismi e istituzioni di potere» (Foucault, 2011, 8). Beer chiarisce questo punto come segue:

algorithms are also a notional presence in discourse. We might look at how that term or notion is deployed to create or perpetuate certain truths about

² Si veda www.che-fare.com/facebook-lalgoritmo-e-lo-tsunami.

social orders and the like, or how certain truths are cultivated through discussions or evocations of the algorithm. This would be to suggest that the notion of the algorithm is itself doing some work in these discursive framings. It is a notion that carries some persuasive weight and is likely to be suggestive of wider power claims and rationalities (2017, 9).

Pensiamo al caso Cambridge Analytica³ e all'enfasi mediatica intorno al *micro-targeting* automatizzato di utenti Facebook sulla base delle loro caratteristiche psicografiche. La mitizzazione di questa strategia di analisi, rappresentata come un'arma magica di persuasione di massa, testimonia la potenza evocativa di algoritmi e *big data* nell'immaginario sociale contemporaneo, cui nemmeno i media più critici restano immuni⁴.

La tesi principale portata avanti da questo articolo è che al centro del regime discorsivo che circonda la connotazione dominante data all'algoritmo vi sia una nozione chiave: quella di neutralità algoritmica. In breve, l'idea che la disintermediazione tecnica, la rimozione (apparente) dell'umano, abbia come salvifica conseguenza la scomparsa della fallibilità e scarsa oggettività che, ahimè, contraddistinguono i nostri processi decisionali (cfr. Christian, Griffith, 2016). La manifestazione concreta di questo discorso datista nel tardo capitalismo informazionale è l'applicazione sempre più ubiqua di tecniche algoritmiche ad ambiti tradizionalmente lontani dalle competenze degli ingegneri informatici. Il caso più evidente è quello dei sistemi di raccomandazione utilizzati in piattaforme come Amazon, YouTube o Spotify (cfr. Airoldi, 2015), i quali prendono sostanzialmente il posto dei tradizionali intermediari culturali umani (ad esempio, critici professionisti) nell'orientare i gusti artistici dei consumatori digitali (cfr. Morris, 2015; Barile, Sugiyama, 2015). Non c'è da stupirsi dunque se nel 1995 l'autorevole *Harvard Business Review* introduceva ai lettori un esempio pionieristico di algoritmo di raccomandazione applicato al marketing con il titolo «Do you want to keep your customers forever?». L'articolo proseguiva come segue: «Customers, whether consumers or businesses, do not want more choices. They want exactly what they want – when, where, and how they want it – and technology now makes it possible for companies to give it to them» (Pine II, Peppers, Rogers, 1995).

³ La Cambridge Analytica (CA) è stata una società di consulenza britannica che progettava campagne elettorali e di comunicazione tramite l'analisi di *big data*. Il 2 maggio 2018 la società ha dichiarato bancarotta a causa dello scandalo dovuto al ruolo dell'azienda nella campagna elettorale di Donald Trump, basata su dati estratti illegalmente da milioni di utenti Facebook.

⁴ Si veda theconversation.com/cambridge-analyticas-secret-psychographic-tool-is-a-ghost-from-the-past-94143.

Da allora, il mito di una macchina neutrale ed efficiente cui delegare attività e decisioni complesse ha colonizzato quasi ogni mercato e ambito del sociale, grazie alla valanga di dati forniti gratuitamente dai *prosumer* del Web 2.0 (Ritzer, 2013), ai rapidi avanzamenti tecnici nei campi del *machine learning* e del *deep learning*, alla promessa di soluzioni di business più competitive e, di conseguenza, alla sovrabbondanza di finanziamenti per la ricerca privata (Gambetta, 2018). Questo discorso interessato intorno alla neutralità del calcolo matematico ha tuttavia radici più profonde. Il paragrafo seguente rappresenta un tentativo di ripercorrerne la storia e mitologia, in breve e senza pretese di esaustività.

3. *QUO FACTO, CALCULEMUS*

Nel diciassettesimo secolo il filosofo e matematico Gottfried Leibniz, ancora giovane, sperava di poter escogitare «una certa scrittura universale, mediante la quale possiamo eseguire calcoli su ogni genere di argomenti e trovare dimostrazioni come nell'algebra e nell'aritmetica». Il suo tentativo, quindi, era quello di elaborare un modello logico capace di risolvere ogni dibattito in un mero calcolo, capace di fornire risposte precise e inconfutabili. Nel suo *Dissertatio De Arte Combinatoria* del 1666, il matematico si interroga sulla cosiddetta «arte delle combinazioni», un metodo che consentirebbe di generare idee e invenzioni inedite, oltre che scomporre e analizzare idee complesse in elementi più semplici. Descritta come «la madre di tutte le invenzioni», questa scienza, secondo Leibniz, avrebbe portato progressi in ambiti più disparati della conoscenza, dalla musica alla medicina, dalla fisica alla politica. Ispirato dal lavoro di Pascal, Leibniz realizza inoltre nel 1671 la *Stepped Reckoner* (calcolatrice a scatti), un dispositivo meccanico capace di eseguire le quattro operazioni aritmetiche.

Oltre a contribuire allo sviluppo del calcolo integrale e differenziale, Leibniz ha anticipato l'arrivo di un immaginario computazionale. Il mito del calcolo di Leibniz consisteva nell'utopia di un linguaggio del pensiero composto da concetti «puri» che, come gli elementi semplici dell'aritmetica, potessero essere utilizzati in metodi formali, portando alla meccanizzazione e automazione della ragione stessa. Esso fu ispirato da molte linee di pensiero e da diversi autori – dalle opere di Aristotele al tentativo di Thomas Hobbes di equiparare ragionamento e calcolo, oltre che dal lavoro di Raimondo Lullo, che nella sua *Ars Magna* del 1308 proponeva forme argomentative basate su permutazioni di attributi logici fondamentali, e di cui Leibniz venne a

conoscenza grazie al gruppo *Herborn Encyclopaedists*⁵.

Le idee di Leibniz saranno poi a loro volta determinanti nel guidare gli sforzi dell'inglese Charles Babbage, considerato uno dei «bissonni» del moderno computer (Campbell-Lelly *et al.*, 2013). Nel 1823 il progetto del suo *Difference Engine* ricevette copiosi finanziamenti da parte del governo britannico, ingolosito dalla promessa di migliorare i commerci via mare eliminando in automatico gli errori presenti nelle tavole logaritmiche di navigazione, grazie alla «unerring certainty of mechanism» (Ivi, 7). Il mito del calcolo già allora solleticava le mire espansionistiche di un capitalismo ansioso di trasformare l'incertezza in rischio calcolabile. Non a caso, a inizio Novecento Max Weber individuava nella spinta verso la razionalizzazione la caratteristica chiave di una modernità ossessionata dal sogno di applicare il calcolo matematico a «qualsiasi cosa» (1946, 139).

Anche nella narrativa la percezione di una datificazione o matematizzazione dell'umano e del sociale è riscontrabile da tempo immemore. Nel 1726, ne *I viaggi di Gulliver*, Johnatan Swift immagina che nella città di Lagado il protagonista incontri un dispositivo noto come «il motore», capace di «scrivere libri di filosofia, poesia, politica, leggi, matematica e teologia, senza la minima assistenza del genio o dello studio». L'intento è provocatorio, la sua è una caricatura di un metodo di sperimentazione produttiva che avrebbe trovato anche tentativi concreti di applicazione – pensiamo agli esperimenti poetici generativi del gruppo letterario francese *Oulipo* nonché, ovviamente, a casi più recenti di scrittura collettiva e digitale⁶. Nel saggio *Cibernetica e fantasmi*, Italo Calvino riconosceva che

nel modo in cui la cultura d'oggi vede il mondo, c'è una tendenza che affiora contemporaneamente da varie parti: [...]il pensiero, che fino a ieri ci appariva come qualcosa di fluido, evocava in noi immagini lineari come un fiume che scorre o un filo che si dipana [...], oggi tendiamo a vederlo come una serie di stati discontinui, di combinazioni di impulsi su un numero finito di organi sensori e di controllo (1980, 209).

Nel testo si riscontra quindi da parte dell'autore sia un passaggio dall'umano all'ibrido-macchina che, più simbolicamente, una transizione dal continuo al discreto. La fusione dell'umano con il macchinico, e di conseguenza i timori e i dubbi rispetto la personalizzazione e l'umanizzazione del *cyborg*, sono temi che non a caso la letteratura – in

⁵ Si veda publicdomainreview.org/2016/11/10/let-us-calculate-leibniz-llull-and-computational-imagination.

⁶ Si veda www.prismomag.com/fine-mito-autore-copyright.

particolare fantascientifica – ha affrontato in numerose direzioni nell'ultimo mezzo secolo. Al tempo stesso, con la messa a punto di metodi scientifici sempre più sofisticati, affini alle cosiddette scienze dure, nei vari campi del sapere si è fatta strada la convinzione che la matematizzazione dei processi fosse un mezzo – talvolta necessario – per dare rispettabilità ad una teoria. Per usare le parole del teorico dei media Tarleton Gillespie: «conclusions described as having been generated by an algorithm wear a powerful legitimacy, much the way statistical data bolsters scientific claims» (2016, 23-24).

Non c'è da stupirsi dunque se, con il rapido sviluppo dei computer, la capillare diffusione del digitale e la convergenza del dibattito pubblico sul tema dei *big data*, il mito del calcolo abbia trovato una ghiotta occasione per ripresentarsi. Nel giugno 2008 su *Wired* venne pubblicato il noto *The End of Theory* di Chris Anderson. L'autore proclamava l'inizio dell'«Era dei Petabyte», durante la quale il metodo scientifico sarebbe divenuto obsoleto grazie all'analisi di grandi moli di dati, e quindi l'elaborazione di modelli teorici non avrebbe avuto più alcuna ragion d'essere⁷. Dall'articolo emerge una fiducia cieca verso algoritmi e correlazioni statistiche, non solo in termini di efficienza e velocità, ma anche di neutralità ed efficacia decisionale. Come ricorda Eleonora Priori nel suo contributo al volume *Datacrazia*:

Alla base delle convinzioni di Anderson, che da un positivismo scientifico già sorpassato scivolano pericolosamente verso un incondizionato fideismo tecnologico (che ha poi finito per prendere il nome di datismo), vi è l'idea che, nell'era del petabyte, la nozione di correlazione sostituisca quella di causalità, consentendo alla scienza di progredire senza la necessità di confrontarsi con modelli coerenti, teorie unificanti o spiegazioni meccanicistiche (2018, 126).

Delegando alla macchina decisioni in tempo reale, le domande al centro dello sguardo scientifico sul mondo – come? perché? – vengono sostituite dal laconico «basta che funzioni» proprio del pragmatismo ingegneristico⁸. Gli ultimi dieci anni di storia hanno visto affermarsi senza grossi intoppi e su scala globale la dottrina del *laissez faire* made in Silicon Valley e, con essa, il discorso dominante sulla (presunta) neutralità di algoritmi e piattaforme (Pasquale, 2016).

⁷ Si veda www.wired.com/2008/06/pb-theory.

⁸ Si veda logicmag.io/03-dont-be-evil.

4. OLTRE IL MITO: CASI DI NON-NEUTRALITÀ ALGORITMICA

Storicamente, le scienze sociali e il dibattito pubblico hanno prestato molta più attenzione a come la tecnologia influenza la società che al rapporto inverso, spesso adottando quella postura intellettuale semplicistica nota come determinismo tecnologico. Questa posizione, assumendo la tecnologia come indipendente dalla società nel suo complesso, va logicamente a sostegno del discorso dominante intorno alla neutralità del calcolo e all'ineluttabilità del progresso, in quanto la società non può far altro che rassegnarsi ed adattarsi all'inevitabile cambiamento tecnologico. Durante gli anni Ottanta, una contro-narrazione⁹ partita da lontano (cfr. il Paragrafo 5) si fece largo grazie all'espansione del filone di ricerca divenuto poi noto come *STS* (*Science and Technology Studies*): la tecnologia, lungi dall'essere una sfera autonoma, è una costruzione sociale che dipende tanto da interessi economici e politici contestuali quanto dal percorso scientifico e tecnologico avvenuto in precedenza – anch'esso, a sua volta, socialmente condizionato (cfr. MacKenzie, Wajcman, 1999). Ad esempio, la lampadina di Edison fu studiata come parte di un preciso sistema di distribuzione dell'elettricità e progettata per essere competitiva in termini di costi rispetto alla già esistente luce a gas. Gli «intenti» e «obiettivi» (Foucault, 2011, 84) del discorso imprenditoriale di Edison entrarono direttamente nei brevetti, volti a minimizzare la quantità di costoso rame utilizzato nei cavi elettrici (MacKenzie, Wajcman, 1999, 16-19). In altre parole, nessuna tecnologia è mai stata neutrale.

Esattamente come la lampadina di Edison, gli algoritmi sono costruzioni sociali che riflettono interessi, discorsi di verità, assunti arbitrari sul mondo sociale (Rieder, 2017; Pasquinelli, 2014). Allo scopo di illustrare concretamente questo punto, nei prossimi paragrafi presenteremo in breve degli esempi di sistemi algoritmici dedicati alle funzioni più diverse, dalla classificazione automatica di immagini alla prevenzione del crimine, concentrandoci perlopiù sui loro errori. Questi casi serviranno a mostrare in pratica la fallacia del mito della neutralità algoritmica, così come a esaminarne le varie sfaccettature discorsive. In particolare, tre aspetti cardine del mito verranno discussi e problematizzati attraverso gli esempi seguenti: primo, l'assunto che nella fase di progettazione e/o addestramento dell'algoritmo non abbiano interferito distorsioni sistematiche e che, pertanto, l'*output*

⁹ In questo testo, usiamo i termini «narrazione» e «contronarrazione» nell'accezione data da Pasquale nella sua discussione intorno al platform capitalism (2016), vicina a quella Foucaultiana di «discorso» (2011).

restituisca un punto di vista oggettivo sulla realtà (Paragrafo 4.1); secondo, l'idea che l'automatizzazione e la disintermediazione di processi complessi siano di per se stesse garanzie di neutralità (Paragrafo 4.2); terzo, la pretesa che i risultati del calcolo algoritmico, narrati come accurati e infallibili, non alterino la realtà secondo logiche socialmente, storicamente e politicamente determinate (Paragrafo 4.3).

4.1. Lupi, husky e oggettività

Nel senso più ampio del termine, per algoritmo si intende una sequenza di operazioni da eseguire in un ordine prestabilito, un po' come la ricetta per cucinare una torta. Nel caso in cui un algoritmo-procedura sia progettato per l'analisi e la risoluzione di questioni umane e sociali, alla base della sua fase di creazione è necessaria, a maggior ragione, l'elaborazione teorica di un modello – la formalizzazione in termini matematici di un problema e della sua soluzione (Gillespie, 2016, 19).

Dietro la genericità di una nozione che, come sottolinea Gillespie, è sineddoche di sistemi tecnici differenti ma strettamente collegati, esiste un ampio spettro di tecniche e approcci differenti. Mentre la presenza della teoria e delle sue implicazioni è più evidente nel caso di un algoritmo classico, deterministico, poiché esso risulta essere niente di più che l'automatizzazione di un processo manuale – si pensi a una macchina che cucina torte alla crema –, questa fase umana di teorizzazione e modellizzazione sembra venire a mancare qualora si tratti di algoritmi utilizzati per simulare capacità intuitive, progettati in modo *data-driven* grazie a informazioni «digerite» in fase di apprendimento. Nel caso degli algoritmi di *machine learning*, tra il momento della scrittura del codice e l'applicazione finale, c'è una fase durante la quale – secondo l'interpretazione scienziata – l'algoritmo apprenderebbe la «realtà» in modo «oggettivo» da una grossa mole di dati classificati. Tuttavia, sia la scelta del dataset che la generazione dei dati stessi sono il prodotto di azioni umane culturalmente e socialmente situate, le quali possono introdurre distorsioni sistematiche nel modello. Come mostrato da vari testi – ad esempio *Armi di distruzione matematica* di Cathy O'Neil (2016) e *Algorithms of oppression* di Safiya Umoja Noble (2018) – questi *bias* possono portare a conseguenze sociali indesiderabili, specialmente quando riguardano piattaforme utilizzate da milioni di utenti (Airoldi, 2018).

Per decostruire il discorso intorno all'oggettività degli algoritmi, consideriamo l'esperimento di un gruppo di ricercatori dell'Università di Washington (Ribeiro *et al.*, 2016), pensato per mettere in discussione

l'accuratezza di predizioni effettuate tramite *machine learning*. Ipotizziamo di voler costruire un classificatore capace di distinguere foto di husky da foto di lupi. Per quanto ampio sia il nostro *dataset*, se tutti gli husky sono ritratti (involontariamente o volontariamente, come in questo studio) in presenza di neve, si ottiene un algoritmo che classifica come husky qualunque animale (lupo o husky) fotografato in presenza di neve. Questo esempio è utile a formalizzare la critica all'oggettività degli algoritmi. La macchina, infatti, non sta sbagliando, nel senso che effettivamente ha trovato una *feature* capace di discernere, nei casi analizzati, le due classi di oggetti. L'inadeguatezza dello strumento, allora, deriva dall'assunto che i dati di partenza siano adeguati a rappresentare il reale e le sue peculiarità. Per quanto questo esempio non sia legato a scelte socialmente rilevanti, quando si tratta di applicazioni più concrete il discorso rimane il medesimo: lo scienziato – *data scientist* – ipotizza che un *dataset* di apprendimento e un'eventuale classificazione manuale costituiscano un *input* adeguato a rappresentare fedelmente il fenomeno analizzato. La presunta oggettività del calcolo è, perciò, «fanzionale» (Gillespie, 2014). Il funzionamento dei sistemi algoritmici, anche nel caso di *machine learning* e intelligenza artificiale, non è mai totalmente indipendente da scelte umane, errori, e distorsioni culturalmente indotte.

4.2. Amazon, bitcoin e disintermediazione

Gli algoritmi sono perlopiù descritti da sviluppatori e aziende come strumenti capaci di rendere efficienti processi che, altrimenti, richiederebbero intermediari umani (Morris, 2015). È il caso, ad esempio, delle raccomandazioni automatiche proposte da piattaforme come Amazon e di criptovalute come Bitcoin. Fino a che punto quella della disintermediazione algoritmica è una retorica discorsiva che maschera una molteplicità di interventi umani?

Il 18 settembre 2017, i giornalisti dell'emittente inglese Channel 4 firmavano un'inchiesta con al centro un imputato *sui generis*: il sistema di raccomandazione di Amazon¹⁰. La notizia, poi ripresa da testate quali *Guardian*, *Independent* e *NY Times*, suonava particolarmente appetitosa in mesi costellati da diversi attacchi terroristici. Cercando uno specifico agente chimico sulla piattaforma, il sistema di raccomandazione suggeriva tra gli articoli «spesso comprati insieme» altri ingredienti che, se combinati, avrebbero potuto essere utilizzati per fabbricare una rudimentale bomba artigianale. L'algoritmo in questione funziona

¹⁰ Si veda www.channel4.com/news/potentially-deadly-bomb-ingredients-on-amazon.

secondo una logica nota come *collaborative filtering*: se un numero di persone considerevole compra sia X che Y, i due prodotti saranno con grande probabilità correlati (cfr. Airoidi, 2015). A fare scalpore non fu tanto la possibilità di acquistare facilmente agenti chimici potenzialmente pericolosi online, quanto piuttosto il carattere automatico di suggerimenti moralmente tabù – completi persino di un *set* di pallini d'acciaio, utile a rendere ancora più letale una ricetta già esplosiva. Di chi è la colpa? L'algoritmo, da un certo punto di vista, ha eseguito il suo compito in modo oggettivamente corretto, tracciando correlazioni nei *pattern* d'acquisto degli utenti. I risultati, però, tanto neutrali non sono. Seguendo la linea argomentativa dell'ex *executive* di Facebook intervistato dal Guardian nel 2017¹¹, la colpa è, in ultima istanza, della società. L'algoritmo è innocente, in quanto non ha fatto altro che riprodurre una verità statistica sul mondo sociale («statistical truth») difficile da digerire. Il discorso sulla presunta neutralità dell'algoritmo così, però, vacilla: non solo perché l'*output* della disintermediazione algoritmica è il prodotto delle discutibili attività degli utenti della piattaforma, ma anche e soprattutto perché Amazon, per assicurare i media, ha dovuto promettere un intervento umano sul codice. Il portavoce ha successivamente dichiarato alla BBC che i suggerimenti automatici sono sistematicamente filtrati tramite una *blacklist* di parole chiave, la quale sarebbe stata aggiornata di conseguenza¹². La mitica disintermediazione algoritmica, nella pratica, sembra perciò essere a geometria variabile.

Se decostruire il discorso intorno alla neutralità dell'algoritmo risulta essere relativamente semplice considerando gli errori di app e piattaforme, i sistemi algoritmici di criptovalute come Bitcoin sono raramente oggetto di critiche o scandali. La tecnologia *blockchain* alla base di Bitcoin consente una disintermediazione totale, grazie alla registrazione automatica e in tempo reale di ogni transazione in un registro pubblico condiviso. Qui ci viene però in aiuto la tesi di dottorato del noto hacker, attivista e *media artist* Jaromil Rojo, intitolata *Algorithmic Sovereignty* (2018), la quale offre una preziosa analisi di prima mano – e di stampo foucaultiano – sulla genesi del progetto Bitcoin. «Bitcoin has no single monetary authority, but a shared pact and the underlying rationality of a mathematical algorithm – the intangible dream of neutrality», scrive l'autore (Ivi, 37). Dalla scelta del logo al mito fondativo del misterioso Satoshi Nakamoto, passando per

¹¹ Si veda www.theguardian.com/technology/2017/may/02/facebook-executive-advertising-data-comment.

¹² Si veda www.bbc.com/news/technology-41320375.

le storie personali degli sviluppatori-attivisti (Ivi, 38-41), il percorso sociopolitico che ha portato allo sviluppo della criptovaluta mostra, secondo Rojo, che quella della neutralità algoritmica, anche nel caso di Bitcoin, rimane un'utopia. La tecnologia alla base della disintermediazione totale è un prodotto dell'azione collettiva umana, fatto di idee e scelte arbitrarie, oggettificate nel codice e perpetuate attraverso di esso. Così conclude l'autore:

there is nothing inherently neutral in an algorithm. To the contrary, an algorithm is there to implement visions, ideas, beliefs and to satisfy needs and desires [...]. To perceive the presence of a decentralised algorithm as guarantee of neutrality, to consider its existence per-se as the basis of a constituency, is an error made by many. This error is also evident on the surface of Bitcoin, a project whose foundations are social, political and technical (Ivi, 13).

4.3. Compas, Lombroso e infallibilità

Il crescente consenso scientifico e istituzionale intorno all'accuratezza – per non dire infallibilità – di decisioni esternalizzate a tecniche di intelligenza artificiale (Natale e Ballatore, 2018) ha portato ad applicazioni eticamente discutibili di sistemi algoritmici ad ambiti delicati come, ad esempio, quello giudiziario.

Nel noto racconto *Rapporto di minoranza* di Philip K. Dick, la divisione Precrimine è capace di prevedere crimini futuri. Non molto diversamente, Compas (*Correctional Offender Management Profiling for Alternative Sanctions*) è un software che promette di analizzare la fedina penale e i dati personali di un imputato per stabilire la possibilità che commetta nuovamente un crimine¹³. In uno studio pubblicato su *Science Advances*, Julia Dressel e Hany Farid (2018) scrivono: «i sostenitori di questi sistemi dicono che i big data e l'uso del machine learning avanzato rendono queste analisi più accurate e con meno pregiudizi rispetto a quelle svolte dagli uomini». Ma, in verità, «Compas non è più accurato e più imparziale delle previsioni di persone con poca o senza esperienza di diritto penale» (Dressel, Farid, 2018, 1). La ricerca mostra innanzitutto che piccoli gruppi di persone scelte a caso sono in grado di prevedere se un imputato commetterà un crimine con un'accuratezza del 67,0%, percentuale identica a quella di Compas. Inoltre, l'algoritmo del Compas ha evidenti pregiudizi razziali. Già due anni prima dello studio, ProPublica aveva elaborato una lunga inchiesta su questo software: a parità di condizioni, una persona di colore veniva

¹³ Si veda www.equivant.com/solutions/inmate-classification.

sistematicamente classificata come ad alto rischio¹⁴.

L'applicazione di tecniche di *machine learning* alla prevenzione della criminalità si è spinta fino al punto di rivalutare il pensiero di Cesare Lombroso. Secondo una controversa ricerca (Wu, Zhang, 2016), è possibile distinguere tra criminali e non criminali con il 90,0% di accuratezza semplicemente attraverso l'analisi automatica di immagini di volti umani. Le parole degli autori sono imbevute del mito della neutralità algoritmica: «unlike a human examiner/judge, a computer vision algorithm or classifier has absolutely no subjective baggages, having no emotions, no biases whatsoever due to past experience, race, religion, political doctrine, gender, age, etc., no mental fatigue, no preconditioning of a bad sleep or meal» (Ivi, 2).

Basta poco a decostruire un discorso così ingenuo: come svelano i tipi di Callingbullshit.org¹⁵, il *dataset* utilizzato da Wu e Zhang consiste in un misto di foto segnaletiche di criminali incarcerati – certo non molto felici di esserlo – e immagini di volti estratti automaticamente dal Web, probabilmente da pagine professionali, spesso con tanto di camicia e cravatta. Se nel caso Compas è la riproduzione algoritmica delle disuguaglianze sociali a confutare il mito della neutralità, le caratteristiche delle immagini utilizzate per addestrare la macchina-Lombroso minano alla radice l'apparente infallibilità del modello – un po' come nel caso illustrato nel Paragrafo 4.1.

Ovviamente sappiamo che la comunità scientifica è ben lontana dall'approvare le tesi e i metodi lombrosiani, ma qui vogliamo mostrare come in alcuni ambienti di ricerca (pubblica o privata), anche a causa dell'assenza di un adeguato dibattito epistemologico, l'idea che la correlazione sia sufficiente e la teoria superflua (Priori, 2018) può portare persino ad avvalorare una pseudo-frenologia. Come sottolineano Ribeiro e colleghi, anziché valutare gli algoritmi esclusivamente sulla base della (presunta) accuratezza dell'*output* – solitamente misurata come la percentuale di elementi del *test set* correttamente classificati dalla macchina – bisognerebbe mettere al centro la sua «interpretabilità» (2016, 2). Solo in questo modo, aprendo la scatola nera e affermando la necessità di una «spiegazione» del modello, è possibile avere fiducia nei risultati.

¹⁴ Si veda www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

¹⁵ Si veda callingbullshit.org/case_studies/case_study_criminal_machine_learning.html.

5. L'APE, L'ARCHITETTO E IL LABORATORIO: SULLA NON-NEUTRALITÀ DELLA SCIENZA

«Ciò che fin da principio distingue il peggior architetto dall'ape migliore» – scriveva Marx nel *Capitale* - «è il fatto che egli ha costruito la celletta nella sua testa prima di costruirla in cera». Il processo lavorativo porta alla luce un risultato già immaginato come idea dal lavoratore. Al tempo stesso, i mezzi utilizzati per realizzare il lavoro sono «indici dei rapporti sociali nel quadro in cui viene compiuto il lavoro» oltre che elementi di distinzione dell'epoca economica in corso. Prendendo ispirazione da questo passaggio di Marx, quattro fisici della Sapienza – Giovanni Ciccotti, Marcello Cini, Michelangelo de Maria e Giovanni Jona-Lasinio – si misero a riflettere su come la ricerca scientifica e il suo decorso siano storicamente soggetti a logiche economiche, e quindi su come le stesse rivoluzioni scientifiche mettano in crisi l'immagine della scienza come superiore alla storia. «L'idea di autonomia delle teorie scientifiche rispetto alla società [...] non ha fondamento nella realtà», scrivevano ne *L'ape e l'architetto* (1976).

Possiamo ricondurre il discorso sulla neutralità algoritmica a quello sulla neutralità della scienza, e con questo metterlo a critica. L'algoritmo prodotto dagli architetti/ingegneri realizza nell'elemento macchina un'idea già presente nella mente degli architetti/ingegneri – intesi, nel caso dei *big data*, come la massa dei produttori di informazione. «Non-neutralità è un concetto formalmente negativo» – affermavano Ciccotti e colleghi – «ma in quella formulazione è implicita una considerevole articolazione e un arricchimento nell'idea di scienza e di scientificità».

Mentre *L'ape e l'architetto* andava in stampa, l'antropologo e filosofo Bruno Latour e il sociologo britannico Steve Woolgar si trovavano in un laboratorio scientifico californiano specializzato in ricerca biologica, per condurre uno studio etnografico i cui risultati furono pubblicati tre anni dopo in un volume dal titolo *Laboratory Life: The Construction of Scientific Facts* (1986). La loro tesi principale – suggestiva all'epoca come oggi, nonché certamente più radicale di quella dei fisici italiani – è che ciò che chiamiamo scienza non sia altro che una costruzione sociale. L'osservazione partecipante del laboratorio e dei suoi strani nativi rivelava qualcosa che gli accademici conoscono bene: le attività convulse di un «paperificio», una fabbrica di *output* scientifici dove le carriere individuali dei ricercatori e i parametri delle riviste contano più dei sedicenti fatti. Persino strumenti come lo spettrometro di massa, lungi dall'essere neutrali, agli occhi di Latour e

Woolgar apparivano come la materializzazione di asserti scientifici risultanti da tradizioni disciplinari anch'esse socialmente condizionate, trasformati in stabilizzatori di realtà volti a «ricavare l'ordine dal disordine» (Ivi, 242). In sintesi, ciò che gli scienziati oggetto di studio facevano per la maggior parte del tempo era incapsulare conoscenza in “scatole nere”:

The activity of creating black boxes, of rendering items of knowledge distinct from the circumstances of their creation, is precisely what occupies scientists the majority of the time. The way in which black boxing is done in science is thus an important focus for sociological investigation. Once an item of apparatus or a set of gestures is established in the laboratory, it becomes very difficult to effect the retransformation into a sociological object (Ivi, 259-260)

Lo sforzo sociologico necessario a svelare l'“in scatolamento” delle attività del laboratorio ricorda da vicino la fatica, omologa, per aprire la *black box* del codice algoritmico (Pedreschi *et al.*, 2018) o scassinare la gabbia discorsiva del mito della neutralità (Rojo, 2018; Foucault, 2011). Sarebbe da chiedersi, allora, come sia stato possibile passare in poco più di trent'anni dalla riflessività epistemologica dei saggi di cui sopra alla fede iper-ottimistica in correlazioni statistiche e tecnologie opache (Priori, 2018). Quasi sicuramente la risposta ha a che fare con il mito del calcolo nell'era dei *big data* e i tanti meccanismi di potere a cui questo discorso si collega.

6. CONCLUSIONI

L'algoritmo, come lo spettrometro di massa del laboratorio studiato da Latour e Woolgar (1986), è essenzialmente uno «stabilizzatore di fiducia» (Gillespie, 2014, 179), un produttore di discorsi, cristallizzati negli *output* restituiti dalla *black box* del *machine learning*. Al contempo, è il prodotto di discorsi, iscritti nelle regole generative del codice, nell'“inconscio tecnologico” (Beer, 2009) che processa le nostre esistenze digitali dietro il volto rassicurante dell'interfaccia utente. Infine, l'algoritmo è esso stesso discorso (Beer, 2017; Seaver, 2017) – giustificazione per licenziamenti e finanziamenti, soluzione ottimale ai complessi problemi della vita quotidiana, garanzia di efficienza e neutralità (cfr. Christian, Griffith, 2016). Tuttavia, come il “Turco Meccanico” creato da von Kempelen nel 1769 per Maria Teresa d'Austria illudeva di giocare a scacchi autonomamente mentre era in realtà controllato da un umano, così le più recenti tecnologie di

intelligenza artificiale traggono risultati sulla base dell'apprendimento compiuto su dati prodotti da migliaia di utenti connessi e delle scelte ingegneristiche nascoste nel codice software – riguardanti, ad esempio, l'operativizzazione di concetti arbitrari quali quelli di rilevanza, significatività, similarità, soddisfazione (Gillespie, 2014; Rieder, 2017). Gran parte delle considerazioni critiche riguardo a calcolabilità statistica e neutralità scientifica, fiorite durante il Novecento nelle scienze sociali come in quelle dure, sono applicabili anche alla odierna ingegnerizzazione algoritmica della vita quotidiana (Latour, Woolgar, 1986; Ciccotti *et al.*, 1976). Ciononostante, complice la relativa novità dell'argomento, i dibattiti pubblico e accademico rimangono in larga misura dominati dal positivismo ingenuo implicito nella narrazione tecno-ottimista e pro *big data*.

Aldilà del livello del discorso e delle sue sfaccettature, una domanda molto concreta sorge spontanea: che fare? Come criticare il mito della neutralità algoritmica senza diventare «tecnici dell'Apocalisse»? – «specializzati nel dimostrare che il nuovo orizzonte di problemi è radicalmente equivoco, antiumano, e che occorre rifarsi al culto dei valori di un tempo per garantire all'umanità la sopravvivenza» (Eco, 2017, 368-369). Qualche considerazione conclusiva ci viene suggerita dalla letteratura, nonché dagli esempi presentati nelle pagine precedenti.

In primo luogo, l'opacità di intelligenza artificiale e *machine learning*, il loro essere delle *black box* (Pasquale, 2015) il cui funzionamento è imperscrutabile non soltanto per l'utente finale, ma anche per l'ingegnere informatico che ha contribuito all'addestramento della macchina, non preoccupa soltanto per una semplice mancanza di trasparenza, o perché l'alone di mistero alimenterebbe il mito della neutralità algoritmica. Il problema principale, sottolineato da Pedreschi e colleghi (2018), è rappresentato invece dall'impossibilità di controllare la presenza di *bias* nascosti nel modello. Se casi come quello di Compas, menzionato sopra, sono venuti a galla, molte altre applicazioni distorte del *machine learning* in ambiti come quello bancario, assicurativo o militare potrebbero restare sottotraccia per molto tempo prima che gli errori vengano scoperti e corretti. La questione dell'opacità degli algoritmi e dei loro eventuali *bias* (Noble, 2018) porta ulteriore supporto alla tesi di una necessaria interpretabilità delle scatole nere algoritmiche. Ciò che diversi autori suggeriscono è di investire in tecnologie e metodologie per la spiegazione delle attività degli algoritmi, allo scopo di allinearle con i valori e le aspettative della società, salvaguardando così l'autonomia e la consapevolezza umana in ambito decisionale (Ribeiro *et al.*, 2016; Pedreschi *et al.*, 2018).

Un secondo punto è rappresentato dal mancato coinvolgimento della collettività nella progettazione e nel perfezionamento degli algoritmi di rilevanza pubblica (Gillespie, 2014; Boccia Artieri, 2014). Sebbene gli effetti di algoritmi come quelli di Facebook o Google sulla vita quotidiana di miliardi di persone siano incommensurabili, il coinvolgimento dell'utente-*prosumer* nel processo computazionale – al di là dell'estrazione più o meno consenziente dei suoi dati personali – è praticamente nullo, e i margini di personalizzazione dell'algoritmo sono di norma molto limitati. Rojo (2018, 135), riferendosi *in primis* alle comunità di sviluppatori, parla a questo proposito di un bisogno di sovranità, da contrapporsi alla disintermediazione tecnica e alla sua sedicente neutralità: «sovereignty relates to the way a community can influence an algorithm, appropriate it, distribute it, share it and create new ones». Questa sovranità, aggiungiamo noi, andrebbe idealmente allargata alle masse di cittadini-utenti finali. Perché questo sia possibile e auspicabile in pratica si dovrebbe mettere in discussione l'iperspecializzazione accademica – la quale ostacola la formazione e circolazione di un sapere critico multidisciplinare –, nonché sfidare l'idea dominante dell'algoritmo come *asset* commerciale, proprietà intellettuale dalla formulazione inaccessibile e dalle conseguenze imprevedibili (Pasquale, 2015; Airoidi, 2018). Un maggiore controllo decentralizzato sui codici con cui interagiamo quotidianamente rappresenterebbe, crediamo, un antidoto pratico e potenzialmente efficace al proliferare dei miti del calcolo e della neutralità algoritmica.

La nostra speranza è che emerga un controdiscorso capace di re-immaginare il ruolo degli algoritmi nella società senza scivolare nelle fallacie epistemologiche del «datismo» (Priori, 2018) o nella retorica caricaturale dei «tecnici dell'Apocalisse» (Eco, 2017). Perché ciò accada, è utile guardare all'Ape e all'Architetto (Ciccotti *et al.*, 1976), così come al laboratorio di Latour e Woolgar (1986) e ai filoni di ricerca che ne sono scaturiti (MacKenzie, Wajcman, 1999). Questo non per rassegnarsi all'impossibilità di produrre una qualsivoglia forma di sapere scientifico sul mondo, ma per guardare all'utopia della neutralità algoritmica come a un equilibrio instabile, contestuale e partecipativo, piuttosto che come a un dogma discorsivo volto a disinnescare la critica prima ancora che essa abbia luogo.

RIFERIMENTI BIBLIOGRAFICI

- AIROLDI, M. (2015). Potrebbe interessarti anche: recommender algorithms e immaginario, il caso YouTube. *IM@GO*, 1(6), 132–
-

- 150.
- (2018). L'output non calcolabile. Verso una cultura algoritmica. In D. Gambetta (a cura di) *Datacrasia, Politica cultura algoritmica e conflitti al tempo dei big data*. Ladispoli: D Editore.
- AMOORE, L., Piotukh, V. (a cura di) (2016). *Algorithmic Life. Calculative devices in the age of Big Data*. London-New York: Routledge.
- BARILE, N., SUGIYAMA, S. (2015), The Automation of Taste: A Theoretical Exploration of Mobile ICTs and Social Robots in the Context of Music Consumption. *International Journal of Social Robotics*, 7(3), 407-416.
- BEER, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, 11(6), 985-1002.
- (2017). The social power of algorithms. *Information, Communication and Society*, 20(1), 1-13.
- (2018). *The Data Gaze: Capitalism, Power and Perception*. London: Sage.
- BOCCIA ARTIERI, G. (2014). La rete dopo L'overload informativo. La realtà dell'algoritmo da macchia cieca a bene comune. *Paradoxa*, 2, 100-113.
- CALVINO, I. (1980). Cibernetica e fantasmi. *Una pietra sopra*. Torino: Einaudi.
- CAMPBELL-KELLY, M., ASPRAY, W., ENSMINGER, N., YOST, J.R. (2013³). *Computer: a history of the information machine*. Boulder, Colorado: Westview Press.
- CARAH, N. (2017). Algorithmic brands: A decade of brand experiments with mobile and social media. *New Media & Society*, 19(3), 384-400.
- CHENEY-LIPPOLD, J. (2011). A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society*, 28(6), 164-181.
- CHRISTIAN, B., GRIFFITHS, T. (2016). *Algorithms to live by: The computer science of human decisions*. London: William Collins.
- CICCOTTI, G., CINI, M., DE MARIA, M., JONA-LASINIO, G. (1976). *L'Ape e l'architetto: paradigmi scientifici e materialismo storico*. Roma: Feltrinelli.
- DRESSEL, J., FARID, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), 1-5.
- ECO, U. (2017 [1964]) *Apocalittici e integrati*. Firenze: Giunti Editore.
- ESPOSITO, E. (2017). Algorithmic memory and the right to be forgotten
-

- on the web. *Big Data & Society*, 4(1), First Published.
- FOUCAULT, M. (2011 [1976]). *La volontà di sapere. Storia della sessualità I*. Milano: Feltrinelli.
- GAMBETTA, D. (a cura di) (2018). *Datacrazia, Politica cultura algoritmica e conflitti al tempo dei big data*. Ladispoli: D Editore.
- GILLESPIE, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, K. Foot (a cura di) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge (MA): MIT Press.
- GILLESPIE, T. (2016). Algorithm. In B. Peters (a cura di) *Digital Keywords*. Princeton: Princeton University Press, pp. 18-30.
- HALLINAN, B., STRIPHAS, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society*, 18(1), 117-137.
- LATOUR, B., WOOLGAR, S. (1986). *Laboratory Life. The Construction of Scientific Facts*. Princeton (NJ): Princeton University Press.
- MACKENZIE, D., WAJCMAN, J. (a cura di) (1999²). *The social shaping of technology*. Buckingham: Open University Press.
- MCQUILLAN, D. (2015). Algorithmic states of exception. *European Journal of Cultural Studies*, 18(4-5), 564-576.
- MORRIS, J. W. (2015). Curation by code: Infomediaries and the data mining of taste. *European Journal of Cultural Studies*, 18(4-5), 446-463.
- NATALE, S., BALLATORE, A. (2017). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence*, First Published.
- NOBLE, S. U. (2018). *Algorithms of Oppression: How search engines reinforce racism*. New York: NYU Press.
- O'NEIL, C. (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. Danvers: Crown Publishing Group.
- PASQUALE, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge-London: Harvard University Press.
- (2016). Two narratives of platform capitalism. *Yale Law and Policy Review*, 35, 309-319.
- PASQUINELLI, M. (a cura di) (2014). *Gli algoritmi del capitale*. Verona: Ombre Corte.
- PEDRESCHI, D., GIANNOTTI, F., GUIDOTTI, R., MONREALE, A., PAPPALARDO, L., RUGGIERI, S., TURINI, F. (2018). Open the Black Box: Data-Driven Explanation of Black Box Decision Systems.
-

- arXiv*, Preprint arXiv:1806.09936.
- PINE II, J., PEPPERS, D., ROGERS, M. (1995). Do You Want to Keep Your Customers Forever?. *Harvard Business Review*, 73, 103-114.
- PRIORI, E. (2018). Is correlation enough? (spoiler: No!). In D. Gambetta (a cura di) *Datacrazia, Politica cultura algoritmica e conflitti al tempo dei big data*, Ladispoli: D Editore.
- RIBEIRO, M.T., SINGH, S., GUESTIN, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier”. *arXiv*, Preprint arXiv:1602.04938.
- RIEDER, B. (2017). Scrutinizing an algorithmic technique: the Bayes classifier as interested reading of reality. *Information Communication and Society*, 20(1), 100-117.
- RITZER, G. (2013). Prosumption: Evolution, revolution, or eternal return of the same?. *Journal of Consumer Culture*, 14(1), 3-24.
- ROJO, J. (2018). Algorithmic Sovereignty. *Tesi di Dottorato*, University of Plymouth.
- SEAVER, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1-12.
- STRIPHAS, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, 18(4-5), 395-41
- WEBER, M. (1946 [1919]). Science as a vocation. In H. H. Gerth, C. W. Mills (a cura di), *From Max Weber: Essays in sociology* (pp. 129-156). London: Routledge.
- WU, X., Zhang, X. (2016). Automated inference on criminality using face images. *arXiv*, Preprint arXiv:1611.04135, 4038-4052.
- ZARSKY, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science Technology and Human Values*, 41(1), 118-132.
-

IL POTERE RAZIONALE DEGLI ALGORITMI TRA BUROCRAZIA E NUOVI IDEALTPI

di *Chiara Visentin**

Abstract

At first sight algorithms and bureaucracies may seem to have little in common. However, as systems of rules and procedures, both of them can govern aspects of life and society and even become institutionalized components of them. Several academic studies have deployed this association to demystify the power and functioning of algorithms, to show the bureaucratic nature of the governance by means of them or to highlight their novelty with respect to past forms of organization. After examining their claims, this article observes how there are few doubts that some general features of bureaucracy such as extension and intensification of control and inhibition of human discretion are accentuated by algorithms. On a more concrete level, on the other hand, conflicting theses and observations have been formulated on whether algorithmization and bureaucratization bring about similar or different effects on fields and organizations. Thus, it is concluded that more empirical research is needed to settle the issue of whether a new ideal-type of algorithmic power can be built and how it would look like.

Keywords

Algorithm, Bureaucracy, Organization, Rationalization, Technology

* CHIARA VISENTIN è iscritta al corso di laurea magistrale in Sociologia e ricerca sociale dell'Università di Pisa, ha lavorato come Knowledge Engineer presso Amazon Alexa, ed è stata allieva della Scuola Normale Superiore di Pisa, dove ha frequentato la Classe di Lettere e Filosofia.

Email: chiara.visentin@sns.it

1. INTRODUZIONE

Algoritmo e burocrazia a prima vista hanno poco in comune. Secondo il Dizionario online Treccani (2018) “algoritmo” nella sua accezione più generale significa «qualunque schema o procedimento matematico di calcolo; più precisamente, un procedimento di calcolo esplicito e descrivibile con un numero finito di regole che conduce al risultato dopo un numero finito di operazioni, cioè di applicazioni delle regole» (algoritmo 2018). “Burocrazia”, invece, designa «Il complesso dei funzionari che, articolati in vari gradi gerarchici, svolgono nello stato le funzioni della pubblica amministrazione» o in astratto il potere assunto da essi, oppure una «Osservanza esagerata dei regolamenti, spec. nella forma esteriore» (burocrazia 2018). Dunque, da un lato il nome di un tipo formale di istruzioni per arrivare a un risultato tendenzialmente nell’area del calcolo, dall’altro un gruppo di esseri umani con alcune caratteristiche organizzative, sociali, politiche e culturali specifiche.

Tuttavia, già riflettendo su queste nozioni di base, si intuiscono alcune aree di sovrapposizione. Entrambi, ad esempio, sono tipi di tecniche, entrambi mirano a definire e portare a termine procedure. In quanto tali, hanno alcune caratteristiche in comune, come il fatto che richiedono l’esplicitazione dei fattori coinvolti, e che la loro essenza è inseparabile dalle regole che vi vengono applicate, anzi si potrebbe dire che l’applicazione di regole sia proprio la loro ragion d’essere. Resta da riflettere su se e perché approfondire il contenuto e il significato di tali aree di sovrapposizione possa aiutare il ricercatore a gettar luce su problemi o aspetti rilevanti per la comprensione del presente.

Il rapporto tra tecnica e istituzioni sociali è stato oggetto di numerose riflessioni sociologiche e antropologiche. Classica è ormai, ad esempio, la linea di pensiero che si dipana attraverso il pensiero (tra gli altri) dell’antropologo Arnold Gehlen (1983, 1984; Berger, Kellner, 1965) e approda nella nota sintesi di Peter Berger e Thomas Luckmann *La realtà come costruzione sociale* (1991), secondo cui la tecnica è radicata profondamente e inestricabilmente nella costituzione fisiologica, psichica e sociale dell’uomo, e dunque assai meno autonoma e puramente razionale di quanto si tenda a immaginarla e assumerla. Inoltre, istituzionalizzazione, abitudine e tecnica procedono insieme verso l’imposizione di struttura - di una struttura intersoggettiva - a un mondo altrimenti troppo “aperto” per via della mancanza, nell’uomo, di istinti imperativi come quelli degli altri animali.

Fin dal sorgere della moderna società industriale la burocrazia è stata un tema importante non solo nello studio di sociologi e scienziati

politici, ma anche nell'immaginario artistico che ne ha dipinto a tinte per lo più fosche l'incombente potere impersonale e opaco che opprime e schiaccia la libertà e la vitalità individuale. La burocrazia è un'istituzione consolidata assunta talora, sulla scia dell'opera di Max Weber, a simbolo ed elemento essenziale della modernità stessa (Kallinikos, 2004).

Più recentemente, l'irruzione delle tecnologie dell'informazione in sempre più vaste porzioni del mondo del lavoro, dell'amministrazione, ma anche della cultura fin nelle sue componenti più intime ha suscitato e sta suscitando entusiasmi e timori che spaziano dall'accademia alle arti e all'intrattenimento. I timori, soprattutto, talvolta si esprimono riprendendo *topoi* distopici già applicati alla burocrazia intorno alla prima metà del secolo scorso. Quel che è certo è che nello svolgere un grandissimo numero di attività individuali e sociali è oggi sempre più impossibile non interfacciarsi con le tecnologie digitali e quindi, in un modo o nell'altro, essere soggetti all'applicazione delle loro regole, le regole dell'algoritmo. Regole nuove, di un tipo nuovo, stanno entrando a far parte delle nostre vite e delle nostre società, e non è chiaro con che effetti e che implicazioni. Alcuni studiosi, partendo da osservazioni come questa, ritengono che almeno alcuni complessi algoritmici, i più pervasivi, possano e debbano essere pensati come istituzioni, capaci di plasmare le pratiche organizzative dei gruppi e degli enti nel loro raggio d'influenza, oltre, naturalmente, all'agire e al pensare degli individui (Ananny, 2016; Caplan, boyd, 2018; Napoli, 2014).

Dunque, algoritmo e burocrazia non solo sono due tipologie di tecniche aventi come carattere essenziale l'applicazione di regole, ma occupano anche entrambi un posto di grande rilievo nel panorama istituzionale delle società in cui viviamo. Sia le regole di certe burocrazie sia le regole di certi algoritmi ci governano, sono dotate di potere, in sensi da indagare e chiarire. Ci si chiede allora in che misura questi due tipi di insiemi di regole siano eterogenei, in che modo si configuri la loro coesistenza, se parallelamente senza toccarsi impongano le loro regole su ambiti della vita e della società distinti esercitando così forme di potere del tutto diverse, oppure se, toccandosi in qualche punto, si integrino o entrino in competizione e conflitto. Ciascuno di questi scenari, a sua volta, aprirebbe una miriade di domande sul significato socio-politico della relazione tra i due meccanismi.

L'idea di pensare l'algoritmo servendosi di nozioni sviluppatesi intorno allo studio della burocrazia, e di tracciare tra i due fenomeni analogie, linee di continuità o linee di rottura è stata proposta

recentemente in molti studi e interventi (Berry, 2014; Birkbak, Carlsen, 2016; Caplan, boyd, 2018; Clair, 2017; Muellerleile, Robertson, 2018; Peeters, Schuilenburg, 2018; Powell, 2016). Anche prima che il termine 'algoritmo' diventasse una parola chiave imprescindibile per parlare dell'espandersi della sfera d'azione delle tecnologie informatiche nella società, la riflessione sul rapporto tra queste ultime e la burocrazia non era mancata (Aneesh, 2009; Angell, Samonas, 2009; Beniger, 1995; Bovens, Zouridis, 2002; Zuurmond, 1994, 1998).

Nel seguito di questo articolo le principali posizioni espresse in queste ricerche verranno esaminate. Si tenterà poi di ricavare un insieme il più possibile coerente dei tratti che secondo i vari studiosi che hanno toccato la questione caratterizzano la relazione tra algoritmo e burocrazia. Sulla base delle contraddizioni che sfuggiranno a questo tentativo sarà possibile individuare i punti condivisi e le questioni controverse su cui un'ulteriore ricerca potrà essere impostata.

2. NUOVI MODI DI PENSARE L'ALGORITMO

Recentemente la figura dell'algoritmo ha assunto un rilievo notevole al di fuori del campo strettamente scientifico/tecnologico, ed è stata rivestita di significati altri rispetto a quello ivi canonico di «finite set of rules which gives a sequence of operations for solving a specific type of problem» (Knuth, 1997, 27). Alla base di questo sconfinamento giacciono esigenze critiche, da parte tanto del discorso pubblico quanto di accademici che mirano, nella prospettiva delle scienze umane e sociali, a mettere in luce il sempre maggior potere di cui sono investiti applicativi e piattaforme digitali operanti sulla base di codice e algoritmi (Beer, 2017), denunciarne i rischi (Danaher, 2014), indagare su come questo abbia conseguenze in sfere diverse dalla tecnologia come ad esempio la politica (Woolley, Howard, 2016), la cultura (Gilbert, 2018), le professioni (Napoli, 2014).

2.1. *Demistificare l'algoritmo*

Una strategia che alcuni di loro hanno adottato a tal fine è quella che qui abbiamo scelto di etichettare con il termine demistificazione, nel senso di «Evidenziazione dell'aspetto reale di un comportamento, di un'ideologia ecc., che viene così sfrondata dei suoi tratti esteriori, spesso ingannevoli» (demistificazione 2018), «sottoporre a critica radicale un'idea, un fenomeno, un personaggio in cui altri credono incondizionatamente e in modo irrazionale» (demistificare 2018).

Esistono, sostengono questi studiosi, concezioni distorte dell'algoritmo, in cui aspetti importanti sono occultati. La mancata considerazione di questi aspetti conduce a forme di eccessiva fiducia nelle operazioni degli algoritmi, che inibisce l'autonomia dell'individuo e della società con conseguenze negative.

L'assunto comune è che l'algoritmo ha subito una "mistificazione", l'occultamento di certi caratteri o effetti e l'esaltazione di altri, almeno potenzialmente a vantaggio di determinati interessi a detrimento di altri. Ad esempio, vari autori hanno osservato come talora il destinatario di operazioni algoritmiche tenda ad attribuirvi fideisticamente caratteristiche di perfetta e infallibile razionalità, che suscitano sentimenti e modi di pensare analoghi a un'esperienza religiosa (Ames, 2018; Bogost, 2015; Finn, 2018)¹. Un altro esempio sono le denunce dell'*ex Design Ethicist* di Google e co-fondatore del Center for Humane Technology Tristan Harris, che ha spiegato come gli applicativi che quasi tutti usiamo per fare sempre più cose per sempre più tempo ritenendoli utili, divertenti e comunque innocui sono in realtà progettati al preciso fine di massimizzare il profitto delle compagnie che li rilasciano, deliberatamente sacrificando il benessere e il sano sviluppo psicologico e cognitivo dell'utente (Harris, 2017).

L'accostamento in chiave analogica dell'algoritmo alla burocrazia è in questa prospettiva promettente, in quanto la burocrazia è stata in passato oggetto di numerose analisi critiche che vi hanno visto di più o di meno di quello che essa dichiarava di essere e che il (supposto) senso comune la riteneva essere. Vedremo di seguito quali tra esse sono state sfruttate ai fini della critica dell'algoritmo, e in che modo.

Herbert Marcuse (1941) in *Some social implications of modern technology*, sostenendo uno stretto apparentamento tra burocrazia e tecnologia moderne, afferma che un meccanismo risultante dall'"apparato"² ch'esse compongono fa sì che «The objective and impersonal character of technological rationality bestows upon the bureaucratic groups the universal dignity of reason» (1998, 57). Inoltre, poiché talora si parla di burocrazie private, viene in questo modo trasmessa l'idea illusoria dell'armonia tra gli interessi, inevitabilmente particolari, che le guidano, e l'interesse comune³.

¹ Simili osservazioni sono state fatte anche in stadi precedenti dello sviluppo e della diffusione delle tecnologie informatiche (per esempio Parnas (1993)).

² Una simile interconnessione organica tra tecnologia e industria burocratizzata, includente anche le scienze naturali, è presente nell'analisi della modernità di Gehlen (1984, 58).

³ Merita di essere menzionato che nello stesso articolo Marcuse riconosce anche i risvolti progressivi della burocratizzazione: «The social character of bureaucratization is

Ad applicare questo ragionamento anche a quelli ch'egli chiama «code-objects» o «computal objects» è David Berry (2014, 35), in un volume che mira a proporre un approccio al digitale nello spirito della teoria critica. A tal fine si occupa di individuare la particolarità, e quindi la limitatezza⁴, della razionalità sottesa al computazionale, e tra le vie che segue a tal fine vi è quella dell'accostamento con la burocrazia, portatrice della ragione strumentale denunciata dai francofortesi. La tecnologia serve a darci un senso di controllo attraverso l'elencazione e la classificazione, operazioni analoghe a quelle svolte dalla burocrazia (Ivi, 199)⁵. Analogamente alla burocrazia, pretende – in virtù della sua impersonale oggettività – di incarnare una ragione universale, assimilata, senza rendersene conto, dagli utenti, in quanto si trova incorporata nel codice e nelle interfacce; e resa così una seconda natura (Ivi, 38). Inoltre, oltre a trattarsi di un caso di ragione strumentale e strumentalizzata, si tratta anche di una ragione privatizzata, visto il carattere proprietario della produzione degli oggetti e dei servizi che introducono il computazionale nelle nostre vite (*Ibidem*).

Jürgen Habermas, in *Teoria dell'agire comunicativo* (1986), interpretando entro il suo modello esplicativo la teoria weberiana della burocratizzazione, intende quest'ultima come l'istituzionalizzazione di un *medium* di regolazione e controllo che, formalizzando un ambito, lo separa dal mondo della vita con l'integrazione sociale immediata e spontanea che lo presiede, trasferendolo nel raggio d'azione di un'integrazione sistemica inerentemente mediata. Una simile mediatizzazione è necessaria alla gestione dei sistemi sociali complessi tipici del mondo moderno. In quanto tale non è di per sé patologica, ma può diventarlo, rovesciandosi in una «colonializzazione» (Ivi, 973), quando «La monetarizzazione e la burocratizzazione [...] strumentalizzano gli apporti dal mondo vitale strutturati in modo autonomo» (Ivi, 979), cioè quelli inerenti alla riproduzione simbolica del mondo vitale. Questa,

largely determined by the extent to which it allows for a democratization of functions that tends to close the gap between the governing bureaucracy and the governed population» (1998, 58).

⁴ Nel senso di Foucault: «the critique of what we are is at one and the same time the historical analysis of the limits that are imposed on us and an experiment with the possibility of going beyond them» (1984, 49).

⁵ Si veda anche Vismann e Winthrop-Young (2008). La linea di pensiero che permette di screditare la tecnologia accostandola alla burocrazia si chiarisce nell'ottica proposta da Berger *et al.* (1974), che analizzano da un punto di vista fenomenologico la differenza tra i due campi concludendo che mentre la formalizzazione della tecnologia è finalizzata a qualcosa di altro (l'efficienza), quella della burocrazia è idealmente tassonomia fine a sé stessa. Un analogo argomento è presente nella critica alla tecnica di Heidegger (Dreyus, Spinosa 1997, 161).

infatti, deve avvenire in modo comunicativo, orientato all'intesa e imprescindibilmente immerso in «una tradizione culturale in tutta la sua ampiezza» (Ivi, 985).

Andrew Gilbert coglie in questa visione una chiave per l'interpretazione dei processi di «computerizzazione della cultura» (2018, 87). Come la burocrazia e il mercato, anche l'algoritmo è un sistema non discorsivo, che formalizza un ambito e lo separa dal mondo della vita e dell'agire comunicativo: «do our speaking for us» (Ivi, 89), riducendo la fatica di raggiungere intese su tutto – e in questo senso è necessario – ma al costo, e al rischio, di comprimere, fino a comprimerlo troppo e in modo patologico, il mondo del senso soggettivamente compreso e interiorizzato nella sua interezza implicita e immediata, della libertà, della «partecipazione alla riproduzione di significati culturali condivisi» (*Ibidem*) e, con ciò, anche le basi stesse di una sfera pubblica democratica.

Le interpretazioni sopra descritte puntano l'attenzione sull'eccesso di potenza tecnico-razionale della burocratizzazione: lo straripare della sua razionalità strumentale in campi che possono venirne danneggiati, la percezione diffusa che la sua logica sia una logica universale e l'(ingiusto) assoggettamento delle altre logiche che ne deriva. Altre interpretazioni della burocrazia che hanno trovato echi nella critica dell'algoritmo ne enfatizzano invece il difetto: si tratta di un insieme di pratiche molto meno efficiente di quanto si sia inclini a pensare.

Una simile lettura è alla base del celebre lavoro *Il fenomeno burocratico* (1969) di Michel Crozier che, opponendosi esplicitamente alla tesi weberiana della imbattibile superiorità tecnica del metodo burocratico di organizzazione, imposta la sua analisi a partire dalla concezione popolare e tendenzialmente peggiorativa della burocrazia legata a inutili e frustranti formalismi, e giunge a una diagnosi degli interessi di gruppo conflittuali che vi si esprimono.

È questo il senso di burocrazia che Airi Lampinen (2017) propone di sfruttare per pensare l'interazione con l'algoritmo, per mettere in luce gli sforzi strategici di diversi *stakeholder* nella costruzione e nella ricezione delle tecnologie.

Una simile linea di pensiero è stata in seguito sviluppata dalla sociologia neoistituzionalista, i cui autori hanno messo in luce l'irrazionalità inerente nella stessa struttura formale di organizzazioni sulla carta votate all'efficienza, ma in pratica talora influenzate ben più fortemente da «interorganizational influences, conformity, and the persuasiveness of cultural accounts» (Powell, DiMaggio, 1991, 13). In particolare, John W. Meyer e Brian Rowan (1977) hanno notato come

molte pratiche e procedure vengono adottate non sulla base di considerazioni razionali sulla loro efficienza nel contesto organizzativo di arrivo, ma perché esse assumono uno statuto mitico nel contesto sociale più ampio in cui l'organizzazione opera essendovi in qualche modo istituzionalizzati. L'adeguamento a esse è perciò fonte di legittimità e il non adeguamento è segno di devianza ed espone ai rischi conseguenti (fallimento, ostilità esterna e defezioni interne). Sulla stessa linea, Paul J. DiMaggio e Walter W. Powell (1983) hanno osservato i processi di omogeneizzazione che si innescano quando un insieme di organizzazioni si struttura in un campo legato da interdipendenze e contatti, che hanno chiamato "isomorfismo istituzionale": le organizzazioni cambiano non per funzionare meglio, ma per essere più simili, perché sono costrette, o per imitazione, o perché indotte da pressioni normative.

Robyn Caplan e danah boyd (2018) costruiscono l'accostamento tra algoritmo e burocrazia per poter applicare questi concetti allo studio dell'influenza di algoritmi sui vari campi e settori produttivi e intellettuali, sviluppando in prima persona il caso del giornalismo. Secondo loro, essendo ormai acquisito che l'algoritmo rappresenta una potente fonte razionalizzante nel mondo contemporaneo⁶, è opportuno guardare a metodi di razionalizzazione del passato tanto per capire come impostare l'analisi dell'algoritmo quanto per "de-mitizzarlo" (Ivi, 2). L'algoritmo dunque, secondo una tesi già esplorata da Philip Napoli (2014), assumerebbe il ruolo di istituzione, imponendosi quindi alle organizzazioni come fattore di legittimità talora dal carattere prettamente o prevalentemente rituale, col risultato di spingerle verso una sempre maggiore omogeneizzazione nell'adozione delle pratiche e logiche da esso dettate e implicate. La burocratizzazione non è che un processo simile, in cui il modo burocratico di organizzare e valutare, come oggi il modo algoritmico, diviene istituzionalizzato in un numero sempre maggiore di settori e conduce le organizzazioni che operano in questi settori ad adottarlo pena l'esclusione.

Un'analoga visione dell'algoritmo come istituzione traspare dalla trattazione di Andreas Birkbak e Hjalmar B. Carlsen (2016), che nell'ottica del pragmatismo americano assimilano l'algoritmo alla figura del pubblico ufficiale, deputato a fornire un ordinamento cognitivo universalmente accessibile, pragmaticamente essenziale per il pubblico, tramite l'emanazione di «coarse signs»⁷.

⁶ Come nota anche Pasquale (2015).

⁷ Letteralmente, segnali grossolani, grazie a cui i membri di una comunità, in generale ignoranti e con poco tempo a disposizione, possono avere accesso a questioni altrimenti

Merita infine di essere menzionato un altro spunto interessante di accostamento dell'algoritmo alla burocrazia in chiave demistificante presente in un lavoro meno recente, il celebre *Computer power and human reason: from judgment to calculation* del computer scientist Joseph Weizenbaum (1976). Il noto informatico si interrogava sui limiti dell'"intelligenza" codificata algebricamente e, in un capitolo eloquentemente intitolato *Incomprehensible programs*, affermava: «Program formulation is [...] rather more like the creation of a bureaucracy than like the construction of a machine» (Ibid., 234). Intendeva dire che, come anche Marvin L. Minsky (1967) aveva osservato in un passo che cita lungamente, la costruzione di un articolato programma non è una concatenazione di azioni e reazioni perfettamente coerente e prevedibile, ma assomiglia di più a una rete di corti giuridiche (le "sub-routine"), che valutano informazioni e le trasmettono ad altre corti, includendo nei loro "verdicti" una serie di decisioni su quali sotto-unità hanno "giurisdizione" sui risultati intermedi da manipolare. Spessissimo si tratta di processi basati non su teorie rigorose, ma su *rules of thumb*, stratagemmi che sembrano funzionare e altri meccanismi ad hoc aggiunti man mano, e il programmatore stesso non conosce con esattezza il percorso dell'informazione attraverso il programma, né i suoi risultati.

2.2. La perfezione della burocrazia

Nella panoramica sopra delineata c'è, per così dire, un grande assente, ovvero Max Weber, la cui teoria della burocrazia rappresenta una pietra miliare imprescindibile per pensare questa forma organizzativa. Non ch'egli non sia mai nominato, ma la sua concezione della burocrazia è vista più come un punto di partenza da superare o come una cornice descrittiva contestuale per arrivare a inquadrare il punto, che come l'ossatura stessa su cui sviluppare l'accostamento tra algoritmo e burocrazia.

Quest'ultima direzione è invece stata percorsa in due recentissimi studi che sostengono tesi di sostanziale continuità tra le logiche burocratiche e quelle algoritmiche, elaborati l'uno da Christopher Muellerleile e Susan Robertson (2018) e l'altro da Rik Peeters e Marc Schuilenburg (2018), entrambi entro un ambito legato all'applicazione dell'algoritmo alla *governance* e all'amministrazione della giustizia.

Il primo studio individua nella teoria weberiana della burocrazia

troppo complesse o confuse. Concetto introdotto dal celebre scrittore e giornalista statunitense Walter Lippmann nel 1925 (1993, 54).

caratteristiche chiave che oggi formano l'essenza di una burocrazia digitale meno visibile ma non meno potente. Uno dei punti della teoria di Weber (1968) che i due autori sollevano a tal fine è il fatto che il metodo amministrativo di tipo burocratico è «capace di universale applicazione» (Ivi, 214), non essendo intrinseco a nessuno scopo particolare (Muellerleile, Robertson, 2018, 192), ma essendo piuttosto «uno strumento di precisione che può porsi a disposizione di interessi di potere molto diversi, sia puramente politici che puramente economici, come di qualsiasi altro tipo» (Weber, 1968, 292). Un altro è il fatto che la burocrazia poggia sul, e al tempo stesso serve al, controllo su informazione e comunicazione – «un potere esercitato in virtù del sapere» (Ivi, 219).

Emergono in ultima analisi nella teoria weberiana, secondo l'interpretazione dei due autori, tre caratteristiche chiave dell'ordinamento burocratico.

1. Efficienza: «Il fondamento decisivo per il procedere dell'organizzazione burocratica è [...] la sua superiorità puramente tecnica su ogni altra forma. Un meccanismo burocratico pienamente sviluppato è rispetto ad esso nello stesso rapporto in cui si trova una macchina nei confronti dei mezzi non meccanici di produzione dei beni. Nell'amministrazione burocratica [...] la precisione, la rapidità, l'univocità, la pubblicità degli atti, la continuità, la discrezione, la coesione, la rigida subordinazione, la riduzione dei contrasti, le spese oggettive e personali sono recati alla misura migliore» (Ivi, 276-277).

2. Oggettività: «[...] l'adempimento «oggettivo» significa in primo luogo un adempimento «senza riguardo alla persona», in base a regole prevedibili. «Senza riguardo alla persona» è però anche la parola d'ordine del «mercato» [...]. La sua [della burocrazia] specifica caratteristica, gradita al capitalismo, ne promuove lo sviluppo in modo tanto più perfetto quanto più essa si «disumanizza» – e ciò vuol dire che consegue la sua struttura propria, ad essa attribuita come virtù, che comporta la esclusione dell'amore e dell'odio, di tutti gli elementi affettivi puramente personali, in genere irrazionali e non calcolabili, nell'adempimento degli affari d'ufficio (Ivi, 278)⁸.

3. Razionalità: «Formal rationality is the depersonalized, even dehumanized, set of codified rules, laws, and regulations. This sort of rationality is based on scientific and economic calculation, is historically predominant in capitalistic societies, and is most closely associated with

⁸ Alcune delle citazioni riportate dall'opera di Weber sono suggerite da Muellerleile e Robertson, ma la maggior parte di esse sono state reperite dall'autrice di questo articolo seguendo il senso dell'argomentazione dei due studiosi.

bureaucratic order» (Muellerleile, Robertson, 2018, 195)

Questi principi rappresentano la logica organizzativa della burocrazia, che può rimanere immutata anche al variare delle sue manifestazioni visibili. Per mostrare che questo è il caso, gli autori evidenziano come tali principi siano saldamente incorporati nella *governance* digitale sempre più diffusa. Innanzitutto, per quanto riguarda l'efficienza, essa è alla base di gran parte delle giustificazioni offerte per l'espansione delle nuove tecnologie, oltre che chiaramente discernibile nello spirito di progetti come le «data driven smart city» (Ivi, 201). Inoltre, l'automazione dei processi di raccolta, elaborazione e presentazione dei dati accelera esponenzialmente i processi rendendoli più economici grazie alla rimozione o la facilitazione della decisione umana. Per quanto riguarda poi l'oggettività, essa funge da principio di legittimità per le nuove burocrazie digitali sotto forma di «*apparent data and algorithmic neutrality*» (Ivi, 202), di quantificazione pervasiva, e della sua «affinità elettiva» con la società di mercato. Infine, per quanto riguarda la razionalità «tecno-scientifica» tipica della burocrazia, formalizzazione ed astrazione hanno trovato nel codice informatico un nuovo e più potente supporto, eliminando le barriere spazio-temporali che limitavano «l'intensità e l'estensione della prestazione» (Weber, 1968, 217). Da tutto questo risulta che «*while the conventional bureaucratic structure appears to be disintegrated and dispersed, the social order engendered by information technologies transports the logic of bureaucracy more deeply into society*» (Muellerleile, Robertson, 2018, 209), come dimostra in modo emblematico il caso del progetto pilota di credito sociale del governo cinese.

Il tema della burocrazia weberiana come «potere esercitato in virtù del sapere» (Weber, 1968, 219) è centrale anche nel secondo degli studi nominati all'inizio di questo paragrafo (Peeters, Schuilenburg, 2018) e lo è anche l'immagine della «gabbia di [...] acciaio» (Weber, 2014, 240) come sistema che disciplina gli uomini come ingranaggi di una macchina espungendo la riflessione sui valori. La burocrazia, secondo gli autori, incarna la razionalità strumentale, in quanto segue regole coerenti prestabilite, chiarendo in modo univoco procedure, fini, mezzi, misure del successo, tempi e aspettative – il che non mira primariamente, come ha mostrato György Gajduschek (2004), a perseguire l'efficienza nel senso contemporaneo, ma a un obiettivo «informativale» di riduzione dell'incertezza.

Sulla base di studi precedenti sull'applicazione di tecnologie informatiche in contesti burocratici, i due studiosi concludono che «*Information and communication technology is not just a tool that*

bureaucracies can use, but is, in itself, fundamentally bureaucratic» (Peeters, Schuilenburg, 2018, 270). Come ha notato Jannis Kallinikos (2005), infatti, si tratta di meccanismi di semplificazione, la scomposizione di un compito o problema in insiemi di operazioni da svolgere sequenzialmente, e chiusura, l'isolamento delle singole operazioni per proteggerne l'esecuzione da interferenze esterne. Il risultato, come notato da Helen Margetts (1999), è una ulteriore centralizzazione e razionalizzazione dei processi di decisione a opera delle nuove tecnologie, realizzata, ad esempio, tramite la riduzione dello spazio di autonomia e discrezione del personale osservata da Mark Bovens e Stavros Zouridis (2002).

I casi su cui gli autori si soffermano per supportare queste tesi sono le forme di *predictive policing* e di *predictive justice*. L'introduzione di algoritmi in campi quali la sorveglianza e la giustizia tende a renderli «more machine-like»⁹ (Peeters, Schuilenburg, 2018, 276), stimolando processi di razionalizzazione, formalizzazione e standardizzazione nel trattamento dei casi, sottraendo valutazioni al giudizio del professionista per renderle l'esito di sequenze di operazioni automatizzate, e introducendo così procedure di tipo burocratico in pratiche che precedentemente ne erano libere. Inoltre, su un piano più astratto, gli autori suggeriscono che i nuovi metodi algoritmici riducano la rilevanza dello status dell'individuo in favore delle sue caratteristiche oggettive e misurabili, anche questa una caratteristica della burocrazia. Dunque, è legittimo affermare che l'algoritmo (in questi campi) è burocratico, perché struttura il comportamento e la decisione in modo simile all'organizzazione burocratica idealtipica se non marcandone ancora di più i tratti caratterizzanti di standardizzazione, formalizzazione e razionalizzazione.

Prima che, per usare le parole di Gilbert, il termine algoritmo diventasse la «buzzword» (2018, 87) per indicare in senso lato la computerizzazione, anche Ian O. Angell e Spyridon Samonas hanno sostenuto che «computers introduce bureaucratic mechanisms into

⁹ L'assimilazione della burocrazia alla macchina ha una storia molto lunga e radicata, come testimonia il celeberrimo passo in cui Weber afferma che «spirito rappreso è pure quella macchina vivente che è costituita dall'organizzazione burocratica, con la sua specializzazione del lavoro tecnico, la sua delimitazione delle competenze, i suoi regolamenti e i suoi rapporti di obbedienza ordinati gerarchicamente. In unione con la macchina inanimata, essa è all'opera per preparare la struttura di quella servitù futura» (Weber, 1968, 703); e, per citare solo un altro tra i punti salienti nella storia di questa associazione, Gareth Morgan nel suo classico studio delle organizzazioni attraverso metafore scrive «Organizations that are designed and operated as if they were machines are now usually called bureaucracies» (2006, 13).

organisations» (2009, 4) e che «The algorithm is itself a bureaucracy» (Ivi, 12), ritenendo, in un'ottica analoga a quella di Peeters e Schuilenburg (2018), che la burocrazia vada intesa anche come un certo tipo di regole e norme comportamentali. Il tratto principale per cui queste si ritrovano nell'algoritmo è l'inibizione della discrezione, soprattutto tramite le categorizzazioni, tipizzazioni e istruzioni preventivamente imposte sulle situazioni da gestire, e tramite l'intensificazione del controllo sull'operato dell'individuo. Ne deriva una perdita di flessibilità e di capacità di adattamento all'imprevisto e all'extra-ordinario.

Il tema dell'inibizione della discrezione, nelle burocrazie come negli algoritmi e in generale nell'automazione, è vasto e molto ricco di implicazioni e meriterebbe una trattazione a sé stante. Qui ci si limiterà a menzionare che il processo ha suscitato, a fianco di critiche (Carr, 2015), anche apprezzamenti per via delle sue potenzialità: dal punto di vista tecnico, per via del miglioramento della qualità di decisioni prima spesso prese d'istinto e quasi a caso da professionisti con pochi dati a disposizione (Mayer-Schönberger, Cukier, 2013), e dal punto di vista morale, per via della possibilità di eliminare i *bias* dei decisori umani, che tendono generalmente nella direzione di sfavorire ulteriormente gruppi sociali già svantaggiati (Zarsky, 2012, 35).

2.3. Nuovi idealtipi

Di contro a queste interpretazioni che enfatizzano la continuità, ne sono state formulate altre secondo cui l'accento va posto sulla rottura che l'algoritmo introduce rispetto al modello della burocrazia. Due studiosi che hanno sviluppato tesi del genere, A. Aneesh (2009) e Arre Zuurmond (1994, 1998), si sono posti esplicitamente la questione dell'appropriatezza di parlare di "iper-burocratizzazione" per i processi di digitalizzazione che investono il mondo delle organizzazioni e del lavoro, arrivando a una risposta negativa.

Zuurmond (1998) studia questi fenomeni in una fase che se confrontata con il presente si può definire incipiente. Egli nota che le organizzazioni di tipo burocratico investite dalla digitalizzazione subiscono una serie di mutamenti che le allontanano dall'idealtipo della burocrazia fino a rendere quest'ultimo sterile ai fini della loro comprensione: parallelizzazione dei processi, automazione del coordinamento di routine e in generale di gran parte del controllo verticale, perdita del monopolio informativo da parte dello *street-level*

*bureaucrat*¹⁰, *deskilling* e trasferimento della conoscenza in *expert systems*, strutture più piatte, più organiche e meno meccaniche. Zuurmond propone di introdurre un nuovo idealtipo, chiamato “infocrazia”, per cogliere le peculiarità di questa nuova tecnica di controllo, basata su strumenti – le tecnologie informatiche – nuovi e altri rispetto alla struttura burocratica. Ciò non significa, è inteso, che il controllo si sia affievolito e la razionalizzazione si sia arrestata, anzi essi sono diventati più pervasivi: «We may have left the iron cage [...] just to switch to a virtual fortress» (Ivi, 271).

“Algocrazia” è invece il termine scelto da Aneesh (2009) per il sistema di *governance* da lui individuato, da considerarsi un terzo modello a fianco degli altri due, la burocrazia e il mercato¹¹. Ciascuno di essi è identificato da uno specifico «ruling principle» (Ivi, 347): il legale-razionale per la burocrazia, il sistema dei prezzi per il mercato, e il programma o l'algoritmo per l'algocrazia.

L'algocrazia non va confusa con una forma di “iper-burocrazia”, un errore che Aneesh stesso dichiara di aver commesso in passato (Aneesh, 1999). I sistemi algocratici, infatti, a differenza di quelli burocratici, strutturano il campo delle azioni possibili senza bisogno che gli agenti interiorizzano il rispetto per regole e leggi, né vi siano indotti dalla cognizione di punizioni: la loro azione è controllata dando forma all'ambiente in cui si svolge, e facendo in modo che siano presenti solo alternative programmate.

La realtà concreta che Aneesh osserva per giungere a questa conclusione sono le operazioni di sviluppo di software distribuite su scala globale: in questi processi la produzione di un servizio si svolge in parti del mondo lontanissime tra loro e in contesti organizzativi sostanzialmente l'uno all'altro estranei, ed è coordinata essenzialmente

¹⁰ La questione dell'impatto delle tecnologie informatiche sulla cosiddetta *street-level bureaucracy* è stata approfondita da Mark Bovens e Stavros Zouridis, che nella loro analisi tendono a trasformarsi dapprima in *screen-level bureaucracies*, in cui molte decisioni sono sottratte all'ufficiale che si prende cura del caso personalmente e programmate nel software, e infine in *system-level bureaucracies*, in cui tutte le transazioni sono automatizzate, connesse e controllate attraverso un sistema informatico unico (2002).

¹¹ Già il porre questa distinzione rivela che Aneesh adotta una definizione di burocrazia più ristretta rispetto agli studi che abbiamo visto nei paragrafi 2.1 e 2.2, che vedevano in essa la forma organizzativa caratteristica della modernità tanto nella sfera pubblica quanto in quella privata, apparentata strettamente alle logiche del mercato più che alternativa ad esse. Queste diverse interpretazioni in un certo senso riflettono quella che Gajduschek (2003) ha identificato come una dicotomia che percorre gli studi sulla burocrazia: da un lato c'è chi la intende come un fenomeno organizzativo (principalmente nel campo della teoria dell'organizzazione e del *management*), dall'altro chi la intende come un ambito al di fuori del mercato e legato al governo (principalmente economisti neoclassici e teorici della pubblica amministrazione).

attraverso strumenti di comunicazione e controllo informatici. In un simile caso mancano i presupposti imprescindibili per l'esistenza di un coordinamento burocratico o legale-razionale, in quanto mancano una supervisione burocratica transnazionale legalmente vincolante, l'interazione di persona e una singola cultura organizzazionale su cui fare affidamento. Né ha senso dire che l'algocrazia è una «software version of bureaucracy» (Ivi, 355), perché non esiste un meta-linguaggio comune tra il codice legale che presiede alla *governance* burocratica e il codice binario che presiede a quella algocratica.

Il termine “algocrazia” è stato più recentemente recuperato da John Danaher in un articolo che ne denuncia i rischi legati a un'eccessiva compressione delle possibilità per gli esseri umani di partecipare alla decisioni pubbliche e di comprenderle (2016). Sebbene il pericolo sia simile a pericoli indicati in passato in relazione alla burocrazia (Habermas 1986), per Danaher è chiaro che si tratta di dispositivi di organizzazione e decisione di natura diversa.

3. TRA CONTINUITÀ E DISCONTINUITÀ

Non è facile confrontare tra loro le posizioni di tutti questi autori perché essi si rifanno a diverse interpretazioni di cosa si intende con burocrazia. Possiamo tuttavia provare a schematizzare i tratti da loro individuati e ricavare una cornice concettuale coerente per interrogarci sulla continuità o sulla rottura che un particolare processo di algoritmizzazione rappresenta rispetto al processo di burocratizzazione inteso a un dato livello di astrazione ed entro una data cornice teorica.

A un livello di astrazione piuttosto alto, seguendo un'impostazione che fa capo alla teoria weberiana della burocrazia, è possibile ragionare intorno all'estensione e all'intensificazione del controllo che le forme di amministrazione (nel senso più ampio del termine) algoritmiche, o meglio idealtipi di esse, apportano rispetto all'idealtipo dell'«apparato amministrativo di tipo razionalmente legale» (1968, 214).

Le nozioni di estensione e intensificazione del controllo, presenti anche solo implicitamente e intuitivamente in molti degli studi su algoritmo e burocrazia, sono esplicitamente trattate nella discussione weberiana della burocrazia. Esse sono componenti essenziali della sua superiorità tecnica o maggiore efficienza rispetto agli altri tipi di esercizio del potere. Lo «sviluppo quantitativo» (Ivi, 272) e «l'espansione intensiva e qualitativa [...] dei compiti amministrativi» (Ivi, 275) sono fattori essenziali a monte e, secondo un circolo di feedback positivo, a valle, della loro burocratizzazione, ed è normale che

«l'“intensità” dell'azione statale verso l'esterno – la forza di espansione – e verso l'interno – l'influenza statale sulla cultura – siano in diretta proporzione con il grado di burocratizzazione» (Ivi, 274). Un'osservazione, questa, peraltro in linea con la tesi habermasiana della colonializzazione del mondo della vita a opera della burocrazia (Gilbert, 2018; Habermas, 1986).

L'idea che, in una scala di tipi di organizzazione ed esercizio del potere ordinata secondo estensione e intensificazione (almeno potenziali) crescenti, quello attraverso l'algoritmo si collochi sul gradino superiore rispetto a quello attraverso l'ufficio, ereditandone ed accrescendone le virtù tecniche in quanto a estensione e a intensità, è presente in diverse delle analisi che abbiamo visto. La pensano così tanto sostenitori della tesi della continuità quanto i sostenitori della tesi della discontinuità. Nell'ottica dei primi, si tratta di un segno che algoritmo e burocrazia condividono la stessa natura e sono espressioni delle stesse tendenze pur differendo nel grado di potenza, efficienza ecc. Per i secondi, invece, il potenziamento rappresenta il sintomo di una sorta di salto quantico per cui non è più utile o legittimo parlare dei due oggetti come se fossero in fondo dello stesso tipo. Citando la prefazione di un saggio di David Berlinski (2000) intitolata *The Digital Bureaucrat*, «A digital computer may well do what a bureaucracy has done, but it does it with astonishing speed [...]. This has made all the difference in the world.» (Ivi, xiii)¹².

Queste caratteristiche sono il fulcro dell'interpretazione della relazione tra algoritmo e burocrazia sviluppata negli anni Ottanta da James C. Beniger (1995), secondo cui la burocrazia dev'essere intesa essenzialmente come una tecnologia di controllo, la trasformazione di un'organizzazione in un efficiente processore di informazioni capace di guidare l'azione collettiva verso un obiettivo prestabilito. La sua applicazione diffusa tanto nel settore pubblico quanto in quello privato a partire dalla seconda metà del XIX secolo rappresentò secondo lo studioso il primo atto della Rivoluzione del controllo, che seguì la Rivoluzione industriale adattando le capacità dei sistemi sociali di governare sé stessi e il loro ambiente al passo delle loro enormemente accresciute capacità di processare e muovere materia ed energia. La computerizzazione rappresenta secondo Beniger niente più e niente

¹² Questo esempio è utile a esprimere l'idea di salto quantico per cui l'amministrazione algoritmica accentua una caratteristica di quella burocratica al punto che il fatto che la caratteristica sia la stessa – in questo caso della velocità, essa era stata individuata da Weber come una delle virtù tecniche della burocrazia (1968, 278) – non è più rilevante ai fini della classificazione.

meno che un atto successivo della medesima rivoluzione, che incorpora i processi svolti dalle organizzazioni burocratiche in circuiti che li portano a termine in modo sempre più automatico e sempre meno dipendente dal fattore umano.

La tesi di Beniger, utile nel sottolineare la continuità storica e concettuale tra l'amministrazione burocratica e quella algoritmica (che spesso rappresenta direttamente un'algoritmizzazione della prima, come sottolineato anche da Caplan e Boyd (2018), si basa su caratteristiche che, come abbiamo osservato prima, valgono per la burocrazia e per l'algoritmo a un elevato livello d'astrazione. Sono entrambi tecnologie di controllo, e sono tecnologie di controllo sempre più potenti che rispondono sempre meglio ai medesimi bisogni sociali di accrescimento delle capacità dei gruppi di governare sé e il loro ambiente. Per questo motivo, come diversi recensori hanno notato, un punto debole dell'impostazione di Beniger è la sua eccessiva generalità, che rischia di renderne le implicazioni troppo poco concrete per avere un elevato valore euristico (John, 1988; Porter, 1987).

Sembra quindi che questi due tratti, l'estensione e l'intensificazione del controllo, pur assai interessanti e ricchi di implicazioni, non siano molto utili per decidere della continuità o della discontinuità tra algoritmo e burocrazia, a causa della loro eccessiva generalità, che fa sì che il loro rinvenimento (molto diffuso) sia compatibile sia con l'una che con l'altra tesi.

Sempre su un piano astratto ma di diverso genere, diversi autori hanno indagato la relazione tra algoritmo e burocrazia sulla base della razionalità sottostante. Anche questa impostazione si rifà a Weber, sebbene in alcune analisi, come abbiamo visto, filtrato attraverso le teorie della Scuola di Francoforte. Weber non stabilisce un collegamento chiaro e univoco tra la burocrazia e una determinata forma di razionalità (Brubaker, 1991), ragion per cui è possibile, anche tentando di attenersi alla sua opera, arrivare a conclusioni diverse a questo riguardo.

Alcuni studiosi, infatti, non esitano a vedere nella burocrazia un'espressione della razionalità strumentale, del tutto analoga a quella che presiede il mondo del calcolo algoritmico (Berry, 2014; Gilbert, 2018; Muellerleile, Robertson, 2018). In queste concezioni l'idea del calcolo e della calcolabilità¹³, della modernità come trionfo della

¹³ Che, ricordiamolo, è ben presente nella teoria weberiana della burocrazia: «per la burocrazia moderna, questo [...] elemento – le “regole prevedibili” – ha [...] un'importanza veramente predominante. La caratteristica della civiltà moderna, e specialmente della sua struttura tecno-economica, esige proprio questa “calcolabilità” dell'effetto» (1968, 278).

ragione strumentale sulla ragione materiale e le denunce della perdita di umanità a cui questi processi conducono¹⁴ fanno tutt'uno, e consentono di vedere l'attribuzione sociale di decisioni e potere a una procedura burocratica e a un processo algoritmico come due esempi del trionfo del medesimo tipo di razionalità¹⁵. David Golumbia, nel suo saggio sulla logica culturale della computazione (2009), tenta di fondare filosoficamente questo insieme di intuizioni servendosi delle nozioni di spazio liscio e spazio striato elaborate dai filosofi francesi Gilles Deleuze e Felix Guattari (2003). Lo spazio striato è lo spazio degli ordini burocratici e governamentali dello stato e dell'impresa, simboleggiato da griglie, mappe, coordinate, organizzazioni ad albero (cioè, gerarchiche): rappresenta quel che c'è di comune nelle varie logiche razionalizzanti tipiche della modernità occidentale, logiche che sono sia intellettuali sia sociali (Scaff, 2017), e che talora nel nostro immaginario si sovrappongono al concetto stesso di civiltà o civilizzazione. A questo spazio, piuttosto che allo spazio liscio della spontaneità, appartengono la burocrazia e l'algoritmo, essendo fonti importanti di "striatura"¹⁶.

Su questo punto, oltre a esistere un'intera famiglia di opposizioni implicite a simili tesi – quelle derivanti dal ben più ampio e generale dibattito sulla modernità e sulla radicale novità del principio organizzativo della rete proposta da Castells (2003) – è stata anche formulata da Aneesh (2009) un'esplicita presa di posizione discordante. Secondo il sociologo, come abbiamo visto nel paragrafo precedente, il *ruling principle* legale e quello algoritmico sono radicalmente diversi: tra il linguaggio normativo dell'uno e quello logico-matematico dell'altro non esiste un meta-linguaggio comune, il loro modo di funzionare è diverso perché il primo richiede come componente

¹⁴ Alla weberiana gabbia d'acciaio (Weber, 240) alle varie concettualizzazioni offerte dalla Scuola di Francoforte (Horkheimer 2002; Horkheimer, Adorno 2010, Marcuse, 2013) e dal suo più recente erede Habermas (1988).

¹⁵ Analogamente, ma attraverso una più analitica lettura del testo di Max Weber, Muellerleile e Robertson (2018) costruiscono il concetto di "razionalità tecno-scientifica" di cui vedono l'algoritmo, nelle sue applicazioni amministrative come il credito sociale cinese, come l'ultimo e più sofisticato esponente.

¹⁶ Golumbia nota come questo deve metterci in guardia contro tesi come quelle di Manuel Castells (2003) secondo cui un cambiamento epocale si è verificato con la sostituzione della centralizzazione con la rete come principio logico alla base della società, della cultura e del sapere – saremmo entrati, quindi, entro un processo di razionalizzazione nuovo e diverso rispetto a quello che ha caratterizzato la modernità. Tesi come questa, fa notare Golumbia, implicano un'immagine eccessivamente semplificata, se non del tutto errata, del presente e soprattutto del passato, che ne appiattisce fino a cancellarle le continuità e omologie.

essenziale l'interiorizzazione della regola da parte dell'individuo mentre il secondo modifica l'ambiente esterno all'individuo guidandone così l'azione. Una prova di questa diversità è che l'algoritmo funziona per amministrare tipi di organizzazioni che non sarebbe possibile amministrare burocraticamente.

Il *ruling principle* di cui parla Aneesh non è la razionalità di cui parlano Muellerleile e Robertson (2018) e Barry (2014), tuttavia è interessante osservare come la radice stessa del potere amministrativo, il linguaggio performativo in cui esso si concretizza, siano visti come radicalmente diversi. Anche ammesso che le leggi e i regolamenti che guidano l'azione dell'ufficio e il codice algoritmico siano entrambi espressioni di una razionalità strumentale e tecno-scientifica tipica della modernità, non si può, se si accetta la tesi di Aneesh, ignorare che questi due tipi di controllo si esercitano secondo principi e modalità differenti e non riducibili l'uno all'altro. Si potrebbe dire che la differenza che intercorre tra potere legale razionale esprimendosi nella burocrazia e potere algoritmico esprimendosi nell'algocrazia è analoga a quella che intercorre tra il primo e il potere tradizionale, o quello carismatico.

Non è facile scartare una simile idea, se si considerano seriamente le trasformazioni che subisce il principio di legalità nella *governance* algoritmica. Parte nella nozione weberiana di potere legale-razionale è infatti quella del diritto di chi vi è soggetto ad appellarsi in caso di impropria applicazione delle leggi, e dunque un certo grado di pubblicità di queste (Weber, 1968). La gran parte degli algoritmi che amministrano sempre più sfere del mondo sociale sono invece proprietari, e dunque non accessibili a chi vi è soggetto, che di conseguenza non può, nel suo rapporto con questi poteri, contare sulle stesse garanzie su cui può contare nel rapporto con le tradizionali burocrazie pubbliche. Sebbene anche queste ultime, come d'altronde tutte le istituzioni, presentino un certo grado di opacità, esse in linea di principio non sono mai del tutto scatole nere per gli amministrati, avendo la loro fonte ultima di legittimità in leggi conoscibili da tutti – e, a livello generalissimo e intuitivo, in parte effettivamente conosciute dalla maggior parte di loro. Lo stesso non accade nel caso degli algoritmi, la cui costruzione avviene per la maggior parte in base a criteri tecnici e segreti, i quali possono sfociare in esiti che contraddicono valori morali consolidati, o più subdolamente creano nuovi privilegi o ne rinforzano di vecchi, o sortiscono altri effetti che qualcuno potrebbe giudicare negativi, senza che chi ne è vittima sia dotato degli strumenti cognitivi o legali necessari per appellarsi, e la società civile di quelli per riflettere e agire politicamente al riguardo

(Caplan, boyd, 2018; O'Neil, 2016; Pasquale, 2015).

Su piani più concreti, sono stati individuati nella letteratura recente altri tratti di convergenza e di divergenza tra burocrazia e algoritmo.

Sembra pacifico che il noto principio burocratico “*sine ira et studio*” (Weber, 1968, 220, dove è denominato anche «impersonalità formalistica») sia condotto dall'algoritmo a un grado di sofisticazione ancor più elevato. In moltissimi studi si nota la crescente inibizione della discrezionalità umana con l'automazione dei processi amministrativi e decisionali (Angell, Samonas, 2009; Mayer-Schönberger, Cukier, 2013; Zarsky, 2012). La mancata completa eliminazione di *ira et studio* è denunciata in quanto *bias* (Algorithmic bias 2018; O'Neil 2016), e il senso di scandalo che ne deriva è segno di quanto forte sia l'aspettativa di oggettività, imparzialità e neutralità riposta nell'algoritmo, cardine dei molti dei discorsi a favore dell'automazione algoritmica della *governance* (Muellerleile, Robertson, 2018; Peeters, Schuilenburg, 2018).

Un altro tratto che algoritmizzazione e burocratizzazione sembrano avere in comune è la spinta che entrambi danno alla standardizzazione e alla formalizzazione (Caplan, boyd, 2018; Peeters, Schuilenburg, 2018)¹⁷. Per essere svolto da una burocrazia come da un algoritmo, un compito deve cessare di essere affidato alla conoscenza tacita (Polanyi, Sen, 2009) e idiosincratica di un particolare esecutore, e venire definito in modo esplicito e secondo certe convenzioni tali da renderlo comprensibile ed eseguibile in linea di principio universalmente. Questo implica l'esplicitazione delle sue varie fasi, la suddivisione di passaggi e azioni prima intrecciati o fusi, la sua più netta separazione da elementi o processi estranei (Kallinikos, 2005; Peeters, Schuilenburg, 2018). Un'altra implicazione notevole è che gli esecutori del processo tenderanno a divenire sempre più simili l'uno all'altro (Caplan, boyd, 2018).

Peeters e Schuilenburg (2018), che, come abbiamo visto, studiano la relazione tra algoritmo e burocrazia nel campo della giustizia e della polizia, traggono da queste osservazioni conclusioni sulla convergenza tra i due meccanismi amministrativi: entrambi tendono ad accrescere la centralizzazione delle organizzazioni e il loro meccanicismo. Proprio il contrario notano tuttavia Aneesh (2009) e Zuurmond (1998), secondo i quali con l'introduzione di algoritmi nei processi amministrativi e decisionali le organizzazioni tendono a diventare più piatte, più organiche e meno centralizzate.

¹⁷ Siano esse frutto di adeguamenti istituzionali o di processi di razionalizzazione.

4. VISIONI ASTRATTE E UN *PUZZLE* CONCRETO CON PEZZI ANCORA DA TROVARE

Sul piano più astratto e generale, alla luce di quanto visto, sembra assodato che esistano molte linee di continuità tra la burocrazia e l'algoritmo: i tre punti individuati da Muellerleile e Robertson (2018) aiutano bene, se opportunamente ampliati e specificati, a evidenziarle. L'efficienza, intesa in senso ampio come potenziamento del controllo sia in estensione sia in intensità, accresciuta dalla burocrazia rispetto alle forme di esercizio del potere prevalenti in epoca premoderna, è ulteriormente accresciuta con l'introduzione di modalità algoritmiche di amministrazione dei vari ambiti dell'esistenza. Parimenti accresciuta è l'oggettività, intesa come inibizione della discrezione umana.

Per quanto riguarda la razionalità soggiacente, è difficile arrivare a conclusioni finanche provvisorie, in quanto non è facile in primo luogo accordarsi su come distillare da una famiglia di istituzioni una logica di funzionamento tanto generale da poterla dire la loro razionalità. Chi ne parla in genere ha in mente lunghe tradizioni di pensiero filosofico che non si adattano facilmente a essere smentite o confermate dall'osservazione sociologica, storica o antropologica. Tuttavia, che certe analogie, almeno tra il discorso che parla della burocrazia e con cui la burocrazia parla di sé stessa e quello che si va sviluppando intorno all'algoritmo, esistano, è fuor di dubbio: questi due oggetti idealtipici tendono talvolta a giustificarsi appellandosi agli stessi principi, che possono essere ricondotti alla razionalità strumentale weberiana.

Anche la tesi di Aneesh (2009), se mantenuta solo nel suo significato più astratto, può essere letta come l'evidenziazione di una ulteriore linea di continuità: l'esternalizzazione del controllo operata dall'algoritmo rispetto alla burocrazia non sarebbe in fondo troppo diversa dall'esternalizzazione del controllo operata dalla burocrazia stessa rispetto alle forme di esercizio del potere precedenti. L'interiorizzazione delle norme richiesta dalla tradizione dopotutto è maggiore rispetto a quella richiesta dalla burocrazia, che al contrario, separando l'amministratore dai mezzi dell'amministrazione, pone i presupposti per plasmarne l'ambiente fisico, informazionale e morale e rendere così possibile l'esecuzione delle mansioni anche senza una profonda interiorizzazione di tutte le norme che presiedono all'organizzazione.

Tuttavia, l'eterogeneità dell'esercizio del potere attraverso l'algoritmo non deve essere sottovalutata, se non altro perché il collegamento con la legge in senso tradizionale, così importante nella definizione di

potere legale-razionale, sembra venir meno o diventare meno chiaro nel caso del potere dell'algoritmo.

Il tema dell'assenza di un meta-linguaggio comune tra il codice binario e il codice legale sollevato da Aneesh (2009) a questo proposito è interessante, e invita a riflettere sul fatto che, per quando si possano osservare analogie ideali e continuità storiche tra burocrazia e algoritmo, non si può dare per scontato che questi due tipi di istituzioni siano sempre capaci di assimilarsi a vicenda: le tecniche, i messaggi, le istruzioni (e l'istruzione) pensati per progettare gli uni o per interagire con essi, non saranno necessariamente ottimali per gli altri.

Un fiorire di novità che si intrecciano con le vecchie pratiche e logiche, non facile da osservare dal punto di vista della teoria, è probabilmente racchiuso in ogni caso in cui un algoritmo viene investito di potere su una porzione per quanto piccola della realtà sociale, o fa il suo ingresso in un processo precedentemente monopolio di un'organizzazione burocratica. Non è un caso che le contraddizioni più stridenti nella letteratura si riscontrino proprio in contesti del genere: è stato osservato che l'algoritmo rende meno gerarchiche le relazioni e che le rende più gerarchiche, ch'esso rende le organizzazioni più meccaniche e che le rende più organiche, che induce più centralizzazione e più decentramento, ed è probabile che nessuna di queste affermazioni sia errata.

Solo con più osservazioni a disposizione sarà possibile individuare la differenza specifica del potere algoritmico e la sua entità e rilevanza. Le altre forme di potere non dovranno perciò essere espunte dal quadro: la teoria della burocrazia weberiana non si limita a delimitare e isolare il suo oggetto, ma ne mostra i rapporti essenziali con il potere carismatico e con il potere tradizionale. Analogamente, indagando il potere dell'algoritmo non si potrà fare a meno di osservare i punti in cui esso poggia su altri tipi di potere, oltre a quelli in cui li sostituisce.

RIFERIMENTI BIBLIOGRAFICI

- Algorithmic bias. (2018). In *Wikipedia, the free encyclopedia*.
Algoritmo. (2018). In *Vocabolario Online Treccani*.
AMES, M. G. (2018). Deconstructing the algorithmic sublime. *Big Data & Society*, 5(1), First Published.
AMOORE, L., PIOTUKH, V. (2016) (ed.). *Algorithmic life: calculative devices in the age of big data*. London-New York: Routledge-Taylor & Francis Group.
ANANNY, M. (2016). *Toward an Ethics of Algorithms: Convening*,
-

- Observation, Probability, and Timeliness. *Science, Technology, & Human Values*, 41(1), 93-117.
- ANEESH, A. (1999). Technologically Embedded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies. *Sociological Abstracts*, American Sociological Association, August.
- (2009). Global Labor: Algoratic Modes of Organization. *Sociological Theory*, 27(4), 347-370.
- ANGELL, I. O., SAMONAS, S. (2009). The Risk of Computerised Bureaucracy. *Journal of Information System Security*, 5(2), 1-25.
- BEER, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1-13.
- BENIGER, J. R. (1995). *Le origini della società dell'informazione: la rivoluzione del controllo*. Torino: UTET.
- BERGER, P. L., BERGER, B., KELLNER, H. (1974). *The homeless mind: modernization and consciousness*. New York: Vintage Books.
- BERGER, P. L., KELLNER, H. (1965). Arnold Gehlen And The Theory Of Institutions. *Social Research*, 32(1), 110-115.
- BERGER, P. L., LUCKMANN, T. (1991). *The social construction of reality: a treatise in the sociology of knowledge*. Harmondsworth: Penguin.
- BERLINSKI, D. (2000). *The advent of the algorithm: the idea that rules the world*. New York: Harcourt.
- BERRY, D. M. (2014). *Critical theory and the digital*. New York: Bloomsbury.
- BIRKBAK, A., CARLSEN, H. B. (2016). The Public and its Algorithms. Comparing and experimenting with calculated publics. In L. Amore, V. Piotukh (eds.), *Algorithmic life: calculative devices in the age of big data* (pp. 21-34). Oxon: Routledge.
- BOGOST, I. (2015). The Cathedral of Computation. *The Atlantic*, January 15.
- BOVENS, M., ZOURIDIS, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review*, 62(2), 174-184.
- BRUBAKER, R. (1991). *The limits of rationality: an essay on the social and moral thought of Max Weber*. London: Routledge.
- Burocrazia. (2018). In *Vocabolario Online Treccani*.
- CAPLAN, R., BOYD, DANA. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society*, 5(1), 1-12.
- CARR, N. G. (2015). *The glass cage: how our computers are changing us*. New York: W. W. Norton & Company.
-

- CASTELLS, M. (2003). *La nascita della società in rete*. Milano: Università Bocconi.
- CLAIR, A. (2017). Rule by Nobody Algorithms update bureaucracy's long-standing strategy for evasion. *Real Life*. Rule by Nobody Algorithms update bureaucracy's long-standing strategy for evasion. *Real Life*, February 21.
- CROZIER, M. (1969). *Il fenomeno burocratico*. Milano: Etas Kompass.
- DANAYHER, J. (2016). The Threat of Allogracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, 29(3), 245-268.
- DELEUZE, G., GUATTARI, F., M. (2003). *Mille piani: capitalismo e schizofrenia*. Roma: Cooper & Castelvechi.
- Demistificare. (2018). In *Lemmario italiano*, Garzanti linguistica.
- Demistificazione. (2018). In *il Sabatini Coletti - Dizionario della Lingua Italiana*. Edizione online.
- DIMAGGIO, P. J., POWELL, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147.
- DREYUS, H. L., SPINOSA, C. (1997). Highway bridges and feasts: Heidegger and Borgmann on how to affirm technology. *Man and World*, 30, 159-177.
- FINN, E. (2018). *What Algorithms Want: imagination in the age of computing*. Cambridge: The Mit Press.
- FOUCAULT, M. (1984). What is Enlightenment?. In P. Rabinow (ed.), *The Foucault Reader* (pp. 32-50). New York: Pantheon Books.
- GAJDUSCHEK, G. (2003). Bureaucracy: Is It Efficient? Is It Not? Is That The Question?: Uncertainty Reduction: An Ignored Element of Bureaucratic Rationality. *Administration & Society*, 34(6), 700-723.
- GEHLEN, A. (1983). *L'uomo. La sua natura e il suo posto nel mondo*. Milano: Feltrinelli.
- (1984). *L'uomo nell'era della tecnica*. Milano: SugarCo Edizioni.
- GILBERT, A. S. (2018). Algorithmic culture and the colonization of life-worlds. *Thesis Eleven*, 146(1), 87-96.
- GOLUMBIA, D. (2009). *The cultural logic of computation*. Cambridge: Harvard University Press.
- HABERMAS, J. (1986). *Teoria dell'agire comunicativo*. Bologna: Il Mulino.
- HARRIS, T. (2017). *How a handful of companies control billions of minds every day*. Disponibile su https://www.ted.com/talks/tristan_harris_the_manipulative_tricks_tech_companies_use_to_capture_your_attention.
- HORKHEIMER, M. (2002). *Eclisse della ragione*. Torino: Einaudi.
- HORKHEIMER, M., ADORNO, T. W. (2010). *Dialettica dell'illuminismo*. Torino: Einaudi.
-

- JOHN, R. R. (1988). Out of Control. *Isis*, 79(4), 675-679.
- KALLINIKOS, J. (2004). The Social Foundations of the Bureaucratic Order. *Organization*, 11(1), 13-36.
- (2005). The order of technology: Complexity and control in a connected world. *Information and Organization*, 15(3), 185-202.
- KNUTH, D. E. (1997³). *The art of computer programming*. Reading, Mass: Addison-Wesley.
- LAMPINEN, A. (2017). Algorithmic Systems, Strategic Interaction, and Bureaucracy. *Rajapinta*, November 22.
- LIPPMANN, W. (1993). *The phantom public*. New Brunswick (N.J.): Transaction Publishers.
- MARCUSE, H., (1998). *Collected papers of Herbert Marcuse*. Ed. D. Kellner. New York: Routledge.
- (2013). *L'uomo a una dimensione: l'ideologia della società industriale avanzata*. Torino: Einaudi.
- MARGETTS, H. (1999). *Information technology in government: Britain and America*. New York: Routledge.
- MAYER-SCHÖNBERGER, V., CUKIER, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- MEYER, J. W., ROWAN, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340-363.
- MINSKY, M. (1967). Why programming is a good medium for expressing poorly understood ideas. In *Design and Planning II – Computers in Design and Communication*. New York: Hastings House Publishers.
- MORGAN, G. (2006). *Images of organization*. Thousand Oaks: Sage Publications.
- MUELLERLEILE, C., ROBERTSON, S. (2018). Digital Weberianism: Bureaucracy, Information, and the Techno-rationality of Neoliberal Capitalism. *Indiana Journal of Global Legal Studies*, 25(1), 187.
- NAPOLI, P. M. (2014). Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption: Automated Media. *Communication Theory*, 24(3), 340-360.
- O'NEIL, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown.
- PARNAS, D. L. (1993). Foreword. In L. R. Wiener (ed.), *Digital woes: why we should not depend on software* (pp. ix-xiii). Boston: Addison Wesley.
- PASQUALE, F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University
-

Press.

- PEETERS, R., SCHUILENBURG, M. (2018). Machine justice: Governing security through the bureaucracy of algorithms. *Information Polity*, 23(3), 267-280.
- POLANYI, M., SEN, A. (2009). *The tacit dimension*. Chicago: University of Chicago Press.
- PORTER, G. (1987). Socioeconomic Transformations: The Control Revolution. *Science*, 236(4804), 970-972.
- POWELL, A. (2016). Accountable machines: bureaucratic cybernetics?. *Media Policy Blog*, March 24.
- POWELL, W. W., DiMAGGIO, P. (1991) (ed.). *The New institutionalism in organizational analysis*. Chicago: University of Chicago Press.
- SCAFF, L. A. (2017). Rationalization. In B. S. Turner (ed.), *The Wiley-Blackwell Encyclopedia of Social Theory* (pp. 1-3). Oxford (UK): John Wiley & Sons.
- VISMANN, C., WINTHROP-YOUNG, G. (2008). *Files: law and media technology*. Stanford: Stanford University Press.
- WEBER, M. (1968). *Economia e società*. Milano: Edizioni di Comunità.
- (2014). *L'etica protestante e lo spirito del capitalismo*. Milano: Rizzoli.
- WEIZENBAUM, J. (1976). *Computer power and human reason: from judgment to calculation*. San Francisco: W. H. Freeman.
- WOOLLEY, W., HOWARD, P. N. (2016). Automation, algorithms, and politics| Political communication, computational propaganda, and autonomous agents – Introduction. *International Journal of Communication*, 10, 4882-4890.
- ZARSKY, T. Z. (2012). Automated prediction: perception, law, and policy. *Communications of the ACM*, 55(9), 33.
- ZUURMOND, A. (1994). *De infocratie: een theoretische en empirische heroriëntatie op Weber's ideaaltype in het informatietijdperk*. Den Haag: Phaedrus.
- (1998). From Bureaucracy to Infocracy: Are Democratic Institutions Lagging Behind? In *Public Administration in an Information Age: A Handbook* (pp. 259-271). Amsterdam: IOS Press.
-

DISCRIMINAZIONE E ALGORITMI

Incontri e scontri tra diverse idee di fairness

di *Mattia Galeotti**

Abstract

The growing spread of Machine Learning algorithms in our society, was accompanied in the last years by various cases of discrimination that have been bred by automatized procedures, in particular with respect to race, gender, religious affiliation or sexual orientation. For this reason the subject of algorithmic fairness became central, but finding a solution appears to be a difficult problem. In this work we intend to show different mathematical models that are needed to define various notions of fairness, to investigate the conditions of compatibility and incompatibility between the models, and the relations and frictions with the material conditions. A feature of today governmental apparatuses that are based on automatized processes, is the reduction of political and social problems to the optimization of some functions; this is also the case when we talk about discrimination, that in this approach can be solved by measuring and deleting the bias encoded in algorithms. In this work we try instead to read from a political and sociological point of view the mechanisms of numerical optimization, and we investigate the relations between the subject of discrimination as it has historically appeared, and the multiple statistical formalizations developed with the purpose of governing this problem.

Keywords

Bias; fairness; statistics; movements; algorithms

* MATTIA GALEOTTI è post-doc presso Università degli Studi di Trento.
Email: galeotti.mattia.work@gmail.com

1. INTRODUCTION

Sempre più processi decisionali funzionano oggi attraverso la costruzione di modelli predittivi matematici che, basandosi su database storici, cercano di automatizzare i meccanismi di selezione; esempi molto conosciuti sono la pubblicità in rete personalizzata sull'utente, i punteggi individuali ai clienti di istituzioni bancarie per decidere se concedere o meno un prestito, lo screening automatico dei curricula dei candidati per un posto di lavoro, i punteggi che nel sistema giudiziario americano determinano il rischio di recidività per individui condannati, e molti altri. Questi strumenti di *Machine Learning* stanno dunque divenendo essenziali in sempre più funzioni della vita collettiva, e parallelamente alla loro diffusione si afferma come centrale il problema delle discriminazioni che possono accompagnare questi processi o direttamente essere veicolate tramite essi. In particolare, come vedremo meglio nella seconda sezione, specifici casi di studio hanno riscontrato la riproduzione di discriminazioni rispetto alla razza¹, al genere, all'appartenenza religiosa o all'orientamento sessuale, veicolati attraverso processi valutativi di questo tipo, ponendo la questione della discriminazione strutturale, oggi centrale nel dibattito politico, anche nello specifico contesto delle decisioni automatizzate. Più in generale il problema della discriminazione algoritmica si inserisce nel dibattito sul rapporto tra saperi tecnoscientifici e sistemi sociali, un terreno molto fertile per la sociologia della scienza e gli *Science and Technology Studies*.

Ogni processo di apprendimento, valutazione e/o selezione basato su dati statistici, funziona attraverso una formalizzazione matematica, la definizione di specifici strumenti di calcolo e quindi un calcolo effettivo tramite questi strumenti. Sebbene nella narrazione e nel marketing dell'Intelligenza Artificiale si tenda a identificare gli algoritmi con il semplice momento del calcolo, la formalizzazione e la definizione degli strumenti sono essenziali per organizzare e dare senso ai dati su cui si vuol far operare un processo algoritmico, e quindi al processo stesso. Riprendendo Mazzotti (2015) ci soffermiamo sull'aspetto di mutua costituzione tra scienza e società, un aspetto che è stato studiato soprattutto dal punto di vista dei sistemi tecnologici, e molto meno per quanto riguarda le procedure formali e deduttive. Nelle scelte assiomatiche proprie di ogni modello matematico sono codificate delle

¹ In tutto il testo utilizzeremo il termine "razza" riferendoci ad un costrutto storico che ha carattere performante al livello sociale, dunque come elemento reale ma non biologico di auto-riconoscimento degli individui, e potenzialmente di discriminazione. Per questo uso del termine vedere ad esempio Curcio, Mellino (2012).

ipotesi sociali e delle visioni del mondo, e la scelta di un modello per una specifica task deriva dall'incontro tra queste ipotesi, gli obiettivi contingenti e specifici interessi di gruppo; in più, queste cornici di senso vengono riprodotte dall'algoritmo stesso, tramite la sua azione di valutazione e selezione.

Ogni popolazione è caratterizzata da alcuni attributi sensibili come la razza, il genere, l'appartenenza religiosa o l'orientamento sessuale, cioè attributi rispetto ai quali potrebbe attuarsi una discriminazione, la *fairness* di un algoritmo è la proprietà di non discriminare rispetto a questi attributi; seguendo Dwork, Hardt, Pitassi, Reingold, Zemel, (2012) e Friedler, Scheidegger, Venkatasubramanian (2016) possiamo individuare due filoni negli approcci alla formalizzazione del concetto di *fairness*: da una parte la *fairness* come descrizione corretta del contesto reale, in cui ogni individuo riceve una "giusta" valutazione senza discriminazioni; dall'altra la *fairness* come parità statistica di valutazione tra i differenti gruppi sociali, con le stesse percentuali di successo per ogni valore dell'attributo sensibile.

La linea di studio che ci interessa approfondire è quella delle relazioni tra i modelli di *fairness* e le scelte politiche e organizzative connesse con i processi di selezione. Come abbiamo detto i modelli matematici codificano delle visioni del mondo, in particolare le diverse definizioni di *fairness* codificano idee distinte di discriminazione, tra queste diverse impostazioni esistono linee di incompatibilità e di frizione, ed ogni modello formale deve confrontarsi con i vincoli matematici e quelli del contesto storico. L'analisi delle incompatibilità e dei limiti di ogni formalizzazione permette di mettere in luce in che maniera l'algoritmo opera, ed in quali dinamiche strutturali viene ad inserirsi. Un esempio chiarificatore è descritto nel recente articolo di Rachel Courtland (2018): l'autrice sottolinea che è matematicamente impossibile attuare nel sistema giudiziario delle forme di *fairness* individuale e contemporaneamente di parità statistica tra gruppi, se la probabilità di arresto rimane più elevata per alcuni gruppi razziali; per questo se l'algoritmo viene utilizzato nel contesto giudiziario senza interferire col lavoro di polizia, allora dovrà scegliere una tra le due forme di *fairness* incompatibili.

Per valutare il tipo di azione degli algoritmi sul contesto sociale in cui agiscono, risulta particolarmente utile il concetto di "*governance by the numbers*" per com'è utilizzato in Katz (2017): attraverso la creazione di graduatorie e soglie di inclusione, le modellizzazioni statistiche partecipano ad una forma di disciplina della vita sociale. In questo senso gli algoritmi divengono dei veri e propri dispositivi di

governo, secondo l'utilizzo introdotto da Foucault di questo termine². La "governance by the numbers" è quindi da intendersi come messa in pratica di una certa governamentalità, diffusa e legittimata in particolare con una narrazione secondo cui gli algoritmi sono processi in grado di descrivere oggettivamente alcuni comportamenti sociali: gli algoritmi diventano quindi portatori di una "*vision from nowhere*". Quest'ultima nozione, come abbiamo già detto, dissimula completamente la contingenza storica e sociale in cui il dispositivo è stato concepito ed è messo in azione.

In quest'ottica la ricerca di un algoritmo privo di *bias* (di razza, di genere, etc.) risulta epistemologicamente poco interessante, la nozione stessa di *bias* appare inadeguata, in quanto suggerisce l'esistenza di un processo oggettivo a cui i diversi algoritmi si avvicinano per approssimazione. Piuttosto la nostra trattazione suggerisce che l'analisi matematica dei modelli può permettere di capire quali forme di discriminazione sono proprie ad ogni algoritmo, aprendo ad una descrizione degli algoritmi come fattori operanti in un determinato periodo storico ed all'interno di specifiche ipotesi di normazione dello spazio sociale.

Nella sezione 2 mostreremo alcuni noti e documentati casi di discriminazione riscontrati in vari campi d'utilizzo di algoritmi decisionali. Nelle sezioni 3 e 4 introdurremo gli strumenti matematici necessari alla nostra trattazione, ed approfondiremo alcuni quadri assiomatici entro i quali è possibile approcciare il problema della *fairness*, esplorando le compatibilità ed incompatibilità tra questi differenti approcci. Infine, nella sezione 5 mostreremo in che modo gli algoritmi organizzano costruzioni di senso e divengono strumenti di *governance*.

2. IL PROBLEMA DELLA FAIRNESS NELL'UTILIZZO SOCIALE DEGLI ALGORITMI

La diffusione di algoritmi decisionali basati sul Machine Learning in ogni ambito della vita collettiva, si è accompagnata negli ultimi anni a un sempre maggior numero di casi in cui quegli stessi algoritmi hanno dimostrato di veicolare o riprodurre discriminazioni basate sulla razza, il genere, l'appartenenza religiosa, l'orientamento sessuale e altre caratteristiche; il tema della giustizia algoritmica si è dunque affermato come centrale. L'analisi del problema ha immediatamente rivelato la necessità di approfondire il contesto di applicazione, i dati di apprendimento e le funzioni di questi processi automatizzati, e di considerare gli algoritmi come strumenti dentro contesti storici contin-

² Per il concetto di dispositivo in Foucault si veda Agamben (2006).

genti, non nettamente separabili dalle condizioni materiali in cui operano e dai soggetti sociali che li utilizzano. Seguendo il punto di vista degli “Science and Technology Studies”, non soltanto gli algoritmi fanno parte di un’infrastruttura tecnoscientifica che opera dentro il contesto sociale, ma è vero anche l’inverso, i dispositivi scientifici e tecnici, tra cui gli algoritmi, sono costituiti da processi di natura sociale.

Nel 2017 l’articolo della rivista online Quartz (Sonnad, 2017) osservava uno strano comportamento di *Google Translate*: nelle traduzioni dal turco all’inglese, la parola turca “o”, pronomi corrispondenti alla terza persona singolare e di genere neutro, veniva tradotta nell’inglese “he” oppure “she” in base alle parole che la accompagnavano, rivelando una chiara discriminazione di genere. Parole come “soldier”, “doctor” oppure “hardworking” portavano ad una traduzione maschile, mentre “teacher”, “nurse” e “lazy” portavano ad una traduzione femminile. Facile immaginare in questo contesto che sia direttamente l’insieme di dati a cui Google ha accesso a contenere una discriminazione di genere, perché quei dati corrispondono al linguaggio utilizzato dagli utenti in rete. Allo stesso tempo la pervasività di questo strumento fa temere per un effetto di rinforzo della discriminazione linguistica.

Ancora più problematica è la questione degli algoritmi che aiutano nella selezione dei candidati per un posto di lavoro fornendo un punteggio sulla base dei curricula forniti. Nel recente articolo di Chen, Ma, Hannak, Wilson (2018), è stato studiato l’impatto del genere in questo tipo di selezioni in venti città degli Stati Uniti, ottenendo una vasta gamma di esempi di discriminazione. Chiaramente dispositivi di questo tipo non soltanto svantaggiano ingiustamente alcuni individui, ma più in generale rischiano di riprodurre condizioni di svantaggio sistemico, come un minor tasso di impiego o salari più bassi per uno specifico gruppo.

Il caso forse più conosciuto è quello di uno strumento denominato COMPAS, per *Correctional Offender Management Profiling for Alternative Sanctions*, sempre più utilizzato nelle corti giudiziarie degli Stati Uniti con l’obiettivo di determinare il rischio che individui condannati per un crimine divengano recidivi. Nel 2016 il sito giornalistico ProPublica pubblicava l’articolo di Angwin, Larson, Mattu, Kirchner (2016), dimostrando che il COMPAS era chiaramente discriminante verso gli individui neri. In particolare, l’inchiesta metteva in luce che questo strumento ha in alcuni casi tassi di accuratezza molto bassi: un alto tasso di falsi positivi (cioè individui falsamente indicati come a rischio di recidività) nel caso di individui neri, ed un maggior

tasso di falsi negativi (individui falsamente indicati come non a rischio) tra individui bianchi rispetto ai non-bianchi. Il COMPAS ed altri strumenti di polizia preventiva stanno subendo numerose critiche, in particolare dai *community group* degli Stati Uniti che si organizzano contro le discriminazioni della polizia, ma la loro diffusione sembra al momento più rapida degli strumenti in grado di regolarli³.

Le istituzioni locali, nazionali ed internazionali sono da alcuni anni alle prese con la costruzione di strumenti legislativi e prassi amministrative in grado di affrontare questi temi.⁴ Negli ultimi anni si è comunque osservata una rinnovata spinta per “responsabilizzare” gli algoritmi e renderli più trasparenti. Come riportato nell’articolo di Courtland (2018), il consiglio cittadino di New York ha messo in piedi una *task force* per incentivare la condivisione pubblica degli algoritmi, ed indagare i loro funzionamenti discriminanti; recentemente il governo francese si è impegnato a rendere *open* tutti gli algoritmi utilizzati nella burocrazia di selezione statale, ed il governo inglese ha invitato alla responsabilizzazione e trasparenza dei *data* nel settore pubblico.

L’indagine della discriminazione rimane comunque un campo difficile da regolamentare, proprio perché i confini tra l’algoritmo e gli altri comportamenti sociali sono sfumati ed imprecisi. Nel seguito approfondiremo in che modo gli algoritmi entrano in relazione con i fenomeni discriminatori e li interpretano, apprendono e/o riproducono, a partire da specifici modelli matematici.

3. SCHEMATIZZAZIONE DELL’APPRENDIMENTO E PROBLEMA DELLA AWARENESS

Esiste una vasta letteratura scientifica in cui i processi di Machine Learning e di selezione tramite algoritmi sono descritti come raccolta e sintesi di alcuni dati di per sé già presenti nel contesto “naturale”, in questa visione il *bias* è semplicemente la misura di un errore commesso nel processo di apprendimento, una differenza numerica tra il risultato

³ Nei nostri esempi non abbiamo trattato, e non approfondiremo nel seguito, il carattere performante e discriminante presente anche nella definizione degli spazi di possibilità di ogni attributo, cioè l’organizzazione della raccolta di dati secondo griglie che prevedono specifiche risposte possibili per ogni attributo. Chiaramente anche questi aspetti contribuiscono ad influenzare il contesto sociale a cui l’algoritmo viene applicato, e quindi costituiscono uno degli elementi della governamentalità tramite algoritmi, sebbene nella nostra analisi ci soffermeremo su altri aspetti.

⁴ Per una panoramica del rapporto tra strumenti matematici e vari settori di applicazione, tra cui le istituzioni legislative, sono referenze importanti gli studi Žliobaitė (2015), Romei, Ruggieri (2014) e Barocas, Selbst (2016).

dell'algoritmo ed il dato "vero". Dal nostro punto di vista invece le procedure di apprendimento e misura statistica partecipano alla costruzione di senso ed alla visione del mondo in cui gli stessi fenomeni di discriminazione vengono rilevati. La nostra analisi approfondirà dunque in che modo gli specifici quadri analitici e statistici contribuiscono a queste costruzioni di senso.

In questa sezione introdurremo alcuni elementi di teoria statistica centrali nella nostra trattazione. Successivamente descriveremo una prima formalizzazione della fairness di un algoritmo tramite la nozione di *unawareness*, spiegando il motivo della inadeguatezza di questo concetto rispetto ai contesti concreti di applicazione dei processi algoritmici. Concentreremo la nostra analisi sui quadri assiomatici necessari per definire diverse nozioni di fairness e discriminazione, anche se non entreremo nel dettaglio dei diversi strumenti di calcolo (regressioni lineari, reti neurali, *deep learning*, etc.), i modelli assiomatici da noi trattati permetteranno un'analisi approfondita del rapporto tra algoritmi e discriminazione.

Vediamo di seguito alcuni concetti statistici fondamentali. Un attributo sarà una variabile aleatoria discreta. Come esempio pensiamo ad un attributo X corrispondente alla razza di un individuo di una certa popolazione. Alla variabile sarà associato un insieme $V=\{v_1, v_2, \dots\}$ di possibili valori ed una distribuzione di probabilità tale che $p(X=v_i)$ sia un numero reale compreso tra 0 ed 1 per ogni v_i nell'insieme V , ed inoltre

$$\sum_{v_i \in V} p(X = v_i) = 1.$$

Quest'ultima condizione corrisponde alla certezza che l'attributo assuma un valore nell'insieme V . Nel nostro esempio, V è l'insieme delle razze presenti nella popolazione di riferimento e per ogni razza v_i , $p(X=v_i)$ è la probabilità che un individuo scelto casualmente sia di razza v_i , cioè equivalentemente $p(X=v_i)$ è la frazione di popolazione di razza v_i . Quando non c'è rischio di confusione, faremo un piccolo abuso di notazione indicando con $p(X)$ la distribuzione di probabilità dell'attributo X .

Dati due attributi X_1, X_2 , indichiamo con $X_1|X_2$ la variabile aleatoria X_1 condizionata alla variabile X_2 : resta invariato l'insieme dei possibili valori di X_1 , ma le distribuzioni $p(X_1|X_2)$ e $p(X_1)$ possono essere differenti, in particolare $p(X_1|X_2)$ dipende dal valore assunto dall'attributo X_2 . Ad esempio, se consideriamo la popolazione di una

città, indichiamo con X_1 il solito attributo razza, e con X_2 l'attributo corrispondente al quartiere di residenza di un individuo della popolazione, allora $p(X_1=v_i|X_2=q)$ indica la probabilità che X_1 assuma valore v_i sapendo che X_2 è il quartiere q . Chiaramente questo valore potrebbe variare molto marcatamente, per ogni razza v_i , al variare del quartiere X_2 . Nel caso in cui $p(X_1|X_2)=p(X_1)$ diciamo che i due attributi sono indipendenti.

Un *dataset* è il dato di:

- un insieme I che indicizza gli individui di una popolazione. Nel nostro caso si tratterà semplicemente della numerazione progressiva degli individui;
- una serie di attributi X_1, X_2, \dots per ogni individuo.

Possiamo immaginare ad esempio un dataset con quattro attributi X_1, X_2, X_3, X_4 corrispondenti a razza, quartiere di residenza, età e reddito. In questo caso $X_1^{(i)}, X_2^{(i)}, X_3^{(i)}, X_4^{(i)}$ saranno rispettivamente la razza, il quartiere, l'età ed il reddito dell' i -esimo individuo, dove i è un indice nell'insieme I . Nelle nostre analisi avremo uno o più attributi la cui previsione è l'obiettivo del dispositivo algoritmico e li chiameremo attributi di classificazione, denotandoli usualmente con la lettera Y . Un attributo non di classificazione è detto attributo descrittivo. Ad esempio, nel caso precedente Y potrebbe essere la probabilità di vincere un concorso di ammissione universitario, o di commettere un crimine.

Introduciamo un ultimo elemento di notazione: come già detto la nozione di attributo di classificazione indica gli attributi la cui previsione è l'obiettivo dell'algoritmo; indichiamo con \hat{Y} il predittore di un tale attributo di classificazione Y , cioè il risultato della predizione algoritmica, in quanto distinto dal "vero" valore Y . L'accuratezza, cioè la proprietà del predittore di essere statisticamente prossimo al vero valore, è una proprietà centrale di un processo algoritmico. Ovviamente la correlazione tra i due valori (quello stimato e quello vero), ed il senso di questa distinzione, dipende fortemente dal modello utilizzato, come vedremo meglio nel seguito.

Il funzionamento di un algoritmo di apprendimento e previsione può essere schematizzato come segue. A partire da un dataset, l'algoritmo stabilisce una correlazione tra gli attributi descrittivi e l'attributo di classificazione. L'algoritmo viene quindi utilizzato su individui (della medesima popolazione) di cui sono conosciuti gli attributi descrittivi, per prevedere l'attributo di classificazione. L'accuratezza del risultato può essere misurata, come vedremo, in molti modi. Questa schematizzazione generale non distingue gli specifici tipi di apprendimento, che non saranno oggetto della nostra trattazione.

Tra gli attributi descrittivi del dataset saranno presenti degli attributi indicati come sensibili (ad esempio razza, genere, appartenenza religiosa, orientamento sessuale, etc.), la nozione di discriminazione e quella di fairness saranno introdotte a partire dalla correlazione tra questi attributi e l'attributo classificatore. È importante sottolineare, come già detto nell'introduzione, che lo studio della discriminazione si concentra sia sul dataset di apprendimento che sul risultato dell'algoritmo, perché per dare senso al procedimento valutativo, e quindi anche alla discriminazione da esso veicolata, è necessario studiare la relazione che viene a crearsi tra questi due oggetti.

Un'idea intuitiva (non formalizzata matematicamente) di discriminazione ad opera di un processo algoritmico si fonda sul concetto di informazione: si ha discriminazione quando l'esito della valutazione di un individuo fornisce informazione sull'attributo sensibile (e viceversa). A partire da questo punto di vista nasce la proposta di oscuramento degli attributi sensibili come rimedio alle pratiche discriminatorie, ad esempio attraverso la creazione di database che non prevedono la registrazione di questi attributi: l'ignoranza del decisore dovrebbe cancellare anche il rischio di discriminazione. La questione della consapevolezza (*awareness*) degli attributi sensibili si può ritrovare anche in ambito legislativo e istituzionale, in particolare il comportamento rispetto agli attributi razziali e religiosi costituisce una grande differenza tra il sistema anglosassone e quello francese: nel primo la razza e la religione sono attributi registrati nei processi di censimento della popolazione, mentre le leggi della Repubblica Francese si muovono in senso diametralmente opposto.

Questa idea di eliminazione della discriminazione attraverso la unawareness è oggi però ritenuta totalmente inadeguata per comprendere i fenomeni discriminatori codificati nei dispositivi algoritmici, ed un semplice esempio permette di capirne il motivo: supponiamo che un attributo protetto X_1 sia correlato con un attributo non protetto (o più di uno) X_2 , cioè che la distribuzione di probabilità $p(X_1|X_2)$ dipenda in maniera non trascurabile dal valore del secondo attributo X_2 ; la cancellazione dell'attributo sensibile allora non corrisponde alla cancellazione totale dell'informazione, qui utilizzata in senso intuitivo, su quello stesso attributo⁵.

⁵ Una nozione di informazione che corrisponde a questo utilizzo intuitivo del termine si può ritrovare nella Teoria dell'Informazione di Shannon. Dati due attributi X_1 , X_2 si definisce una grandezza positiva $H_{X_1|X_2}$ denominata entropia condizionale di X_1 rispetto a X_2 : se V_{X_1} e V_{X_2} sono gli insiemi dei valori assunti da X_1 e X_2 rispettivamente, e denotiamo $p(x_1|x_2)$ la probabilità $p(X_1=x_1|X_2=x_2)$ per tutti i valori x_1 , x_2 in V_{X_1}, V_{X_2} rispettivamente, allora l'entropia condizionale è definita come

Un caso classico è quello di un processo di selezione, pensiamo alla valutazione necessaria per la concessione di un prestito bancario, in cui uno degli attributi descrittivi è il quartiere di residenza: in molte situazioni l'attributo razziale è fortemente dipendente dal quartiere di residenza, pertanto cancellare il primo non elimina la possibilità che individui di una specifica razza siano avvantaggiati o svantaggiati. Il processo di selezione dipenderà ancora dall'attributo sensibile, in maniera potenzialmente discriminatoria. In aggiunta, la cancellazione dell'attributo dal dataset non permetterà nessuna forma di controllo della discriminazione. Questo fenomeno prende il nome di *redlining* (o codificazione ridondante). Le origini del termine sono da ricercare nella pratica di alcuni istituti bancari ed assicurativi americani di negare servizi finanziari a specifici quartieri; in particolare negli anni del *New Deal* venne istituita una vera e propria mappatura dei territori cittadini in cui alcune zone a prevalenza afroamericana e di basso reddito erano contrassegnate col colore rosso. Lo studio di questo fenomeno ha permesso di evidenziarne non soltanto il carattere chiaramente discriminatorio, ma anche gli effetti performati nel rafforzare le condizioni di svantaggio delle zone marcate in rosso. Risulta chiara la natura governamentale, in senso foucaultiano, di una tale mappatura: l'organizzazione della popolazione non è ottenuta tramite l'obbedienza, ma attraverso l'utilizzo di alcuni saperi sulla popolazione, in particolare grazie alla definizione di una "natura" del corpo sociale iscritta nella cartografia (Vagnarelli, 2017) sul concetto di governamentalità in Foucault). Come esposto nello stesso articolo di Vagnarelli, i dispositivi governamentali funzionano in modo tale per cui «Non vi sarà più la natura da un lato e il sovrano dall'altro ma la "natura" della popolazione farà il suo ingresso all'interno delle tecniche di potere» (Vagnarelli, 2016, 150). Questa descrizione del *redlining* si applicherà anche ai dispositivi algoritmici che affronteremo più avanti.

Nel seguito rispetteremo sempre il principio per cui gli attributi sensibili, potenzialmente soggetti a discriminazione, vengono registrati nel dataset di apprendimento. Sottolineiamo che questa scelta, largamente accettata nella letteratura scientifica, è comprensibile solo a partire da una nozione intuitiva di discriminazione che mette in evidenza l'inadeguatezza della "fairness come unawareness".

$$H_{X_1|X_2} := \sum_{x_2 \in V_{X_2}} p(X_2 = x_2) \cdot \left(\sum_{x_1 \in V_{X_1}} p(x_1 | x_2) \log(p(x_1 | x_2)) \right).$$

Questa grandezza viene considerata una misura dell'informazione portata dall'attributo X_2 sull'attributo X_1 : quanto più $H_{X_1|X_2}$ è prossima allo zero, tanto più la conoscenza di X_2 dà informazione su X_1 . In particolare, $H_{X_1|X_2}$ vale zero quando la conoscenza del valore di X_2 determina esattamente il valore di X_1 , ed è massima quando X_1 e X_2 sono indipendenti.

Osserviamo insomma con questo primo esempio che la definizione di *fairness* entra necessariamente in relazione con il significato storicizzato della discriminazione.

In più vedremo in seguito che a partire da altre nozioni di *fairness*, il principio della *unawareness* può essere anche completamente ribaltato. Con questo intendiamo dire che dentro altre modellizzazioni, un algoritmo fair può potenzialmente applicare delle forme di “azione positiva”, o *affirmative action*, andando cioè a compensare direttamente alcune categorie svantaggiate.

4. DUE DIVERSE IDEE DI FAIRNESS IN UN PARADIGMA DI AWARE DATA MINING

In questa sezione analizzeremo due impostazioni del problema della *fairness* all'interno di un paradigma di *aware data mining*: con questa locuzione intendiamo che il processo di costruzione del dataset di apprendimento, la raccolta dei dati anche detta “data mining”, registra gli attributi sensibili. Le due definizioni che proponiamo non sono da intendersi come contrapposte o divergenti, ci interessa però mostrare le diverse implicazioni sociali codificate dai due approcci e provare a leggere alla luce di queste il rapporto tra gli algoritmi e altri dispositivi sociali di valutazione, controllo e selezione.

Con l'obiettivo di semplificare la nostra trattazione, considereremo un attributo sensibile X di tipo binario: X ha come valori possibili 1 e 0, e 1 corrisponderà alla classe svantaggiata, ad esempio se X è il genere di un individuo scelto tra due possibili, il valore 1 dovrebbe corrispondere al genere potenzialmente discriminato. Anche l'attributo di classificazione che cercheremo di prevedere, indicato con \hat{Y} , sarà di tipo binario: in generale indicheremo con $\hat{Y}=1$ l'esito positivo nel processo di selezione, e con $\hat{Y}=0$ l'esito negativo; nel caso di un processo di valutazione per la concessione di un prestito bancario, $\hat{Y}=1$ corrisponderà alla scelta di concedere il prestito, mentre $\hat{Y}=0$ a quella di non concederlo. La scelta della notazione \hat{Y} è dovuta al fatto che in molti casi pratici questo attributo è considerato un predittore (vedi sezione precedente) per il “vero” attributo Y : nel caso del prestito bancario si può ad esempio considerare che esista un attributo Y corrispondente alla capacità di solvenza dell'individuo, che l'algoritmo cerca di predire⁶.

⁶ Ovviamente la possibilità di ben definire l'attributo Y varia per ogni specifico procedimento, nel caso di un prestito bancario è in realtà molto difficile individuare una definizione per la “vera” capacità di ripagare un prestito quando questo prestito non è stato

Denominiamo accuratezza la precisione con la quale il predittore \hat{Y} predice correttamente il valore Y . L'accuratezza si misura d'abitudine considerando la probabilità che \hat{Y} sia corretto quando $Y=1$ e quando $Y=0$, cioè tramite i due valori⁷

$$p(\hat{Y} = 1 \mid Y = 1), p(\hat{Y} = 0 \mid Y = 0).$$

Il primo approccio che prendiamo in considerazione definisce come non-discriminanti delle procedure di classificazione per le quali l'accuratezza è indipendente dall'attributo protetto X . I due lavori che utilizzeremo come punto di riferimento sono Dwork *et al.* (2012) e Hardt, Price, Srebro (2016); in particolare nel secondo abbiamo due definizioni formali di predittori non discriminanti:

- un predittore \hat{Y} rispetta l'*uguaglianza delle opportunità* se per un individuo con $Y=1$, la probabilità che venga classificato correttamente è indipendente dall'attributo protetto X , cioè

$$p(\hat{Y} = 1 \mid X = 1, Y = 1) = p(\hat{Y} = 1 \mid X = 0, Y = 1)$$

- un predittore \hat{Y} rispetta l'*uguaglianza delle probabilità* se oltre alla condizione precedente, la stessa cosa vale per gli individui con $Y=0$, cioè

$$p(\hat{Y} = 0 \mid X = 1, Y = 0) = p(\hat{Y} = 0 \mid X = 0, Y = 0).$$

Le condizioni di uguaglianza delle opportunità ed uguaglianza delle probabilità attestano che l'accuratezza nella descrizione, cioè la probabilità di uguaglianza tra \hat{Y} e Y , sia indipendente dal fattore sensibile X . La dipendenza tra X ed il valore "vero" Y , non viene considerata in questo senso un fattore discriminante. Secondo questo approccio dunque a partire dal contesto su cui sono raccolti i dati, l'algoritmo apprende la correlazione tra gli attributi descrittivi (tra cui X) e Y , e la fairness corrisponde alla precisione nella predizione di Y

concesso; diversamente nel caso di un algoritmo che cerca di prevedere l'interessamento di un individuo per un particolare prodotto venduto in un supermercato, è sempre ben definito un attributo che indichi se quel prodotto è stato acquistato o meno. I diversi quadri epistemologici della nozione di predittore non sono comunque oggetto della nostra trattazione.

⁷ Ricordiamo che la probabilità condizionale $p(\hat{Y}|Y)$ è definita come la probabilità che \hat{Y} assuma uno specifico valore quando è noto il valore di Y .

(tramite il predittore \hat{Y}).

Nei casi pratici, una simile nozione di fairness permetterà di individuare tra diversi algoritmi applicati su uno stesso dataset, quello meno discriminante; allo stesso tempo questi strumenti d'analisi non permettono di fare una valutazione della discriminazione intrinseca al dataset.

Prima di introdurre il secondo approccio consideriamo un'altra misura della discriminazione denominata differenza media

$$d_1 := p(\hat{Y} = 1 \mid X = 0) - p(\hat{Y} = 1 \mid X = 1).$$

Questa grandezza⁸ esprime la differenza tra la probabilità di avere un esito positivo condizionato al valore sensibile X nella classe 0, e la stessa probabilità con valore sensibile X nella classe 1. La differenza media è una misura della discriminazione che non è condizionata a nessun altro attributo oltre che quello protetto, e la condizione $d_1=0$ descrive una parità statistica, rispetto all'algoritmo, tra i vari gruppi caratterizzati da uno stesso valore dell'attributo sensibile. Un'altra misura⁹ definita in modo analogo è l'*impact ratio*

$$I_1 = \frac{p(\hat{Y} = 1 \mid X = 0)}{p(\hat{Y} = 1 \mid X = 1)}.$$

Il secondo approccio che introduciamo utilizza queste misure “di gruppo” per quantificare la fairness dell'algoritmo: un algoritmo è ritenuto non-discriminante se la differenza media d_1 è prossima allo zero (o l'*impact ratio* I_1 è prossimo all'unità). Nei casi pratici queste misure della discriminazione possono essere utilizzate per valutare il risultato di un algoritmo ma anche il dataset di apprendimento. In Pedreschi, Ruggieri, Turini (2008; 2009) il rapporto tra la fairness del dataset e quella del risultato dell'algoritmo è indagata approfonditamente, in questo senso il dataset iniziale diventa una componente dell'algoritmo stesso, poiché è possibile valutare se il processo di apprendimento ha corretto la discriminazione presente nei dati iniziali.

⁸ Ovviamente è possibile definire la stessa differenza per l'esito negativo, ma la grandezza così ottenuta

$$d_0 := p(\hat{Y} = 0 \mid X = 0) - p(\hat{Y} = 0 \mid X = 1)$$

è semplicemente l'inverso di d_1 , cioè $d_0 = -d_1$

⁹ Per una presentazione generale di queste ed altre misure della discriminazione utilizzate in letteratura, vedere lo studio Žliobaitė (2015).

In Kamiran, Calders (2009; 2010) vengono studiate alcune tecniche per aumentare la fairness di gruppo di un algoritmo, in particolare modificando il dataset di apprendimento o i risultati dell'algoritmo in modo da favorire il gruppo svantaggiato, attuando in questo modo una forma di "azione positiva". Senza sviluppare una descrizione di queste tecniche, risulta importante sottolineare che la riduzione della discriminazione di gruppo potrebbe entrare in conflitto con l'altra nozione di fairness che abbiamo dato sopra. Gran parte degli articoli che abbiamo citato si occupano infatti anche di individuare le condizioni necessarie per avere un algoritmo accurato e fair sotto entrambi i punti di vista, esistono però dei casi di incompatibilità, in particolare ci soffermeremo qui sull'esempio riportato dall'articolo di Courtland (2018). Consideriamo una popolazione di individui arrestati ed un algoritmo che valuta la possibilità che ogni individuo divenga recidivo subendo un secondo arresto, sia $Y=1$ se l'individuo diviene recidivo e $Y=0$ in caso contrario, e sia \hat{Y} il predittore di Y . Se vogliamo un predittore che rispetti l'ipotesi di uguaglianza delle probabilità descritta sopra, ma allo stesso tempo esiste una discriminazione sistemica in cui gli individui di attributo protetto $X=1$ vengono arrestati con maggior frequenza, allora è impossibile eliminare la discriminazione secondo una metrica di gruppo come l'impact ratio: gli individui di classe $X=1$ verranno considerati più probabilmente a rischio di recidività degli altri, ottenendo con più frequenza una classificazione $\hat{Y}=1$. L'incompatibilità tra i due punti di vista si manifesta proprio dal rapporto con una condizione di discriminazione sistemica, cioè la maggior frequenza assoluta di arresti per alcune categorie di individui. Quantificando con più precisione, se denominiamo d la differenza tra la probabilità di arresto per individui con $X=1$ e per individui con $X=0$, allora esiste una dipendenza lineare tra l'impact ratio I_1 e la differenza d .

Un'interessante prospettiva sui due approcci di cui sopra, viene portata dall'articolo di Friedler, *et al.* (2016). L'idea di fairness come uguaglianza delle opportunità o uguaglianza delle probabilità, viene descritta come un tentativo di cancellare le discriminazioni a partire dalla supposizione che i metodi di raccolta dati (test, esami, analisi a campione) producano una descrizione accurata della realtà, una prospettiva indicata con l'acronimo WYSIWYG, per *What You See Is What You Get*. I metodi basati sul WYSIWYG partono dunque dall'idea che esista una classificazione "giusta" degli individui, ed il compito di un algoritmo fair è quello di ricostruire questa classificazione eliminando le discriminazioni individuali basate sugli attributi protetti.

L'idea di fairness come parità statistica tra gruppi si basa, secondo

Friedler *et al.* (2016), sulla supposizione che in un contesto non discriminatorio, i gruppi di individui caratterizzati da uno stesso valore degli attributi sensibili, otterrebbero percentuali simili di risultati positivi, una prospettiva indicata con l'acronimo WAE, per *We're All Equal*. I metodi basati sul WAE si fondano dunque sull'idea che l'unica classificazione "giusta" è quella in cui si ha una parità statistica dei risultati tra i diversi gruppi ed il compito di un algoritmo fair è quello di operare una selezione che rispetti il più possibile questo valore di parità. Il diverso focus tra discriminazione individuale e discriminazione di gruppo è evidenziato anche in Dwork *et al.* (2012), a partire da un punto di vista WYSIWYG: l'idea di parità statistica tra gruppi viene criticata a partire dalle discriminazioni individuali che potrebbe produrre.

La prospettiva dei due approcci WYSIWYG e WAE non riesce a cogliere la complessità e varietà delle proposte per la costruzione di algoritmi fair, in particolare come abbiamo già detto in diverse situazioni vengono utilizzati entrambi i metodi, laddove esistono condizioni di compatibilità, come mostrato sempre in Dwork *et al.* (2012) e in Friedler *et al.* (2016), ma anche in Kleinberg, Ludwig, Mullainathan, Rambachan (2018) e Corbett-Davies, Pierson, Feller, Goel, Huq (2017). In quest'ultimo lavoro in particolare la fairness viene propriamente analizzata in quanto *constrained optimization*: si lavora cioè in un paradigma WYSIWYG inserendo però alcuni vincoli sulla disparità statistica tra gruppi.

La classificazione di questi due approcci è comunque utile ad evidenziare le differenti visioni del mondo implicite nelle varie definizioni di fairness, cioè come queste visioni vengano codificate dagli specifici strumenti matematici utilizzati. In questo senso ci interesseremo in particolare agli elementi di incompatibilità tra i diversi modelli, in modo da evidenziare gli elementi sostanziali di queste costruzioni di senso:

Anche se abbiamo introdotto questi due assiomi come differenti visioni del mondo o sistemi di credenze, questi possono anche essere una scelta strategica. Qualsiasi sia la motivazione (che è in fin dei conti matematicamente irrilevante), la scelta degli assiomi è fondamentale per un processo di decisione. Gli assiomi scelti determinano il significato di fairness (Friedler *et al.*, 2016, 9)¹⁰.

Possiamo collegare queste analisi con lo studio del concetto di "struttura" sviluppato in (Crammond, Carey, 2016). Gli autori si soffermano sulla relazione tra il concetto di disuguaglianza strutturale e

¹⁰ Traduzione dell'autore.

quello di *habitus* in Bourdieu, nel caso particolare delle disuguaglianze sanitarie, rilevando l'esistenza di alcuni discorsi socialmente pervasivi (com'è il caso della nozione di "rischio" associata con l'abitudine di fumare) che strutturano l'*habitus* individuale senza determinare univocamente i comportamenti.

Nel nostro studio della fairness rileviamo la presenza di diversi discorsi con differenti livelli di pervasività. La fairness, o complementariamente la discriminazione, è un concetto non aggirabile all'interno dello studio di algoritmi; questo primo elemento può venir sviluppato secondo differenti metodi (come quelli WYSIWYG e WAE che abbiamo descritto sopra) che vanno quindi a specificarne il significato; ognuno di questi approcci genera conseguentemente un proprio campo di pervasività che abbiamo caratterizzato con alcune assunzioni fondative: ad esempio WYSIWYG suppone l'esistenza a priori di una classificazione individuale corretta, mentre WAE assume la condizione di parità statistica tra gruppi. Una prospettiva di questo tipo permette di approfondire in che modo il dispositivo algoritmico influisce sull'*habitus*, e quindi anche di rilevare gli aspetti di incompatibilità tra varie e compresenti nozioni pervasive.

5. LA DISCRIMINAZIONE ALL'INTERNO DELLA "GOVERNANCE BY THE NUMBERS"

Yarden Katz (2017) ipotizza che la centralità assunta nell'ultima decade dal termine Intelligenza Artificiale (in ambito accademico, industriale, politico, sui media, etc.) sia spiegabile in particolare attraverso il tentativo di affermare alcune agende politiche di stampo neoliberista. Nella sua trattazione Katz utilizza il concetto di "*governance by the numbers*" ripreso da (Supiot, 2012): questa locuzione sta ad indicare l'introduzione di metriche quantitative in numerosi ambiti sociali, dal lavoro, alla selezione scolastica, al sistema giuridico, fino alle valutazioni di rischio per la concessione di prestiti bancari o la stipula di contratti assicurativi; attraverso queste metriche è così possibile istituire forme di governance che funzionano sulla base di graduatorie e soglie di inclusione. Da questo punto di vista la cultura dell'Intelligenza Artificiale, e la connessa proliferazione di algoritmi valutativi e di selezione, hanno la funzione di rendere efficaci e conosciute queste metriche quantitative, oltre che di disciplinare la popolazione al loro utilizzo.

Una caratteristica fondamentale individuata da Katz nel processo di diffusione di procedure algoritmiche, è la creazione di un punto di vista

esterno (“*vision from nowhere*” nell’articolo) rispetto allo specifico contesto sociale in cui l’algoritmo introduce la sua metrica; da questo punto d’osservazione sarebbe possibile una descrizione oggettiva del terreno sociale su cui l’algoritmo agisce, ed è proprio questa pretesa di oggettività che fornisce al processo di selezione una naturalizzazione e dunque una legittimità sociale e politica. Riconosciamo in questo meccanismo il funzionamento di un dispositivo di governo secondo l’utilizzo del concetto che ne fa Foucault. La “natura” della popolazione diviene una tecnica di potere, e questo meccanismo diviene produttivo della popolazione stessa, cioè «Il termine dispositivo nomina ciò in cui e attraverso cui si realizza una pura attività di governo senza alcun fondamento nell’essere. Per questo i dispositivi devono sempre implicare un processo di soggettivazione, devono, cioè, produrre il loro soggetto» (Agamben, 2006, 16).

In realtà la “*vision from nowhere*” corrisponde ad un processo di invisibilizzazione delle condizioni contingenti in cui l’algoritmo è stato concepito, dei contesti materiali, storici, sociali a cui deve attenersi, degli interessi di gruppo che hanno agito per la sua diffusione, del sistema assiomatico che gli permette di funzionare e della costruzione di senso codificata in questo sistema assiomatico. L’algoritmo è legittimato in quanto oggettivo, ma è allora necessario riconcettualizzare questo termine, come suggerito in Mazzotti (2015), secondo termini durkehimiani: «per un concetto matematico essere oggettivo significa essere istituzionalizzato, inserito in una rete di concetti e pratiche supportate dagli interessi collettivi di un gruppo» (Mazzotti, 2015, 467).

La marginalizzazione del formalismo matematico necessario per la definizione di una procedura valutativa è strettamente necessaria alla costruzione fittizia dell’esternità della “*vision from nowhere*”, perché vengono nascosti proprio i vincoli metodologici ed assiomatici a cui l’algoritmo è interno.

Nel momento in cui il tema della fairness viene inserito tra le caratteristiche formalizzate di un algoritmo, anche il problema della discriminazione rientra potenzialmente all’interno di quel paradigma oggettivista che legittima la governance statistica, e l’eliminazione della discriminazione si trasforma da problema politico a problema di ottimizzazione numerica. Attraverso i punti che abbiamo sollevato nelle sezioni precedenti, vogliamo invece mostrare che la costruzione matematica di un’idea di fairness avviene sempre all’interno di uno specifico modello assiomatico calato in un insieme di condizioni materiali a cui deve adattarsi, e ogni paradigma universalista nasconde inevitabilmente la contingenza di questi vincoli. Mettendo in luce le

frizioni tra le diverse modellizzazioni matematiche della discriminazione, e quelle con l'attuale organizzazione sociale, è possibile evidenziare le forme di discriminazione strutturale soggiacenti ai sistemi di governance.

Ha particolare utilità riprendere le considerazioni di Longo sul rapporto tra matematica e costruzione di conoscenza, e più precisamente tra geometria ed organizzazione dello spazio, per cercare di fare un'analogia con il rapporto tra modelli statistici ed interpretazione sociologica della discriminazione: «ogni discorso sull' "esistenza" di una struttura matematica [...] è fuori luogo» (2014, 6); così come le strutture materiali della visione non spiegano di per sé i concetti geometrici di linea e bordo, allo stesso modo le condizioni storiche della discriminazione non spiegano di per sé i diversi modelli di fairness algoritmica; piuttosto le condizioni di possibilità storiche permettono la definizione nella «comunità simbolica» (Ibidem) di modelli matematici di discriminazione in grado di organizzare lo spazio sociale tramite specifici strumenti, che nel contesto attuale assumono nella maggioranza dei casi la forma di dispositivi governamentali.

L'esempio riportato da Courtland (2018) che abbiamo descritto nella sezione 4 risulta particolarmente chiarificatore di questo rapporto con le condizioni materiali e dei rapporti di reciproca (in)compatibilità e frizione tra modelli matematici distinti. Se un dispositivo algoritmico viene concepito per assistere il processo giudiziario, la retorica che legittima la "governance by the numbers" descriverà il problema dell'eliminazione della discriminazione come l'ottimizzazione di una specifica funzione. La trattazione che abbiamo fatto ci permette però di osservare che separando il processo giudiziario dalle prassi concrete di controllo e repressione, si deve accettare come vincolo strutturale la discriminazione negli arresti, e questo pone una condizione di incompatibilità tra due modelli di fairness. L'applicazione del modello di valutazione si fonda quindi su una serie di scelte politiche multilivello (separazione tra giudiziario e amministrazione delle pratiche di controllo, applicazione dell'algoritmo al campo giudiziario, scelta di un modello di fairness tra i due modelli incompatibili) che vengono negate se il problema della discriminazione si pone meramente come problema di ottimizzazione.

Non si possono inoltre trascurare gli effetti di riproduzione della discriminazione inerenti a questa dinamica: come detto nella sezione precedente, se l'algoritmo giudiziario è modellizzato per approssimarsi il più possibile all'uguaglianza delle probabilità (una nozione di fairness che viene preferita in ambito giudiziario), allora la discriminazione di

gruppo riprodotta dall'algoritmo sarà direttamente correlata alla discriminazione di gruppo attuata dalla polizia, e l'algoritmo avrà come effetto quello di propagare le discriminazioni strutturali. Più in generale questo fenomeno si inserisce all'interno di tutti quegli effetti di performatività propri ad ogni procedura formalizzata. Nonostante non sia l'oggetto di questa trattazione, ci preme ricordare che la stessa definizione degli insiemi di valori possibili per ogni attributo codifica una normazione dello spazio sociale: quando all'attributo "genere" vengono assegnati due valori possibili (o comunque un insieme di valori prestabiliti), si escludono dalla trattazione tutte le soggettività che si identificano in maniera non binaria, e questo oltre ad avere degli effetti di invisibilizzazione immediati, produce forme di discriminazione nella misura in cui la raccolta di dati ha concrete conseguenze politiche, sociali, lavorative, etc.

Come abbiamo provato a spiegare fino a qui, è necessario considerare il modello assiomatico alla base di un algoritmo come facente parte dell'algoritmo stesso per dare una cornice di senso alla sua azione. Nella retorica oggi più affermata, i processi automatizzati sono invece dei sistemi di calcolo più o meno interpretabili, il cui obiettivo è ridotto alla descrizione tramite il calcolo di un ambiente sociale autonomamente provvisto di senso. Più precisamente, le metriche che permettono la misurazione della realtà vengono sempre più spesso naturalizzate, nascondendo l'atto del dispiegamento di queste metriche come creativo di informazione. Ad essere oscurato è anche il fatto che la stessa cultura degli algoritmi favorisce in molti casi il dispiegamento di queste metriche, incanalando forme di disciplinamento collettivo e qualificandosi quindi anch'essa come dispositivo governamentale.

Cerchiamo di osservare in che modo gli specifici quadri assiomatici aderiscono al contesto materiale nel caso dei due modelli di fairness che abbiamo visto nella sezione 4:

- la nozione di fairness individuale formalizzata con l'uguaglianza delle possibilità, e l'uguaglianza delle probabilità, si fonda su un'idea di ranking individuale "giusto": la metrica specifica di questo ranking dipende dalle condizioni del dataset e dall'algoritmo utilizzato, e una condizione del tipo descritto (l'indipendenza tra \hat{Y} e l'attributo sensibile X , condizionata alla conoscenza del vero valore Y) permette l'esistenza di una classificazione veritiera basata sull'attributo Y . Nel caso descritto in Courtland (2018), sono le pratiche di controllo della polizia a determinare l'attributo Y , e quindi appare chiaramente come una condizione frutto di contingenze potenzialmente discriminanti sia incorporata nella metrica;

- nel caso della parità statistica, invece, la metrica utilizzata è quella della differenza media (o dell'impact ratio) tra le probabilità di successo per i differenti gruppi (ogni gruppo è caratterizzato dal valore dell'attributo sensibile). In questo caso le condizioni materiali in cui l'algoritmo si inserisce sono descritte dal dataset, e il processo di valutazione cerca di approssimare una condizione di parità statistica. Nel caso di Courtland (2018) un meccanismo di parità statistica in campo giudiziario può essere approssimato anche in presenza di dataset con discriminazione (prodotta dalle pratiche di controllo e repressione) a costo di rinunciare alle forme di fairness come uguaglianza delle probabilità. In questo senso l'algoritmo ha la funzione di correggere in sede giudiziaria la discriminazione del dataset creato dalle prassi della polizia.

Da entrambi questi esempi si ricava che il tema della fairness, se posto come semplice ottimizzazione di una funzione di discriminazione, rientra completamente dentro il paradigma del dispositivo di "governance by the numbers". Inoltre, vista la centralità assunta dalle battaglie contro la discriminazione, la qualificazione di un algoritmo come "algoritmo fair", fornisce una legittimazione degli stessi dispositivi di governo anche quando questi dispositivi operano a partire da condizioni strutturali in cui sono presenti forti discriminazioni, riproducendole.

L'emergenza storica del concetto di discriminazione, frutto di un percorso tortuoso ed in particolare di conflitti attorno ai temi del razzismo e delle discriminazioni di genere, non è oggetto di questa trattazione, ci preme però ripetere che questo tema raggruppa una serie di significati originati storicamente, che non possono essere ridotti alla loro formalizzazione aritmetica. In questo senso è utile ricordare l'esempio della "fairness come unawareness" di cui abbiamo parlato nella sezione 3: se assumiamo assiomaticamente questa definizione di fairness, chiaramente è possibile eliminare completamente la discriminazione algoritmica, semplicemente costruendo database che non contengono gli attributi sensibili; questo avviene a costo di una separazione radicale tra il significato storico della discriminazione e la sua formalizzazione. Questo processo di separazione è presente, in diverse maniere, dentro ogni modellizzazione formale, e quindi le interpretazioni degli algoritmi come dispositivi neutri di analisi oggettiva della realtà, attuano effettivamente un'amputazione del processo storico contenuto nel concetto di discriminazione. Guardare il tema della discriminazione attraverso strumenti come gli algoritmi di decisione, dal nostro punto di vista significa piuttosto capire in che

modo le forme di discriminazione strutturale interagiscono con questi sempre più diffusi strumenti di governo.

Ci sembra importante porre l'accento anche su alcune scelte terminologiche che indirizzano la percezione collettiva del problema: è inadeguato riferirsi alla ricerca di algoritmi privi di bias o *unbiased*, perché queste locuzioni suggeriscono la presenza di contesti di selezione completamente "giusti" a cui ogni algoritmo può avvicinarsi per approssimazione. Il nostro interesse è piuttosto orientato sull'organizzazione di senso (la visione del mondo) specifica che viene codificata in ogni definizione di fairness, nei rapporti concreti di relazione e frizione che vengono a formarsi tra modelli distinti, e con la realtà sociale.

Oggi ogni opzione di governo iscrive le sue motivazioni ideologiche anche all'interno delle modellizzazioni matematiche proprie alla "governance by the numbers", lo studio della discriminazione strutturale deve dunque porre il problema del significato politico di queste modellizzazioni per poter portare una critica profonda al ruolo degli algoritmi. Per lo stesso motivo queste analisi sono anche uno strumento irrinunciabile di quei movimenti politici e culturali che portano una critica generale ad ogni sistema fondato su forme di discriminazione strutturale.

6. CONCLUSIONI

La narrazione oggi più diffusa attorno ai processi di decisione algoritmica pone il problema della fairness come l'insieme degli strumenti di calcolo necessari ad ottimizzare alcune funzioni che misurerebbero il livello di discriminazione oggettiva. La nostra proposta di ricerca, che abbiamo provato embrionalmente a sviluppare in questo lavoro, suggerisce invece di ricercare il rapporto tra fairness ed algoritmi nella complessa rete di compatibilità ed incompatibilità che esistono tra i modelli matematici con cui è possibile descrivere la discriminazione, e le strutture sociali in cui la discriminazione strutturale è emersa storicamente e come problema politico, soprattutto in conseguenza delle lotte messe in atto dai soggetti oppressi. Ci siamo qui concentrati in particolare su un caso di studio che evidenzia il rapporto tra diversi modelli di fairness applicabili al campo degli algoritmi giuridici, e la presenza di una discriminazione razziale strutturale nelle pratiche di polizia; questo stesso approccio potrebbe essere generalizzato a numerosi campi come la selezione scolastica e lavorativa, l'accesso al credito bancario, l'analisi del linguaggio, il

gender salary gap, etc.

Il nostro lavoro si fonda sull'assunzione che ogni analisi efficace della discriminazione debba saper individuare in che modo i dispositivi di governo vigenti si iscrivono in un quadro strutturalmente discriminante, ed in che modo lo riproducono. In un'epoca in cui i modelli statistici e gli algoritmi sono centrali per le pratiche di governo, la nostra ricerca vuole anche essere uno strumento per portare nel campo delle decisioni automatizzate una critica ai meccanismi che veicolano e riproducono ogni forma di discriminazione.

RIFERIMENTI BIBLIOGRAFICI

- AGAMBEN, G. (2006). *Che cos'è un dispositivo?*. Roma: Nottetempo.
- ANGWIN, J., LARSON, J., MATTU, S., KIRCHNER, L. (2016). *Machine bias. there's software used across the country to predict future criminals. and it's biased against blacks*. *Propublica*, May 23.
- BAROCAS, S., SELBST, A. (2016). Big data's disparate impact. *California Law Review*, 104, 671-732.
- CHEN, L., MA, R., HANNAK, A., WILSON, C. (2018). Investigating the impact of gender on rank in resume search engines. *Proceedings of the 2018 chi conference on human factors in computing systems*, 651, 1-14.
- CORBETT-DAVIES, S., PIERSON, E., FELLER, A., GOEL, S., HUQ, A. (2017). Algorithmic decision making and the cost of fairness. *KDD '17 Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797-806.
- COURTLAND, R. (2018). Bias detectives: the researchers striving to make algorithms fair. *Nature*, 558, 357-360.
- CRAMMOND, B., CAREY, G. (2016). What do we mean by 'structure' when we talk about structural influences on the social determinants of health inequalities?. *Social Theory & Health*, 15(1), 1-15.
- CURCIO, A., MELLINO, M. (2012). *La razza al lavoro*. Roma: Manifestolibri.
- DWORK, C., HARDT, M., PITASSI, T., REINGOLD, O., ZEMEL, R. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214-226.
- FRIEDLER, S. A., SCHEIDEGGER, C. E., VENKATASUBRAMANIAN, S. (2016). On the (im)possibility of fairness. *arXiv.org*, 1609.07236
- HARDT, M., PRICE, E., SREBRO, N. (2016). Equality of opportunity in supervised learning. *Proceedings of the 30th international conference*
-

- on neural information processing systems*, 3323-3331.
- KAMIRAN, F., CALDERS, T. (2009, Feb). Classifying without discriminating. *2009 2nd international conference on computer, control and communication*, 1-6.
- KAMIRAN, F., CALDERS, T. (2010). Classification with no discrimination by preferential sampling. *Informal proceedings of the 19th annual machine learning conference of belgium and the netherlands (Benelearn '10, Leuven, Belgium, may 27-28, 2010)*, 1-6.
- KATZ, H. (2017). *Manufacturing an Artificial Intelligence Revolution: Neoliberalism and the 'new' big data* Yarden Katz. Harvard: Harvard University.
- KLEINBERG, J., LUDWIG, J., Mullainathan, S., Rambachan, A. (2018). Algorithmic fairness. *AEA Papers and Proceedings*, 108, 22-27.
- LONGO, G. (2014). Le conseguenze della filosofia. In R. Lanfredini (a cura di), *A Plea for Balance in Philosophy. Essays in Honour of Paolo Parrini. Vol.2: New Contributions and Replies* (pp. 17-44). Pisa: ETS.
- MAZZOTTI, M. (2015). Per una sociologia degli algoritmi. *Rassegna Italiana di Sociologia*, 3-4, 465-478.
- PEDRESCHI, D., RUGGIERI, S., TURINI, F. (2008). Discrimination-aware data mining. *Proceedings of the 14th acm sigkdd international conference on knowledge discovery and data mining*, 560-568.
- PEDRESCHI, D., RUGGIERI, S., TURINI, F. (2009). Measuring discrimination in socially-sensitive decision records. *Proceedings of the SIAM International Conference on Data Mining Sdm*, 581-592.
- ROMEI, A., RUGGIERI, S. (2014). A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review*, 29(5), 582-638.
- SONNAD, N. (2017). Google translate's gender bias pairs "he" with "hardworking" and "she" with lazy, and other examples. *Quartz*, November 29.
- SUPIOT, A. (2012). *The spirit of Philadelphia: Social justice vs. the total market*. London: Verso.
- VAGNARELLI, G. (2017). Foucault e i confini del governo: la governamentalità. *Heteroglossia*, 15, 145-164.
- ŽLIOBAITÉ, I. (2015). A survey on measuring indirect discrimination in machine learning. *arXiv.org*, 1511.00148.
-

LA COSTRUZIONE SOCIO-TECNICA DEGLI ALGORITMI

Una ricerca nelle infrastrutture di dati

di *Biagio Aragona, Cristiano Felaco**

Abstract

The paper presents an empirical research conducted within four European data infrastructures with the aim to examine the socio-technical data assemblages that produce algorithms. In this respect, we questioned experts and data team working in these infrastructures using focus groups and interviews with the intent to flesh out our understanding of the nature of algorithms, and to identify and deconstruct their own components.

Keywords

Algorithm, Data assemblage, Data infrastructure, Qualitative research

* BIAGIO ARAGONA è Ricercatore di sociologia generale presso il Dipartimento di Scienze Sociali dell'Università degli Studi di Napoli Federico II. Email: aragona@unina.it

CRISTIANO FELACO è Assegnista di ricerca post-doc presso il Dipartimento di Scienze Sociali dell'Università degli Studi di Napoli Federico II. Email: cristiano.felaco@unina.it

1. INTRODUCTION

Di fronte alla disponibilità di una grande quantità di dati sulle più varie attività delle persone, dai viaggi al lavoro, dalla comunicazione ai consumi, gli algoritmi che governano i software sono diventati uno strumento fondamentale in qualsiasi settore della società (Steiner, 2012; Diakopoulos, 2013; Kushner, 2013).

Obiettivo principale di questo lavoro è comprendere la natura degli algoritmi attraverso l'individuazione e la ricostruzione dei tipi di dinamiche e di interazione tra i diversi attori umani e non – esperti, tecnici, ricercatori e apparati tecnici e strumentazioni tecnologiche – che intervengono nella costruzione dell'algoritmo stesso. È una posizione, questa, che va oltre la considerazione dell'algoritmo come un mero strumento tecnico, ma che, identificandosi nel filone “critical algorithm studies” (Seaver, 2013; Gillespie, Seaver, 2016), riconosce l'algoritmo come l'esito dell'azione di più elementi che, interagendo tra loro, vanno a formare un sistema più complesso. A riguardo, Kitchin (2017) osserva che uno dei modi possibili per condurre studi empirici sugli algoritmi è esaminare il loro più complesso assemblaggio socio-tecnico, inteso come un sistema composto da diversi apparati di natura tecnica e sociale che sono inestricabilmente intrecciati e che vanno a definire la produzione dei dati. In questa prospettiva, gli algoritmi sono il prodotto combinato di diversi apparati, di molteplici tecniche analitiche e di varie comunità di esperti in competizione che rende opaca la loro origine e il loro funzionamento (Burrell, 2015).

Le logiche di lavoro di un algoritmo non sono direttamente conoscibili, sono racchiuse nella “scatola nera” che li comprende, ed è possibile accedere, quindi, solo ai dati iniziali e ai risultati, ma non all'intero processo. Per comprendere gli elementi che costituiscono un algoritmo e ricostruire le diverse interrelazioni tra essi, diventa necessario aprire la scatola nera che li governa e scomporre il processo in cui questi algoritmi sono disegnati e usati (Seaver, 2013, 2017). In questo modo si rende possibile la comprensione dei problemi, degli attori coinvolti, ed è possibile ricostruire le scelte, le negoziazioni e vincoli che vanno a contribuire alla formazione dell'algoritmo.

Da questi presupposti muove la ricerca sugli algoritmi, la quale nasce da un progetto più ampio, denominato *B-Data (Big Data Assemblages: Techniques and Actors)*, che ha l'obiettivo di analizzare gli assemblaggi dati che producono e usano open data e big data all'interno delle infrastrutture dati in Europa. In questa sede, viene presentata la parte del progetto dedicata allo studio degli algoritmi e alla

comprensione del complesso di elementi che costituiscono il loro funzionamento. Lo scopo è esaminare gli assemblaggi socio-tecnici che producono algoritmi all'interno di quattro infrastrutture dati in Europa, prestando particolare attenzione alle componenti sia umane (code team e sviluppatori, esperti di dominio, utenti e stakeholders), sia non umane (piattaforme, hardware, software, ecc.) che agiscono e co-partecipano alla formazione degli algoritmi stessi. Poiché gli assemblaggi dati sono fluidi e pertanto è difficile stabilire i loro confini, le infrastrutture dati rappresentano degli spazi fisici in cui ha luogo l'assemblaggio dati e, quindi, punti privilegiati di osservazione degli algoritmi. Nello specifico, le infrastrutture dati europee selezionate per la ricerca sono quattro: Web Science Institute (Wsi), Istituto nazionale di statistica (Istat), Centro norvegese per i dati di ricerca (Nsd) e Consorzio degli archivi dati europei delle scienze sociali (Cessda). All'interno di questi centri, si è scelto di intervistare esperti che si occupano della costruzione degli algoritmi e che, più in generale, sono direttamente coinvolti nel lavoro di assemblaggio dati. Nel dettaglio, sono stati intervistati i direttori e responsabili di settore delle quattro infrastrutture di dati al fine di ottenere una riflessione critica sull'assemblaggio in generale e sui suoi apparati. Oltre alle interviste, sono stati condotti focus group con i membri di data team provenienti da differenti settori e con competenze diverse allo scopo di ottenere una ricostruzione delle dinamiche, delle attività e delle relazioni tra attori umani e non umani che hanno luogo all'interno di queste infrastrutture dati, e che entrano in gioco nel disegno, sviluppo e impiego di un algoritmo.

2. L'ALGORITMO COME SEQUENZA DI ISTRUZIONI

All'interno della letteratura informatica, gli algoritmi vengono comunemente identificati con macchine astratte, modelli matematici e metodi computazionali (Moschovakis, 2001).

Miyazaki (2012) fa risalire l'origine del termine al latino *algorithmus*, che a sua volta deriva dal nome di un matematico arabo del IX secolo, Muḥammad ibn Mūsā al-Khwārizmī, ritenuto fra i primi a teorizzare il concetto, i cui scritti consistevano in una serie di operazioni di calcolo tradotte in formule matematiche. Adottato nella matematica e nella logica moderna, l'algoritmo ha assunto il significato di un procedimento di calcolo descrivibile attraverso un certo numero di formule (Steiner, 2012). Soltanto intorno alla metà del XX secolo, con l'affermazione dell'informatica e lo sviluppo dei primi linguaggi di programmazione, l'algoritmo verrà concepito nell'accezione moderna di

insieme specifico di istruzioni che, a partire da determinati input, porta alla creazione di risultati desiderati (Miyazaki, 2012) e, quindi, al raggiungimento di un determinato risultato o alla risoluzione di un dato problema.

Ogni tentativo di definire in modo univoco gli algoritmi (Gurevich, 2011) è stato contrastato dalla loro evoluzione. Si è passati nel corso del tempo da quelli sequenziali a quelli interattivi, real-time, ibridi, ecc. Dal punto di vista meramente tecnico, questi strumenti sono definiti come una raccolta di semplici istruzioni per svolgere alcune attività (Sipser, 2006) o, più nello specifico, una procedura governata da precise istruzioni e definita da una serie di passaggi (Berlinski, 2000), nonché un modo per descrivere in maniera astratta e formalizzata una procedura computazionale (Dourish, 2016). In questa prospettiva, un algoritmo nasce successivamente alla creazione di un modello, vale a dire dopo aver formalizzato il problema e tradotto gli obiettivi in termini computazionali (Gillespie, 2014). Kowalski (1979) argomenta a riguardo che l'algoritmo deriva dalla separazione tra logica e controllo: la componente logica specifica il dominio di conoscenza da utilizzare per la formulazione di una soluzione a determinati problemi, mentre la componente di controllo determina le strategie di *problem-solving* attraverso cui tale conoscenza viene applicata. In altri termini, la logica specifica sul piano teorico ciò che deve essere fatto, mentre il controllo definisce come sarà operativamente fatto. Da ciò ne deriva che l'efficienza di un algoritmo può esser migliorata modificando gli aspetti di controllo, sia prevedendo ulteriori passaggi e includendo nuovi dati strutturati (input) per elaborare l'impianto logico su cui esso si basa, sia agendo sul paradigma logico sottostante (Kowalski, 1979). Quest'ultimo caso implica non tanto un cambiamento di strategia operativa, quanto piuttosto un approccio completamente nuovo per risolvere algoritmicamente il problema in questione.

Almeno dal punto di vista teorico, la formulazione di un algoritmo ha un'esistenza autonoma rispetto ai linguaggi di programmazione e alle macchine usate per poterlo eseguire (Goffey, 2008). Tuttavia, il rapporto tra logica e controllo non è sempre diretto. Alcuni costrutti teorici possono essere espressi come algoritmi nella forma di equazioni matematiche, mentre, in casi più complessi è necessario dapprima tradurre un problema in una serie di istruzioni (pseudo-codici) che devono essere poi codificate (Goffey, 2008). Si tratta, in quest'ultimo caso, di operare una duplice traduzione: una formalizzazione del problema (piano logico) e una traduzione di tale logica in una serie strutturata di istruzioni e di passaggi che vanno a determinare il

funzionamento dell'algoritmo stesso (piano del controllo). Nell'ambito dell'informatica, quindi, i processi di traduzione sono descritti principalmente come tecnici e formali, in cui il focus è incentrato sulla progettazione dell'algoritmo e sulla ricerca di una maggiore efficienza dal punto di vista tecnico (Seaver, 2013).

3. L'ALGORITMO COME STRUMENTO SOCIO-TECNICO

Diversi autori si discostano dalla visione tecnicista, a favore di un approccio più complesso che riconosce il ruolo cruciale che gli algoritmi svolgono nelle tradizionali istituzioni sociali, politiche ed economiche – come, ad esempio, per prevenire e contrastare la criminalità nell'ambito della *predictive policing* o per supportare le scelte riguardo le assunzioni e i licenziamenti nel mondo lavorativo – e, più in generale, la loro influenza sulla realtà sociale (O'Neil, 2006; boyd, Crawford, 2012; Nakamura, 2013; Grosser, 2014; Tufekci, 2015; Pasquale, 2015). Gli algoritmi non possono, perciò, essere concepiti esclusivamente come strumenti tecnici che lavorano in modo autonomo e isolato e, quindi, analizzati in maniera separata dal contesto in cui sono sviluppati e impiegati, escludendo così il più ampio assemblaggio di sistemi di pensiero, finanza, politica, infrastrutture, istituzioni, relazioni interpersonali, che danno forma alla loro produzione (Kitchin, 2014). Come è stato anticipato, nell'ambito dell'emergente filone di studi che va sotto l'etichetta “critical algorithm studies” vi è un diffuso accordo nel considerare l'algoritmo come un insieme più complesso di passaggi definiti per produrre specifici risultati in cui vanno ad intrecciarsi pratiche sociali e materiali che hanno una propria natura culturale, storica e istituzionale (Montfort *et al.*, 2012; Takhteyev, 2012; Napoli, 2013; Dourish, 2016; De Rosa, Aragona, 2017). Tali studiosi contestano l'obiettività, l'imparzialità e la conseguente pretesa di affidabilità degli algoritmi, sottolineando che gli stessi codici non sono mere astrazioni, ma hanno una valenza anche sociale e politica in quanto un programmatore, per quanto si sforzi a rimanere distaccato e imparziale nella creazione di un algoritmo, porterà con sé il proprio bagaglio personale di conoscenze e i propri modelli sia di pensiero sia culturali (Porter, 1995; Gillespie, 2014); quest'ultimi a loro volta, andranno inevitabilmente a fondersi nei processi di calcolo o di traduzione. E non solo, la progettazione di un algoritmo è anche vincolata alle risorse disponibili e alla qualità dei dati, dal complesso di requisiti che ne regolamentano l'impiego (standard, protocolli e leggi in vigore), oltre che dagli strumenti strettamente tecnologici, come

piattaforme, hardware, software e infrastrutture (Kitchin, Dodge, 2011; Diakopoulos, 2013; Drucker, 2013; Neyland, 2015). In altri termini, la programmazione di un algoritmo è un processo non lineare e in continuo divenire in quanto è l'esito di più tentativi ed errori, discussioni e negoziazioni che possono aver luogo in diversi momenti storici e spaziali (Kitchin, Dodge, 2011), e i cui risultati non sono infallibili (Gillespie, 2014; Neyland, 2015).

In tale prospettiva, l'azione degli algoritmi viene interpretata come parte di assemblaggi molto più ampi e complessi di azioni e decisioni, definiti come «Complessi sistemi socio-tecnici composti da più apparati ed elementi che sono profondamente intrecciati, il cui scopo principale è la produzione dei dati» (Kitchin, Lauriault, 2014, p.6). Gillespie (2014) ricorre alla figura retorica della *sineddoche* per mostrare con efficacia la natura socio-tecnica degli algoritmi, visti come sistemi più complessi di assemblaggio in cui l'algoritmo stesso, i modelli di riferimento, gli obiettivi e i dati, gli strumenti tecnologici, sono collegati e intrecciati con il lavoro dei tecnici e degli esperti che discutono sui modelli utilizzati, che sono impegnati nel trattamento e nella gestione dei dati, che progettano e applicano gli algoritmi in specifici contesti. Ne deriva pertanto che tali strumenti non sono neutrali, la loro produzione ha un impatto anche sulla formazione della percezione pubblica e solleva importanti questioni sul piano delle implicazioni normative (Anderson, 2011; Barocas, Hood, Ziewitz, 2013). Una parte importante delle informazioni che si sperimentano in rete sono infatti generate da un algoritmo, si pensi all'uso dei motori di ricerca che è diventata una routine per rintracciare notizie o informazioni, oppure i video e i contenuti dei social media che sono quasi sempre ordinati in modo personalizzato attraverso un algoritmo che analizza i dati demografici e il comportamento degli utenti (Sandvig, 2015). Questi strumenti hanno profondamente influenzato il modo di operare di specifici settori: dal mondo della finanza, stravolgendo i criteri tradizionali di valutazione del rischio e classificazione dei clienti delle agenzie di credito (Pasquale, 2015; Arnoldi, 2016), a quello, ad esempio, della ricerca e del lavoro di catalogazione, orientando, come mostra Geiger (2017) nel suo studio su Wikipedia, la creazione di nuovi articoli o la modifica di quelli già esistenti.

Oltre a questa funzione di classificazione, gli algoritmi sono anche utilizzati per regolare specifici interventi. Un esempio a riguardo è l'*anticipatory governance*, una forma di profilazione che viene impiegata per prevedere comportamenti futuri e guidare gli interventi da effettuare (Kitchin, Lauriault, 2014). Una procedura, questa, sovente

utilizzata per regolare i controlli negli aeroporti, che produce una classificazione dei passeggeri per i livelli di pericolosità in base alla quale viene stabilito il tipo di trattamento da riservar loro ai diversi controlli (Dodge, Kitchin, 2004; Amoore, 2006). Un altro esempio di delega del ruolo decisionale ad un algoritmo è relativo alla procedura di assegnazione delle cattedre ai docenti in Italia (De Rosa, Aragona, 2017). Il Ministero dell'istruzione, dell'università e della ricerca (Miur) ha implementato un algoritmo per stabilire la destinazione finale degli insegnanti tenendo conto delle posizioni disponibili, delle preferenze e dei punteggi degli insegnanti (titoli, situazioni speciali come familiari con disabilità, mobilità professionale, ecc.) e una serie di variabili contestuali (posti vacanti nella propria scuola, livello scolastico, età, ecc.). Al termine del processo, i risultati della mobilità si rivelarono preoccupanti, in quanto migliaia di insegnanti furono trasferiti a centinaia di chilometri dalla propria provincia di residenza. Questo rappresenta anche un esempio di quanto questi strumenti non siano infallibili e del rischio in cui si può incorrere nell'affidare completamente ad essi gli esiti decisionali.

Sono solo alcuni esempi, questi, che testimoniano l'uso sempre più massiccio degli algoritmi per disciplinare, regolare e controllare diversi aspetti della vita sociale, un cambiamento che ha spinto numerosi studiosi a sostenere l'avvicendamento di una nuova era, quella della "governance algoritmica" (Musiani, 2013), che si basa su processi automatizzati, automatici e autonomi che escludono l'azione umana (Kitchin, Dodge, 2011). È un tipo di potere non sempre trasparente e lineare sia perché gli algoritmi operano in un sistema più ampio di relazioni in cui agiscono contemporaneamente diversi elementi, sia perché possono verificarsi effetti collaterali dovuti alla mancanza di sorveglianza umana o, ancora, errori umani nella scrittura del codice (Diakopoulos, 2013; Kitchin, 2017). A questi effetti va aggiunta anche l'azione degli utenti finali che fanno un uso personalizzato di questi strumenti, incorporando e modellando nella propria vita quotidiana la tecnologia (Gillespie, 2014).

Non sempre accessibili, inoltre, sono le operazioni e i codici alla base degli algoritmi (Geiger, 2017). C'è da chiedersi, quindi, fino a che punto sia possibile accedere alle operazioni effettuate da questi strumenti che sembrano essenzialmente opache (Burrell, 2015) e, quindi, ricostruire il complesso di dinamiche e di interazioni tra i diversi attori (umani e non umani) che co-partecipano alla costruzione dell'algoritmo stesso. A tale scopo, diventa necessario aprire la "scatola nera" che li governa e scomporre il processo in cui sono disegnati e

sviluppati (Seaver, 2013). Il termine “scatola nera” è usato nell’ambito della cibernetica per indicare una conoscenza parziale di un processo, di cui si conoscono soltanto input e output. Secondo Pasquale (2015), il termine “scatola nera” ha una doppia valenza: può riferirsi ad un dispositivo di registrazione dati, come ad esempio i sistemi di monitoraggio dei dati sulla mobilità attraverso aerei, treni e automobili, oppure può indicare un sistema le cui logiche di lavoro sono opache, si conosce solo l’insieme di dati iniziali e i risultati, ma non il processo. Aprire la scatola nera permette di avere accesso ai processi di elaborazione e di comprendere problemi, attori coinvolti, e ricostruire scelte, compromessi, conflitti che hanno contribuito alla produzione di un algoritmo. In risposta a ciò, vari studiosi provenienti da diversi campi di ricerca hanno iniziato a porre attenzione sulla natura degli algoritmi più in generale (Geiger, 2014; Montfort *et al.*, 2012), sugli effetti che producono in specifici settori della società (Amoore, 2009; Pasquale, 2015) o, ancora più nello specifico, sulle procedure di costruzione dei codici e sul modo in cui operano (Gillespie, 2014; Seaver, 2013) al fine di decostruire la natura degli algoritmi, il loro modo di lavorare e di esercitare potere (Kitchin, 2017).

4. LA RICERCA NELLE INFRASTRUTTURE DATI

Studiare gli algoritmi nel più ampio assemblaggio dati implica l’analisi di diversi domini: gli oggetti materiali (le infrastrutture, piattaforme e, più in generale, il sistema tecnologico), la sfera del linguaggio (codici, regolamenti, istruzioni, ecc.) e gli attori sociali che sono direttamente o indirettamente coinvolti nell’assemblaggio (tecnici, utenti, committenti, ecc.). L’esame di queste tre componenti comporta delle difficoltà a livello empirico per via della fluidità e dei confini sfumati degli assemblaggi dati, oltre al fatto che i vari assemblaggi sono collegati fra loro in molti modi. Al fine di poter afferrare e analizzare gli assemblaggi dati in tutta la loro natura multiforme, una possibile soluzione è guardare ai luoghi in cui gli algoritmi sono creati. Come mostrano Aragona, Felaco e Marino (2018) in una recente ricerca sullo studio degli assemblaggi di big data, i centri di ricerca possono rappresentare uno spazio fisico in cui poter osservare il modo in cui convergono i diversi apparati ed elementi, compresi gli algoritmi. A tale scopo, sono state selezionate quattro infrastrutture di dati in Europa, in cui la produzione degli algoritmi è un’attività centrale per la produzione, utilizzo e condivisione di dati digitali: il Web Science Institute (Wsi), l’Istituto nazionale di statistica (Istat), il Centro norvegese per i dati di

ricerca (Nsd) e il Consorzio degli archivi dati europee delle scienze sociali (Cessda).

Le attività dei quattro centri selezionati sono influenzate da diversi fattori, che vanno dal governo, economia e politica, alla finanza e mercato. L'Istat introduce all'interno del programma annuale quelle attività statistiche richieste dagli apparati del governo locale e nazionale per la progettazione delle politiche pubbliche, allo stesso modo, il Wsi ha rapporti con il governo ed è sostenuto da partner nel governo, nelle imprese e nell'industria; l'Nsd è fondato dal Consiglio di ricerca norvegese allo scopo di facilitare l'accesso ai dati per la ricerca e, infine, il Cessda riunisce gli archivi di dati sulle scienze sociali in tutta Europa con l'intento di fornire servizi di dati su larga scala. Sebbene le *mission* siano diverse, questi centri condividono alcune caratteristiche comuni. Essi si servono di una nutrita schiera di tecnici ed esperti provenienti da diversi campi di ricerca, principalmente dal settore informatico, statistico, scienze sociali e giuridico; si trovano quotidianamente ad affrontare problemi tecnici ed etici nella produzione, gestione e analisi dei dati e, infine, si trovano costantemente a confrontare le proprie attività con le amministrazioni politiche, i cittadini, i ricercatori, e, quindi, con i vari stakeholder e utenti finali.

Al fine di ricostruire gli obiettivi conoscitivi e il complesso di interessi degli attori interni ed esterni che partecipano alla costruzione degli algoritmi, nella ricerca sono stati coinvolti non solo gli sviluppatori di software e il team di informatici e ingegneri impegnati nello sviluppo degli algoritmi, ma anche coloro che sono direttamente coinvolti nell'assemblaggio dati. Nel periodo tra ottobre 2017 e luglio 2018 sono state condotte 9 interviste tra direttori (2) e responsabili di sezione (7). La scelta è ricaduta sulla figura del direttore quando l'infrastruttura dati era caratterizzata al suo interno da una struttura meno complessa - cioè meno gerarchica e organizzata in poche sezioni operative - e sui responsabili di sezione, invece, quando il centro era di maggiori dimensioni e strutturato in maniera più gerarchica. I colloqui sono stati condotti attraverso interviste con un livello medio-basso di direttività e di standardizzazione al fine di incoraggiare una riflessione critica sui vari aspetti che compongono l'assemblaggio degli algoritmi e favorire la comprensione dei significati che gli attori intervistati attribuiscono alle proprie attività in base al ruolo svolto all'interno della struttura, alle proprie conoscenze e alla propria formazione. Insieme alle interviste, sono stati condotti tre focus group con i membri del data team che non ricoprivano ruoli apicali. Con l'obiettivo di avere gruppi eterogenei e con formazioni differenti (informatici, statistici, sociologi,

esperti legali in materia di protezione dati ecc.), sono stati invitati a partecipare ai focus gli esperti appartenenti alle sezioni di informatica, metodologia e aspetti legali e protezione dei dati. Nello specifico, all'interno dei focus group sono stati sottoposti diversi quesiti per far emergere i racconti esperienziali dei singoli attori riguardo diversi aspetti che co-partecipano all'assemblaggio di un algoritmo, stimolando, laddove necessario, l'intervento di alcuni partecipanti a fornire informazioni rispetto al proprio lavoro. Ciò ha permesso di ottenere una gamma ampia di punti di vista e di esplorare varie procedure da diverse prospettive, oltre a cogliere quelle dinamiche relazionali tra diverse comunità di esperti che si verificano nella progettazione di un algoritmo.

5. RISULTATI

La progettazione e la costruzione degli algoritmi è descritta come l'esito dell'azione di più elementi che, interagendo e a volte sovrapponendosi, vanno a definire un complesso sistema socio-tecnico. In generale, dall'analisi delle interviste è possibile isolare due tematiche principali a cui gli interlocutori fanno riferimento nel raccontare la propria esperienza all'interno delle infrastrutture dati selezionate: il co-design, cioè la partecipazione comune tra più figure nella realizzazione dell'algoritmo, e il rapporto tra algoritmo e conoscenza¹.

La costruzione dell'algoritmo non segue un percorso lineare, né è confinata nel dominio di conoscenza degli sviluppatori informatici, quanto piuttosto è influenzata dall'azione, dalle esigenze e dagli interessi degli attori che appartengono in maniera diretta o indiretta all'assemblaggio nel suo complesso. Il co-design è innanzitutto un'attività interdisciplinare. La scelta delle istruzioni alla base della progettazione di un algoritmo che sottende un software o una piattaforma, richiede il confronto tra diversi profili di esperti all'interno dell'infrastruttura, in riferimento all'oggetto in questione e alle specifiche priorità:

W: Il lavoro interdisciplinare implica attività di co-design [...] Nell'ambito del progetto [*per la creazione di una piattaforma*], gran parte del tempo lavorativo lo abbiamo trascorso seduti in una stanza per affrontare insieme i problemi. [...] Gran parte delle nostre discussioni hanno riguardato le priorità dei vari aspetti [...] c'è stata quindi tanta negoziazione.

¹ Per motivi di privacy e di confidenzialità sono stati omessi volutamente informazioni che permettessero l'identificazione dell'ente e dei partecipanti.

Non è facile avere a disposizione esperti con una formazione mista in grado di proporre una soluzione ad un problema che non si limiti all'ambito tecnologico, ma che inglobi una riflessione più ampia sui diversi aspetti del fenomeno da più prospettive. In tale ottica, il confronto tra esperti di diversi settori e con specifiche competenze si rivela una condizione necessaria soprattutto in assenza di risorse interne all'ente con una formazione mista, tale da inglobare le diverse competenze (informatiche, statistiche, metodologiche, sociologiche, giuridiche, ecc.) necessarie per la costruzione di un algoritmo. La copresenza necessaria di diverse figure include tuttavia anche la gestione delle dinamiche interne relative più strettamente ai rapporti interpersonali; tali dinamiche diventano cruciali soprattutto in assenza di esperti con una formazione sia teorica sia tecnica al punto che, di fronte a conflitti interni, si preferisce cambiare la "composizione dei gruppi" piuttosto che rinunciare al confronto tra più esperti per dare una soluzione algoritmicamente adeguata ad un problema:

V: Quando sviluppi le ontologie nel caso di *linked data* di un qualunque sistema, il significato dei dati lo conosce lo statistico e il sociologo che ha creato i dati. Io esperto di IT posso rappresentarlo usando un certo linguaggio, ma non conosco qual è il significato [...]. Molto spesso i nostri algoritmi hanno bisogno di tecniche avanzate di statistica, tant'è che noi abbiamo pochi, ma preziosi "centauri" che sono in grado di analizzare dal punto di vista metodologico un problema e, al contempo, proporre una soluzione che è anche tecnologica [...] Il confronto tra diverse figure diventa un elemento cruciale di successo e anche se, a volte, possono capitare dei problemi nei rapporti interpersonali, si cambia la composizione dei gruppi ma non si può far a meno di questo confronto.

Un altro punto che afferisce alla sfera del co-design è caratterizzato dalla figura dei committenti, il cui complesso di esigenze e interessi va ad influenzare indirettamente il disegno dell'algoritmo. A riguardo, interessante è l'episodio raccontato da un ricercatore in cui durante la creazione di una piattaforma commissionata da psicologi, si è trovato a dover calibrare e, in alcuni casi, intervenire sulle istruzioni dell'algoritmo affinché si avvicinasse alle richieste dei committenti in termini di usabilità e di risultati attesi:

W: Parte dell'obiettivo del progetto dal mio punto di vista è stato quello di mettere gli strumenti nelle mani degli psicologi. Quello che abbiamo cercato di fare è costruire una piattaforma, in modo che gli psicologi fossero in grado di progettare i vari interventi e specificare le variabili e i dati che

intendevano catturare. [...] Nel corso del tempo abbiamo sviluppato una serie di diverse funzioni di esportazione e abbiamo affrontato una serie di processi di co-design con gli psicologi per stabilire quali tipi di dati volevano e in che modo volevano estrarli. [...] A volte richiedevano una riga separata nel foglio di calcolo per ogni sessione eseguita da un partecipante, altre, invece, una riga per ogni partecipante e quindi costruire sessioni multiple. [...] Nel caso dei dati e della loro esportazione, dovevano utilizzare procedure semplici, quindi file .csv, che possono essere facilmente utilizzati e convertiti o formati SPSS perché loro hanno una tradizione nell'uso di SPSS. [...] Gli psicologi sono stati i co-progettisti per tutta la costruzione della piattaforma.

Il processo che a partire dai dati iniziali porta alla produzione dei risultati desiderati è caratterizzato, dunque, da diverse dinamiche che interagiscono fra loro. Oltre alle esigenze dei committenti, il disegno di un algoritmo deve fare i conti con la presenza e l'influenza di attori non umani. In particolare, la creazione di un algoritmo spesso si scontra con i limiti delle tecnologie e delle risorse a disposizione. È il caso di un ricercatore che, in occasione di uno studio su reti dinamiche, si trova a dover fare i conti con gli strumenti a disposizione, rivelando che l'uso di un software comunemente utilizzato per l'analisi delle reti è governato da proprie logiche interne che non si sposano bene con i risultati che si aspettava di ottenere. La soluzione adottata è stata di creare un proprio software che rispondesse meglio ai risultati desiderati:

L: Avevo bisogno di uno strumento di visualizzazione dinamica che mostrasse la crescita della rete nel tempo. E c'è un software comunemente usato per l'analisi di rete di grandi dimensioni che si chiama Gephi. Sebbene sia molto usato, non funzionava bene con le grandi reti. Ho trascorso alcune settimane a cercare di far funzionare il software Gephi esaminando tutta la documentazione degli sviluppatori [...]. Quindi, ho provato a far funzionare il software in base ai risultati che mi aspettavo e ho dedicato ben tre settimane a questo lavoro. [...] Alla fine, ho capito che avrei dovuto scrivere il mio software per ottenere esattamente il tipo di visualizzazione che stavo cercando.

Pur riconoscendo la presenza dell'azione concomitante di più attori e strumenti tecnologici nella scrittura di un algoritmo, nella dinamica dei focus group emergono interessanti punti di contrasto. Alcuni intervistati sono portati a minimizzare l'azione umana enfatizzando l'aspetto meramente tecnico e sequenziale del processo. Per alcuni, infatti, gli algoritmi si configurerebbero più semplicemente come una serie di passaggi, strumenti che non sempre sono ideati per scopi specifici o che, posti in un dato ordine, producono un determinato

output:

R: L'algoritmo è l'espressione della formulazione di un problema. [...] Gli algoritmi possono esser visti come una serie di passaggi che trasformano input in output.

C: L'algoritmo è pura matematica, cioè un modo per guardare le informazioni, un modo per scegliere come guardarle insieme a specifiche variabili.

Per altri, invece, l'azione di più attori, sia umani sia non umani, va ad incidere sulla creazione dei risultati finali e sulla loro interpretazione:

S: Si sceglie la struttura più vicina al cervello, cioè al proprio modo di processare le informazioni. L'algoritmo è il motivo per cui prendiamo decisioni con l'obiettivo di rimuovere il pregiudizio umano, ma in realtà non è possibile poiché lo stesso algoritmo è progettato dalla mente umana e dipende dai vari soggetti che lavorano...le skill, la formazione...

La co-partecipazione di più attori richiama un altro tema, quello del rapporto tra algoritmo e conoscenza che ha sollevato in particolar modo due questioni principali. Un primo aspetto riguarda la discussione intorno all'oggettività delle procedure e dei risultati degli algoritmi. Dall'analisi delle interviste emerge che la conoscenza dei domini che viene applicata agli algoritmi è parte integrante degli algoritmi stessi. Ciò indica, dunque, che la conoscenza scientifica delle discipline, che già esiste, viene incorporata negli algoritmi nella fase di progettazione; pertanto, tali strumenti dipendono anche dalle scelte soggettive di chi è direttamente coinvolto in questo processo che, a loro volta, saranno condizionate dalle competenze in materia, dall'appartenenza ad uno specifico settore scientifico, quindi, dalla propria formazione e da quell'insieme di valori, credenze ed esperienze:

A: Quindi il problema è come dai dati passiamo alla conoscenza [...] Qualcuno può trovare una teoria, applicarla sapendo che funzionerebbe per produrre della conoscenza e, nel caso contrario, usare una tecnica diversa, o usare un algoritmo che porti a quel tipo di dati con un altro tipo di conoscenza, una conoscenza che deriva dai dati. Quindi bisogna chiedersi che tipo di conoscenza, il tipo d'informazioni che si sta cercando.

Anche se il processo è automatizzato, gli algoritmi usati per processare i dati sono carichi di valore e prospettive che contraddistinguono un particolare approccio scientifico. Questa

questione ne richiama un'altra più generale, cioè se gli algoritmi possano essere visti come regimi di conoscenza. A riguardo, gli intervistati evidenziano che questi strumenti non rappresentano conoscenza di per sé, quanto piuttosto siano strumenti per produrre conoscenza. La conoscenza appartiene alla sfera umana, l'algoritmo può essere descritto come "un'espressione di conoscenza", ma non può rappresentare di per sé conoscenza. Ne deriva che gli algoritmi non sono neutrali né è ineccepibile l'interpretazione dei risultati prodotti in quanto una stessa conclusione può essere messa in discussione dalle riflessioni di altri analisti:

P: Penso che tu debba essere umano per avere conoscenza e quindi la conoscenza è del progettista dell'algoritmo. L'algoritmo per me non è conoscenza in sé; l'algoritmo è l'output di quella conoscenza. [...] L'algoritmo è, in qualche modo, l'espressione di quella conoscenza o la soluzione per risolvere alcuni problemi all'interno del dominio di quella conoscenza.

N: Per passare dagli input agli output tramite un algoritmo è necessario un qualche tipo di conoscenza nell'area di ricerca in cui si sta intervenendo.

La "azione umana", infine, è ancora più evidente per gli algoritmi elaborati per produrre e gestire dati testuali poiché il linguaggio verbale può avere molteplici valenze rispetto a quello numerico, soprattutto se al contenuto di un messaggio scritto è associato un'immagine o una emoticon come nel caso dei social network:

F: Penso che ci sia bisogno dell'azione umana soprattutto quando lavoriamo, ad esempio, con i dati di Twitter e vogliamo elaborare una *sentiment analysis* [...] abbiamo davvero bisogno dell'azione umana per interpretare gli stati d'animo soprattutto quando le persone usano le emoticon in Twitter le quali possono avere anche un significato opposto rispetto al contenuto del messaggio testuale. Quindi, penso che gli umani siano necessari.

M: Se osservi la linguistica, la semantica o qualsiasi altra logica è necessaria una sorta di interpretazione umana del significato del testo. [...] gli esseri umani hanno bisogno infatti di più input in senso linguistico che in senso numerico, perché i numeri hanno una logica intrinseca che rende meno ambigua la loro interpretazione.

Da quanto è emerso, dunque, gli algoritmi risultano essere composti da competenze, bagagli di conoscenze, collaborazioni, negoziazioni e tecnologia, elementi, materiali e non, che vanno ad intrecciarsi

formando un sistema più complesso.

6. CONCLUSIONI

La costruzione di un algoritmo è, senza dubbio, una questione tecnica e, come tale, è imprescindibilmente legata alla strumentazione tecnologica disponibile. La ricerca nelle infrastrutture dati mostra, infatti, come la produzione dei risultati desiderati attraverso un algoritmo è legata alla qualità degli strumenti tecnologici in possesso. Ma la sua creazione è anche il risultato della co-progettazione tra più figure professionali, che spesso si trovano a negoziare circa le priorità e le strategie da seguire. E non solo. I committenti che con il loro bagaglio di interessi ed aspettative vanno in qualche modo ad indirizzare il disegno dell'algoritmo che sottende uno specifico prodotto tecnologico.

Alla luce di quanto emerso dalla ricerca, è possibile trarre due considerazioni conclusive, una di tipo metodologico e l'altra di carattere epistemologico.

Dal punto di vista metodologico, esaminare gli algoritmi, senza considerare il contesto più ampio in cui vengono utilizzati, rischia di "feticizzarli" e quindi di tralasciare alcuni aspetti cruciali che ne determinano il loro funzionamento (Thomas, Nafus, Sherman, 2017). Per tale motivo, l'analisi del più ampio assemblaggio dei dati può rappresentare un solido percorso metodologico per lo studio degli algoritmi, soprattutto di fronte alla crescente proliferazione dei dati digitali presenti nelle varie piattaforme e supporti che permettono la loro circolazione e la loro diffusione. La destrutturazione dell'assemblaggio socio-tecnico degli algoritmi può rappresentare una valida strategia metodologica per cogliere la loro natura contingente, relazionale e contestuale.

Un'altra considerazione riguarda gli aspetti epistemologici che hanno a che fare con i regimi di conoscenza e i limiti della conoscenza algoritmica. Un qualsiasi algoritmo esprime una specifica visione di un modello o di una regolarità, incorpora il "pregiudizio" di coloro che sviluppano i modelli matematici e che successivamente li traducono in codice. Campi disciplinari di affiliazione e teorie di riferimento entrano, dunque, nella costruzione dell'algoritmo stesso. In tale prospettiva, gli algoritmi e i dati prodotti non sono neutrali, non parlano da soli, ma attraverso qualcuno o qualcosa (Silver, 2012; Bowker, 2013; Hales, 2013). Allo stesso tempo, la loro costruzione deve fare i conti con i limiti tecnologici e con l'insieme di aspettative degli attori che sono collegati all'assemblaggio socio-tecnico. Pertanto, la conoscenza

prodotta nei contesti in cui gli algoritmi vengono applicati è orientata piuttosto che neutra e imparziale.

Il collegamento tra gli elementi evidenziati, ossia tra la costruzione socio-tecnica degli algoritmi e la loro performatività negli ambiti decisionali, andrebbe ulteriormente approfondito attraverso studi che adottino un approccio qualitativo alla destrutturazione degli algoritmi e all'analisi dei contesti in cui vengono usati.

RIFERIMENTI BIBLIOGRAFICI

- AMOOORE, L. (2006). Biometric borders: Governing mobilities in the war on terror. *Political Geography*, 25, 336-351.
- (2009). Algorithmic war: Everyday geographies of the war on terror. *Antipode*, 41, 49-69.
- ANDERSON, C. W. (2011). Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience. *Journal of Communication*, 5, 529-547.
- ARAGONA, B., FELACO, C., MARINO, M. (2018). The Politics of Big Data Assemblages. *Partecipazione e Conflitto*, 11(2), 448-471.
- ARNOLDI, J. (2016). Computer algorithms, market manipulation and the institutionalization of high frequency trading. *Theory, Culture & Society*, 33(1), 29-52.
- BAROCAS, S., HOOD, S., ZIEWITZ, M. (2013). *Governing algorithms: A provocation piece*. New York: New York University.
- BERLINSKI, D. (2000). *The Advent of the Algorithm: The 300-Year Journey from an Idea to the Computer*. San Diego: Harcourt.
- BOWKER, G.C. (2013). Data flakes: An afterword to 'Raw Data' is an oxymoron. In L. Gitelman (Ed.), *"Raw data" is an oxymoron* (pp. 167-171), Cambridge (Mass.): MIT Press.
- BOYD, D., CRAWFORD, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, *Information, communication & society*, 15(5), 662-679.
- BURRELL, J. (2015). *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*. Big data & Society, 3(1), 1-12.
- DE ROSA, R., ARAGONA, B. (2017). Unpacking Big Data in Education. A Research Framework. *Statistics, Politics and Policy*, 8(2), 123-137.
- DIAKOPOULOS, N. (2013). *Algorithmic accountability reporting: On the investigation of black boxes*. A Tow/Knight Brief. Tow Center for Digital Journalism, Columbia Journalism School.
- DODGE, M., KITCHIN, R. (2004). Flying through code/space: the real
-

- virtuality of air travel. *Environment and Planning A.*, 36, 195-211.
- DOURISH, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3(2), 1-11.
- DRUCKER, J. (2013). Performative materiality and theoretical approaches to interface. *Digital Humanities Quarterly*, 7(1).
- GEIGER, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, 17(3), 342-356.
- (2017). Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture. *Big Data & Society*, 4(2), 1-14.
- GILLESPIE, T. (2014). Algorithm [draft] [#digitalkeyword]. *Culture Digitally*, June 25.
- , SEAVER, N. (2016). Critical algorithm studies: A reading list. *Social Media Collective*, December 15.
- GOFFEY, A. (2008). Algorithm. In M. Fuller (ed.), *Software studies – A lexicon* (pp. 15-20). Cambridge: MIT Press.
- GROSSER, B. (2014). What do metrics want? How quantification prescribes social interaction on Facebook. *Computational Culture*, 1(4).
- GUREVICH, Y. (2011). *What is an algorithm*. Technical Report. Microsoft Research, Redmond, WA.
- HALES, D. (2013). Lies, damned lies and big data. *Aid on the Edge of Chaos*, February 1.
- HALFORD, S., POPE, C., WEAL, M. (2013). “Digital futures? Sociological challenges and opportunities in the emergent semantic web”. *Sociology*, 47(1), 173-189.
- KITCHIN, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14-29.
- , DODGE, M. (2011). *Code/space: Software and everyday life*. Cambridge: MIT Press.
- , LAURIAULT, T. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. *Social Science Research Network*. Disponibile su http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- KOWALSKI, R. (1979). Algorithm = Logic + Control. *Communications of the ACM*, 22(7), 424-436.
- KUSHNER, S. (2013). The freelance translation machine: Algorithmic culture and the invisible industry. *New Media & Society*, 15(8), 1241-1258.
-

- MIYAZAKI, S. (2012). Algorithmic: Understanding micro-temporality in computational cultures. *Computational Culture*, 2. Disponibile su <http://computationalculture.net/article/algorithmic-understanding-micro-temporality-in-computational-cultures>
- MONTFORT, N., BAUDOUIN, P., BELL, J., BOGOST, I., DOUGLASS, J., MARINO, M.C., MATEAS, M., REAS, C., SAMPLE, M., VAWTER, N. (2012). *10 PRINT CHR\$(205.5 + RND (1)): GOTO 10*. Cambridge: MIT Press.
- MOSCHOVAKIS, Y. N. (2001). What is an algorithm? In B. Engquist, W. Schmid (Eds.) *Mathematics Unlimited – 2001 and beyond* (pp. 919-936). New York, NY: Springer.
- MUSIANI, F. (2013). Governance by algorithms. *Internet Policy Review*, 2(3). Disponibile su <http://policyreview.info/articles/analysis/governance-algorithms>
- NAKAMURA, L. (2013). The social algorithms of race: Sorting it out in jihad worlds. In M. Shoshana, G. Kelly (Eds.), *The New Media of Surveillance*. London: Routledge, pp. 159–162.
- NAPOLI, P. M. (2013, May). The algorithm as institution: Toward a theoretical framework for automated media production and consumption. Paper presented at the Media in Transition Conference, Massachusetts Institute of Technology, Cambridge, MA. Disponibile su ssrn.com/abstract=2260923
- NEYLAND, D. (2015). On organizing algorithms. *Theory, Culture & Society*, 32(1), 119-132.
- O'NEIL, C. (2006). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Crown.
- PASQUALE, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
- PORTER, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- SANDVIG, C. (2015). Seeing the Sort: The Aesthetic and Industrial Defense of 'The Algorithm.' *Journal of the New Media Caucus*. Disponibile su <http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/>
- SEEVER, N. (2013). Knowing algorithms. *Media in Transition* 8. April.
- (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1-13.
- SILVER, N. (2012). *The signal and the noise: why so many predictions fail--but some don't*. New York: The Penguin Press.
-

- SIPSER, M. (2006²). *Introduction to the Theory of Computation*. Boston, MA: Thompson Course Technology div. of Thompson Learning.
- STEINER, C. (2012). *Automate this: How algorithms took over our markets, our jobs, and the world*. New York, NY: Portfolio.
- TAKHTEYEV, Y. (2012). *Coding places: Software practice in a South American City*. Cambridge: MIT Press.
- THOMAS, S. L., NAFUS, D., SHERMAN, J. (2017). Algorithms as fetish: Faith and possibility in algorithmic work. *Big Data & Society*, 5(1), 1-11.
- TUFEKCI, Z. (2015). Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency, *Colorado Technology Law Journal*, 13(2), 203.
-

LA (NON) NEUTRALITÀ DELLA SCIENZA E DEGLI ALGORITMI

Il caso del machine learning tra fisica fondamentale e società

di Aniello Lampo, Michele Mancarella, Angelo Piga*

Abstract

The impact of Machine Learning (ML) algorithms in the age of big data and platform capitalism has not spared scientific research in academia. In this work, we will analyze the use of ML in fundamental physics and its relationship to other cases that directly affect society. We will deal with different aspects of the issue, from a bibliometric analysis of the publications, to a detailed discussion of the literature, to an overview on the productive and working context inside and outside academia. The analysis will be conducted on the basis of three key elements: the non-neutrality of science, understood as the intrinsic relationship with history and society; the non-neutrality of the algorithms, in the sense of the presence of elements that depend on the choices of the programmer, which cannot be eliminated whatever the technological progress is; the problematic nature of a paradigm shift in favour of a science and a society dominated by data. The deconstruction of the presumed universality of scientific thought from the inside becomes in this perspective a necessary first step also for any social and political discussion. This is the subject of this work in the case study of ML.

Keywords

Neutrality, science, algorithms, machine learning, scientific paradigm, physics

* ANIELLO LAMPO lavora all'Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya (UOC), Barcelona (Spain). Email: nello.lampo@gmail.com

MICHELE MANCARELLA è Ricercatore indipendente. Email: mancarell@gmail.com

ANGELO PIGA lavora all'ICFO - Institut de Ciències Fotòniques di Barcelona. Email: angelo.piga@gmail.com

1. INTRODUCTION

«It's time to ask: what can science learn from Google?» si chiedeva provocatoriamente Chris Anderson, editor di Wired, in un celebre articolo del 2008 (Anderson, 2008). La domanda suggerisce di guardare al cuore del metodo scientifico. Costruire modelli teorici, falsificarli, eseguire dei test, è lo schema che ogni scienziato introietta ed utilizza, come risultato di secoli di epistemologia. L'idea di Anderson è che nell'era dei petabyte questo approccio risulterebbe obsoleto: nel XXI secolo, infatti, la mole di dati a disposizione sarebbe tale da poter estrarre previsioni da essi senza chiamare in causa modelli, dichiarando dunque “la fine della teoria”. Una “rivoluzione copernicana”, un cambio di paradigma, il cui principale perno tecnico-teorico risiederebbe nel *Machine Learning*, ovvero «la capacità di una macchina di imparare senza essere stata programmata per farlo» (Arthur, 1959). Quasi 10 anni dopo l'articolo di Anderson, Google ha investito nel solo 2016 tra i 20 e i 30 miliardi di dollari in ML ed in generale in Intelligenza Artificiale¹ (*Artificial Intelligence*, d'ora in poi AI), di cui il 90% in ricerca e sviluppo. Nello stesso periodo di tempo, assistiamo ad una crescita delle pubblicazioni scientifiche in questo campo².

La fisica non è esente da questa esplosione. Se l'interesse per tecniche avanzate di estrazione di informazione dai dati non dovrebbe sorprendere, sarebbe sbagliato provare a spiegare le ragioni del fenomeno appellandosi solo a criteri interni alla scienza, come ad esempio l'efficienza tecnica e la convenienza formale.

Al contrario, il boom del ML in fisica esemplifica il rapporto inestricabile tra il sapere scientifico ed il contesto storico nel quale esso viene prodotto. Una delle posizioni principali in accademia rispetto all'analisi di tale legame può essere condensata in un'unica parola: neutralità. Quest'espressione sottintende che l'attività dello scienziato, e dunque i suoi metodi ed i contenuti che produce, seguirebbero una progressione lineare, sconnessa dal contesto sociale e politico. Al contrario, in questo lavoro partiremo dal principio secondo cui la

¹ In letteratura, esistono diversi termini che si riferiscono alla possibilità di programmare una macchina per prendere decisioni autonomamente (cioè senza un modello predittivo). Le differenze sono sottili e spesso i termini sono usati come sinonimi, anche a sproposito. In particolare, il Machine Learning è una branca dell'Intelligenza Artificiale, la quale include altri settori come ad esempio la robotica. Nella sezione II introdurremo nel dettaglio altre differenze tecniche e terminologiche che ci serviranno nel corso dell'articolo. Quando non diversamente specificato ci riferiremo all'insieme di queste tecniche col nome Machine Learning.

² I dati verranno discussi in dettaglio nella sezione III.

scienza è a tutti gli effetti un'attività umana, ed in quanto tale viene esercitata in una data epoca storica, all'interno di un preciso sistema di pensiero, di produzione e, per arrivare alla quotidianità dello scienziato, nel quadro di un certo rapporto lavorativo. Di conseguenza la ricerca scientifica nel suo complesso riflette necessariamente questi aspetti del mondo circostante: in questo senso diremo che essa ha un carattere *non neutrale*.

La discussione intorno alla non-neutralità della scienza è stato uno degli argomenti principali del dibattito epistemologico tra fine ottocento e novecento ed oscilla fra l'idea estrema di una completa indipendenza della scienza dalla società, come ad esempio sostenuto dai convenzionalisti (Poincaré, 1905) e quella opposta dell'anarchismo epistemologico di Feyerabend (2002), passando per tutta una serie di posizioni intermedie rappresentate ad esempio da Popper (Popper, 1991), Lakatos (Lakatos, 1995) e Kuhn (Kuhn, 1973). In particolare, quest'ultimo propose un'analisi storica che evidenziava come la scienza si sviluppi secondo bruschi cambi di paradigma la cui affermazione dipende anche da fattori sociali.

L'obiettivo principale di questo lavoro è declinare tale dibattito nell'attualità assumendo come *case-study* l'utilizzo del ML nelle scienze pure. Proveremo a dimostrare che l'impiego di tali strumenti nell'ambito della fisica non può essere motivato prescindendo da una collocazione storica in quella che si sta configurando come "l'era dei dati e degli algoritmi".

Ciò richiede di quantificare se e in che misura un trend all'interno della ricerca in fisica si possa riscontrare; capirne le ragioni; discuterne i legami con il contesto produttivo e lavorativo. Parleremo in questo caso di "non-neutralità della scienza", nel senso chiarito poco sopra, mostrando che questo fenomeno non dipende da dinamiche interne alla scienza stessa, ma da fattori economici, sociali e politici.

In particolare, prenderemo in esame il contesto lavorativo accademico, pesantemente sovradeterminato dai tagli e le riforme imposte dopo la crisi del 2008, che portano a precarietà e riduzione delle posizioni permanenti. Questa situazione può condurre il ricercatore a scelte dei suoi programmi di ricerca dettate dalla necessità di acquisire competenze spendibili fuori dall'accademia piuttosto che dalle prospettive scientifiche. Vedremo come il boom del ML in accademia vada interpretato infatti alla luce della grande espansione del settore dei big data e dell'AI, che rappresenta oggi un settore industriale capace di attrarre forza lavoro con alti salari e investimenti.

Il rapporto della scienza "pura" con le dinamiche sociali e politiche

del tempo va letto però in un senso ulteriore oltre quello appena discusso. Ad esempio, ne *L'Ape e l'Architetto* (Ciccotti et al., 1976) si evidenzia «il ruolo sovrastrutturale che la produzione di scienza pura svolge in quanto forma specifica di cultura». Intendiamo riferirci in particolare al fatto che le scienze pure³ contribuiscono alla creazione di un sistema di linguaggio, di nozioni, di metodi, aspettative, legati alla possibilità di descrivere il mondo in termini di modelli matematici⁴ anche al di fuori di esse⁵.

Ciò costituisce una questione determinante al giorno d'oggi, in quanto spesso si considerano il “dato”, il modello matematico o l'algoritmo come elementi oggettivi, portatori di una “verità” imparziale, il cui impiego servirebbe a giustificare decisioni politiche come conseguenze logiche di un ordine naturale. Decostruire la pretesa universalità del “sistema scienza” *dall'interno delle scienze pure* diventa quindi un compito cardine per ogni discussione sociologica o politica che ruoti attorno all'utilizzo di strumenti scientifici. Un secondo obiettivo del lavoro è proprio l'inizio di tale decostruzione per quanto riguarda il ML e l'AI. Parleremo in questo caso, più specificamente, di “non-neutralità dell'algoritmo”, intendendo con questo concetto la presenza *all'interno dell'algoritmo stesso* di elementi che implicano scelte a discrezione del programmatore.

Tale discorso assume una rilevanza ancora maggiore accettando l'ipotesi di Anderson. Se rinunciamo alla teoria, cioè alla costruzione di *framework* predittivi che guidino le osservazioni, falsificabili o migliorabili, non resta infatti che ridursi all'idea secondo cui “correlation is enough”. Nell'uso di algoritmi per regolare la vita sociale e politica, questa affermazione va molto al di là del piano epistemologico - che pure è problematico di per sé. Elevare correlazione a strumento di previsione significa infatti amplificare, rafforzare e perpetuare lo *status quo*, in particolare i rapporti di forza e di potere all'interno della società.

Un ultimo obiettivo del lavoro è quindi problematizzare l'idea di un

³ Qui e in seguito useremo il termine “scienze pure” per indicare in particolare fisica e matematica. In altri casi, specificheremo la differenza.

⁴ Parliamo di “modelli” riferendoci a modelli predittivi, come nel metodo scientifico. In questo contesto, il claim di Anderson è volto a sostituire modelli con dati grezzi. Discuteremo in seguito la rilevanza di questo aspetto.

⁵ Un caso rilevante per la presente discussione viene dalla sociologia, di cui ad esempio Goldthorpe promuove una formulazione statistica in termini di teoria delle popolazioni (Goldthorpe, 2016), mentre Latour (Latour, 2010) denuncia come «Sociology has been obsessed by the goal of becoming a quantitative science» considerando irraggiungibile, se non inutile, questo obiettivo. Per una critica simile che prende direttamente piede dall'ipotesi di Anderson si veda anche (Crawford, 2011).

cambio di paradigma verso un'epoca completamente *data-driven*. Innanzitutto, l'analisi della letteratura nella fisica ci permetterà di valutare dall'interno se questo sia effettivamente il caso. D'altra parte, abbiamo appena visto che una prospettiva prettamente interna alla scienza non è sufficiente. Il concetto stesso di paradigma ci impone di mettere in relazione la scienza con il contesto storico. In altre parole, prenderemo *sul serio* l'affermazione di Anderson, mostrando come, sebbene non giustificata limitando lo sguardo alla sola fisica, essa acquisisce un senso diverso prendendo in considerazione "ciò che possiamo imparare da Google", cioè l'aura di universalità e legittimità che gli algoritmi (e il ML in particolare) acquisiscono nell'era dei big data. Concluderemo che questa prospettiva mette lo scienziato di fronte al dovere di assumere e comunicare con chiarezza come la scienza e i suoi strumenti siano attività umane e per questo criticabili e mai del tutto neutrali. La discussione del caso della fisica presentata in questo lavoro vuole essere un passo in questa direzione.

La struttura dell'articolo è la seguente. Nella sezione I, discuteremo il quadro alla base di questo studio: la non-neutralità della scienza. Nella sezione II, introdurremo il concetto di ML e alcuni elementi tecnici e storici che motivano l'esplosione del suo utilizzo negli ultimi 10 anni. Passeremo quindi a studiare l'utilizzo di algoritmi di ML in fisica, nella sezione III. Presenteremo innanzitutto una stima del fenomeno sulla base di un'analisi bibliometrica, per poi discutere nel dettaglio alcuni casi concreti: la fisica delle alte energie (sezione IIIa), l'astrofisica (sezione IIIb) e la fisica delle basse energie (sezione IIIc). Vedremo che l'analisi storica, bibliometrica e della letteratura conducono a considerare il ruolo delle grandi piattaforme e del mercato del lavoro. Questo sarà l'oggetto della sezione IV. Infine, nella sezione V saremo in grado di sviluppare un parallelo tra la fisica e altri campi, e di commentare i concetti di non-neutralità della scienza e degli algoritmi e il problema del cambio di paradigma.

2. LA NON-NEUTRALITÀ DELLA SCIENZA

Con non-neutralità della scienza intendiamo l'impossibilità di una demarcazione netta fra la struttura della scienza – ovvero le sue teorie, il cosiddetto metodo scientifico e la sua organizzazione interna – da un lato e il contesto storico, la struttura della società e i rapporti di forza interni ad essa, dall'altro.

Il lavoro di molti filosofi della scienza si concentra sulla possibilità di isolare e definire norme secondo cui lo scienziato dovrebbe

procedere: è il problema del “metodo”. Ancora oggi la pretesa *oggettività* della scienza, il suo valore e la sua funzione sociale sono spesso difese sulla base di una pretesa universalità del metodo scientifico. L'attuale visione scientifica è costruita sulla base di tale metodo, a partire da Galileo e Bacone, passando per Newton e poi Einstein, fino alle moderne misure che completano la conoscenza del modello standard delle particelle elementari.

Già Galileo era consapevole dell'importanza della presenza di una teoria speculativa che guidasse l'osservazione e nel *Dialogo* esprime la sua ammirazione per gli scienziati capaci di «anteporre quello che il discorso gli dettava, a quello che le sensate esperienze gli mostravano apertissimamente in contrario». Secoli dopo, passando attraverso una lunga storia che vede protagonisti filosofi come Descartes, Locke, Kant, i “positivisti” dell'Ottocento, i “convenzionalisti” di inizio Novecento, fino al neopositivismo, sono probabilmente Karl Popper e Thomas Kuhn ad aver affermato un lessico e un approccio divenuti standard per il lavoro degli scienziati.

In particolare, nel “razionalismo critico” popperiano (Popper, 1991) il procedere della scienza avviene per “prova ed errore”, avanzando ipotesi e teorie e testandone le previsioni. Tali test possono condurre unicamente a falsificazioni e mai a una verifica, dato che lo scienziato non può essere certo che non esistano altri fatti, ancora sconosciuti, che invalidano la teoria. Il falsificazionismo popperiano ha avuto un impatto talmente forte nel dibattito epistemologico da diventare il discriminante che tuttora definisce la scienza, e in particolare l'elemento che la eleva al di sopra di altre discipline che non portano in sé la possibilità di correggere i propri errori. In questo concetto oggi si cela la supposta oggettività della scienza, che si configura come prototipo esemplare dell'attività razionale umana, la strada più sicura, se non verso la certezza, almeno verso la verità secondo un progresso lineare. Nella visione di Popper le idee cambiano secondo criteri tutti interni alla scienza, il cui sviluppo risulta dunque cumulativo e svincolato dalla storia della società: in questo senso potremmo dire *neutrale*.

Un elemento importante consiste appunto nello sfondare la barriera tra storia e scienza, tra società, politica e progresso lineare del pensiero scientifico. È soprattutto a Lakatos e Kuhn che si deve una svolta in questa direzione. Ogni standard o modello che ha dominato la cultura scientifica nel corso della storia è in realtà il prodotto di una “visione del mondo” che trascende la sola scienza ed è dettato dall'epoca storica e da un insieme di credenze e di sistemi di pensiero. È questo che Kuhn definisce un *paradigma* (Kuhn, 1972). La conoscenza scientifica segue

un andamento alternato tra lunghi periodi di “scienza normale”, fatti di accumulazione di dati seguendo criteri interni alla scienza e a un paradigma vigente, e “rivoluzioni” che portano alla sostituzione di un paradigma con un altro, in cui la frontiera tra il “fuori” e il “dentro” la scienza cade. Diversi paradigmi sono incommensurabili e non è possibile scartarne uno in base alle prescrizioni del falsificazionismo.

C'è di più: i dati sperimentali candidati alla verifica logica di una teoria sono sempre contaminati dalle assunzioni insite in un paradigma, e i dati stessi sono sempre *theory-laden* (carichi di teoria).

È Paul Feyerabend a spingersi oltre (Feyerabend, 2002). In effetti, nessuna teoria scientifica è mai in accordo con *tutti* i fatti noti del suo tempo, e i canoni di “razionalità scientifica” non esistono se svincolati dal contesto storico. L'unica soluzione è allora scrivere “contro il metodo”, che significa riconoscere che gli standard sono «una guida che a sua volta è parte di un'attività guidata ed è mutata da questa»: cioè l'osmosi tra scienza, epoca, scelte politiche è continua, totale e inevitabile. È in questi termini che si deve parlare di scienza.

In questo quadro, il risultato principale è che la scienza va considerata come un'attività umana, il cui sviluppo non si può scindere dal percorso storico regolato da fattori sociali e politici. Nel resto del testo specializzeremo questo punto di vista al lavoro del fisico teorico degli ultimi anni, ed in particolare analizzeremo il caso dell'exploit nell'uso del ML.

3. DAL MACHINE LEARNING ALLE DEEP NEURAL NETWORKS

In questa sezione, introdurremo alcuni concetti e definizioni che ci saranno utili per inquadrare il problema e per la discussione seguente.

Come già accennato nell'introduzione, il termine *Machine Learning* (ML) si riferisce alla «capacità di una macchina di apprendere da sola dai dati, senza essere stata esplicitamente programmata per farlo»⁶. I primi lavori di ML risalgono agli anni 50, con semplici modelli ad output binario (attivo/inattivo) per spiegare il funzionamento dei neuroni. Uno di questi modelli è il *Perceptron* (Rosenblatt, 1958). Venne rapidamente realizzato che diversi Perceptrons (anche chiamati “neuroni”) possono essere combinati in un *layer* per fornire un output a più classi – il “layer di output”. Tale struttura è l'esempio più semplice di *rete neurale* o *Neural Network* (NN). Nel caso delle NN, i layers

⁶A S. Arthur (Arthur, 1959) viene attribuito il conio del termine Machine Learning. Per un'introduzione storica, si veda l'introduzione in (Goodfellow 2016). Per un'introduzione formale e rigorosa del problema, si veda ad es. (Mostafa, 2012).

possono inoltre essere combinati in successione per raggiungere livelli più alti di precisione e astrazione (si parla di “*hidden layers*”, “layer nascosti”, per i layers più interni). Per quanto quello delle NN sia un esempio di ML, esistono molti algoritmi di ML che non si basano su reti neurali. In seguito, ci riferiremo con ML al secondo tipo di algoritmi e con NN al primo.

È solo negli anni 2000 che la possibilità di sviluppare reti con molti layers viene sfruttata fino in fondo, per tre motivi. Innanzitutto, la disponibilità di dati diventa esponenzialmente più grande e molto più varia⁷. Il secondo motivo è l’aumento, anche questo di diversi ordini di grandezza, della potenza di calcolo, in particolare grazie all’utilizzo delle cosiddette *Graphic Processing Unit* (GPU) (Raina 2009). Questi due progressi hanno inoltre portato – terzo motivo – alla necessità di affinare alcuni algoritmi noti per renderli più performanti (Hinton, 2012; Maas, 2013).

Grazie a questi tre elementi chiave (volume di dati-potenza di calcolo-nuovi algoritmi), si inizia a parlare di *Deep Learning* (DL), o di *Deep Neural Networks* (DNN, reti neurali profonde), riferendosi proprio alla possibilità di costruire reti a molti layer.

Nella prossima sezione, studieremo innanzitutto la presenza dei tre termini sopra introdotti nella letteratura scientifica, e commenteremo in dettaglio dei casi specifici.

4. IL MACHINE LEARNING NELLA FISICA FONDAMENTALE

In questa sezione discuteremo l’utilizzo del ML nella ricerca. Cercheremo di dare una stima della portata del fenomeno, e mostreremo che si riscontra un’impennata di pubblicazioni a riguardo in tutti i settori dopo il 2010 circa. Ci concentreremo poi sulla fisica e su alcune branche particolarmente significative, analizzando in dettaglio dei casi specifici.

4.1. *L’impennata delle pubblicazioni: i dati*

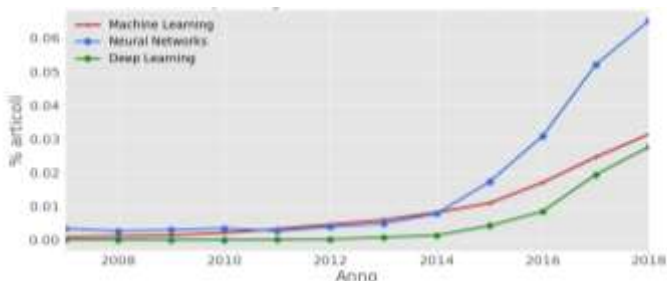
Nella letteratura scientifica, l’archivio open-access online

⁷ Consideriamo alcuni dataset di riferimento: “Modified National Institute of Standards and Technology” (MNIST) è il più largo degli anni 90 e conta 60 mila cifre scritte a mano. Oggi lo Street View House Numbers (SVHN) – una raccolta di immagini di numeri messa a disposizione da Google Street View – è di un ordine di grandezza più largo. Il dataset ImageNet conta oltre 14 milioni di immagini, contenenti gli oggetti più diversi. Infine, il dataset rilasciato dal “Workshop of Machine Translation” (WMT) per la traduzione automatica arriva a 1 miliardo di dati.

www.arxiv.org è lo spazio più largamente utilizzato per condividere articoli o preprint. Il sito conta, nel 2018, più di 10mila nuovi articoli al mese⁸. Dato il suo largo utilizzo, una ricerca tra gli articoli contenuti nel database rappresenta un mezzo utile per avere una stima della presenza di un trend. Abbiamo effettuato quindi una ricerca degli articoli che contengano le espressioni “Machine Learning”, “Neural Networks” o “Deep Learning” nel titolo o nell’abstract. I dati presentati nelle figure che seguono sono stati raccolti il 14 Dicembre 2018.

Il grafico in Figura 1 mostra la percentuale di articoli contenenti ML (linea rossa), DL (linea verde) e NN (linea blu) nell’abstract o nel titolo. È immediato notare una netta impennata a partire dal 2014, preceduta da un lieve incremento a partire dal 2010 circa. Il numero di articoli a tema NN è passato da 113 nel 2008 a 4575 nel 2017, pari ad un incremento totale di 40 volte e relativo di poco più di 30. Nel 2018 il numero di articoli supera gli 8000. Simili stime valgono per gli altri due termini, con un incremento relativo pari a circa 55 volte per ML. Il termine DL è addirittura assente prima del 2008, mentre conta più di 1500 articoli nel 2017. Come si capisce dal grafico, tale cambiamento di tendenza nel numero assoluto di articoli non è spiegabile con il generale incremento degli articoli inviati ad Arxiv. che è invece aumentato di solamente il doppio, da 4970 articoli inviati nel mese di gennaio 2008 a 10332 articoli nel mese di dicembre 2017.

Fig. 1. Percentuale di articoli Arxiv con “Machine Learning”, “Deep Learning” o “Neural Networks”) in titolo o abstract



È interessante chiedersi anche quali campi di ricerca risentono maggiormente di questa tendenza e in seguito concentrarsi sulla fisica. Arxiv divide gli articoli in 8 categorie: Fisica, Informatica (Computer

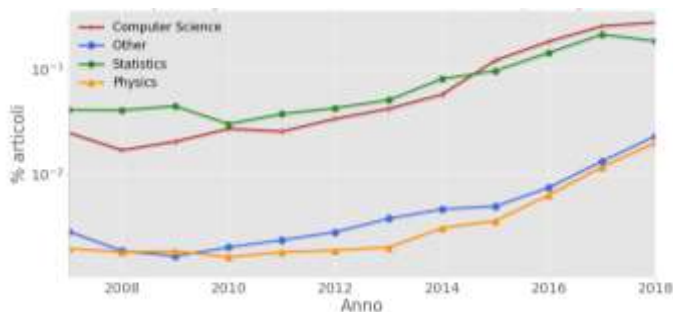
⁸ https://arxiv.org/stats/monthly_submissions

Science), Matematica, Finanza Quantitativa, Statistica, Biologia Quantitativa, Economia, Energia Elettrica e Scienza dei Sistemi.

In Figura 2, mostriamo il numero di articoli su Arxiv con una qualsiasi delle espressioni ML, DL e NN nel titolo o nell'abstract, divise per categorie. I dati sono normalizzati rispetto al numero totale di articoli usciti nello stesso anno nella stessa categoria e sono visualizzati in scala logaritmica. È chiara quindi una crescita esponenziale delle ricerche con queste parole chiave. Com'era prevedibile, la "Computer Science" (linea rossa), domina il trend nei valori assoluti degli articoli pubblicati, con una significativa impennata dopo il 2014. Tra gli altri campi, seguono la "Statistica" (linea verde), e proprio la "Fisica" (linea arancione), per cui l'incremento è ancor più significativo (circa 10 volte tra il 2008 e il 2018). La linea blu tratteggiata rappresenta la somma di tutte le altre sezioni tematiche dell'archivio ed anche in questo caso si nota una forte crescita che dimostra la trasversalità nell'utilizzo di queste nuove tecniche pressoché in tutti gli ambiti della fisica.

Il periodo dell'impennata, in particolare in Informatica, segue un momento di svolta nei progressi nel campo del riconoscimento di immagini. Nel 2012, una DNN ha vinto una celebre competizione, la *ImageNet Large Scale Visual Recognition Challenge*, con un margine di oltre il 10% in più di precisione (Krizhevsky, 2012) rispetto al secondo classificato. Questo avvenimento è oggi considerato come il vero e proprio inizio della "rivoluzione del Deep Learning" ed ha convinto definitivamente le maggiori piattaforme (Google, Facebook, Microsoft, etc.) a investire massicciamente in progetti di ricerca e sviluppo sulle reti neurali. L'ipotesi, che confermeremo in seguito con esempi specifici, è che anche la fisica abbia risentito di questa svolta.

Fig. 2. Percentuale di articoli Arxiv con "Machine Learning", "Deep Learning" o "Neural Networks" in titolo o abstract, divisi per categoria

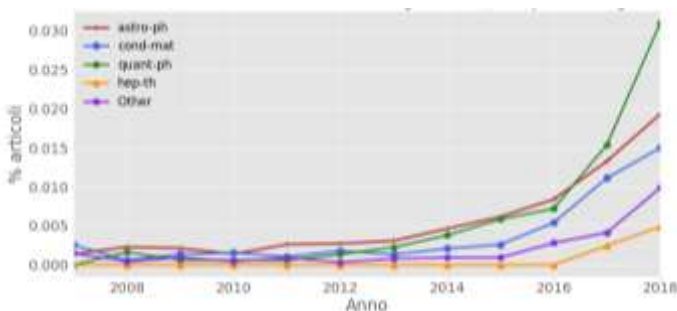


Restringiamoci ora alla sola categoria “Fisica”. Questa è a sua volta ripartita in 32 diverse sottocategorie⁹. In Figura 3, riportiamo il grafico delle percentuali di articoli nella categoria “Fisica”, contenenti una delle tre espressioni ML, DL e NN in titolo o abstract, e divisi per sottocategoria.

I dati mostrano che le categorie dominanti sono “Astrofisica” (astro-ph), “Fisica della Materia Condensata” (cond-mat), “Fisica Quantistica” (quant-ph) e “Fisica teorica delle Alte Energie” (hep-th), con un significativo incremento post-2014¹⁰. Discuteremo questi tre casi in dettaglio in seguito.

Anche se non analizzeremo tutti i casi in dettaglio in questo lavoro, è significativo notare l’incremento nei campi più disparati. Nella figura riassumiamo tutti i campi minori in un’unica curva indicata come “other”, che sembrano segnalare, come anticipato in precedenza, un interesse crescente per l’argomento che sembra partire da ambiti dove la sua applicazione è più immediata (come ad esempio l’astrofisica, come spiegheremo), ad ambiti in cui l’applicazione è più pionieristica (come la fisica teorica o la diagnosi di tumori in fisica medica). In effetti, tutti i settori ad eccezione dei primi due hanno un numero di pubblicazioni praticamente nullo fino a dopo il 2010. I dati di Arxiv presentati in questa sezione confermano, come ipotizzato, l’esistenza di un trend che eccede notevolmente la crescita del numero totale di articoli.

Fig. 3. Percentuale di articoli Arxiv con “Machine Learning”, “Deep Learning” o “Neural Networks”) in titolo o abstract nella categoria “Fisica”, divisi per sottocategoria



⁹ Una lista completa può essere consultata a <https://arxiv.org/archive/physics>.

¹⁰ Per rendere il grafico leggibile ed escludere informazione poco significativa, abbiamo scelto di riportare nel grafico solo le sottocategorie che hanno avuto un incremento di almeno un fattore 5 tra il 2008 e il 2018, e che abbiano avuto almeno 10 articoli nel 2018.

4.2. La fisica delle alte energie

Tecniche statistiche e algoritmi di previsione avanzati sono fondamentali nell'analisi dati. La fisica sperimentale non fa eccezione. In questa sezione discuteremo il caso più noto di una grande collaborazione in Fisica delle Alte Energie: l'acceleratore di particelle *Large Hadron Collider* (LHC) del CERN di Ginevra.

Una review apparsa su *Nature* (Radovic, 2018) spiega come il ML sia usato da lungo tempo nella collaborazione, ma segnala una svolta importante a partire dal 2014, con l'investimento massiccio nello studio di tecniche di DL. Tra le ragioni si citano esplicitamente il progresso nel riconoscimento immagini da parte di Google e la competizione *ImageNet* del 2012.

Del 2014 è anche il lancio di una competizione sulla piattaforma *Kaggle.com*, la *Higgs Boson Machine Learning Challenge* (Adam-Bourdarios, 2014), volta allo sviluppo di algoritmi per rivelare il bosone di Higgs (la famosa particella alla base del meccanismo che dà massa alla materia, al centro del premio nobel del 2013), vinta da una combinazione di 70 reti neurali, ciascuna con 3 layers di 600 neuroni¹¹.

Ancora del 2014 è un articolo che introduce a tutti gli effetti il DL in fisica delle alte energie, analizzandone alcune differenze sostanziali con le tecniche di ML già utilizzate (Baldi, *et al.*, 2014). Gli autori mettono l'accento sulla profonda trasformazione concettuale che il DL apporta. Gli algoritmi di ML infatti funzionano efficacemente solo con dati già preparati dai ricercatori sulla base della conoscenza del processo fisico in gioco. Questa procedura è nota come *feature engineering*. Ad esempio, l'impulso delle particelle rivelato a LHC permette di ricostruire la "massa invariante" del processo, una grandezza fisica ad alto potere discriminante (*high-level feature*), che è usata come input per il ML. Questa fase può essere estremamente complicata o computazionalmente costosa. Con una rete neurale profonda, si possono invece usare direttamente i dati "grezzi" del rivelatore (*low-level features*), raggiungendo una precisione equivalente: è la rete stessa che indipendentemente recupera l'informazione che prima veniva ottenuta tramite la conoscenza dei processi fisici in gioco.

Tuttavia, il *reverse engineering* di una NN, cioè la ricostruzione del processo decisionale che ha condotto all'output, è ad oggi impossibile: nel secondo caso, non possiamo quindi estrarre l'equivalente delle *high-level features*, e rinunciamo alla conoscenza del processo fisico in favore del potere predittivo. La transizione qui esemplificata, verso

¹¹ <https://github.com/melisgl/higgsml/blob/master/doc/model.md>

un'analisi "cieca", è la cifra della rivoluzione insita nel DL e la ragione per cui il concetto di "cambiamento di paradigma" viene spesso evocato recentemente anche in fisica. In realtà il problema è più complesso e un modello falsificabile rimane comunque in gioco. Una rete neurale ha infatti bisogno di una fase cosiddetta di *training* (il suo "allenamento"), in cui l'algoritmo impara da una porzione di dati (*training set*) in cui è noto (in questo caso) se un segnale sia presente o no, per poi essere in grado di fornire previsioni. Il *training* richiede centinaia di migliaia di dati per essere efficace: per allenare la rete, il segnale cercato viene quindi calcolato sulla base del modello teorico da testare (nel caso di LHC, può essere ad esempio il Modello Standard – ad ora il più importante e preciso modello che prevede e cataloga le particelle fondamentali della natura – come sue estensioni), e i dati di allenamento generati attraverso simulazioni numeriche. Alla base del processo rimangono dunque un modello predittivo e la sua capacità di fornire previsioni falsificabili.

Un altro caso rilevante è l'esperimento NOvA per la rivelazione di neutrini (Ayres, 2007). In particolare, il tipo di rete utilizzata è basata sul modello *GoogLeNet* (Szegedy, 2014), vincitore della competizione *ImageNet* 2015, come esplicitamente spiegato dagli autori.

Per finire, si può citare lo studio delle proprietà di un'altra particella elementare, il quark beauty, a LHCb, in cui si è iniziato ad utilizzare modelli basati sull'algoritmo usato da Google per il suo traduttore (CERN, 2017).

4.3. *L'astrofisica*

In Figura 3 si vede che astro-ph è una delle sottocategorie con il maggior incremento in numero di pubblicazioni.

Questo dato non è sorprendente, visto che un telescopio fornisce immagini ottiche, e il riconoscimento di immagini è cruciale per estrarre efficacemente informazione. Come caso esplicito, consideriamo la rivelazione di *lenti gravitazionali forti*¹². Ad oggi l'analisi accurata di una sola lente gravitazionale può richiedere anche settimane di lavoro e una conoscenza molto approfondita dei processi fisici che intervengono nella sua osservazione. Un recente articolo apparso su *Nature* (Hezaveh,

¹² Le lenti gravitazionali forti sono un fenomeno previsto dalla teoria della relatività generale, che consiste nella formazione in un telescopio di più immagini della stessa sorgente luminosa, a causa della deflessione della luce da parte di strutture gravitanti lungo il percorso che essa compie per arrivare all'osservatore.

2017) mostra che usando delle NN¹³, il processo di rivelazione è circa 10 milioni di volte più veloce che nel caso standard ed ha una precisione paragonabile. Soprattutto, però, in modo analogo all'esempio di LHC, questo approccio non richiede alcuna conoscenza specifica dei processi fisici in gioco nella misura: i dati delle immagini del telescopio passano direttamente all'algoritmo. Precedentemente, la misura richiedeva l'inclusione di centinaia di migliaia di parametri supplementari per modellare il segnale e complesse tecniche statistiche per estrarre il risultato finale.

Un altro esempio è la rivelazione di *onde gravitazionali* (Abbott, 2016), in cui studi recenti (Gabbard, 2018) mostrano come sia possibile rivelare un segnale a partire dai dati grezzi, raggiungendo la stessa performance (o addirittura migliorandola) della lunga e complessa tecnica standard di analisi dati.

Altri casi di applicazioni alla frontiera dell'astrofisica moderna sono lo studio di *esopianeti* (Lam, 2018; Shallue, 2018) e gli esperimenti cosmologici di prossima generazione (Gillet, 2018).

Gli esempi discussi, come nel caso delle alte energie, vanno nella direzione di una scienza sempre più data-driven, rispetto a una fase precedente in cui la modellizzazione di tutte le componenti di un esperimento era determinante.

Tuttavia, anche in questo caso l'allenamento della rete richiede dei dataset di allenamento provenienti da simulazioni numeriche risultato di complessi codici che forniscono previsioni sulla base dei modelli che sono da testare nell'esperimento¹⁴. In ogni caso, tale processo è determinante per l'utilizzo delle NN che vengono così ridimensionate a efficaci *tool* per l'analisi dei dati.

4.4. *La meccanica quantistica e la fisica della materia condensata*

Gli ultimi due ambiti di ricerca su cui ci soffermiamo sono la fisica della materia condensata (cond-mat) e la meccanica quantistica (quant-ph).

A differenza della fisica delle particelle e dell'astrofisica, in cui i casi che abbiamo analizzato si riferiscono prevalentemente all'uso di tecniche di DL per l'analisi dati, la meccanica quantistica e la materia

¹³ Ancora una volta, una di queste reti è AlexNet, vincitrice della competizione ImageNet del 2012.

¹⁴ Nel caso delle onde gravitazionali, ad esempio, occorre risolvere le equazioni della relatività generale per un sistema di oggetti compatti orbitanti, il che è risultato di tecniche numeriche combinate con accurate approssimazioni analitiche. Nel caso del lensing gravitazionale, esistono codici di simulazione di immagini di galassie che combinano modelli a immagini già esistenti.

condensata sono il campo di sperimentazione di nuove frontiere. Elenchiamo le più importanti¹⁵.

Una prima menzione va fatta per il cosiddetto *quantum machine learning* (QML), cioè l'uso di algoritmi che sfruttano le enormi potenzialità del calcolo quantistico per risolvere problemi classici di ML in tempi e costi computazionali esponenzialmente minori. Il QML è un campo nato di fatto nel 2013 da una collaborazione tra MIT e Google¹⁶. Nel 2013 Google (in collaborazione con la NASA) acquista anche un esemplare del prototipo di computer quantistico D-Wave, giustificando l'investimento di 10 milioni di dollari proprio per le possibili applicazioni in ML¹⁷. Va notato tuttavia come il computer quantistico sia lontano dall'essere realizzato e il QML rimanga un ambito prettamente teorico. D'altra parte, tecniche di ML stanno venendo utilizzate proprio per migliorare alcuni processi sperimentali utili per la costruzione del computer quantistico. In questo senso, il ML è un utile strumento più che un cambio di paradigma (Biamonte, 2017).

Concludiamo la sottosezione con uno sguardo alle importanti applicazioni allo studio delle *transizioni di fase quantistiche*. Una delle categorie di problemi tradizionali è studiare le proprietà di atomi o gas di atomi interagenti a una temperatura ambiente spesso vicina allo zero assoluto. Le equazioni da risolvere per avere una completa conoscenza delle sue fasi e dell'evoluzione dinamica hanno immensi costi computazionali e metodi numerici o approssimazioni sono necessari. Si cerca quindi di usare il ML per classificare le fasi del sistema. L'approccio è recentissimo, e risale solo agli ultimi tre anni. Nel caso di algoritmi di *supervised learning* (Carrasquilla, 2017) – cioè in cui sia disponibile un training set in cui le fasi siano note – il ricercatore sa già quali sono le fasi del sistema e cerca di ottenere l'esatto punto della transizione di fase (ad esempio una temperatura).

Il metodo cosiddetto di *unsupervised learning* (Van Nieuwenburg, 2017) è più interessante nel senso di un possibile cambio di paradigma, perché non richiede la conoscenza pregressa delle fasi. In questo caso l'algoritmo è in grado di dividere automaticamente i dati in gruppi di affinità. Tuttavia, è solitamente necessaria una *feature engineering* altamente sofisticata che da sola potrebbe condurre al risultato voluto, senza necessariamente ricorrere al ML (Huembeli, 2018). Un altro

¹⁵ Per una rassegna tecnica si veda la recente review (Dunjko, 2018)

¹⁶ Vedi (Lloyd, 2013). Gli autori sono Seth Lloyd, guru mondiale dell'informazione quantistica e del computer quantistico, Patrick Rebentrost (un suo giovane collaboratore), entrambi affiliati MIT, e Masoud Mohseni, senior research del team Google AI.

¹⁷ Google and NASA Launch Quantum Computing AI Lab: <https://www.technology-review.com/s/514846/google-and-nasa-launch-quantum-computing-ai-lab/>

problema risiede nel fatto che spesso risulta necessario sapere in anticipo quante sono le fasi del sistema. Questo impedisce quindi di trovare fasi nuove e rende comunque necessaria una conoscenza qualitativa, ma profonda, del sistema stesso.

5. PIATTAFORME E ACCADEMIA

In più di un caso discusso nelle sezioni precedenti è emerso esplicitamente un legame tra la ricerca sviluppata da grandi piattaforme come Google e il suo successivo impiego in fisica. Questo è indicativo della necessità di guardare al di fuori della fisica per comprendere fino in fondo il fenomeno. In questa sezione daremo una misura quantitativa di quale sia il trend degli investimenti privati nell'ambito dei big data, mostrando un exploit che fa da attrattore nel mercato del lavoro. Successivamente mostreremo come nell'ambito accademico si assista da diversi anni ad una costante diminuzione di fondi e possibilità di accesso a posizioni permanenti: una situazione che porta naturalmente i giovani ricercatori a spostarsi spontaneamente su ambiti che in termini di competenze acquisite hanno maggiori prospettive lavorative in settori privati esterni all'accademia e alla ricerca di base.

Le grandi piattaforme, tra i maggiori attori del capitalismo mondiale, basano oggi i loro profitti quasi esclusivamente su dati, algoritmi e risorse di calcolo: ad esempio, oltre il 95% dei profitti di Facebook proviene dalla vendita di spazi pubblicitari "adattati" ai gusti dell'utente¹⁸; l'80% degli streaming di Netflix proviene da raccomandazioni piuttosto che da ricerche dirette (Gomez-Urbe, 2016). Al di là delle maggiori piattaforme, questo trend si estende ai settori più disparati. Secondo un rapporto di J.P. Morgan's¹⁹, tra quelli più rilevanti in cui il ML si rivela essenziale ci sono banche e sicurezza digitale, finanza²⁰, usi governativi, comunicazione e servizi, assicurazioni, trasporti, sanità, educazione.

Una stima della dimensione di questo mercato – fornita sempre da J.P. Morgan's – parla di 58 miliardi di dollari entro il 2021, a partire da circa 12 miliardi nel 2017; questo rappresenta uno dei settori tecnologici in più forte crescita, con un tasso di crescita annua composto (CAGR) di quasi 50%. Il 44% delle imprese oggetto dello studio indica il settore di

¹⁸ <https://www.statista.com/statistics/267031/facebook-annual-revenue-by-segment/>

¹⁹ <https://flamingo.ai/wp-content/uploads/2017/11/JPMorganAnInvestorsGuideToArtificialIntelligencev2.pdf>

²⁰ Nel maggio 2017, J.P. Morgan ha pubblicato il più grande report esistente su Big Data e Machine Learning in finanza (J.P. Morgan, 2017)

ML/AI come la tecnologia che avrà il maggior impatto sull'impresa nella prossima decade, mentre un'altra indagine condotta da O'Reilly Media su un diverso campione²¹ dà lo stesso settore al 61%.

Da questo quadro si capisce l'entità dello studio e dell'applicazione del ML negli ultimi 5-10 anni al di fuori dell'accademia. Di tale tendenza si trova una chiara traccia anche nei dati che riguardano il mercato del lavoro.

Secondo il rapporto *LinkedIn's 2017 US Emerging Jobs Report* pubblicato dal sito LinkedIn del Dicembre 2017²², le due occupazioni che hanno visto il più alto incremento dal 2012 sono *Machine Learning Engineer* e *Data Scientist*, con crescite di 9.8 e 6.5 volte rispettivamente. Tra le prime tre posizioni nei lavori svolti precedentemente da chi ha assunto le posizioni di *Machine Learning Engineer* e *Data Scientist* troviamo *Research Assistant* e *Teaching Assistant*, a testimonianza di un flusso che proviene dall'accademia.

Ancora, il rapporto 2018 dell'azienda IBM *The quant crunch: how the demand of data science skills is disrupting the job market*²³ parla di un incremento del 40% dei posti di *Data Scientist* nel 2016 e di una proiezione di crescita uguale di qui al 2020. Tra questi, circa il 40% richiede un titolo di studio dal master in su, mentre il 78% almeno tre anni di esperienza.

Questi dati suggeriscono un forte bisogno di manodopera altamente formata in discipline scientifiche, che sembrano in molti casi provenire dall'accademia. Guardiamo dunque alle condizioni di lavoro proprio nell'accademia.

L'ipotesi che prendiamo in considerazione è che molto dell'interesse dei giovani ricercatori verso il ML risiede nella possibilità di acquisire tecniche spendibili in un mercato del lavoro che, abbiamo visto, è in enorme espansione, al contrario del panorama nel campo dell'università: poche prospettive di posto fisso, alta competizione, precarietà e stress poco sostenibili a fronte di bassi salari.

Negli USA e in Europa, almeno fino al 2008, le università non sono in grado di assorbire i dottori e questo trend è in costante peggioramento

²¹ 2018 Outlook: Machine Learning and Artificial Intelligence, A Survey of 1,600+ Data Professionals (14 pp., PDF, no opt-in).

²² <https://economicgraph.linkedin.com/research/LinkedIns-2017-US-Emerging-Jobs-Report>

²³ <https://public.dhe.ibm.com/common/ssi/ecm/im/en/iml14576usen/analytics-analytics-platform-im-analyst-paper-or-report-iml14576usen-20171229.pdf>

in tutti i paesi avanzati²⁴ (Cyranoski, 2011). K. Powell (Powell, 2015) fa notare come negli USA il numero di postdoc nelle scienze è cresciuto del 150% tra il 2000 e il 2012, senza però nessun parallelo incremento delle posizioni fisse e delle *tenure-track*. Come risultato, sarà il settore privato (o pubblico, ma governativo, cioè non accademico) a impiegare i dottori.

Sempre negli USA, l'ultimo report del *National Science Foundation*, il *Science and Engineering indicator 2018*²⁵ mostra come la percentuale di dottori di ricerca con speranza di ottenere una posizione di tenure o tenure-track è solo del 20% circa. Dal 2000 ad oggi si riscontra una netta diminuzione di posizioni da *full-time faculty* e contratti a termine. Restringendosi alle scienze fisiche, quelle che più ci riguardano, se nel 1973 il 21% del totale tra le posizioni in accademia erano occupate da fisici, oggi la percentuale è scesa al 14%.

Se la scarsità di posizioni stabili obbliga probabilmente molti ricercatori ad uscire dall'accademia contro la propria vocazione, i più alti stipendi in ambito privato hanno comunque una forte attrattiva: in media un postdoc americano guadagna 43.000 dollari nell'università e ben 73.000 nel privato come stipendio iniziale (Powell, 2015).

Da più parti (Gould, 2015; Powell, 2015; Cyranoski, 2011) si lancia l'allarme, auspicando come soluzione la drastica riduzione del numero di dottorati, e di una riconfigurazione strategica della struttura dei gruppi di ricerca, con un aumento delle posizioni fisse di ricercatore, a vantaggio della produttività "pro-capite". Per quanto una posizione simile possa apparire pragmatica in un sistema con scarsi finanziamenti, essa risulta insufficiente in quanto non tiene conto del contesto produttivo in cui l'università è inserita.

La ragione sistemica dell'alto numero di postdoc e dottori costretti a lasciare l'accademia è infatti il bisogno di formare, a spese dello Stato, un esercito di tecnici altamente specializzati a disposizione delle grandi e piccole aziende, ovvero risorse che verranno nella maggior parte dei casi trasferite al contesto privato.

Questi dati vanno dunque letti nel contesto del mondo del lavoro e della produzione moderni riassunti sopra: a fronte di un mercato in espansione enorme nel settore dell'AI vi è un bisogno estremo di manodopera nel campo della Data Science, che viene colmato grazie ad un'incapacità sistemica dell'accademia nell'assorbire i ricercatori.

²⁴ Discorsi a parte vanno fatti per paesi come Cina, India, Polonia, Singapore nei quali, per svariati motivi, solo recentemente si è assistito ad un boom di investimenti pubblici in ricerca).

²⁵ <https://www.nsf.gov/statistics/2018/nsb20181/report>

A riprova di questo fatto, nel già citato rapporto di IBM si stima che entro il 2020 la domanda di tecnici dei dati nel settore privato aumenterà del 28%, con un picco del 39% in settori altamente qualificati dove è necessario un dottorato (e dove gli stipendi superano i 100.000\$). A fronte di questo boom di richieste IBM denuncia una carenza di profili specifici e chiede esplicitamente che “il sistema educativo” si faccia carico della loro produzione²⁶.

6. IL MACHINE LEARNING DALLA FISICA ALLA SOCIETÀ

In questa sezione commenteremo alcuni elementi critici legati all'uso del ML in ambito sociale, paragonandole al caso della fisica. Evidenzieremo in particolare che molti di questi elementi sono parte del processo già *nelle scienze pure*, per ragioni intrinseche.

6.1. Precisione, falsi positivi e fairness

Quando si fa uso di modelli e strumenti matematici, un concetto chiave è la loro capacità di fornire previsioni “precise” in modo quantificabile.

Tuttavia, già all'interno della scienza, senza riferirsi al ML, il concetto o la soglia di “precisione” accettati nelle rispettive comunità variano significativamente. Un caso paradigmatico risale al 2015: la Società Italiana di Fisica (SIF) non sottoscrisse la *Dichiarazione sui cambiamenti climatici* per la Conferenza di Parigi sul clima COP 21, obiettando che il grado di precisione degli studi sul clima non era abbastanza elevato per affermare “incontrovertibilmente” l'impronta antropica sul clima. La definizione di “incontrovertibile”, o la soglia oltre la quale si accetta una misura come evidenza, vengono tuttavia decise dalla comunità e non sono “contenute” nei dati o nella misura; sono il risultato di tecniche statistiche che possono variare significativamente a seconda del campo.

Anche la definizione stessa di “precisione” di un algoritmo può variare in modo sostanziale. Vediamo il caso del ML con un esempio dalla fisica: un rivelatore di particelle. L'algoritmo che regola il *trigger* dell'esperimento può essere programmato per minimizzare, ad esempio, la perdita di “veri segnali” a scapito di avere un maggior numero di

²⁶ Le conclusioni di questa sezione poggiano su dati provenienti dal mercato del lavoro combinati con l'analisi delle pubblicazioni scientifiche delle sezioni precedenti. D'altra parte, sarebbe utile e necessario corredare queste conclusioni con un lavoro di inchiesta nella comunità accademica, e in particolare tra i giovani ricercatori.

“falsi allarmi” (cioè eventi erroneamente classificati come segnale), oppure per minimizzare il *rate* di falsi allarmi, a costo di rischiare di mancare alcuni eventi reali. È importante sottolineare che è la stessa definizione matematica di questi due casi (“Falso positivo” e “Falso negativo”) a far sì che un compromesso tra i due sia necessario²⁷. Esistono addirittura altre possibili scelte oltre alle due qui citate. Questo tipo di scelte è essenziale nella scienza (e nell’uso degli algoritmi in generale), fa parte del suo stesso statuto e non è superabile con nessun avanzamento tecnologico. Se nel caso di un esperimento di fisica delle particelle una simile decisione potrebbe avere conseguenze tutto sommato rimediabili (magari un tempo più lungo per arrivare ad una scoperta), non lo stesso si può dire di un software che abbia un ruolo decisionale o consultivo nella vita pubblica. Ciò ci porta a riformulare tali concetti in termini di *fairness* (equità) dell’algoritmo. Un esempio famoso è il software COMPAS²⁸, usato negli Stati Uniti per predire il rischio nel rilasciare un detenuto su cauzione prima del processo (Courtland, 2018). Un team di giornalisti ha mostrato come il software non fosse “equo” – al contrario di quanto sostenuto dall’azienda produttrice –, in quanto prediceva un maggior numero di falsi positivi tra la comunità afroamericana. Alla radice della controversia vi è proprio il fatto che i due team usavano definizioni di equità diverse. È importante capire che questa differenza si dà *in primis* a livello matematico: esistono infatti diverse definizioni di *fairness*²⁹ che sono statisticamente impossibili da conciliare, e la cui scelta è in mano al programmatore. Il prof. A. Narayanan, *computer scientist* a Princeton, ha recentemente tenuto un seminario dal titolo *21 fairness definitions and their politic*³⁰ dove mostra appunto che tante possono essere le definizioni di questo concetto applicabili in un algoritmo.

Si potrebbe obiettare che molti classificatori hanno differenze che possono essere minimali (pochi punti percentuali, o meno) tra i diversi tipi di errore. Tuttavia, nel caso di vite umane ogni dettaglio può essere determinante, come nel caso di SKYNET, un algoritmo usato dalla

²⁷ Ad esempio, l’*f-score*, una media ponderata tra le due, o l’*accuracy*.

²⁸ COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), prodotto dall’azienda Northpointe, ora Equivant (<http://www.equivant.com/>)

²⁹ In questo caso la, *predictive parity* – cioè il fatto che la probabilità di essere ri-arrestati tra bianchi e neri non dipenda dal gruppo di appartenenza – e il *false positive rate*, cioè la misidentificazione come soggetto a rischio. A sua volta un minimo *false positive rate* è inconciliabile con un minimo *false negative rate* - in questo caso, la misidentificazione come soggetto non a rischio.

³⁰ Qui il video della conferenza: <https://www.youtube.com/watch?v=jIXluYdnyyk>. Per il report: <https://docs.google.com/document/d/1bnQKzFAzCTcBcNvW5tsPuSDje8WWWY-SSF4wQm6TLvQ/edit>

NSA per monitorare la popolazione pakistana assegnando un rischio di collusione col terrorismo. Una fuga di notizie³¹ ha rivelato che il livello di falsi positivi poteva oscillare tra lo 0.008% e il 0.18%. Su una popolazione di 55 milioni di abitanti, questo significa che fino a circa 100mila pakistani potrebbero essere erroneamente messi sotto sorveglianza come sospetti terroristi (Hosni, 2017). Al di là dei numeri, questo esempio mostra anche la necessità di una decisione a monte sull'utilizzabilità dell'algoritmo stesso: è accettabile applicare uno strumento simile al monitoraggio di un gruppo etnico (per quanto piccolo, addirittura nullo, possa essere l'errore)?

6.2. *Bias e ipotesi autoavveranti*

Nella prima sezione abbiamo parlato del concetto di “dati carichi di teoria” introdotto da Kuhn. Questo concetto è legato a quello di *bias*. Nell'applicazione degli algoritmi di DL al contesto sociale abbiamo già molti esempi che dovrebbero fare scuola. Vediamone uno eclatante: lo scorso anno in uno studio dell'università di Stanford (Wang, 2017) una DNN è stata usata per distinguere, tramite riconoscimento facciale, tra eterosessuali e omosessuali. A detta degli autori la precisione oscillava tra l'80% e il 70% – molto maggiore dei risultati di un umano. Molte testate giornalistiche rilanciarono lo studio, che fece scalpore perché giustificava possibili ipotesi neo-lombrosiane e cause genetiche dell'omosessualità. Qualche mese dopo i ricercatori di Google³² mostrarono con analisi più sofisticate che l'algoritmo era affetto da grossolani stereotipi di genere nell'allenamento e più che distinguere la fisionomia, discriminava in base al trucco, gli occhiali e altre caratteristiche riguardanti il look e contesto sociale più che la genetica. Resta un altro grande problema: la domanda posta all'algoritmo e la divisione binaria tra etero e omosessuali sono di per sé controverse e frutto di stereotipi. In questo caso, la non-neutralità della scienza e dell'algoritmo sono palesi. Anche qui possiamo trovare analogie con la fisica. Nel 1572, Tycho Brahe osservò un nuovo oggetto celeste molto luminoso, che interpretò come la formazione di una nuova stella. Oggi sappiamo che l'evento è una Supernova Ia, un evento astronomico che corrisponde alla fine del ciclo di vita di una stella, non conosciuto all'epoca di Tycho Brahe. Lo stesso errore di classificazione potrebbe

³¹ SKYNET: Applying Advanced Cloud-based Behavior Analytics. The Intercept, 8 May 2005.

³²<https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477>

avvenire nel caso di un moderno algoritmo di ML allenato, poniamo il caso, a classificare l'immagine di un oggetto celeste come "stella" o "galassia". Il problema è di nuovo che il numero di classi è determinato dall'esterno e guidato dalla teoria.

Un ulteriore spunto di riflessione viene ancora dal software COMPAS. Il fatto che la predizione di rischio tra gli afroamericani sia più elevata e che questo risulti da un bias pregresso, innesci un processo per cui tali pregiudizi vengono confermati e probabilmente enfatizzati.

Un altro esempio sono i casi dei cosiddetti software di *predictive policing* usati negli Stati Uniti per predire in quali aree è più probabile che avvenga un crimine (senza sapere, sottolineiamo, per quale ragione), come *PredPol*³³. L'uso indiscriminato dei dati a disposizione può creare un feedback per cui, se gli agenti sono spinti ad essere presenti in quartieri precedentemente segnati da più reati, compiranno più interventi per il semplice fatto di essere già sul posto, e non perché il rischio sia realmente più elevato (Ensign, 2017). Questo indurrebbe un circolo vizioso paradossale per cui la pretesa "pericolosità" verrebbe confermata! Alla base vi è una doppia fallacia: l'assumere che i dati possano rivelare tendenze valide in futuro in assenza di una spiegazione (quello che in fisica avevamo chiamato "modello"), e il fatto che più interventi polizieschi siano dovuti a maggiore insicurezza (classico caso di correlazione spuria). Non è del resto questa una situazione nuova. In economia la polemica corre da anni su binari simili: l'economia neoclassica non solo fallisce nelle sue previsioni (Sylos Labini, 2016), ma è l'uso stesso dei suoi strumenti, ricavati all'interno di un paradigma teorico senza basi empiriche, che porta gli attori in gioco a comportarsi in maniera tale da far scoppiare le crisi (Bouchaud, 2008). Ancora una volta, è determinante capire come in fisica il fatto che un modello sia presente per fornire previsioni non è abbastanza enfatizzato nella discussione sul ruolo degli algoritmi.

6.3. *La proprietà dei dati.*

In fisica, la replicabilità degli esperimenti e l'accesso ai dati sono considerati un principio inviolabile e condizioni necessarie per trarre qualsiasi conclusione.

Inoltre, come abbiamo visto, il tipo di dati da raccogliere (in altri termini, la progettazione dell'esperimento) sono guidati dalla teoria e parte del processo di ricerca. La stessa cosa non si può dire delle

³³ PredPol, Predict Prevent Crime. Predictive Policing Software: www.predpol.com/

moderne *data warehouse* e dell'uso di modelli data-driven al di fuori dell'accademia. Particolarmente grave è la questione della proprietà dei dati. Se infatti molti strumenti alla base dell'estrazione di valore tramite previsioni sono oggi pubblici³⁴, la vera risorsa alla base dell'accumulazione sono proprio i dati su cui allenare gli algoritmi e questi sono spesso proprietari: "data is the new oil", recita una battuta, e tanto basta per farne capire il valore. È importante menzionare che questo fenomeno ha un impatto sul processo di migrazione al di fuori dell'accademia, in quanto spesso ricercatori in istituzioni pubbliche hanno serie difficoltà di accesso ai dati per condurre le loro ricerche, mentre il problema non si pone per dipendenti di grandi piattaforme private³⁵. La privatizzazione dei dati genera quindi delle gerarchie di potere rigidissime e violente, aggravate dal fatto che la proprietà tende a concentrarsi nelle mani di poche grandi piattaforme.

6.4. *Il ruolo dello scienziato nella società.*

Abbiamo sottolineato più volte come gli algoritmi di ML prevedano, nel loro utilizzo, una serie di passaggi in cui l'utilizzatore è tenuto a prendere delle decisioni che influenzeranno l'output, come la selezione delle feature o la scelta di una soglia di probabilità³⁶ e che questa arbitrarietà è presente a vari livelli. I moderni algoritmi di DL, invece, funzionano più come *black boxes* difficilmente decifrabili, in cui il processo che porta alla decisione finale può addirittura non essere ricostruibile, ma che comporta comunque scelte come il criterio con cui valutarne la precisione. In entrambi i casi, inoltre, si assume la possibilità di applicare a casi non conosciuti uno strumento predittivo allenato solo su altri esempi, rinunciando completamente ad un modello. L'adottare questo modo di procedere "induttivista" comporta già una profonda presa di posizione epistemologica, spesso ignorata al di fuori della fisica.

L'uso della modellizzazione e della matematica in contesto sociale ha una doppia implicazione: da un lato rende falsamente oggettivo (o, si potrebbe dire, naturale, neutrale) un processo "social-driven" (parafrasando il "data-driven"), in un certo senso "normalizzandolo".

³⁴ Come nel caso delle librerie python TensorFlow e PyTorch, sviluppate rispettivamente da Google e Facebook per implementare reti neurali profonde.

³⁵ L'esempio più eclatante è il caso di Facebook, che non permette l'accesso ai suoi contenuti tramite API - nemmeno in forma limitata, come nel caso di Twitter. Per una discussione di alcuni casi specifici legati a Google e Facebook, si veda Cozzo (2018).

³⁶ Anche se in questo lavoro siamo interessati al caso del ML, il discorso vale in generale per ogni tipo di algoritmo.

Dall'altro fa ricadere sullo scienziato, tramite la sua ricerca, la decisione politica, ergendolo a tecnico-veicolo del sistema culturale ed economico egemone. Se per la prima implicazione abbiamo già fatto obiezioni sui rischi di una applicazione naïve in campo sociale di tecniche matematiche, forzando le ipotesi in cui erano state inizialmente sviluppate (qual è il dominio, come retroagisce il sistema, etc.), ben più sottile, ma altrettanto importante e rischiosa è la seconda implicazione. Lo scienziato, in un contesto di divisione tecnica del lavoro, è spesso convinto o è spinto a credere di avere un ruolo puramente tecnico e di non avere un ruolo politico oppure, paradossalmente, crede di averne uno ancora più alto: essere il disvelatore della verità, davanti alla quale la politica non conta. Feticismo della scienza, alienazione e precarietà lavorativa si mischiano per rafforzare in maniera inconsapevole il sistema socioeconomico e ideologico, a maggior ragione quando nemmeno lo scienziato è pienamente consapevole del lavoro svolto dall'algoritmo, data la sua opacità e *bias*.

7. CONCLUSIONI

Questo lavoro è dedicato allo studio dell'utilizzo del ML nella fisica contemporanea. La scelta di questo caso parte dal riconoscimento della natura storicamente e politicamente determinata della scienza e il suo legame con la materialità dell'epoca in cui essa si sviluppa. Il capitalismo contemporaneo si fonda sull'estrazione di valore da grandi moli di dati tramite algoritmi avanzati di ML e grande potenza di calcolo. Questi tre elementi hanno segnato un momento di svolta nel ruolo delle grandi piattaforme, mutando quindi la produzione e il mercato del lavoro. È naturale dunque chiedersi quale sia l'impatto di tale processo sullo sviluppo della scienza, e allo stesso tempo quale ruolo la scienza abbia nell'alimentare e legittimare un certo paradigma nei suoi aspetti politici e sociali, in questo caso il ruolo determinante che gli algoritmi giocano nel mondo contemporaneo.

L'analisi si è sviluppata attorno a tre elementi:

1. *la scienza non è neutrale*: la ricerca scientifica, anche nelle scienze "pure", evolve in modo non lineare seguendo inestricabilmente le trasformazioni del modello produttivo egemone nel resto della società.

2. *gli algoritmi non sono neutrali*, sia in quanto prodotti in un ambito scientifico che non è neutrale, sia perché contengono elementi soggetti a decisioni esterne non eliminabili con alcun progresso tecnologico.

3. Nella prospettiva di un paradigma “data-driven”, diventa tanto più importante affermare l’impossibilità di avere strumenti oggettivi e neutrali addirittura all’interno delle scienze pure.

Abbiamo affrontato il problema secondo linee complementari.

Una analisi storica e bibliometrica dell’utilizzo del ML nella fisica (sezione IIIa) ci ha permesso innanzitutto di mostrare come la sua esplosione coincida con avanzamenti tecnologici avvenuti fuori dall’accademia per ragioni legate a necessità del mercato (estrarre valore da grandi moli di dati) che hanno a loro volta determinato una profonda trasformazione della produzione e del mercato del lavoro. Abbiamo inoltre mostrato come il boom del ML in fisica corrisponda anche a necessità materiali contingenti dei ricercatori, i quali, sottoposti a condizioni di lavoro precarie e a poche prospettive di carriera dentro l’accademia, vengono probabilmente spinti a indirizzare studi e ricerche verso un ambito spendibile nel mercato del lavoro (sezione IV).

In secondo luogo, abbiamo passato in rassegna l’uso del ML nei campi della fisica fondamentale in cui questo è più rilevante: la fisica delle particelle, l’astrofisica, la meccanica quantistica e la fisica della materia condensata (sezioni IIIb-IIIId). Significativamente, l’impatto dei fattori esterni all’accademia sopra menzionato si riscontra direttamente nella letteratura scientifica. Allo stesso tempo, abbiamo mostrato come ipotesi *à la Anderson* che paventino una prossima “fine della teoria” e una scienza che faccia a meno di modelli non sembrano reggere un confronto approfondito con la realtà, al di là di ogni entusiasmo tecnofuturistico³⁷.

Infine, il fatto che problemi di interpretazione dei risultati emergano chiaramente anche in un contesto teorico e sperimentale fortemente controllabile come quello della fisica fondamentale, in cui le variabili in gioco si presumono ben definite, ci ha permesso di trasferire le nostre critiche all’utilizzo degli algoritmi in ambito sociale, dove *bias*, pregiudizi, incontrollabilità del sistema in oggetto e *feedback* assumono un ruolo dirimente e in cui in gioco non ci sono variabili di un esperimento, ma persone (sezione V).

Tuttavia, tornando ad Anderson e valutando l’impatto che gli algoritmi hanno oggi, il loro impiego ed una sempre più diffusa retorica basata su presunte “verità scientifiche” e in particolare sui dati, è forte il sospetto che la sua previsione, anche se basata su presupposti falsi, si possa avverare. La possibilità di predizioni sempre più accurate, la

³⁷ Per altri commenti critici interni alla comunità scientifica si veda anche (Zdeborová, 2017).

potenza di calcolo, e i progressi stupefacenti nel DL sembrano poter riaprire acriticamente la strada al ritorno di una sorta di nuovo positivismo, scientifico e sociale, fondato sui dati, e trainato non tanto dagli scienziati, che restano consapevoli della posta in gioco nel metodo scientifico, ma da fattori esterni alla scienza.

Un vero cambio di paradigma, questo sì, nel senso di Kuhn. Non si tratta più di un positivismo basato sull'oggettività del metodo, ma sulla presunta neutralità dell'algoritmo, sulla possibilità di spiegare tutto attraverso i dati e sulla non-necessità di costruire modelli predittivi (nella scienza) e in ultima istanza immaginarne di migliori nella società.

Se restiamo fedeli all'esercizio di prendere Anderson sul serio, lo scenario futuro è uno sviluppo dominato dai dati, una distopia destinata a riprodurre in maniera circolare lo stato di cose esistenti.

Tocca allora anche al mondo scientifico riportare il problema alla giusta dimensione, ricontestualizzando la scienza stessa e contribuendo a decostruirne l'aura di assoluta oggettività, neutralità ed astoricità che oggi ne caratterizza dati ed algoritmi. Con questo lavoro abbiamo cercato di fare un passo in questa direzione.

RIFERIMENTI BIBLIOGRAFICI

- ABBOTT, B. P. *et al.* (2016). LIGO Scientific Collaboration and Virgo Collaboration. *Physical Review Letters*. 116, 061102.
- ADAM-BOURDARIOS, C. *et al.* (2014). The Higgs boson machine learning challenge., *Journal of Machine Learning Research*, 42, 19-55.
- ANDERSON, C. (2008). The End of Theory, Will the Data Deluge Makes the Scientific Method Obsolete?. *Wired Magazine*, June 23.
- AYRES, D.S. *et al.* (2007), *The NOvA technical design report*. NOvA Collaboration.
- BALDI, P., SADOWSKI, P., WHITESON, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, 5, 4308.
- BIAMONTE, J. *et al.* (2017), Quantum Machine Learning, *Nature* 549, 195-202.
- BOUCHAUD, J.-P. (2008). Economics needs a scientific revolution. *Nature*, 455, 1181-1181.
- Carrasquilla, J., Melko, R. G. (2017). Machine learning phases of matter. *Nature Physics*, 13, 431-434.
- CERN (2017). ATLAS Collaboration. *Identification of Jets Containing*
-

- b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Report No. ATL-PHYS- PUB-2017-003.
- CICCOTTI, G., CINI, M., DE MARIA, M., JONA-LASINIO, G. (1976). *L'Ape e l'architetto: paradigmi scientifici e materialismo storico*. Roma: Feltrinelli.
- COURTLAND, R. (2018). Bias detectives: the researchers striving to make algorithms fair. *Nature* 558, 357-360.
- COZZO, E. (2018). *Pratiche scientifiche ai tempi del capitalismo di piattaforma*. In D. Gambetta (a cura di), *Datacrazia. Politica, cultura algoritmica e conflitti al tempo dei big data*. Roma: Escahon.
- CRAWFORD, K., BOYD, D. (2011). *Six provocation for big data*. Oxford Internet Institute's "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society".
- CYRANOSKI, D., GILBERT, N., LEDFORD, H., NAYAR, A. YAHIA, M (2011). The Ph.D. factory. The world is producing more PhDs than ever before. Is it time to stop?. *Nature*, 472, 276-279.
- DUNJKO, V., BRIEGEL, H. J. (2018). Machine learning & artificial intelligence in the quantum domain: a review of recent progres. *Reports on Progress in Physics*, 81, 074001.
- ENSIGN, D. FRIEDLER, S. A., NEVILLE, S., SCHEIDEGGER, C., VENKATASUBRAMANIAN, S. (2017). Runaway feedback loops in predictive policing. *arXiv*, arXiv:1706.09847
- FEYERABEND, P. K. (2002). *Contro il metodo. Abbozzo di una teoria anarchica della conoscenza*. Milano: Feltrinelli.
- GABBARD, H., WILLIAMS, M., HAYES, F., MESSENGER, C., (2018). *Physical Review Letters*, 120, 141103
- GILLET, N. et al., (2018). Deep learning from 21-cm images of the Cosmic Dawn. *arXiv*, arXiv:1805.02699
- GOLDTHORPE, J. H., (2016). *Sociology as a population science*. Cambridge University Press.
- GOMEZ-URIBE, C. A., HUNT, N. (2016). The Netflix recommender system: Algorithms, Business value, and Innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6 (4), 13.
- GOODFELLOW, I., BENGIO, Y., COURVILLE, A. (2016). *Deep learning*. Cambridge: MIT press.
- Gould, J. (2015). How to build a better PhD, *Nature* 528, 22-25
- HEZAVEH, Y. D., LEVASSEUR, L. P., MARSHALL, P. J. (2017). Fast automated analysis of strong gravitational lenses with convolutional neural networks, *Nature*, 548, 555-557
-

- HINTON, G. E., SRIVASTAVA, N., KRIZHEVSKY, A., SUTSKEVER, I., SALAKHUTDINOV, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv*, arXiv:1207.0580.
- HOSNI, H., VULPIANI, A. (2017). Forecasting in the light of Big Data. *arXiv*, arXiv:1705.11186
- HUEMBELI, P., DAUPHIN, A., WITTEK, P. (2018). Identifying quantum phase transitions with adversarial neural networks. *Physical Review B*, 97(13), 134109.
- KOLANOVIC, M., KRISHNAMACHARI, R. T. (2017). *Big Data and AI Strategies Machine Learning and Alternative Data Approach to Investing*. JP Morgan Global Quantitative & Derivatives Strategy Report.
- KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. (2012). ImageNet classification with deep convolutional neural networks. *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, 1097-1105.
- KUHN, T. S. (1972). *La struttura delle rivoluzioni scientifiche*. Torino: Einaudi.
- LAKATOS, I., FEYERABEND, P., MOTTERLINI, M. (1995). *Sull'orlo della scienza. Pro e contro il metodo*. Milano: Raffaello Cortina.
- LAM, C., KIPPING, D., (2018). Transit clairvoyance: enhancing *TESS* follow-up using artificial neural networks, *Monthly Notices of the Royal Astronomical Society*, 476(4), 5692-5697.
- LATOUR, B. (2010). *Tarde's idea of quantification*. In M. Candea (Eds.) *The Social After Gabriel Tarde: Debates and Assessments* (pp. 145-162). New York: Routledge.
- LLOYD, S., MOHSENI, M., REBENTROST, P. (2013). Quantum algorithms for supervised and unsupervised machine learning. *arXiv*, arXiv:1307.0411
- MAAS, A. L., HANNUN, A. Y., NG, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. Proceedings of the 30 th International Conference on Machine Learning, 1-6.
- MOSTAFA, A., Y. S., MALIK M.-I., LIN, H.-T. (2012), *Learning from data*. Vol. 4. New York: AMLBook.
- POINCARÉ, H., (1905). *Le valeur de la science*. Paris: Flammarion
- POPPER, K. R. (1991). *Scienza e filosofia*. Torino: Einaudi.
- POWELL, K. (2015). The future of the postdoc. *Nature*, 520, 144-147.
- RADOVIC, A. *et al.* (2018). Machine learning at the energy and intensity frontiers of particle physics, *Nature* 560, 41-48.
- RAINA, R., MADHAVAN, A., NG, A. Y. (2009, June). Large-scale deep
-

- unsupervised learning using graphics processors. *Proceedings of the 26th annual international conference on machine learning*, 873-880).
- ROSENBLATT, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- SAMUEL, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210-229.
- SHALLUE, C. J., VANDERBURG, A. (2018). Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *The Astronomical Journal*, 155(2), 1-22.
- SYLOS LABINI, F. (2016). *Rischio e previsione. Cosa può dirci la scienza sulla crisi*. Roma-Bari: Laterza.
- SZEGEDY, C. *et al.*, (2014). Going deeper with convolutions. *arXiv*, arXiv:1409.4842.
- VAN NIEUWENBURG, E. P. L., Liu, Y. H., Huber, S. D. (2017). Learning phase transitions by confusion. *Nature Physics*, 13(5), 435.
- WANG, Y., KOSINSKI, M. (2017) Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of personality and social psychology*, 114(2), 246.
- ZDEBOROVÁ, L. (2017). Machine learning: new tool in the box. *Nature Physics*, 13(5), 420.
-

OLTRE LE BOLLE DEI FILTRI E LE TRIBÙ ONLINE

Come creare comunità “estetiche” informate attraverso gli algoritmi

di *Luca Serafini**

Abstract

This paper aims at demonstrating how, starting from the aesthetic and playful modalities of social relationship fostered by the Internet, it is possible to overturn the dominant relational patterns imposed by the algorithms of the main players of the digital world like Facebook and Google. If the latter promote the formation of tribal, like-minded communities whose narratives look like the “grand récits” of modern epistemology, the media can go beyond this paradigm by implementing technological and cultural alternatives. In this way, starting from the “social” modalities of spreading information online, it is possible to create well-informed communities open to the confrontation with diversity, overcoming the hyper-emotional and self-referential model produced by the “filter bubbles” and the “echo chambers”.

Keywords

Aesthetics, algorithms, community, serendipity

* LUCA SERAFINI è dottore di ricerca in filosofia, è docente a contratto presso la Libera Università Maria Santissima Assunta (Lumsa) di Roma e giornalista professionista. Lavora presso la testata The Post Internazionale e collabora con Il Foglio e Rivista Studio. È autore di due monografie e di saggi pubblicati su riviste italiane e internazionali.

Email: luca86.serafini@gmail.com

1. INTRODUZIONE

L'obiettivo di questo lavoro è quello di dimostrare come gli algoritmi e l'intelligenza artificiale, inserendosi in un contesto in cui l'informazione coincide sempre di più con il concetto di "relazione sociale", riescano a programmare la relazione umana a partire da modalità ludiche ed estetiche di fruizione dell'informazione stessa.

Dimosteremo come, proprio prendendo come modello di riferimento quello estetico-ludico dell'epoca dei social network, sia possibile creare, anche attraverso l'utilizzo degli algoritmi e dell'intelligenza artificiale, comunità informate, aperte alla relazione con il diverso e all'etica della differenza, in grado di approcciare il singolo contenuto informativo in maniera riflessiva e mediata, e non esclusivamente emozionale e irriflessa.

Mostreremo, attraverso esempi teorici e pratici, come l'utilizzo degli algoritmi da parte dei grandi *players* dell'informazione, in particolare gli Over The Top come Facebook e Google, produca la formazione di "comunità estetiche" in cui l'elemento emozionale dà vita a forme di aggregazione di tipo tribale. Ciò avviene attraverso una modalità anch'essa estetica e ludica di diffusione delle informazioni facilitata, anche qui, dall'utilizzo degli algoritmi.

Sosterremo poi come questa modalità di diffusione e fruizione dell'informazione rappresenti un ritorno ad un'epistemologia moderna, quella dei *grand récits*, il cui obiettivo è quello di ridurre la complessità del reale a schemi onnicomprensivi, producendo comunità omogenee e chiuse alla relazione con l'altro, il diverso, comunità del tutto refrattarie alla messa in discussione della propria meta-narrazione.

Sosterremo poi come per ribaltare questo paradigma non si possa prescindere dalla modalità estetica e ludica con cui le persone si informano nell'epoca dei social network. Partiremo da quella che viene definita la "sfera pubblica estetica", una sfera cioè che considera la dimensione emozionale, ludica e finanche frivola parte del discorso pubblico e potenziale fonte di innesco dell'impegno civico.

Verranno analizzati alcuni tentativi di favorire la "serendipità online", ovvero l'esperienza dell'incontro col l'inatteso, con ciò che non era programmato, attraverso soluzioni tecniche (ad esempio estensioni di browser) che favoriscono la fruizione di contenuti non in linea con le nostre ricerche abituali, nonché i tentativi di alcune grandi testate anglosassoni (come il *The Guardian* e il *Washington Post*, che stanno investendo su nuove soluzioni tecnologiche) di sottrarre a Facebook e Google il dominio sugli algoritmi per proporre diversi modelli di

fruizione delle notizie, attraverso i quali l'esperienza online diventi meno programmata, personalizzata e tribale e maggiormente diversificata.

Infine, verranno proposte alcune soluzioni per sottrarre agli Over The Top e al modo in cui utilizzano gli algoritmi il "monopolio della relazione" e per creare comunità di utenti più riflessive, informate e aperte alla contaminazione con la diversità, sempre a partire dalle modalità estetiche e ludiche di fruizione dei contenuti informativi fin qui richiamate.

2. LE COMUNITÀ ESTETICHE

La nascita della società in rete avviene in un periodo storico segnato dalla progressiva intensificazione di alcuni processi culturali iniziati con il passaggio alla società post-industriale e postmoderna. Ciò significa che il nuovo concetto di identità che viene a formarsi con la nascita delle comunità virtuali si inserisce all'interno di fenomeni culturali che avevano già preso piede almeno dagli anni '70 del Novecento.

In alcuni casi, la rete tende ad intensificare alcune dinamiche intra e intersoggettive che vengono comunemente associate alla società postmoderna, come la disseminazione dell'identità o, detto in altri termini, la scissione del soggetto in identità multiple che cambiano a seconda dei contesti.

La possibilità, in rete, di interagire attraverso avatar, profili social, e di appartenere a diverse comunità online con differenti identità (Turkle, 1995), amplifica questo processo di scissione del soggetto che, nella società postmoderna, è determinato anche da altri fenomeni come la prevalenza della sfera estetica e figurale, dell'immediatezza sensoriale e dei processi primari.

In un articolo pubblicato alla fine degli anni Ottanta, Scott Lash (1988) definiva il postmodernismo proprio come il paradigma culturale che porta a compimento la prevalenza del figurale sul discorsivo. Si tratta di una tesi già contenuta nei primi scritti del teorico per eccellenza del postmodernismo, Jean-François Lyotard (1971). Per Lyotard, infatti, se la sensibilità moderna poteva essere considerata discorsiva, quella postmoderna si caratterizza certamente come figurale, poiché vi opererebbero una serie di meccanismi legati all'immediatezza sensoriale, al visivo, all'inconscio, che si rafforzano anche grazie alla proliferazione di immagini e segni tipica della società post-industriale, in cui avviene una estetizzazione della realtà.

Col termine "estetizzazione" si intende la prevalenza delle

immagini, l'investimento immediato del desiderio negli oggetti che è favorito proprio dal figurale, e che si oppone al mantenimento di una distanza cognitivo-razionale. Cinema, televisione, pubblicità, la cultura del consumo in generale, sono tutti elementi che favoriscono questa immersione nel desiderio e nelle sensazioni irriflesse.

L'origine di questa svolta emozionale può già essere rintracciata nella cultura del consumo che si sviluppa nelle grandi città dell'Ottocento (Campbell, 1987), che assurgono a simboli dello stile di vita promosso dal capitalismo industriale (Featherstone, 1992). Già negli scritti sulle metropoli, la moda, le merci, di autori prettamente moderni come Simmel e Baudelaire, si può rintracciare questo concetto di estetizzazione basato sui processi primari e sul figurale. Tuttavia, è con la società post-industriale che questa estetizzazione della realtà viene portata a compimento.

La società post-industriale (Rose, 1992, 20-25) viene generalmente identificata con la transizione da un'economia basata sulla produzione di beni, tipica appunto della società industriale, ad un'economia centrata sui servizi, che diventa progressivamente dominante nei decenni successivi al secondo conflitto mondiale.

Nella fase post-industriale dell'economia, la tecnologia è ormai diventata quella dell'informatizzazione, delle reti, dei networks, che anche nel funzionamento del sistema economico generano processi del tutto nuovi rispetto al capitalismo industriale come quello del lavoro immateriale, del movimento anche questo immateriale di beni, capitali, di quella che viene definita "fine dello spazio" e che si accompagna ai processi legati alla globalizzazione. In questo scenario, la sempre più rapida e globale circolazione degli oggetti diventa soprattutto una circolazione di artefatti culturali e di segni, veicolata da strumenti come le televisioni e i computer.

La realtà quindi diviene, in maniera progressiva, sovraccaricata esteticamente, e ciò genera un predominio di una cultura figurale incentrata sui processi primari, quelli legati alle emozioni, agli istinti, alle sensazioni.

Il già citato Scott Lash parla anche, a questo proposito di "de-differenziazione", intendendo con questo termine, sulla scia delle riflessioni di Walter Benjamin, il passaggio dalla separazione delle varie sfere della società in epoca moderna al loro rispettivo compenetrarsi in epoca postmoderna. Come sottolinea Lash, è proprio il postmodernismo a determinare l'avvento della de-differenziazione. Ciò avviene come conseguenza della disgregazione dei confini tra cultura alta (che perde la sua aura e non è più separata dall'ambito sociale) e bassa, della

colonizzazione che l'estetica opera sull'etico, il teoretico, il sociale. Ciò conduce appunto a quell'estetizzazione della vita individuale e dei meccanismi di produzione e consumo, nonché dei criteri in base ai quali avviene la valutazione morale, che è propria della società post-industriale o del capitalismo avanzato (Lash, 1990, tr. it. 2000, 16-25).

Questo predominio dell'estetica, così come è stato tratteggiato, determina una sostanziale riconfigurazione anche delle relazioni sociali. Prenderemo qui in considerazione tre autori che hanno teorizzato l'avvento di una modalità estetica nei rapporti interpersonali e nella sfera relazionale *tout court*, con particolare riferimento a quei rapporti veicolati dalla rete e dai social network.

Questa panoramica ci servirà a comprendere a quale specifica modalità relazionale prettamente estetica rimandi l'ingegneria sociale portata avanti dagli algoritmi, e dall'utilizzo che ne fanno gli Over The Top. Il primo autore che prenderemo in considerazione è Zygmunt Bauman. In molte sue opere, Bauman utilizza la contrapposizione tra "spazio estetico" e "spazio morale" per caratterizzare alcuni dei tratti peculiari della modernità liquida (Bauman, 1993). Per il sociologo polacco, la rottura con un sistema di convivenza tra gli individui basato sui principi ultra-razionali del moderno (Bauman, 1989) ha progressivamente aperto lo spazio per una disseminazione di rapporti interpersonali basati sull'episodicità, la frammentarietà.

Ne consegue la sostanziale impossibilità di costruire legami profondi, duraturi, la fuga dalle responsabilità morali che questi legami comportano, e il ripiegamento su modalità relazionali di tipo estetico. Le comunità estetiche, per Bauman, sono raggruppamenti di persone tenuti insieme da interessi di natura ludica: possibili esempi sono le comunità di fans, che spesso si aggregano attorno ad un simbolo, come un personaggio dello spettacolo. Bauman le definisce anche "comunità gruccia" per il loro carattere effimero: una volta venuto meno il motivo che ha portato alla costituzione di quella comunità, la stessa si sfalda alla ricerca di nuovi divertimenti (Bauman, 2001). Per Bauman lo spazio estetico della postmodernità è totalmente contrapposto a quello etico, che implica legami profondi, duraturi, responsabilità nei confronti degli altri.

Il carattere estetico delle relazioni interpersonali si ritrova per Bauman anche nell'atteggiamento assunto dalle persone nei confronti del diverso, dello straniero. Quest'ultimo è ormai percepito solo come una fonte di *divertissement* in quanto nuovo, stravagante, inconsueto. Seguendo il significato etimologico del termine estetica, ovvero quello di *aisthesis* (percezione, sensazione), ecco che «gli stranieri, con i loro

modi sconosciuti e imprevedibili, con la loro caleidoscopica varietà di apparenze e azioni, con la loro capacità di sorprendere, sono una fonte particolarmente ricca di piacere per gli spettatori. Sul piano estetico, lo spazio urbano è uno spettacolo il cui valore ludico annulla tutte le altre considerazioni» (Bauman, 1993, tr. it. 2010, 172-173).

Tutto ciò va a scapito dell'accoglienza dello straniero in uno spazio etico, che implica la disponibilità ad includerlo nella propria comunità, a sentirsi responsabile nei suoi confronti in maniera stabile e duratura.

Quanto visto finora, in Bauman, ha molto a che fare con il modello di relazionalità che si sviluppa nella società delle reti e dei social network, in cui basta un clic per stringere amicizia con qualcuno e un altro clic per porvi fine. Le relazioni interpersonali, su Facebook e sugli altri principali canali di socializzazione presenti in rete, si basano proprio su questa modalità estetica fatta di episodicità, istantaneità, di legami che si attivano e si disattivano in un istante e che sono guidati dalla necessità ludica di provare piacere e divertimento attraverso la relazione con l'altro, senza mai assumere nei suoi confronti impegni morali stabili e duraturi.

Questa è una prima importante accezione della modalità relazionale "estetica" che si sviluppa anche e soprattutto attraverso la rete e i social network.

Un altro sociologo come Michel Maffesoli considera invece le comunità estetiche della rete in un modo parzialmente diverso che però, come vedremo, si integra perfettamente con quanto evidenziato da Bauman. Per il sociologo francese, infatti, le comunità estetiche sono delle comunità in cui si cerca di superare l'atomizzazione e l'individualismo che hanno segnato l'epoca moderna, nell'ottica di riaffermare un senso forte del comune. Egli le definisce "neo-tribalismi" (Maffesoli, 1988), suggerendo che in aggregazioni come quelle di fans per un concerto o di tifosi per una partita di calcio, si manifesti un ritorno a forme pre-razionali e arcaiche di fusione comunitaria.

L'estetica delle tribù, sempre a partire dal significato etimologico di *aishtesis* (percezione, sensazione), è quella che spinge a fondersi in comunità sovra-individuali in cui l'unico scopo è quello di provare emozioni con gli altri, abbandonare ogni tipo di impegno e progetto a lungo termine per godere dell'istante (Maffesoli, 1990). Per Maffesoli, la rete e i social network rafforzano queste tendenze tribali, poiché aumentano la connettività tra gli individui, permettendo la formazione di comunità di persone affini per gusti, passioni, le cui aggregazioni in rete possono poi trasferirsi nel mondo reale.

La tecnologia determina per Maffesoli un "reincantamento",

diffondendo simboli e immagini in ogni parte del globo e contribuendo a caricare esteticamente la realtà, tanto quella materiale quanto quella virtuale, attivando così quei meccanismi legati al figurale, all'immediatezza sensibile e all'aspetto emozionale tipici della cultura postmoderna. La rete e i social network favoriscono insomma una circolazione affettiva che dà vita alle comunità estetiche e neo-tribali (Maffesoli, 2008).

Come nel caso di Bauman, queste comunità non hanno nulla di solido, di duraturo. Le persone sentono l'esigenza di godere l'istante, di non preoccuparsi per ciò che verrà domani, di dare valore alla gratuità dello stare insieme senza inquadralo in progetti socio-politici quali che siano. Mettendo insieme l'accezione baumaniana di comunità estetica e quella di Michel Maffesoli, ne viene fuori che, in rete, si formano delle aggregazioni in cui è preponderante l'aspetto emozionale, emotivo, affettivo, a totale discapito di quello razionale.

Si tratta di comunità che, come richiama l'immagine della tribù, hanno un alto tasso di omogeneità tra gli individui che le compongono, ma che allo stesso tempo si definiscono "estetiche" anche per il carattere frivolo e ludico che ha determinato la loro formazione.

Come vedremo, l'ingegneria sociale favorita dall'utilizzo che gli Over The Top fanno degli algoritmi rimanda proprio a questo legame tra comunità tribali ed omogenee, a fortissima carica affettiva ed emozionale e del tutto refrattarie alla riflessione e all'apertura razionale all'altro, al diverso, e il fatto che queste stesse comunità si formino sulla base di forme di interazione ludiche, frivole, in uno spazio estetico che è quello del *divertissement*, in cui tutto ciò che è impegno e riflessione viene bandito.

Tutto questo, come vedremo, ha molto a che fare con la maniera in cui gli Over The Top, diventati ormai i principali editori del globo, condizionano il modo in cui le persone si informano. Prima di arrivare a questo, però, è importante introdurre una terza accezione di "comunità estetica", molto diversa sia da quella di Bauman che da quella di Maffesoli.

Si tratta di una teoria che ha preso forma tra gli anni '90 e gli anni 2000, nel periodo in cui si riteneva che Internet potesse realmente favorire il contatto tra persone e culture diverse dando vita al cosiddetto "villaggio globale". Il già citato Scott Lash e John Urry parlarono a questo proposito, in diversi scritti, di "cosmopolitismo estetico", intendendo con questa espressione la possibilità che si apre, nell'era del web e della globalizzazione, di entrare in contatto con altre culture anche e soprattutto attraverso immagini, segni, artefatti culturali,

attraverso cioè quell'elemento figurale e affettivo che è tipico della cultura postmoderna. Per Lash e Urry la presa di contatto estetica può favorire un'ibridazione culturale che conduce alla responsabilità morale verso l'altro e alla relativizzazione della propria identità (Lash, Urry, 1994; Lash, 1999; Urry, 1995; Urry, Szerszynski, 2006).

Ciò significa che immagini, segni e artefatti culturali sono gli elementi che mi permettono di entrare in contatto con la diversità, magari in una forma leggera e ludica come quella della cultura popolare. Questo primo contatto emozionale, immediato, può poi generare nell'individuo un interesse verso quella specifica cultura che da estetico diventa morale, da irriflesso diventa mediato. Una comunità estetica, in quest'ottica, può diventare anche una comunità etica, e ciò verrebbe favorito proprio dalla circolazione globale di contenuti permessa dalla rete.

Fenomeni come quello delle "camere dell'eco" e delle "bolle dei filtri", che analizzeremo più in dettaglio nelle sezioni successive, dimostrano come questo paradigma si sia rivelato illusorio. L'utilizzo degli algoritmi da parte dei principali *players* del mondo dell'informazione e della socializzazione ha infatti determinato la formazione di comunità estetiche nell'accezione data da Bauman e Maffesoli. Il villaggio globale si è trasformato in un insieme di nicchie auto-referenziali, comunità composte da persone affini che non entrano quasi mai in contatto con opinioni e visioni del mondo diverse dalle proprie. L'estetizzazione della realtà, insomma, ha preso una direzione neotribale e allo stesso tempo ludica.

Nonostante questo, come vedremo, è ancora possibile piegare questa estetizzazione nel senso suggerito da Lash e Urry, permettendo la formazione di comunità eterogenee, aperte alla riflessione e alla responsabilità nei confronti degli altri, comunità allo stesso tempo estetiche ed etiche. Ciò passa per un diverso utilizzo degli algoritmi, che può modificare il modo in cui le persone si informano.

3. L'ESTETIZZAZIONE DELLE RELAZIONI SOCIALI IN RETE

Le ultime statistiche del *Pew Research Centre* indicano che, negli Stati Uniti, più del 50 per cento della popolazione si informa in maniera prevalente attraverso i social network (Meloni, 2017). Come ha scritto Michele Mezza, la disintermediazione prodotta dallo spostamento progressivo delle testate sui social network fa sì che l'informazione diventi conversazione, fino a coincidere in tutto e per tutto con il concetto di relazione interpersonale (Mezza, 2015, 8). Facebook

trasforma la notizia in una relazione amicale: sul Newsfeed appaiono indistintamente post degli amici e articoli delle testate giornalistiche (o articoli delle testate giornalistiche condivisi dagli amici), in un miscuglio che rende sostanzialmente indistinguibili i primi dai secondi.

La trasformazione della notizia in contenuto relazionale, dell'informazione in relazione sociale e amicale, ha anche a che fare con quella che Ruggero Eugeni ha definito la "condizione postmediale", ovvero una situazione in cui i vari media sono non solo non sono più distinti l'uno dall'altro (Jenkins, 2006), ma non sono nemmeno più percepiti come differenti rispetto alle situazioni non medialità di vita sociale (Eugenì, 2015).

Siamo in un'epoca, insomma, in cui non si può più distinguere tra cosa è mediale e cosa no, in cui i media sono ovunque e pervadono ogni ambito delle relazioni sociali. Come sottolinea opportunamente Eugeni citando le tesi di Peppino Ortoleva, la svolta postmediale si accompagna ad una svolta ludica: siamo infatti nell'epoca della *gamification* (Ortoleva, 2009), in cui meccanismi tipici del gioco vengono trasferiti anche a situazioni teoricamente non ludiche.

Per quanto riguarda il mondo dell'informazione, ciò è un diretto esito della progressiva socializzazione della notizia, conseguente alla sempre maggiore preminenza dei social network come mezzi e piattaforme attraverso cui e su cui le persone si informano.

È quindi evidente che, nella misura in cui Facebook e Google diventano i principali intermediari nella diffusione delle notizie, i loro algoritmi concorrano anche a plasmare i rapporti sociali tra le persone. Se notizia e relazione sociale sono la stessa cosa, le notizie che gli algoritmi decidono di farci leggere determinano il tipo di comunità a cui mi affilierò nonché il tipo di relazione che instaurerò con i membri di quella comunità. Come osservava già Marshall McLuhan, del resto, "il messaggio di un medium o di una tecnologia è nel mutamento di proporzioni, di ritmo, e di schemi che introduce nei rapporti umani" (McLuhan, 1964).

Rispetto a quanto evidenziato nella precedente sezione, si può allora dire che il modo in cui agiscono gli algoritmi degli Over The Top, in particolare di Facebook e Google, orienta le relazioni sociali nella direzione delle comunità estetiche tribali e ludiche di cui parlavano Bauman e Maffesoli. Ciò avviene, in primo luogo, attraverso la selezione delle notizie che gli algoritmi decidono di mostrarci, nel Newsfeed nel caso di Facebook, come risultato delle chiavi inserite nel motore di ricerca per quanto riguarda Google.

Si tratta del noto fenomeno delle "bolle dei filtri", messo in luce

compiutamente da Eli Pariser nel suo *Il filtro* (2011). Gli algoritmi di Facebook e Google personalizzano i risultati delle nostre ricerche o i contenuti che ci vengono proposti sulla base delle ricerche che abbiamo effettuato in precedenza (Google) e delle persone con cui interagiamo più spesso (Facebook).

Le bolle dei filtri, insomma, ci impediscono di entrare in contatto con contenuti che possano mettere in discussione le nostre idee, i nostri preconcezioni, la nostra visione del mondo, dando vita a una sorta di determinismo per cui ciò che vedremo in rete domani sarà il prodotto di ciò che abbiamo visto ieri.

Si formano così le cosiddette “camere dell’eco”, nicchie virtuali in cui si interagisce solo con chi già la pensa come noi, in cui le nostre idee sul mondo vengono continuamente confermate e rafforzate, comunità del tutto omogenee in cui non si sviluppa alcun senso critico e riflessivo, ma che sono guidate dalla carica affettiva ed emozionale prodotta dall’interazione tra simili.

Come sottolinea Cass Sunstein, questa dinamica di personalizzazione estrema ingenera una progressiva polarizzazione all’interno delle comunità di simili che si vengono a creare, rafforzando sempre più l’identità di gruppo sulla base di meccanismi di natura emozionale e non razionale (Sunstein, 2017).

Gli algoritmi, in particolare quello di Facebook, tendono poi a fondere l’elemento tribale con quello ludico. L’estetizzazione delle relazioni sociali avviene quindi in primo luogo attraverso il rafforzamento dei legami emotivi con chi già condivide la mia idea e la mia visione del mondo. A questo si aggiunge il fatto che Facebook promuove la fruizione di contenuti leggeri, frivoli. Le recenti modifiche all’algoritmo hanno infatti determinato una maggiore esposizione ai contenuti condivisi dagli amici rispetto agli articoli postati sul social network direttamente dalle testate.

Non solo, ma l’algoritmo di Facebook privilegia di default i contenuti con il maggior numero di like e condivisioni: sono questi ultimi che ci appaiono più facilmente davanti sul Newsfeed, ma si tratta nella maggior parte dei casi di contenuti leggeri che diventano facilmente virali, nonché contenuti ad alto potenziale di eccitazione, ad altissima carica affettiva ed emozionale.

Come spiegano Emanuele Arielli e Paolo Bottazzini in *Idee virali* (2018), i meccanismi alla base della condivisione dei post sono di natura emotiva e affettiva molto più che razionale. Nel momento in cui l’algoritmo di Facebook privilegia la visualizzazione dei contenuti maggiormente virali, ecco quindi che quello stesso algoritmo sta

promuovendo una “forma” delle relazioni sociali basata sul *loisir*, sull’aspetto ludico ed estetizzante.

Quest’ultimo però, come detto, si fonde con quello tribale: siamo esposti a contenuti in linea con la nostra visione del mondo, e per giunta si tratta di contenuti fortemente caricati dal punto di vista emotivo ed affettivo.

Un esempio è rappresentato dai meme satirici che irridono, ad esempio, un personaggio politico, mettendone in ridicolo le idee attraverso una composizione icastica di parole e immagini, che veicola il messaggio in maniera diretta e immediata. I meme hanno una capacità di diventare virali superiore a qualsiasi altro contenuto diffuso sui social, in gran parte a causa del funzionamento degli algoritmi sopra richiamato.

Si tratta, per l’appunto, di contenuti leggeri, ironici, che rimandano a una fruizione del social network di tipo ludico, ma che allo stesso tempo possono trasmettere un messaggio di natura politica e/o sociale. Proprio perché, nella loro composizione di testo e immagini, fanno leva sulle emozioni molto più che sulla razionalità e la riflessività, i meme contribuiscono a rafforzare il legame affettivo e tribale all’interno della camera dell’eco in cui circolano e lo fanno, per l’appunto, sulla base di un contenuto di per sé satirico, giocoso.

Questo contenuto giocoso, però, ha un potere molto maggiore di qualsiasi long-form o testo argomentativo di polarizzare gli utenti rafforzandone posizioni e pregiudizi. Il risultato non potrà che essere quello della formazione su base ludica di comunità omogenee, polarizzate, incapaci di entrare in contatto con l’estraneo, il diverso.

Sono comunità estetico-tribali per come le definiva Maffesoli, ma anche nel senso dato a questa espressione da Bauman. Come sottolinea Mariano Livolsi, questa modalità estetizzante di interazione con la propria comunità in rete può compromettere la capacità di selezionare in maniera autonoma i contenuti informativi nonché di saperli decodificare sulla base di un lavoro critico e riflessivo (Livolsi, 2013, 38-57). Attraverso, il sistema di like e condivisioni, gli individui stigmatizzano o approvano dei contenuti attraverso un carico emozionale di pulsioni, desideri, frustrazioni all’interno delle loro reti personali e personalizzate di contatti.

La nostra idea è che la formazione delle bolle dei filtri e delle camere dell’eco favorita dagli algoritmi tenda ad estremizzare queste tendenze, rendendo più marginali altre forme di interazione che presuppongono invece una maggiore autonomia riflessiva da parte degli individui.

Come emerge dai lavori di Anthony Giddens, infatti, nella tarda modernità (che confluisce nella società post-industriale e postmoderna) prende forma la “de-tradizionalizzazione” (Giddens, 1990), ovvero l’apertura della soggettività ad un’esperienza non più inquadrabile in strutture gerarchiche e spazialmente circoscritte.

La tradizione è infatti per Giddens ciò che teneva insieme le società pre-moderne e pre-capitaliste, società in cui la ritualità, la memoria collettiva, l’ancoraggio ad un sistema di valori predefinito dal contesto di appartenenza, permettevano all’individuo di vivere in una condizione di “sicurezza ontologica” (Giddens, 1991). Tutto ciò favoriva, evidentemente, delle forme di aggregazione basate sulla solidarietà organica tra i membri di comunità ristrette.

Nell’ottica di Giddens, con la tarda modernità si determina il passaggio ad un ordine post-tradizionale in cui le sicurezze derivanti dalla tradizione vengono sostituite da quelle fornite dalla conoscenza razionale. Ecco allora che nello scenario moderno, quello del “dis-ancoraggio” dai contesti locali, la costruzione del sé diventa sempre più un progetto riflessivo.

Ciò che Giddens descrive si colloca nel punto di giuntura tra la categoria di “modernizzazione riflessiva” e quella di “individualizzazione”. Entrambe si riferiscono alla crescente e forzata personalizzazione delle scelte di vita in un ordine post-tradizionale, e la loro interconnessione teorica è stata rilevata non solo da Giddens, ma anche da altri importanti sociologi come Ulrich Beck (2002) e Gilles Lipovetsky (1987, 1989). Tutti questi autori sottolineano come l’autonomia riflessiva del soggetto venga rafforzata, e non intaccata, dalla società globale della rete e dei networks.

Lungi dal negare l’esistenza di questi elementi, la nostra tesi però è che sia proprio il modo in cui i giganti del web utilizzano gli algoritmi a minare, assieme alle possibilità di incontro con l’altro e il diverso, anche le potenzialità riflessive degli individui che, pure, la rete potrebbe invece contribuire a rafforzare (se, appunto, gli algoritmi agissero diversamente).

L’utilizzo dell’intelligenza artificiale, nella misura in cui determina la formazione di comunità estetiche per come sono state delineate finora, dà vita invece a un’epistemologia algoritmica che, pur partendo dalla svolta figurale ed estetizzante della postmodernità, ha in realtà molto a che fare con le grandi narrazioni epistemologiche e metafisiche del Moderno.

L’obiettivo di questa epistemologia algoritmica, che agisce selezionando le informazioni a cui abbiamo accesso sui social network,

è quello di ridurre la complessità del reale a schemi onnicomprensivi che soddisfino il desiderio di senso degli individui. Le comunità estetiche diventano quindi, attraverso la socializzazione e la selezione algoritmica delle notizie, comunità certe della propria meta-narrazione, in cui la personalizzazione si unisce al comunitarismo più esasperato e a una certezza “metafisica” e indiscutibile della propria verità tribale.

4. L'EPISTEMOLOGIA ALGORITMICA E IL RITORNO DEI GRAND RÉCITS

Come si è sostenuto finora, l'informazione, nel momento in cui arriva a coincidere col concetto di relazione umana, è un terreno privilegiato per applicare meccanismi di intelligenza artificiale a quelle stesse relazioni plasmandole in modo da creare comunità iper-affettive, tribali.

Sebbene tutto questo, come visto, alligni nella cultura estetizzante, figurale e iper-emotiva resa possibile dai mutamenti della società post-industriale e della cultura postmoderna, è anche vero però che l'intelligenza artificiale, promuovendo questo tipo di legami, determina un ritorno ad una metafisica moderna in cui le narrazioni delle tribù online sembrano assomigliare ai *grand récits*, le meta-narrazioni che proprio un certo pensiero postmoderno ha contribuito a decostruire (Lyotard, 1979). Nel tracciare questa analogia ci riferiamo al fatto che l'iper-emotività priva di mediazioni razionali delle tribù online risponde ad un'esigenza di riduzione della complessità del reale a schemi semplici, onnicomprensivi, che rispondano all'esigenza di senso e orientamento degli individui. Allo stesso modo, i grandi sistemi filosofici della modernità come marxismo e idealismo contenevano dei principi unitari (la Ragione, lo Spirito, le leggi materialiste) in base ai quali inglobare il senso della realtà, stavolta però su base esclusivamente razionale.

Ma come sottolineava Hannah Arendt (1951), l'ideologia va intesa etimologicamente come “logica dell'idea”, ovvero come un atteggiamento che comprime l'infinita varietà del reale in uno schema logico assoluto che soddisfi il desiderio di senso. Un atteggiamento, insomma, che proprio per la sua base ultra-razionale è impermeabile alla smentita fattuale. Forse per questo Karl Popper, a proposito della teoria marxista della storia e della psicoanalisi freudiana, ebbe a scrivere: «Avvertivo che queste teorie, pur atteggiandosi a scienze, erano di fatto più imparentate con i miti primitivi [...] Esse sembravano in grado di spiegare praticamente tutto ciò che accadeva nei campi in cui si riferivano. [...] Il mondo pullulava di verifiche della teoria. Qualunque cosa accadesse, le confermava sempre [...] e quanto agli increduli, si

trattava chiaramente di persone che non volevano vedere la verità manifesta [...] Un marxista non poteva aprire un giornale senza trovarvi in ogni pagina una testimonianza in grado di confermare la sua interpretazione della storia» (1972, 63-66).

L'epistemologia algoritmica, in questo senso, fa sì che, come nei sistemi filosofici elaborati a partire dai principi della scienza moderna (si pensi al metodo cartesiano), su una base del tutto deduttiva e solo falsamente empirica, ogni dato successivo confermi la tesi di partenza.

Si tratta esattamente di una delle dinamiche maggiormente studiate e analizzate per quel che riguarda il mondo delle comunità online, e che ha a che fare con quello che la sociologia dei media definisce "pregiudizio di conferma" e "effetto *backfire*".

Come ha ben evidenziato Walter Quattrococchi nel suo *Misinformation. Guida alla società dell'informazione e della credulità* attraverso studi condotti su numerose pagine e gruppi Facebook, comunità di utenti raggruppati e fortemente polarizzati su una specifica posizione tendono a prendere per vere notizie poco credibili, purché siano coerenti con la loro narrazione di riferimento (Quattrococchi, Vicini, 2016, 48-50). Si tratta di un processo che ha a che fare con il pregiudizio di conferma, ovvero un meccanismo che porta l'individuo a considerare vere solo informazioni che rientrano nel suo sistema di credenze.

Come se non bastasse, gli studi empirici sui gruppi Facebook dimostrano come le operazioni di *debunking* non solo siano infruttuose, ma tendano addirittura a rafforzare quello stesso sistema di credenze. Tanto più un utente convinto, ad esempio, della veridicità di una teoria complottista verrà esposto a informazioni che mostrano la fallacia della sua posizione, quanto più quello stesso utente rafforzerà la propria convinzione di partenza. Ciò potrà avvenire o ignorando, scartando le informazioni che contraddicono la sua teoria o prendendole come subdoli tentativi di occultare la verità. L'effetto backfire consiste quindi nel paradossale rafforzamento della polarizzazione di individui e gruppi esposti a smentite fattuali delle tesi a cui credono: invece di riconsiderare riflessivamente quelle tesi, questi individui e questi gruppi, guidati dall'iper-emotività tribale prodotta in loro dall'esposizione selettiva a determinati contenuti, rafforzeranno sempre di più la loro identità e il loro senso di appartenenza al gruppo di riferimento.

Per tutti questi motivi, lo stesso Quattrococchi, nel riferirsi al pensiero complottista che prolifera sul web, traccia un'analogia tra quest'ultimo e l'atteggiamento religioso di chi *vuole* credere a qualcosa per potersi orientare e inquadrare la realtà in uno schema unico di senso (Ivi, 104-108).

Ciò che vale per i gruppi complottisti può essere facilmente esteso a numerose altre narrazioni presenti in rete. Come è evidente, l'esposizione selettiva a contenuti e informazioni operata dagli algoritmi costituisce uno dei fattori più importanti per la formazione di queste comunità fortemente omogenee, polarizzate e alla ricerca di una narrazione unica, onnicomprensiva.

Comunità in cui, come detto, l'elemento estetico ed iper-emozionale si fonde con la pretesa ultra-razionale di voler rinchiudere la varietà e la complessità del reale in uno schema unico di senso. Comunità che accettano come tali solo i dati di realtà che confermano il loro sistema di credenze, e che scartano, ignorano o delegittimano tutti gli altri.

Le conseguenze socio-politiche di questi meccanismi possono essere nefaste. Oltre al pesante condizionamento che le fake news hanno avuto rispetto all'esito di alcune elezioni, come le presidenziali statunitensi del 2016, contenuti falsi che girano in rete ma che vengono presi come conferma delle proprie narrazioni di riferimento da parte degli utenti possono favorire addirittura violenze etniche e stragi di minoranze religiose.

È successo con la strage dei Rohingya in Myanmar: Facebook ha infatti ammesso di aver avuto un ruolo (definito per la verità "determinante" da un report delle Nazioni Unite) nel fomentare l'odio nei confronti della minoranza musulmana in quel paese. La scarsa alfabetizzazione digitale, in Myanmar, ha fatto sì che gli utenti fossero privi di strumenti per rispondere criticamente e riflessivamente alla proliferazione di post denigratori e incendiari contro i Rohingya, che Facebook ha reso presto virali (Osnos, 2018).

In un altro paese a scarsa alfabetizzazione digitale come le Filippine, invece, le dinamiche di polarizzazione indotte da Facebook e la proliferazione di fake news hanno favorito la legittimazione del proprio potere autocratico e repressivo da parte di Duterte, come messo in luce da un'inchiesta di BuzzFeed (Alba, 2018).

E ancora, un'inchiesta della BBC ha evidenziato come, in una nazione già fortemente "tribalizzata" e segnata da conflitti etnico-religiosi come la Nigeria, le risposte iper-emozionali degli utenti a contenuti diventati virali sul social (e spesso manipolati) abbiano contribuito a massacri di cristiani da parte di musulmani e viceversa (Adegoke, 2018).

L'ingegneria sociale degli algoritmi riesce a produrre questa epistemologia delle meta-narrazioni tribali proprio grazie al mutamento e alla progressiva socializzazione dei contenuti informativi, ormai del tutto coincidenti, su piattaforme come Facebook, con quelli amicali e

relazionali in senso ampio.

Il collegamento con i sistemi filosofici che si sono sviluppati a partire dalla scienza moderna appare in questo senso piuttosto chiaro. Già un filosofo come Thomas Hobbes, ad esempio, aveva tentato di costruire un sistema insieme politico e sociale in grado di matematizzare le relazioni umane per eliminare l'elemento del conflitto. L'idea del filosofo inglese era che, applicando modelli matematici e scientifici alle relazioni umane, fosse possibile ridurre l'eterogeneità (portatrice di conflitto) a omogeneità, attraverso un sistema artificiale che dall'epistemologia si traducesse alla politica e alle relazioni tra esseri umani. Come ha osservato correttamente Michele Mezza, "non è certo inedito né originale il concetto di una matematizzazione delle relazioni sociali. Sia Thomas Hobbes che Blaise Pascal, nel XVII secolo, affermavano che 'pensare non è altro che calcolare'. Forse è la migliore definizione dell'efficacia degli algoritmi che ordinano oggi la nostra esistenza. La vita è numero" (Mezza, 2018, 26).

L'algoritmo riproduce questa logica di matematizzazione ed eliminazione del conflitto all'interno del singolo gruppo-tribù, agendo sui meccanismi di interazione ed esposizione selettiva, e inasprendo allo stesso tempo il conflitto tra gruppi-tribù rivali, incapaci di confrontarsi all'interno di una sfera pubblica riflessiva come quella teorizzata da Habermas, in cui le opinioni si formano sulla base dell'esposizione non filtrata a tesi contrastanti e dell'analisi riflessiva dei dati di realtà.

L'ingegnerizzazione delle relazioni sociali prodotta dagli algoritmi degli Over The Top tiene quindi insieme l'emoività più esasperata e la metafisica ultra-razionale della filosofia moderna, il tutto a discapito di un'etica della differenza che permetta di approcciarsi criticamente e riflessivamente a contenuti non in linea con il proprio sistema di credenze. Ciò avviene, come detto, agendo principalmente sulle fonti di informazione e i contenuti che queste diffondono: filtrandoli, esponendoli agli utenti sulla base di un sistema di omofilia, nonché socializzandoli e rendendoli quanto più possibile frivoli.

Come ha osservato Cass Sunstein, quando l'informazione era ancora veicolata dagli intermediari tradizionali, come giornali e televisioni, le possibilità di imbattersi in contenuti inattesi, non programmati, magari in contrasto con le nostre convinzioni, era assai più frequente. La personalizzazione operata dagli algoritmi, in particolare quelli di Facebook e Google, fa sì invece che chi ha simpatie progressiste vedrà comparire davanti a sé articoli di giornali progressisti, post di amici che condividono la sua visione del mondo progressista. Tutto ciò avviene, per giunta, nella modalità ludica ed estetizzante già presa in esame. Ci si

informa sui social network, ma si naviga sui social network tendenzialmente per svagarsi (Sunstein, 2017, 13).

La conclusione del sillogismo è che anche l'informazione diventa uno svago, e che l'utente pretende averla a disposizione esattamente in questa modalità. L'estetica tribale dell'esposizione selettiva si salda così con la banalizzazione ludica dei contenuti, il tutto a scapito tanto della riflessività quanto della capacità di avere a che fare con la diversità, con contenuti inattesi, non programmati.

5. COME CREARE LEGAMI FORTI E COMUNITÀ INFORMATE ATTRAVERSO GLI ALGORITMI: SOLUZIONI ALTERNATIVE

Nel suo libro sulle bolle dei filtri, Eli Pariser indica una possibile direzione per superare l'epistemologia algoritmica dominante. Lo studioso americano parla della necessità di andare nella direzione di un'epistemologia falsificazionista simile a quella propugnata da Karl Popper. Cosa significa questo in concreto per quanto riguarda gli algoritmi? Per spiegarlo, Pariser fa l'esempio di Netflix: se sono un accanito consumatore di commedie romantiche, la piattaforma di contenuti streaming mi proporrà, per mezzo del suo algoritmo, altre commedie romantiche, intrappolandomi in un'identità fissa, stabilita una volta per tutte e non suscettibile di cambiamenti.

Diversamente, se l'algoritmo di Netflix iniziasse a propormi sì commedie romantiche, ma anche film come *Blade Runner*, resterebbe aperta la possibilità di falsificare l'idea che la mia identità sia quella di una persona a cui piacciono solo le commedie romantiche. L'identità potrebbe diventare così maggiormente sfaccettata, e per fare ciò basterebbe l'inserimento di alcuni elementi selezionati casualmente e non riconducibili alle mie ricerche precedenti all'interno del pacchetto che mi viene proposto dall'algoritmo.

Questa epistemologia algoritmica di stampo falsificazionista metterebbe in questione quella attualmente predominante e che abbiamo associato ai *grand récits* della modernità filosofica e scientifica, in cui tutto ciò che mi viene proposto in quanto utente ha soltanto la funzione di confermare il mio sistema di credenze e la mia visione del mondo, in cui il dato di realtà non può mai avere la funzione di confutare o anche solo mettere in questione la mia narrativa di riferimento. Come applicare nel concreto questo modello agli algoritmi e al modo in cui vengono utilizzati dai media?

Tra le principali proposte sin qui avanzate, va certamente segnalato il lavoro del Centre for Civic Media del Massachusetts Institute of

Technology (MIT) di Boston. Ethan Zuckerman, direttore del centro, ha spiegato il senso del lavoro del suo team di ricerca nel libro *Rewire. Cosmopoliti digitali nell'era della globalità* (2014).

Partendo da una critica dell'epistemologia algoritmica degli Over The Top, il Centre for Civic Media del MIT ha innanzitutto sperimentato, nel corso degli anni, soluzioni tecnologiche alternative in grado di modificare le modalità attraverso cui si naviga in rete. Tra i vari progetti, si può citare la creazione di estensioni di browser che orientino la navigazione dell'utente verso contenuti eterogenei rispetto a quelli che visita abitualmente. L'algoritmo alla base di questa estensione del browser funziona esattamente al contrario rispetto al meccanismo che regola gli algoritmi di Google e Facebook.

Non si tratta dell'unico tentativo che va in questa direzione. Esistono infatti servizi di navigazione come StumbleUpon e motori di ricerca come DuckDuckGo che agiscono sulla base dello stesso principio, favorendo cioè la navigazione *random* tra pagine web e prevedendo l'assenza di filtri personalizzati come impostazione di default (Veltri, Di Caterino, 2017).

L'idea alla base di tutti questi tentativi è che sia possibile, attraverso una diversa configurazione degli algoritmi, sottrarre agli Over The Top il "monopolio della relazione". Fatto salvo il legame tra innovazione tecnologica e configurazione delle relazioni umane, queste ultime possono essere sottratte a una logica tribale ed essere indirizzate, tramite l'intelligenza artificiale, verso una concezione diversa dell'identità, più aperta alla contaminazione, alla relativizzazione del proprio punto di vista e maggiormente pronta ad approcciarsi riflessivamente a contenuti nuovi, in grado di mettere in questione la propria narrativa di riferimento. Un'identità, quindi, da intendere come qualcosa in continua formazione, ridefinizione, non un nucleo chiuso e definito una volta per tutte.

A tutto questo è riferito il termine "serendipità", utilizzato da Zuckerman per definire questa nuova cultura dei media digitali e del legame tra intelligenza artificiale e relazioni umane. Il concetto di "serendipità" rimanda ad una predisposizione nei confronti delle esperienze non programmate, un atteggiamento aperto all'ibridazione, ad incontri non voluti, non cercati, ma che contribuiscono ad arricchire l'esperienza che abbiamo del mondo. Tutto si gioca, insomma, sul tentativo di trovare l'architettura della rete maggiormente in grado di favorire la serendipità online.

Al momento, il problema principale ha a che fare con il sostanziale oligopolio della rete gestito da pochi grandi *players*. Ciò significa che,

per modificare il modo in cui la gente si informa e in cui vengono pianificati gli algoritmi, molti ritengono che l'unica possibilità sia spingere per una maggiore presa di consapevolezza di giganti come Facebook e Google rispetto al loro ruolo di principali aziende editoriali del globo.

Mark Zuckerberg, ad esempio, dopo aver rifiutato per molto tempo una qualificazione di Facebook che andasse al di là di quella di social network, negli ultimi anni ha avviato diversi progetti che riconoscono una funzione propriamente giornalistica e informativa alla sua piattaforma. Cass Sunstein, proprio in virtù di questo, ritiene che per superare l'epistemologia algoritmica vigente sia necessario che governi, istituzioni e società civile aumentino il loro livello di pressione sui giganti del web, in modo tale che, ad esempio, sul Newsfeed di Facebook possa presto comparire un "serendipity button" che, esponendoci a del materiale informativo inatteso, renda il social più simile allo spazio aperto di una città che a quello chiuso di una tribù virtuale (Sunstein, 2017, 232).

A nostro parere, però, ciò che queste analisi non rilevano a sufficienza è che è impossibile prescindere dal carattere estetico che la svolta postmoderna ha impresso tanto alle relazioni sociali quanto all'esperienza della rete e, con l'ingresso sulla scena dei social network, al legame sempre più stretto tra le prime e la seconda. Anche nel momento in cui si ragiona sulle notizie, sui modi con cui le persone si informano, è necessario comprendere come la serendipità e l'apertura riflessiva a contenuti e incontri inattesi non possano che avvenire attraverso una modalità ludica e leggera di fruizione dei contenuti informativi.

La differenza con il sistema attuale è che questo spazio estetico della rete deve diventare più simile a quello teorizzato da Lash e Urry, e che è stato richiamato nel primo capitolo, ovvero uno spazio virtuale in cui la circolazione di segni, immagini e artefatti culturali determini una presa di contatto con la diversità che tenga insieme la componente emotiva e quella razionale, l'interesse estetico e quello etico-morale.

Lo spazio estetico, insomma, pur essendo ineludibile e ormai coesenziale al mondo digitale, non deve necessariamente produrre le comunità tribali ed iper-affettive di cui parla Maffesoli, né è destinato a promuovere i legami deboli delle comunità estetiche (o comunità-gruccia) a cui si riferiva Bauman. La svolta ludica e social impone però di muovere comunque da quella che Monica Sassatelli ha definito la "sfera pubblica estetica", ovvero uno spazio pubblico che tenga conto dell'elemento emozionale, affettivo, e che valorizzi la cultura popolare

come mezzo per accedere alla conoscenza dell'altro, del diverso, senza ridurla a emblema della mercificazione della cultura stessa (Sassatelli, 2012, 233-244).

Come ha scritto Paul Jones (2007, 88), la sfera pubblica estetica include tutte le forme di produzione estetico-culturale e la loro discussione critica, oltre a considerare integrati l'elemento emozionale e quello riflessivo. Se si assume quindi la sfera pubblica estetica come campo di analisi, è possibile pensare ad una riconfigurazione del discorso pubblico (per mezzo di una riconfigurazione degli algoritmi) che muova da modalità ludiche, affettive e "social" di accesso ai contenuti informativi.

Un primo modo di sottrarre agli Over The Top quello che abbiamo definito il "monopolio della relazione" consiste nella progressiva trasformazione della singola testata giornalistica in un social network, in modo tale però da promuovere legami più forti e riflessivi rispetto a quelli deboli, episodici e puramente emotivi tipici di Facebook. Se su Facebook la sfera pubblica viene fagocitata da quella privata, con la trasformazione social della testata la commistione tra pubblico e privato si mantiene, ma può andare a maggiore vantaggio della partecipazione del singolo alla sfera pubblica.

Gli esempi principali in questo senso sono quelli delle grandi testate del mondo anglosassone, che si stanno trasformando da *newspaper* a *brandpaper*. In tutti questi casi, la tecnologia viene usata per trasformare il rapporto col lettore senza cedere ai legami deboli e alla logica plebiscitaria del "mi piace". Testate come il *The Guardian*, il *New York Times* e il *Washington Post*, organizzano decine di eventi, forum, dibattiti, attraverso cui i lettori interagiscono con la redazione e collaborano proponendo loro stessi dei contenuti. Ciò avviene anche con alcune formule *premium*: il lettore paga un piccolo abbonamento ma ha la possibilità di proporre inchieste alla redazione, interagire direttamente con i giornalisti, avere biglietti scontati per eventi organizzati dal giornale, newsletter dedicate, e così via.

Il *New York Times* sta puntando anche su forme di e-learning: "classi virtuali che si formano attorno a notizie o reportage, dove lettori e giornalisti si mischiano in uno scambio reciproco di informazioni e pareri" (Mezza, 2015, 112). Il *Guardian* ha lanciato addirittura una propria app di incontri per single, accessibile dalla home page del sito, che favorisce la conoscenza tra persone che, in quanto lettrici del giornale, si presuppone abbiano interessi in comune e siano più facilmente compatibili tra loro rispetto a quelle che si possono incontrare su altre piattaforme.

Il questo caso, la grande testata agisce esattamente come un social network, crea legami tra le persone ponendo il rapporto con la testata stessa come centro di gravità di quei legami, e utilizza dei sistemi di intelligenza artificiale (basati sull'individuazione di affinità tra gli utenti) come mezzo per stringerli e rafforzarli. Come afferma Michele Mezza, la testata in questo modo diventa un centro-servizi: usa le informazioni come pretesto per stringere legami di affidabilità con gli utenti, per vendere nuovi servizi, e diventa essa stessa un service provider o un social network come Google, Facebook o Amazon.

Ciò passa anche da un'autonomia dal punto di vista tecnologico: il *Washington Post* ha inserito nella propria redazione numerosi ingegneri e news designers per studiare soluzioni tecnologiche in grado di rendere la testata sempre più indipendente da Facebook.

Sono tentativi attraverso cui queste grandi testate cercano appunto di acquisire un'autonomia sia culturale sia tecnologica che gli permetta di affrancarsi dalla dittatura di Facebook e Google e allo stesso tempo, partendo dalla modalità social, interattiva e ludica, di formare attorno a loro comunità realmente informate.

Ciò passa, infine, anche dalla profilazione degli utenti: il *Guardian*, ad esempio, chiede ai lettori di poter accedere ai loro dati per poter offrire servizi personalizzati. Anche in questo caso, si fa in modo che la profilazione non avvenga tramite Facebook, e che sia finalizzata a una logica non solo consumistica e commerciale, ma anche legata alla creazione di un utente informato. L'algoritmo del *Guardian* personalizza, come quello di Facebook, ma con l'obiettivo di creare legami più forti, riflessivi, stabili, all'interno di una comunità più piccola.

Poiché però, come detto, ciò che deve restare centrale è la dimensione ludico-affettivo-estetica dell'interazione tra gli utenti, intesa come innesco di relazioni più stabili, forti e riflessive, vorremmo proporre alcune ulteriori soluzioni che vanno nella direzione indicata lungo il corso di questo lavoro. Da un punto di vista tecnologico, esistono alcuni software, finora poco utilizzati dai media mainstream e confinati ai media civici, che favoriscono un'interazione più mediata e consapevole tra gli utenti. Tra questi, ad esempio, c'è *Deebase*, un software messo a punto dalla piattaforma online *Partecipedia*, che permette di organizzare delle discussioni online su specifici temi attraverso un sistema di confronto tra le opinioni, con punteggi e votazioni finali. Una democrazia deliberativa online che introduce dei meccanismi tipici del *gaming* (voti, punteggi) per promuovere però una discussione informata.

Lo stesso tipo di meccanismo potrebbe essere esteso dalle testate,

indipendentemente dall'utilizzo di questo software, per i commenti dei lettori (all'interno della testata stessa o su Facebook). Di recente un giornale scandinavo ha studiato un meccanismo per cui prima di commentare un articolo, il lettore deve rispondere ad alcune domande sul suo contenuto, in modo da dimostrare di averlo davvero letto. Solo chi passa il test ha il diritto di commentare.

Se questa soluzione ha un sapore di censura e rischierebbe di far perdere lettori al giornale, perché non pensare però ad un meccanismo per cui chi risponde correttamente alle domande sul contenuto dell'articolo guadagna un punteggio che gli dà maggiore credibilità nei commenti? La competenza verrebbe così premiata e sarebbe immediatamente riconoscibile dagli altri utenti, dando vita ad una competizione al rialzo.

Proprio l'utilizzo di software come *Deebase* potrebbe rendere più facile adottare queste soluzioni da un punto di vista tecnologico. Questo meccanismo potrebbe valere per il contenuto di un articolo così come per quello di una specifica legge. Se si parla del Jobs act, perché non corredare un articolo online con un piccolo questionario che testi il livello di conoscenza della legge, così da dare punti e credibilità ai lettori che rispondono correttamente?

Questo tipo di meccanismo, allo stesso tempo ludico e indirizzato a promuovere la conoscenza di specifici argomenti, esiste già in quelli che vengono definiti dalla letteratura scientifica "media civici" (De Biase, 2013), ovvero piattaforme concepite per favorire, attraverso lo scambio di informazioni, l'impegno civico degli utenti.

Un esempio è costituito da *Uribu*, una piattaforma il cui scopo è quello di denunciare disservizi e soprusi nelle amministrazioni pubbliche. Il meccanismo della piattaforma è il seguente: più si denuncia, più si guadagnano punti e credibilità in una sorta di classifica apposita. Lo stesso principio è alla base della app *Decoro urbano*, in cui gli utenti segnalano tutto ciò che ha a che fare con il degrado di una città, dai rifiuti alle affissioni abusive e così via. Anche qui, più si segnala più si guadagnano punti, in modo tale che la partecipazione sia sempre più incentivata attraverso dei meccanismi di gratificazione e riconoscimento della competenza (Faccioli, 2013, 193-195).

Si tratterebbe, quindi, di utilizzare meccanismi già esistenti nei media civici importandoli nei media mainstream, proprio perché tali meccanismi tengono insieme la dimensione civica e quella estetico-ludica, utilizzando quest'ultima come modalità di accesso all'impegno nella sfera pubblica, secondo le modalità della "sfera pubblica estetica" precedentemente analizzate. Il meccanismo sarebbe, ancora una volta,

quello del *gaming*: un'interazione basata su modalità ludiche, imprescindibile in un'epoca in cui l'informazione si fa sui social, ma che allo stesso tempo favorisce la competenza su alcuni argomenti.

E ancora: così come le grandi testate si stanno trasformando in service provider, potrebbero diventare esse stesse dei forum, organizzando sezioni di discussione tra i lettori su specifici argomenti. Ciò potrebbe avvenire sia sulla home page del sito sia su Facebook, organizzando dei gruppi specifici differenti dalla pagina Facebook principale della testata (come alcuni già stanno facendo).

Se il giornale diventa esso stesso un forum, ma con regole più definite rispetto alla bacheca di Facebook e attraverso meccanismi che promuovano una discussione informata tra i lettori, è possibile salvare allo stesso tempo la dimensione "social" (ivi compresi gli accessi al sito e gli introiti pubblicitari) e quella "sociale", intesa come responsabilità di dar vita a comunità informate, non riducendo la propria funzione esclusivamente a scopi di natura commerciale.

Infine, esistono strumenti già utilizzati da Facebook e che tengono insieme l'impegno civico con forme di auto-promozione e *self-branding* tipiche dei social network. Un esempio è l'inserimento da parte di Facebook del pulsante "ho votato", con rende noto agli amici il fatto di essersi recati alle urne durante le elezioni. Una ricerca dell'Università della California, nel 2010, ha stimato che la sola esistenza di quel bottone aveva spinto 340.000 persone in più ad andare a votare (Nicodemo, 2017, 130). Le testate giornalistiche, che ormai producono gran parte degli accessi e degli introiti tramite i social network, potrebbero quindi utilizzare questo e altri meccanismi tipicamente "social" per promuovere tra i lettori forme di impegno civico, unendo l'elemento affettivo, emozionale e persino narcisistico con l'accesso alla sfera pubblica.

In conclusione, è necessario muovere dal carattere imprescindibile della socializzazione del contenuto giornalistico. Se però le testate riescono a proporre modelli differenti da quelli attualmente predominanti su Facebook, anche attraverso un differente utilizzo degli algoritmi e dell'intelligenza artificiale, è possibile rendere la socializzazione della notizia funzionale alla creazione di comunità informate.

Ciò è necessario per creare un modello culturale che vada oltre quello dei "media populistici" (Dal Lago, 2017) e che favorisca, anche online, mediazione, dibattito e riflessione. Infine, come argomentato, ciò passa per un'autonomia non solo culturale, ma anche tecnologica delle testate. La tecnologia dei media civici può essere trasferita sui media mainstream, in modo tale che il rapporto stesso tra democrazia e

tecnologia non si riduca alla forma plebiscitaria, immediata e iperemotiva rappresentata dal modo in cui Facebook attualmente funziona.

Sebbene si tratti di processi estremamente complessi, vi sono già esempi virtuosi, alcuni dei quali sono stati presi in esame in questo lavoro, che mostrano come la sfera pubblica digitale, pur rimanendo necessariamente estetica, non debba necessariamente sottostare al modello tribale ed iper-affettivo promosso dagli algoritmi degli Over The Top, e che conduce al fenomeno delle camere dell'eco e dell'incomunicabilità tra comunità chiuse nella certezza assoluta ed indiscutibile della propria narrativa. La sfera pubblica digitale può invece essere resa estetica nell'accezione che rimanda a modalità relazionali ludiche ed affettive ma che, allo stesso tempo, promuovono legami più strutturati e riflessivi, nonché l'apertura ad esperienze non programmate e alla serendipità online.

Le varie soluzioni qui analizzate, da quelle esclusivamente tecnologiche (le estensioni di browser e i motori di ricerca che utilizzano algoritmi alternativi) a quelle utilizzate dalle grandi testate anglosassoni per creare comunità di persone affini ma allo stesso tempo informate e tenute insieme da legami forti, fino all'importazione di tecnologie e meccanismi già in uso nei media civici che tengono insieme l'aspetto ludico-emotivo e quello informativo-impegnato, tutto questo insomma configura una possibile epistemologia algoritmica non schiacciata sulle modalità iper-estetiche e tribali attualmente dominanti.

RIFERIMENTI BIBLIOGRAFICI

- ADEGOKE, Y. (2018). Like, Share, Kill. Nigerian police say false information on Facebook is killing people. *BBC World*, November 13.
- ALBA, D. (2018). How Duterte Used Facebook To Fuel the Philippine Drug War. *BuzzFeed*, September 4.
- ARENDT, H. (1951). *Le origini del totalitarismo*. Milano: Edizioni di Comunità, 1967.
- ARIELLI, E., BOTTAZZINI, P. (2018). *Idee virali*. Bologna: Il Mulino.
- BAUMAN, Z. (1989). *Modernità e Olocausto*. Bologna: Il Mulino 1992.
- (1993). *Le sfide dell'etica*. Milano: Feltrinelli 2010.
- (2000). *Il disagio della postmodernità*. Milano: Mondadori, 2002.
- (2001). *Voglia di comunità*. Roma-Bari: Laterza, 2011.
- BECK, U., BECK-GERNSHEIM, E. (2002). *Individualization: Institutionalized Individualism and its Social and Political Consequences*.
-

- London: Sage.
- CAMPBELL, C. (1987). *L'etica romantica e lo spirito del consumismo moderno*. Roma: Edizioni Lavoro 1992.
- DAL LAGO, A. (2017). *Populismo digitale*. Milano: Raffaello Cortina.
- DE BIASE, L. (2013). *I media civici*. Milano: Feltrinelli.
- EUGENI, R. (2015). *La condizione postmediale*. Brescia: La Scuola.
- FACCIOLI, F. (2013). *Comunicazione, sfere pubbliche e processi di civic engagement*. In: Bartoletti, F. Faccioli (a cura di), *Comunicazione e civic engagement* (pp. 174-200). Milano: FrancoAngeli.
- FEATHERSTONE, M. (1992). *Postmodernism and the aesthetization of everyday life*, in *Modernity and Identity*. Oxford-Cambridge: Blackwell.
- GIDDENS, A. (1990). *Le conseguenze della modernità. Fiducia e rischio, sicurezza e pericolo*. Bologna: Il Mulino 1994.
- (1991). *Identità e società moderna*. Napoli: Ipermedium Libri, 1999.
- JENKINS, H. (2006). *Cultura convergente*. Milano: Apogeo 2007.
- JONES, P. (2007). Beyond the Semantic 'Big Bang'. *Cultural Sociology* and an Aesthetic Public Sphere. *Cultural Sociology*, 1(1), 73-95.
- LASH, S. (1988). Discourse or figure? Postmodernism as a regime of signification. *Theory, Culture and Society*, 5(2-3), 311-336.
- (1990). *Modernismo e postmodernismo. I mutamenti culturali delle società complesse*. Roma: Armando, 2000.
- (1999). *Another Modernity. A Different Rationality*. Oxford: Blackwell.
- , URRY J. (1994). *Economies of signs and space*. London: Sage.
- LIPOVETSKY, G. (1987). *L'impero dell'effimero. La moda nelle società moderne*. Milano: Garzanti 1989.
- (1989). *L'era del vuoto. Saggi sull'individualismo contemporaneo*. Milano: Luni Editrice, 1995.
- LIVOLSI, M. (2013). *La partecipazione solitaria*. In R. Bartoletti, F. Faccioli (a cura di), *Comunicazione e civic engagement* (pp. 38-57). Milano: FrancoAngeli.
- LYOTARD, J. F. (1971). *Discorso, figura*. Milano: Mimesis 2008.
- (1979). *La condizione postmoderna. Rapporto sul sapere*, Milano: Feltrinelli 1981.
- MAFFESOLI, M. (1988). *Il tempo delle tribù. Il declino dell'individualismo nelle società contemporanee*. Milano: Guerini, 2004.
- (1990). *Nel vuoto delle apparenze. Per un'etica dell'estetica*. Milano: Garzanti, 1993.
- (2008). *Icone d'oggi. Le nostre idol@trie postmoderne*. Palermo:
-

- Sellerio 2009.
- MCLUHAN, M. (1964). *Gli strumenti del comunicare*. Milano: Il Saggiatore 1967.
- MELONI, V. (2017), *Il crepuscolo dei media*. Roma-Bari: Laterza.
- MEZZA, M. (2015), *Giornalismi nella rete*. Roma: Donzelli.
- (2018). *Algoritmi di libertà*. Roma: Donzelli.
- NICODEMO, F. (2017), *Disinformatia*. Venezia: Marsilio.
- ORTOLEVA, P. (2009) *Il secolo dei media*. Milano: Il Saggiatore.
- OSNOS, D. (2018). Can Mark Zuckerberg Fix Facebook Before It Breaks Democracy?. *New Yorker*, September 17.
- PARISER, E. (2011). *Il filtro*. Milano: Il Saggiatore, 2012.
- POPPER, K. R. (1972). *Congetture e confutazioni*. Bologna: Il Mulino.
- QUATTROCIOCHI, W., VICINI, A. (2016). *Misinformation. Guida alla società dell'informazione e della credulità*. Milano: FrancoAngeli.
- ROSE, M. (1992). *The post-modern and the post-industrial*. Cambridge: Cambridge University Press.
- SASSATELLI, M. (2012). *Festivals, museums, exhibitions. Aesthetic cosmopolitanism in the cultural public sphere*. In G. Delanty (ed.), *Routledge Handbook of Cosmopolitanism Studies* (pp. 233-244). London-New York: Routledge.
- SUNSTEIN, C. (2017). *#Republic. Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press.
- TURKLE, S. (1995). *La vita sullo schermo*. Milano, Apogeo 2005.
- URRY, J. (1995). *Consuming Places*. London: Routledge.
- , SZERSZYNSKI, B. (2006). Visuality, mobility and the cosmopolitan: inhabiting the world from afar. *The British Journal of Sociology*, 57(1), 113-131.
- VELTRI, G., DI CATERINO, G. (2017). *Fuori dalla bolla*. Milano-Udine: Mimesis.
- ZUCKERMAN, E. (2014). *Rewire. Cosmopoliti digitali nell'era della globalità*. Milano: Egea, 2014.
-

ECHO CHAMBERS E POLARIZZAZIONE

Uno sguardo critico sulla diffusione dell'informazione nei social network

di *Costantino Carugno, Tommaso Radicioni**

Abstract

Understanding the algorithms that contribute to the formation of our daily reality requires an in-depth look at how information is disseminated in online social networks (OSN). In this article, we will observe how news propagation is restricted by the presence of virtual borders that limit the interaction between users. This phenomenon, known as "echo chamber" formation, has the effect of polarizing the public debate on conflicting positions. Inside an echo chamber, information is not conveyed through a horizontal exchange between users, but due to the presence of like or follower aggregators, called hubs. This analysis will be carried out considering a case study in two of the main OSNs: Facebook and Twitter. From the study of user interaction networks we will observe how the algorithmic choices made are crucial to the polarization of the debate around a topic of discussion.

Keywords

Social network, opinion dynamics, fake news, echo chamber, polarization

* COSTANTINO CARUGNO lavora all'Università degli Studi di Torino. Email: costa@caos.space

TOMMASO RADICIONI lavora alla Scuola Normale Superiore di Pisa. Email: tommaso.radicioni@sns.it

1. INTRODUCTION

Al giorno d'oggi, l'uso pervasivo degli online social network (OSN) è documentato da numerosi studi che si sono occupati del loro utilizzo, ad esempio nella promozione di campagne sociali (Barisione *et al.*, 2017), nel sostegno alla partecipazione politica attiva (Mercea *et al.*, 2016) e ai movimenti sociali (Bastos, Funk, 2015). Un tema di enorme importanza nell'era della cosiddetta "post-verità" è quello dello studio dell'ecosistema mediatico e della dieta informativa degli utenti all'interno degli OSN. Pur consapevoli dell'impossibilità di circoscrivere l'intero processo di formazione di un'opinione nell'ambito di uno o più OSN, esistono almeno due motivi per limitarci allo studio unicamente in quest'ambito.

Innanzitutto, numerose ricerche sottolineano come un numero sempre maggiore di individui si informi in larga parte sugli OSN. Ad esempio, uno studio basato su un sondaggio di 4971 adulti statunitensi condotto nel 2017 dal Pew Research Center's (Shearer, Gottfried, 2017) mostra come il 67% degli intervistati dichiara di ottenere almeno parte della loro dieta informativa dagli OSN, tendenza in aumento rispetto ad uno studio simile condotto nel 2016. Inoltre, la disponibilità di accesso all'estrazione di dati dagli OSN tramite applicazioni fornite dagli stessi, chiamate in gergo tecnico API (Application Programmable Interface), e l'abbondanza di strumenti di analisi di tali dati permettono di poter realizzare studi su enormi quantità di dati direttamente estratti da uno specifico OSN.

All'interno degli OSN, la dieta informativa degli utenti è definita non solamente a partire dalla rete dei contatti e dai gruppi che l'utente decide di seguire ma da un complesso meccanismo che definisce quale contenuto comparirà nella bacheca di un utente nel mezzo di un flusso incessante di informazioni pubblicate giornalmente. Questo meccanismo è l'algoritmo che determina attraverso un processo decisionale personalizzato, tramite tecniche di "machine learning", quali contenuti avranno più visibilità rispetto ad altri. Comprendere e studiare gli algoritmi che contribuiscono alla formazione della nostra realtà quotidiana necessita di uno sguardo approfondito sulla diffusione di informazioni all'interno della rete di utenti presenti negli OSN. Poiché il funzionamento dell'algoritmo è sconosciuto e secretato dalle aziende che traggono profitto da esso, non vi sono altri modi di indagare come l'algoritmo funziona se non attraverso un'analisi dei suoi effetti.

Lo scopo di questo articolo è dunque quello di indagare quali siano gli effetti di un algoritmo complesso come quello utilizzato dagli OSN tramite l'esplorazione delle interazioni attorno ad uno specifico tema in due degli OSN più diffusi, ossia Facebook e Twitter.

2. IL DOMINIO DELL'ALGORITMO: ESPLORARE GLI OSN E I SUOI MECCANISMI

La presenza sempre più massiccia degli OSN nella dieta informativa di un numero crescente di individui ci interroga profondamente su come l'opinione pubblica, espressione di convinzioni, pregiudizi ed inclinazioni di uno specifico pubblico, si struttura attorno a tematiche ed argomenti di discussione. Nell'ultimo secolo, si è assistito ad una radicale trasformazione della "sfera pubblica" (Habermas, 2005), intesa come insieme dei processi comunicativi formali ed informali, da circuiti culturali più ristretti ai mezzi d'informazione di massa, come la radio o la televisione. Un cambio di paradigma che ha avuto la capacità di raggiungere un numero enorme di nuovi soggetti, ma che si è anche riconfigurato come un modello "top-down" in cui un ristretto numero di canali informativi sono capaci di essere altamente influenti e, dunque, a rappresentare istituzioni capaci di dettare l'agenda politica ed essere determinanti per la formazione di un'opinione pubblica. Al giorno d'oggi, si assiste invece ad una lenta disaffezione verso i media tradizionali: da una parte, il numero sempre maggiore di canali televisivi on-demand e digitali hanno frammentato il pubblico televisivo; dall'altra, un mercato dei quotidiani in declino mostra come la circolazione delle notizie cartacee sia in diminuzione (Bruns, Highfield, 2017b). L'avvento del Web 2.0 e dei social media ha infine creato un canale di diffusione d'informazioni e notizie molto più vario e complesso dei mezzi di comunicazione che erano egemoni fino a qualche decennio fa.

Uno studio approfondito delle dinamiche interne agli OSN è fondamentale per esplorare una delle più grandi rivoluzioni comunicative della nostra epoca. Non ci soffermeremo sul descrivere come i media tradizionali si sono plasmati per aggiornarsi alle nuove tecnologie digitali ma piuttosto sul funzionamento degli OSN che, come già sottolineato, si stanno affermando come uno dei principali media. Come sottolineato da Castells, la peculiarità della comunicazione nell'era del World Wide Web è la nascita di un nuovo modello definito come "auto-comunicazione di massa", ossia una forma di comunicazione che raggiunge un pubblico potenzialmente globale tramite la connessione Internet e che è «auto-generata nel contenuto, auto-diretta nell'emissione, e auto-selezionata nella ricezione di molti che comunicano con molti» (2009, 70).

Con l'avvento degli OSN, si assiste però ad una piccola rivoluzione interna al mondo del Web 2.0. Difatti, al loro interno il flusso di informazioni e notizie non è più auto-diretto ed auto-selezionato, a differenza di blogs, siti web e mailing lists, ma selezionato e

determinato a priori da un processo automatizzato di selezione dei contenuti. Nella sua accezione classica, il termine algoritmo indica qualsiasi procedimento “effettivo” di valutazione o decisione, eseguito secondo un insieme di regole esplicite. Mentre nell’algoritmica classica la macchina ha il compito di eseguire determinati calcoli in un prefissato ordine, all’interno degli OSN l’algoritmo si nutre dei dati prodotti dagli utenti e dunque questi ne costituiscono la materia prima. Da un approccio logico-funzionale consistente nella codificazione formale delle regole per il suo funzionamento, si è passati ad un approccio “data-driven”, ossia in cui l’algoritmo si plasma attorno ai dati che vengono letti, analizzati e sfruttati per rimodellare il suo funzionamento, in un continuo processo di apprendimento e trasformazione. Se la grande rivoluzione degli OSN si traduce dunque in un “capitalismo delle piattaforme” (Vecchi, 2017), nel quale l’estrazione di valore da parte degli OSN avviene attraverso i dati prodotti dagli utenti, il più grande effetto che esso ha sulla società è il principio stesso di funzionamento degli OSN che si alimenta dei dati prodotti dai suoi stessi utenti. Purtroppo, le modalità con le quali l’algoritmo lavora e seleziona i contenuti, attraverso i dati che lo alimentano, non è un’informazione di pubblico dominio ma può essere solo osservato ed analizzato a posteriori. La mancanza di trasparenza nei processi che governano gli OSN genera enormi quesiti: da una parte, l’impossibilità di accesso agli algoritmi degli OSN fa sì che nascano accuse di censura e di copertura mediatica, come è successo nel caso dell’hashtag #occupywallstreet che non riuscì mai ad entrare nei “Trending Topics” di Twitter (Gillespie, 2012) nemmeno durante il picco della mobilitazione. D’altra parte, l’indagine del meccanismo di diffusione dei contenuti all’interno degli OSN, ossia la modellizzazione della “black-box” algoritmica, diventa un’operazione complessa, non essendo studiabili e conoscibili i suoi principi. Ciò ha chiaramente implicazioni profonde sulla formazione di una “sfera pubblica digitale”, ossia uno spazio virtuale dominato dall’algoritmo in cui, più che altrove, si maturano le opinioni su questioni di interesse generale.

Sulle modalità di diffusione delle informazioni e di creazione di spazi di discussione all’interno della sfera pubblica digitale, i ricercatori hanno opinioni molto differenti tra loro (Vaccari, 2018). In particolar modo, il dibattito si concentra sul seguente quesito: il flusso di informazioni si diffonde in maniera fluida e incontrollata nell’oceano di utenti, pagine e gruppi presenti all’interno degli OSN o esistono confini virtuali che ne limitano la diffusione? Quest’ultimo fenomeno, meglio noto come formazione di “echo chamber”, ha come effetto secondario la

polarizzazione del dibattito pubblico su posizioni contrastanti ed incompatibili. La natura e l'effettiva influenza sulla sfera pubblica digitale di tali meccanismi è oggetto di studio e di discussione. Nel seguito, si cercherà di riportare alcuni argomenti per descrivere la complessità della questione per poi passare all'analisi di un "case study" utile a definire nello specifico un esempio concreto di come si configura lo spazio della sfera pubblica digitale.

3. ECHO CHAMBER ED OSN: L'ALGORITMO IN FUNZIONE

L'enorme flusso di informazioni che ogni giorno viene prodotto all'interno degli OSN necessita di metodi di selezione e gestione di una tale mole di dati. Negli ultimi decenni, le maggiori piattaforme virtuali, dai motori di ricerca agli OSN, hanno sviluppato sistemi che valutano l'importanza dei contenuti basandosi su svariati fattori come, ad esempio, l'interesse dell'utente verso tali contenuti calcolato a partire dalle precedenti interazioni dell'utente stesso e quelle di altri utenti simili della rete globale dell'OSN. Tali sistemi di valutazione, o "ranking", sono meccanismi che non trattano allo stesso modo i contenuti, ma creano un regime in cui la visibilità all'interno della "News Feed", ossia il contenuto multimediale che appare nella propria bacheca virtuale, è resa scarsa per costruzione, generando la "minaccia dell'invisibilità" (Bucher, 2012). Come già sottolineato, tali algoritmi sono alimentati dai dati degli utenti che li plasmano e ne alterano il funzionamento. Per fare un esempio concreto, l'algoritmo di Facebook, chiamato "EdgeRank", genera un "Edge" per ogni interazione con un "Oggetto" all'interno della News Feed (Kincaid, 2010). Il valore di questo Edge è determinato da più parametri, resi in parte pubblici dalla piattaforma, come l'affinità con chi ha generato l'oggetto, la sua popolarità e la sua data di pubblicazione. Per questo, il ruolo dominante dei modelli data-driven nel filtraggio e nella personalizzazione dell'accesso all'informazione merita riflessioni profonde su come la sfera pubblica si configuri negli OSN. Tale processo di ranking assume in tal senso un ruolo editoriale nella distribuzione delle notizie, perdendo così la sua presunta funzione neutrale di gestione del flusso di contenuti.

Diversi fattori influenzano l'accesso alle notizie al pubblico ed il modo in cui gli utenti recepiscono il contenuto dell'informazione negli OSN. Alcuni di essi sono la forma con cui viene presentata la notizia, l'interazione che essa ha con la rete sociale dell'utente, la data di pubblicazione e la viralità della notizia stessa. Come già osservato per l'algoritmo EdgeRank, i modelli data-driven identificano quali sono gli

interessi degli utenti e garantiscono che i contenuti ricevuti siano rispondenti ad essi. Tale implementazione genera necessariamente “*bias*” algoritmici, ossia distorsioni provocate dal ranking e dalla selezione di determinati contenuti, ed il cui effetto più rilevante è un differente grado di accesso degli utenti ai mezzi di diffusione ed accesso all’informazione online. Uno dei *bias* maggiormente studiati e discussi è la formazione di “echo chamber”, ossia di gruppi di utenti che comunicano e interagiscono preferibilmente tra loro, escludendo altri membri esterni. All’interno degli OSN, tale concetto è strettamente legato a quello di “*filter bubble*” (Pariser, 2011), ossia l’insieme di scelte algoritmiche che genera la personalizzazione dell’output dei servizi digitali online, dai motori di ricerca agli OSN. Nel seguito, si farà riferimento unicamente al termine echo chamber per indicare come l’algoritmo, personalizzato tramite l’interazione dell’utente con i contenuti interni all’OSN, genera una sfera pubblica su misura per l’utente in questione.

Il dibattito sulla reale influenza delle *echo chamber* sulla dieta informativa degli utenti degli OSN è attualmente in corso. Diversi studi (Bruns, 2017a) mostrano come all’interno della rete australiana di utenti su Twitter vi sono limitate tendenze verso l’emergenza di *echo chamber*. Nonostante esistano gruppi di account che condividono interessi o impianti ideologici comuni, tali studi concludono che non si può parlare di *echo chamber* poiché queste dovrebbero essere molto più disconnesse con l’esterno di quanto siano. Inoltre, un questionario online condotto nel 2017 in Francia, Germania e nel Regno Unito immediatamente dopo ognuna delle rispettive elezioni generali (Vaccari, 2018) mostra come in questi tre paesi gli utenti degli OSN è più probabile che siano in disaccordo con i contenuti politici che visualizzano su queste piattaforme piuttosto di osservare preferibilmente contenuti in linea con le proprie idee. Altre ricerche nell’ambito delle scienze sociali computazionali (Schmidt *et al.*, 2017; Nikolov *et al.*, 2015) sostengono invece che gli utenti degli OSN tendono a confinare la loro attenzione ad un insieme limitato di account di notizie, determinando così una struttura a comunità molto ben definita attorno a poche fonti di informazioni. Inoltre, un recente studio sull’algoritmo EdgeRank di Facebook (Hargreaves *et al.*, 2018) conferma che esso tende a rinforzare l’orientamento dell’utente, in particolare nelle posizioni di maggiori rilevanza nella bacheca, e che anche utenti “neutrali”, in termini di account schierati politicamente che l’utente stesso segue, sono esposti a simili fenomeni di *bias* informativo.

4. DESCRIZIONE DEI METODI D'INDAGINE E DEI DATI

A partire dalle precedenti considerazioni sulla formazione di una sfera pubblica digitale governata dalla presenza di algoritmi di selezione e ranking dei contenuti digitali, si cercherà di rispondere alle seguenti “*research question*”:

RQ1: All'interno degli OSN, è possibile identificare gruppi di utenti che discutono in maniera privilegiata attorno ad uno specifico tema e dunque osservare la formazione di *echo chamber*?

RQ2: Quali sono le caratteristiche principali che permettono di identificare tali gruppi e come si configura la diffusione di contenuti al loro interno?

Al fine di rispondere a tali quesiti, si è considerato un argomento di discussione che ha generato un acceso dibattito all'interno degli OSN, ossia l'efficacia dei vaccini, un tema particolarmente caldo in Italia in prossimità dell'approvazione del DL Lorenzin¹. Come sottolineato anche in articoli di cronaca nazionale (Bocci, 2017), il tema è stato fortemente divisivo e le posizioni espresse all'interno degli OSN possono essere etichettate con maggiore facilità come contrarie o favorevoli all'approvazione di tale decreto. Le reti Twitter e Facebook attorno a tale dibattito sono state studiate prendendo in considerazione l'influenza della loro struttura sulla diffusione di notizie legate alle discussioni sul tema. Negli ultimi anni, il tema della disinformazione negli OSN sta diventando centrale nel dibattito politico globale, in particolare per il rischio di esposizione a notizie volutamente false o di scarsa qualità, considerate nel 2013 dal World Economic Forum tra i “rischi geopolitici e tecnologici” dei nostri tempi (Howell, 2013). Per questo motivo, nel seguito, si osserverà in particolar modo come il dibattito all'interno degli OSN viene influenzato fortemente dalle interazioni che avvengono all'interno della sfera pubblica digitale, creando un ambiente favorevole alla formazione di diete informative molto diversificate tra loro

4.1. Dati

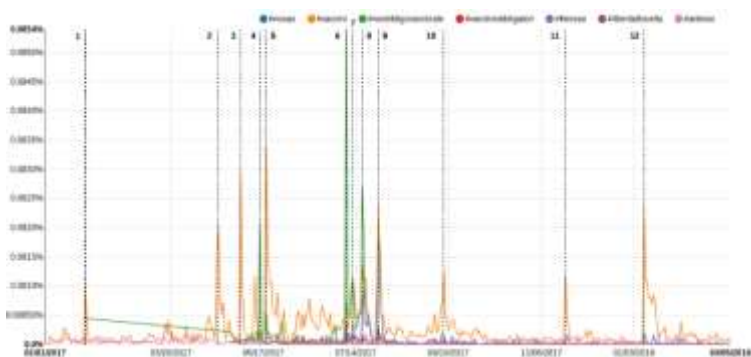
4.1.1. Twitter

I dati utilizzati per l'analisi seguente sono stati raccolti tramite il tool online fornito dal progetto OsOmE (Davis *et al.*, 2016) dell'Università

¹ Il DL Lorenzin in materia di prevenzione vaccinale è un decreto-legge approvato in via definitiva dalla Camera dei Deputati il 28 luglio 2017. Nel decreto è previsto che le vaccinazioni obbligatorie e gratuite passino da quattro a dodici e che esse divengano un requisito d'ammissione all'asilo nido e alle scuole dell'infanzia.

dell'Indiana ed accessibile pubblicamente. I dati estratti da Twitter sono stati selezionati considerando i tweet pubblicati in una data compresa tra il 1° gennaio 2017 e il 4 marzo 2018 contenenti una o più delle seguenti stringhe: *#vaccini*, *#freevax*, *#libertàdiscelta*, *#noobbligovaccinale*, *#vacciniobbligatori*, *#novax*, *#iovaccino*, *#decretolorenzin*, *#vaccino*, *#vaccinare*, *#vaccinarsi*, *#antivax*. Tale lista è stata popolata selezionando alcune stringhe contenenti hashtag delle campagne a favore al DL Lorenzin (*vaccinarsi*, *iovaccino*, *decretolorenzin*) o rilevanti all'interno del dibattito politico in rete sul tema dell'obbligo vaccinale. La presa dati è avvenuta tramite la collezione degli ID dei tweet soddisfacenti le condizioni sopra riportate. Tali ID sono stati poi popolati tramite il software Hydrator che arricchisce i tweet con le seguenti informazioni: autore del tweet, data di pubblicazione del tweet, geo-localizzazione del tweet (se disponibile), contenuto testuale del tweet, link contenuti nel tweet, tipo di tweet (retweet, risposta ad un tweet, tweet originale), hashtag contenuti nel tweet e tutti i dati pubblici sull'autore (numero di followers, numero di following, numero di status pubblicati, etc.). Il dataset finale di Twitter comprende 44034 tweet originati da 12811 utenti differenti. In Figura 1 si può osservare il volume percentuale di tweet pubblicati giornalmente con gli hashtag principali della lista riportata sopra. Come si evince dal grafico, il trend giornaliero degli hashtag segue un andamento dettato dal dibattito pubblico sul tema dei vaccini che si intensifica in corrispondenza di specifici eventi politici riguardanti tale tematica, riportati in Tabella 1. La comparsa dei picchi corrisponde dunque ad un aumento del volume dei tweet in cui si utilizza uno specifico hashtag. In particolare, gli hashtag che presentano i picchi più pronunciati sono *#vaccini* e *#noobbligovaccinale*.

Fig. 1. Volume percentuale di tweet pubblicati giornalmente con gli hashtag principali della lista riportata sopra



Tab. 1. Identificativo degli eventi riportati in Figura 1 insieme alla data e ad una breve descrizione. Come riportato in Figura 1, tali episodi provocano un picco nell'andamento dei trend di Twitter

Numero del picco e data dell'evento	Descrizione
(1) 26 gennaio 2017	Accordo Stato-regioni per una legge nazionale sui vaccini
(2) 19 aprile 2017	Va in onda una puntata di "Report" sul vaccino anti HPV
(3) 3 maggio 2017	Il New York Times pubblica l'editoriale "Populismo, politica e morbillo"
(4) 15 maggio 2017	Il primo progetto di legge per cancellare il decreto Lorenzin viene presentato
(5) 19 maggio 2017	Viene varato il decreto-legge per l'obbligo vaccinale nazionale
(6) 8 luglio 2017	Raduno nazionale no-vax a Pesaro
(7) 12 luglio 2017	Intesa tra Partito Democratico e Forza Italia sul Decreto Vaccini
(8) 18 luglio 2017	Discussione al Senato sul Decreto Vaccini
(9) 28 luglio 2017	Via libera definitivo della Camera al Decreto Vaccini
(10) 7 settembre 2017	Sospensione della moratoria di due anni sulla nuova legge sulle vaccinazioni introdotta dalla regione Veneto
(11) 22 novembre 2017	La Corte Costituzionale respinge i ricorsi presentati dalla Regione Veneto
(12) 10 gennaio 2018	Matteo Salvini dichiara in campagna elettorale di voler abolire il decreto Lorenzin

4.1.2 Facebook

Numerose ricerche sugli OSN hanno mostrato che gli utenti su Facebook tendono a confinare la loro attenzione ad un insieme ristretto di pagine, determinando una struttura comunitaria attorno ad esse, attraverso l'interazione con le notizie che vengono pubblicate (Schmidt *et al.*, 2017). Per questo motivo, sono state selezionate 85 pagine Facebook (appendice A) che pubblicano regolarmente contenuti sul tema del DL Lorenzin e dell'efficacia dei vaccini. Le pagine pubbliche

sono state selezionate sulla base di criteri quantitativi, quali il numero di like, il volume di interazioni (molti commenti e/o condivisioni), l'alta frequenza di pubblicazione, ma anche sulla base di criteri qualitativi nel caso delle pagine di personaggi politici che si sono espressi in materia, pagine ufficiali nella campagna di informazione, pagine menzionate dai quotidiani e/o da altri media, al fine di ottenere un universo il più possibile completo ed esaustivo.

Tramite l'ausilio di due osservatori indipendenti, le pagine selezionate sono state classificate a seconda dei contenuti pubblicati come pagine "pro-vax" (per un totale di 30 pagine), ossia a favore dell'obbligo vaccinale, o "no-vax" (per un totale di 55 pagine), contrarie o scettiche nei confronti di tale obbligo. Di queste pagine, tramite Facebook Graph API (Facebook, 2015), sono stati estratti i post nello stesso periodo di tempo del dataset Twitter, compreso tra il 1 gennaio 2017 e il 4 marzo 2018. In totale, sono stati raccolti 8151 post insieme ad ulteriori informazioni fornite da Facebook, quali il numero di condivisioni, la data di pubblicazione, il testo del post, il numero delle reactions e dei commenti

4.2. *Metodi d'indagine*

Nel seguito, sono stati utilizzati strumenti ed analisi costruiti a partire dai dati che sono estraibili dai due OSN presi in esame. Tramite un procedimento implementato nel linguaggio di programmazione Python, dai dati Twitter sono state dapprima costruite reti di co-occorrenza di hashtag e di retweet. Nel primo caso, ogni elemento della rete, chiamato *nodo*, rappresenta un hashtag, mentre un collegamento tra due hashtag, chiamato *arco*, indica una co-occorrenza dei due hashtag nello stesso tweet. Ogni arco ha un valore, chiamato *peso*, proporzionale al numero di volte in cui due hashtag compaiono insieme all'interno dell'intero dataset. Nel secondo, ogni nodo rappresenta un utente mentre un arco tra due nodi, il cui valore è unitario, indica un retweet dell'utente da cui parte l'arco di uno specifico tweet dell'utente a cui è diretto l'arco. Al fine di visualizzare la rete ed applicarvi un algoritmo di "community detection", ossia di rilevazione di gruppi di nodi maggiormente connessi tra loro, è stato utilizzato Gephi (Bastian *et al.*, 2009), un software open-source sfruttato largamente per l'analisi e il trattamento di reti sociali. Le immagini riportate nel seguito sono state ottenute rimuovendo i nodi la cui somma dei pesi degli archi uscenti è inferiore a 10, e usando l'algoritmo di visualizzazione "Force Atlas 2" (Jacomy *et al.*, 2014). Per identificare le comunità presenti all'interno della rete, è stato utilizzato un algoritmo di community detection chiamato Louvain (Blondel *et al.*, 2008), che consiste nell'ottimizzazione di una funzione, detta

“modularità”, che misura la densità di archi interni di una singola comunità rispetto a quella degli archi esterni.

Per misurare la permeabilità tra i gruppi di utenti, si è usato un parametro chiamato “indice E-I” (Krackhardt, Stern, 1988) che misura la porzione di archi interni, ossia diretti da un membro ad un altro della stessa comunità, rispetto al numero di quelli esterni, ossia diretti da un membro di una comunità ad un membro esterno. A partire da questa misura, si può dunque convertire il numero di archi interni ed esterni in un indice normalizzato che, nella seguente definizione, può rappresentare una stima di quanto le comunità siano chiuse, ossia quanto la loro struttura sia simile a quella di un’echo chamber:

$$\text{Indice } E-I = \frac{\square \text{Link Esterni} - \square \text{Link Interni}}{\square \text{Link Esterni} + \square \text{Link Interni}}$$

Usando questo indice, un nodo i cui archi si connettono unicamente a nodi esterni dalla comunità ha un indice E-I pari a +1 mentre, per un nodo i cui archi si connettono esclusivamente a nodi interni alla comunità, è pari a -1. Poiché questi valori estremi sono molto rari in un contesto reale, all’interno dello spettro di valori compreso tra -1 e +1, l’indice E-I diventa una misura affidabile della tendenza di un nodo a collegarsi a nodi della propria comunità.

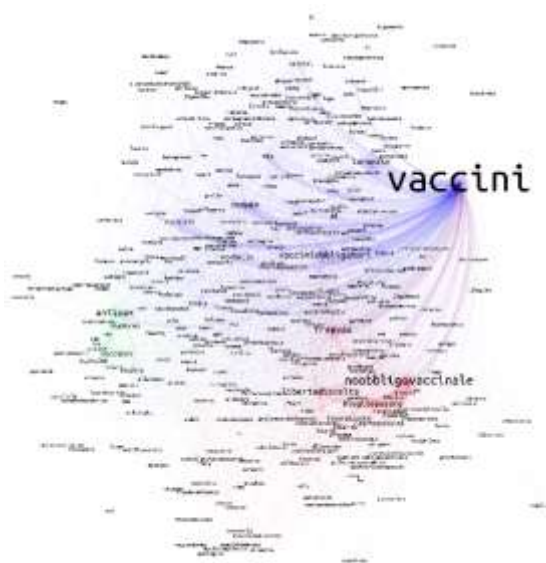
Per quanto riguarda l’analisi delle interazioni di Facebook sul tema dei vaccini, si vuole innanzitutto sottolineare che il numero di post estratti da tale OSN è notevolmente inferiore rispetto a quelli estratti da Twitter, in particolare per via del fatto che una parte consistente di discussioni avviene nei profili personali, ristretti solo agli amici degli utenti, o in gruppi chiusi, accessibili solo ad una ristretta cerchia di utenti. Un’altra criticità è inoltre rappresentata dall’API ufficiale di Facebook, la cui capacità di estrazione dati dall’OSN è stata progressivamente limitata (Facebook, 2018), a seguito dello scandalo di Cambridge Analytica. Per via di tale chiusura delle API ufficiali, non è dunque possibile costruire una rete delle interazioni degli utenti di Facebook. L’analisi presentata in seguito si è invece concentrata sullo studio delle interazioni e del contenuto testuale dei post nelle pagine pubbliche selezionate. Non avendo accesso alla rete sociale attorno alle pagine e alle condivisioni dei post per ragioni legate alla privacy degli utenti, per studiare l’interazione relativa ai post estratti da Facebook, è stato utilizzato un indice ottenuto tramite una “sentiment analysis” effettuata sui commenti riportati sotto i post di tali comunità. Il risultato di tale analisi è un parametro che assume valori compresi tra -1 e +1 e che indica la maggiore presenza di termini etichettati come negativi, che assumono valori inferiori di zero, o positivi, che assumono valori maggiori di zero.

5. RISULTATI E DISCUSSIONE

5.1. *Twitter*

A partire dall'osservazione che la scelta di utilizzare un determinato hashtag è dettata da esigenze comunicative in relazione ad un evento specifico che innesci una discussione all'interno degli OSN, è stata dapprima costruita una rete di co-occorrenze di hashtag. All'interno di Twitter, nel quale i caratteri disponibili per ogni singolo tweet sono poche centinaia, lo studio della rete degli hashtag può essere utile a determinare se le parole chiave utilizzate all'interno di una discussione sono confinate all'interno di un gruppo di utenti o se hanno un utilizzo più trasversale ed ampio. In Figura 2, si può osservare una rappresentazione della rete, composta da 358 nodi e 3243 archi, di co-occorrenze di hashtag.

Fig. 2. Visualizzazione della rete di co-occorrenze di hashtag effettuata tramite il software open-source Gephi. La dimensione del font di ogni hashtag (riportato senza il cancelletto) è proporzionale alla somma dei pesi degli archi uscenti dal nodo



Come si può evincere dalla Figura 2, l'algoritmo di community detection individua tre comunità separate all'interno della rete. Nella comunità più popolata, colorata in blu, si può osservare la compresenza di hashtag appartenenti ad ambiti diversi tra loro, come ad esempio:

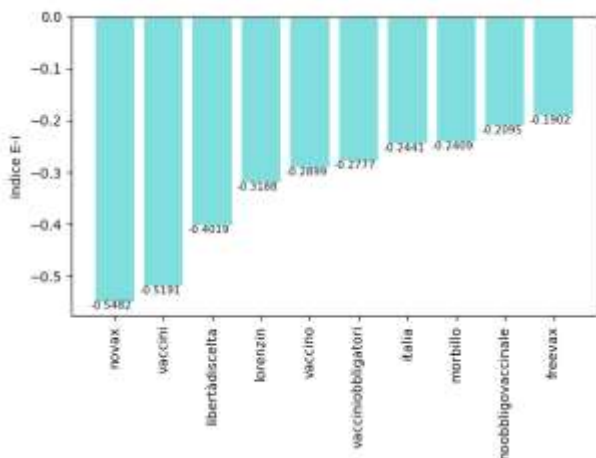
- nomi di eventi politici (*4marzo2018*, *elezionipolitiche2018*), leggi (*jobsact*, *leggeelettorale*), partiti (*m5s*, *forzaitalia*, *pd*) e personaggi della politica italiana (*salvini*, *berlusconi*, *lorenzin*);
- nomi di programmi (*tagadala7*, *dallavostraparte*) o reti televisive (*la7*);
- riferimenti a campagne nazionali a favore della vaccinazione (*iovacchino*, *vaccinarsi*) o a termini legati alla divulgazione sull'efficacia dei vaccini (*morbillo*, *immunitàdigregge*);

Dalla presenza di una tale varietà di tematiche, l'ipotesi avanzata sui temi del dibattito all'interno di questa comunità è che essi spaziano dalla discussione in corrispondenza di una trasmissione televisiva che tratta il tema dei vaccini, fino ad arrivare al supporto alla campagna nazionale a favore dell'approvazione del DL Lorenzin. Questo si può evincere anche dalla centralità all'interno di tale comunità dell'hashtag *#vaccini*, il quale, come si osserva in Figura 1, è massicciamente utilizzato in corrispondenza di svariati episodi politici nazionali. La seconda comunità, colorata in rosso, è caratterizzata dalla centralità di alcuni hashtag (*noobbbligovaccinale*, *8lugliopesaro*, *libertàdiscelta*) attorno ai quali si sviluppano alcune tematiche chiave: la difesa dei bambini e della libertà di scelta vaccinale (*giùlemanidaibambini*, *liberascelta*), il raduno nazionale no-vax a Pesaro (*8luglio*, *pesaro*) e alcune campagne contro l'obbligo vaccinale (*ilsentierodinicola*, *genitoriconsapevoli*). Come riportato in Figura 1, l'utilizzo dell'hashtag *#noobbbligovaccinale* è limitato ad uno specifico arco temporale che è più intenso a partire dalla presentazione del primo progetto di legge per cancellare il DL Lorenzin fino alla sua approvazione definitiva. Da queste osservazioni, si può dedurre che nella seconda comunità si discute principalmente delle campagne contrarie al DL Lorenzin, ed è dunque la comunità che si può identificare con gli utenti che presentano un maggiore scetticismo nei confronti dell'effetto dei vaccini e delle campagne a favore dell'obbligo vaccinale (*nazivax*, *tungsteno*). Nella terza ed ultima comunità, colorata in verde, si può infine osservare la presenza di hashtag in lingua inglese da cui si suppone che essa sia composta principalmente da utenti stranieri che discutono attorno a temi legati alle vaccinazioni (*vaccineswork*, *vaccines*, *autism*).

Delle tre comunità presentate poco sopra, ci concentreremo, nel seguito, sulle prime due, che riflettono maggiormente il dibattito nazionale italiano sul tema del DL Lorenzin e sull'efficacia dei vaccini. Dalla precedente analisi qualitativa, si può innanzitutto concludere che alle due comunità appartengono gruppi di utenti con intenti e parole d'ordine molto differenti tra loro. Si può supporre che alla comunità raffigurata in blu appartengano utenti che strutturano un dibattito attorno ai temi discussi durante una trasmissione televisiva, in supporto ad un partito politico, al DL Lorenzin o all'efficacia dei vaccini. La seconda comunità ha invece alle spalle una partecipazione politica organizzata attraverso precise parole chiave, e interviene nella sfera pubblica digitale in corrispondenza di specifici eventi, come ad esempio quando viene chiamata una mobilitazione di piazza.

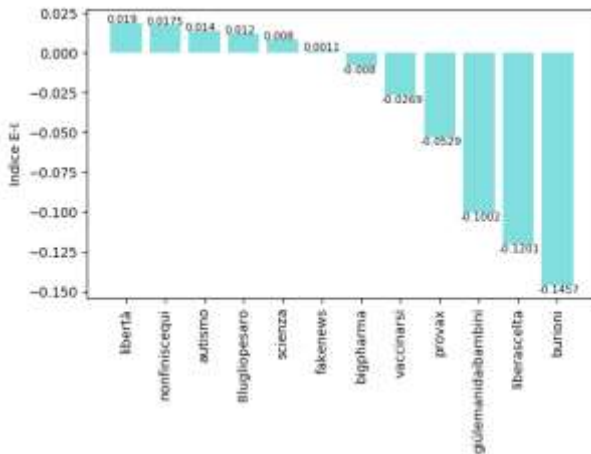
Come si può osservare dal grafico a colonna in Figura 3, tutti i valori dell'indice E-I relativi ai dieci hashtag che hanno il numero maggiore di archi nella rete sono negativi. Poiché valori fortemente negativi dell'indice E-I sono sintomo di una maggiore tendenza dei nodi di una comunità a formare un'echo chamber, si può dedurre che vi è questa tendenza all'interno della rete in esame.

Fig. 3. Grafico a colonna dei valori dell'indice E-I relativi ai dieci hashtag che hanno il numero maggiore di archi nella rete di co-occorrenze di hashtag riportata in Figura 2.



Tuttavia, pur essendo un forte indicatore dell'uso quasi esclusivo di un hashtag da uno dei due gruppi di utenti, ciò vale unicamente per gli hashtag che hanno un numero maggiore di co-occorgenze. Infatti, come si può osservare dalla Figura 4, altri hashtag rilevanti presenti all'interno delle due comunità presentano valori dell'indice E-I prossimi a 0. Tali valori indicano una permeabilità dei due gruppi per un numero consistente di hashtag, i quali vengono utilizzati quasi indistintamente da entrambi i gruppi di utenti. È importante sottolineare che gli hashtag non riportati nel grafico presentano valori compresi tra i due valori estremi riportati in Figura 4, rendendo questo ragionamento valido globalmente per la rete in esame.

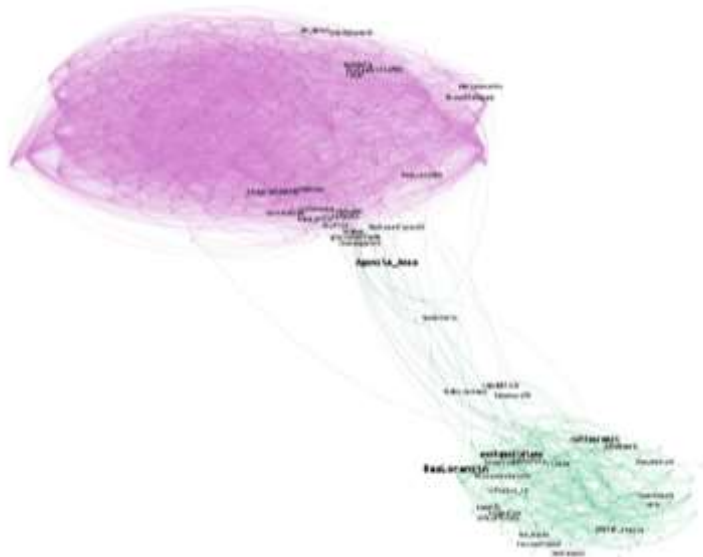
Fig. 4. Grafico a colonna dei valori dell'indice E-I relativi ad hashtag rilevanti nella rete di co-occorgenze di hashtag riportata in Figura 2.



Da questa osservazione, si può dunque concludere che, pur essendoci specifici hashtag utilizzati in maniera privilegiata da un singolo gruppo di utenti, non esistono forti tendenze a formare echo chamber a livello comunicativo. All'interno dei due gruppi di utenti raffigurati in Figura 2, esistono soltanto alcune parole d'ordine utilizzate maggiormente da un singolo gruppo mentre i restanti hashtag formano un bagaglio comune di termini che vengono riempiti di significato a seconda della rete dell'utente. A questo punto, per analizzare con maggiore dettaglio il comportamento dei gruppi di utenti di cui si è discusso, si è costruita la

rete dei retweet effettuati dagli utenti del dataset a disposizione.

Fig. 5. Visualizzazione della rete di retweet effettuata tramite Gephi. La dimensione del font di ogni utente è proporzionale alla somma dei pesi degli archi uscenti dal nodo



In Figura 5, è riportata tale rete tramite una visualizzazione in Gephi ottenuta rimuovendo i nodi la cui somma dei pesi degli archi uscenti è inferiore a 10, e usando l'algoritmo Force Atlas 2. La rete dei retweet analizzata è composta da 541 nodi, che rappresentano gli utenti, e 5285 archi, che indicano un retweet da parte di un utente.

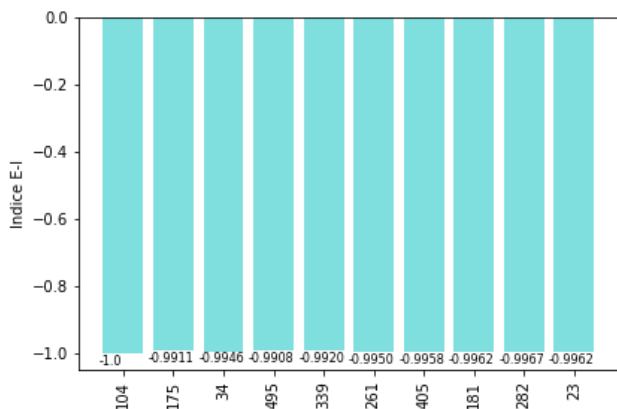
La rete dei retweet rappresenta un indicatore di quali siano i legami tra gli utenti della rete sociale che si struttura attorno alla discussione sui vaccini. All'interno di Twitter, il retweet può avere differenti intenti ma, di certo, il più diffuso è la condivisione del messaggio riportato all'interno del tweet. Le comunità individuate tramite l'algoritmo Louvain di community detection e riportate in Figura 5 sono due. La prima, colorata in verde, consta di account appartenenti a personaggi politici o partiti all'interno della coalizione di Governo durante l'anno

2017 di presa dati (*BeaLorenzin*, *matteoreenzi*, *pdnetwork*, *angealfa*), ad istituzioni politiche nazionali (*Montecitorio*, *MinisteroSalute*), ad istituti sanitari o associazioni (*UNICEF Italia*, *Aifa_ufficiale*, *MSF Italia*), a giornalisti o organi di stampa (*MediasetTgcom24*, *riotta*, *ilfoglio_it*, *RaiNews*) ed a siti di satira (*Labbufala*, *SiamoLaGente*). Questa considerazione iniziale, insieme alla presenza dell'account ufficiale della campagna a favore del DL Lorenzin *IoVaccino*, fa supporre che la prima comunità è costituita in buona parte da chi inserisce la discussione sul tema dei vaccini all'interno di un dialogo con le istituzioni a favore dell'obbligo vaccinale e con gli organi di stampa. Si può notare come la presenza di account che rientrano nelle categorie sopra citate è consistente: dei 178 account appartenenti a tale comunità, 76 account rispondono alle caratteristiche sopra descritte. Alla seconda comunità, colorata in viola, appartengono profili legati alla galassia di associazioni a favore della libertà vaccinale o scettici nei confronti dell'efficacia dei vaccini (*ClivaToscana*, *Vacciniinforma*), a qualche quotidiano o rete televisiva nazionale (*Agenzia_Ansa*, *SkyTG24*, *TgLa7*) e a partiti o politici di opposizione al Governo durante l'anno 2017 di presa dati, in particolare legati al Movimento 5 Stelle (*Mov5Stelle*, *beppe_grillo*, *GiuliaGrilloM5S*). A differenza della prima comunità, la proporzione di tali profili rispetto alla totalità è decisamente inferiore: dei 363 account appartenenti a tale comunità, solamente 34 rientrano nelle categorie elencate sopra. Da questa prima osservazione, si può innanzitutto concludere che i due gruppi di utenti hanno differenti composizioni: al primo gruppo, meno numeroso, appartengono account che interagiscono preferibilmente con account legati alle istituzioni, al Governo o agli organi di stampa, mentre al secondo gruppo appartengono account che non condividono le posizioni istituzionali e dialogano più di frequente con i partiti di opposizione e, in minor misura, con alcuni organi di stampa.

Da un'analisi degli hashtag più usati da tali gruppi si può dedurre che i due gruppi si sovrappongono parzialmente con le comunità descritte dalla rete di hashtag riportata in Figura 2. Infatti, i primi elementi della lista di hashtag più usati dalla comunità colorata in verde in Figura 5 sono i seguenti: *#vaccini*, *#novax*, *#m5s*, *#morbillo*, *#vacciniobbligatori*. Per la comunità colorata in viola tali hashtag sono invece riportati di seguito: *#vaccini*, *#noobbliogvaccinale*, *#libertàdiscelta*, *#8lugliopesaro*, *#liberascelta*. Per i motivi fin qui riportati sulla specifica appartenenza politica degli account di tali comunità e sugli hashtag più usati, da qui in avanti, si può considerare la prima comunità come un gruppo maggiormente a favore del DL Lorenzin e dell'efficacia dei vaccini, o

“pro-vax”, mentre la seconda come un gruppo con un atteggiamento di opposizione al DL Lorenzin e di sfiducia nei vaccini, o “no-vax”. La permeabilità dei due gruppi ad interagire tra loro e dunque la predisposizione al retweet di account appartenenti ad un altro gruppo può essere ancora una volta stimato tramite la misura dell’indice E-I. Come si può osservare in Figura 6, i valori di tale parametro per i cinque account retwittati più volte all’interno dei due gruppi sono prossimi a -1, sintomo evidente di una propensione dei gruppi pro-vax e no-vax a interagire molto raramente tra loro.

Fig. 6. Grafico a colonna dei valori dell’indice E-I relativi ai cinque account retwittati più volte all’interno dei due gruppi riportati nella rete di retweet in Figura 5. Gli “screen-name” degli account sono stati sostituiti con indici relativi al nodo corrispondente.

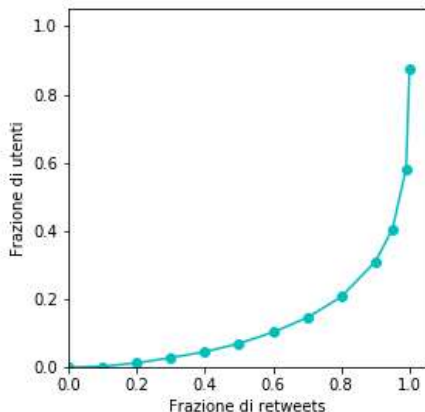


A partire da queste osservazioni, si può entrare più nel dettaglio dei meccanismi di funzionamento della sfera pubblica digitale e dunque di diffusione delle informazioni. Dopo aver caratterizzato e studiato i due gruppi di utenti e gli hashtag più usati da essi, si può infatti osservare come si struttura la comunicazione all’interno dei gruppi pro-vax e no-vax. Dalla stima della percentuale di utenti che genera il maggior numero di retweet, si può innanzitutto notare che soltanto il 7% degli utenti della rete di retweet genera poco più della metà dei retweet totali dell’intera rete. All’interno dei due gruppi, la percentuale di utenti artefice di una determinata percentuale di retweet all’interno della comunità è praticamente la stessa, da cui si conclude che tale

comportamento è uniforme nella rete.

Da qui, si può giungere alla prima importante conclusione: la comunicazione all'interno degli OSN non è orizzontale ma avviene grazie alla presenza di pochi account, chiamati *hubs*, che di fatto sono i principali attori della propagazione delle notizie all'interno della rete sociale che si genera attorno ad una comunità. Il modello top-down dei mezzi di comunicazione di massa si ricrea naturalmente all'interno degli OSN, secondo una legge di potenza che si ritrova in svariati fenomeni sociali ed economici. Tale distribuzione, detta anche distribuzione di Pareto (Pareto, 2013), è alla base della cosiddetta “legge 80/20” per la quale il 20% della popolazione detiene l'80% della ricchezza mondiale. Anche all'interno del dataset Twitter a nostra disposizione, si può osservare che esiste un andamento analogo se al posto della ricchezza pro-capite si considera il numero totale di retweet (Figura 7). È importante sottolineare che non tutti gli utenti del dataset hanno ricevuto retweet per cui il numero totale di retweet si raggiunge considerando circa il 90% degli utenti totali.

Fig. 7. Frazione di utenti che raccolgono il maggior numero di retweet in funzione della frazione dei retweet totali.



L'andamento della legge di Pareto è tipico di sistemi sociali come quello rappresentato dalla rete di retweet di Twitter. In questo contesto, tale fenomeno si può supporre essere riconducibile ad una precisa decisione algoritmica. Privilegiando infatti i contenuti più virali, l'algoritmo permette ad un ristretto numero di contenuti molto popolari di acquisire

ancora più notorietà in un fenomeno auto-alimentato dal procedimento stesso di selezione e personalizzazione della News Feed dell'utente.

La seconda conclusione di rilievo deriva da un'analisi sulla qualità e sulla tipologia delle notizie che fluiscono all'interno delle due comunità riportate in Figura 5 e dunque dei contenuti maggiormente retwittati all'interno dei gruppi pro-vax e no-vax. Come per gli hashtag maggiormente diffusi, anche in questo caso, si ha una rilevante differenza tra le diete informative delle due comunità. Ad esempio, da un'analisi dei link più retwittati e pubblicati dagli account che hanno avuto il numero maggiore di retweet, si ottiene che i link più condivisi dalla comunità pro-vax sono i seguenti:

- Post pubblicato su Facebook da Matteo Renzi², allora leader del PD e primo ministro italiano, in cui viene attaccato Beppe Grillo, esponente del Movimento Cinque Stelle, critico verso l'editoriale del New York Times "Populismo, politica e morbilli";
- Articolo di www.politicaeattualita.it³ sulla risposta della ASL Roma 1 ad una dichiarazione sui vaccini di Alessandro Di Battista, allora parlamentare del Movimento Cinque Stelle;
- Report di www.democratica.com⁴ (sito di informazione dichiaratamente vicino al PD) sulla disinformazione in rete;
- Dichiarazione ufficiale dell'allora ministra Lorenzin⁵ sulla puntata della trasmissione televisiva "Report" a proposito del vaccino anti HPV;
- Dichiarazione a favore dei vaccini pubblicata da www.ansa.it⁶ da parte della campionessa paralimpica Bebe Vio.

Diversamente da quanto visto per la prima comunità, una simile analisi effettuata all'interno del gruppo no-vax ha come risultato la seguente lista:

1. Tweet di Virginia Raggi⁷, sindaca di Roma ed esponente del Movimento 5 Stelle, in cui viene condiviso un comunicato dell'ANCI contro il divieto di inserimento dei bambini non vaccinati nelle scuole;

² Link: <https://www.facebook.com/matteorenziufficiale/posts/10154841689119915>

³ Link: http://www.repubblica.it/salute/2016/09/13/news/usa_1_industria_delle_zuccheri_pago_scientziati_per_mentire-147667151/

⁴ Link disponibile a: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3170075/>

⁵ Link: http://www.salute.gov.it/portale/news/p3_2_4_1_1.jsp?lingua=italiano&menu=salastampa&p=comunicatistampa&id=4884

⁶ Link: http://www.ansa.it/canale_saluteebenessere/notizie/medicina/2017/06/07/bebevio-sui-vaccini-io-chiedo-alla-scienza-non-al-web_70788c0b-af69-4b87-a1a3-bb2b82aa32dd.html

⁷ Link: <https://twitter.com/virginiaraggi/status/959482263607136257>

2. Notizia di www.repubblica.it⁸ sulla scoperta di un documento sulla falsificazione di studi finanziati dall'industria dello zucchero;

3. Articolo scientifico pubblicato su una rivista specializzata⁹ in cui due autori, Gary Goldman e Neil Miller, studiano la correlazione tra la mortalità infantile e il numero di dosi somministrate ai bambini. La bontà dell'articolo in questione è stata successivamente smentita dal sito di debunking "BuTaC - Bufale un Tanto al Chilo" (Coltelli, 2017);

4. Tweet su un'inchiesta parlamentare del Movimento Cinque Stelle sul rapporto Signum sul monitoraggio dei militari inviati in Iraq e le norme vaccinali utilizzate¹⁰. La scientificità di tale inchiesta è stata fortemente criticata da vari quotidiani nazionali (Dotti, 2017);

5. Notizia pubblicata su www.lanazione.it¹¹ su un caso di autismo diagnosticato ad un bambino dopo la somministrazione di nove vaccini. La notizia è stata successivamente smentita dal sito di debunking "BuTaC - Bufale un Tanto al Chilo" (Coltelli, 2015);

Risulta evidente come le diete informative dei due gruppi sono fortemente differenziate. La comunità pro-vax privilegia un'informazione schierata con la formazione al governo, nel periodo di presa dati, composta principalmente da dichiarazioni pubbliche di politici ed altri esponenti pubblici che mirano a confermare una convinzione già presente all'interno del gruppo stesso. Il gruppo di utenti no-vax è schierato fortemente verso posizioni di scetticismo nei confronti dell'efficacia dei vaccini che si concretizzano in una dieta informativa schierata contro l'obbligo vaccinale imposto dal DL Lorenzin. Si può dunque concludere che la mancanza di permeabilità di contenuti tra le due comunità è la causa principale della diffusione virale di notizie spesso non verificate e successivamente smentite. In un quadro simile, mettere in campo pratiche di limitazione nella diffusione di tali notizie, come il cosiddetto "fact-checking", ossia la verifica fattuale delle notizie, è infatti una via difficilmente praticabile a causa della struttura stessa della rete di interazioni all'interno dell'OSN.

5.2. Facebook

Come già rilevato per il dataset di Twitter, in Figura 8 si può osservare

⁸ Link: http://www.repubblica.it/salute/2016/09/13/news/usa_1_industria_delle_zuccheri_pago_scientziati_per_mentire-147667151/

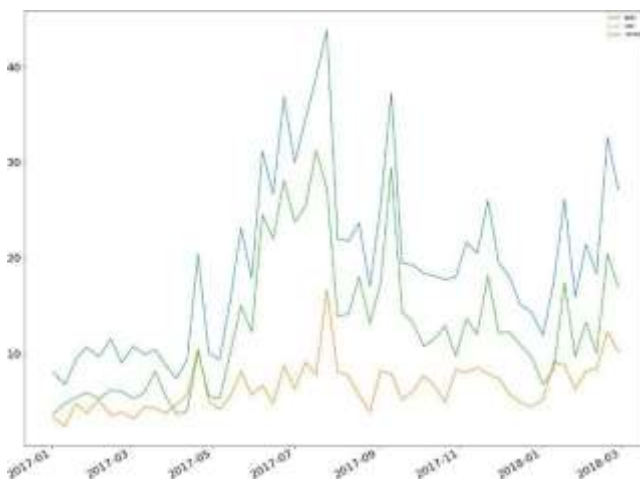
⁹ Link disponibile a: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3170075/>

¹⁰ Link dell'inchiesta parlamentare: http://documenti.camera.it/_dati/leg17/lavori/documentiparlamentari/IndiceETesti/022bis/023/INTERO.pdf

¹¹ Link: http://www.lanazione.it/pistoia/cronaca/vaccini-diagnosi-autismo-1.1173424-?wt_mc=twitteruser

che l'andamento temporale del numero di post pubblicati dalle pagine selezionate presenta dei picchi in corrispondenza di determinati periodi. Da un'analisi qualitativa del trend temporale, si può concludere che tali picchi sono localizzati principalmente nel periodo di discussione del DL Lorenzin, a partire da giugno 2017, legati in particolare all'attività delle pagine no-vax che sono decisive nel determinare l'andamento globale dell'attività delle pagine in esame. Tuttavia, a differenza dell'analisi del trend temporale di Twitter in Figura 1, non si possono ricondurre tali picchi ad una specifica attività giornaliera, probabilmente per la differenza di utilizzo dei due OSN e per la natura delle interazioni al loro interno.

Fig. 8. Andamento temporale del numero dei post pubblicati da tutti i gruppi selezionati (colorato in ciano), il numero dei post dei gruppi pro-vax (colorato in verde) e quelli dei gruppi no-vax (colorato in giallo). Ogni dato è un'informazione aggregata di 6 giorni consecutivi di attività della pagine in esame.

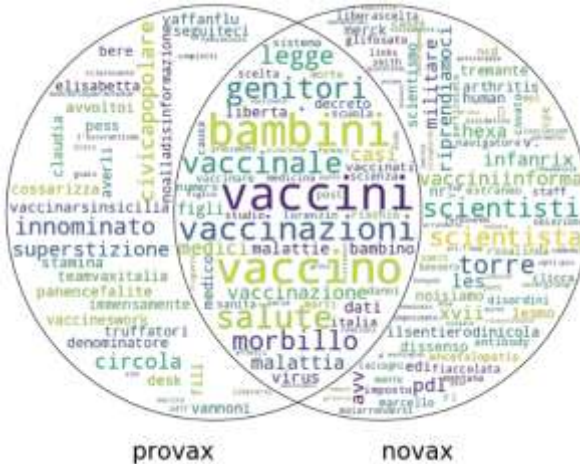


Il numero di post pubblicati è di certo soltanto uno dei possibili indicatori dell'attività all'interno di Facebook. Un trend simile a quello mostrato in Figura 8 si può ritrovare anche nell'andamento temporale del numero di interazione con i post della pagina, ossia il numero di commenti e delle reaction totali accumulate in un determinato periodo temporale. Tale considerazione porta un'informazione aggiuntiva, ossia

il fatto che la comunità di utenti che segue una pagina si attiva maggiormente durante il periodo di maggiore attività. Alla luce del funzionamento dell'algoritmo dietro alla News Feed di Facebook, questo comportamento è quello atteso all'interno dell'OSN in esame.

Come già effettuato per Twitter, si è scelto inoltre di caratterizzare i gruppi di pagine tramite un'analisi testuale dei post pubblicati. All'interno di Facebook, non essendoci limiti sul numero di caratteri, si può ottenere un'analisi più complessa ed approfondita del linguaggio e dei termini maggiormente in uso. Come si può osservare in Figura 10, i termini utilizzati dalle due comunità hanno un'intersezione significativa di vocaboli usati in entrambi i gruppi, quali *vaccino*, *bambini*, *salute*, *genitori*. Altri termini sono invece specifici di una delle comunità: nel gruppo di pagine pro-vax sono maggiormente in uso parole come *superstizione*, *stamina* e *truffatori*; in quelle no-vax, sono utilizzati invece termini come *scienziata*, *glifosato* e *liberascelta*.

Fig.10 Word-cloud dei gruppi di pagine pro-vax e no-vax strutturati in modo da essere inseriti all'interno di un diagramma di Venn. La dimensione del font di ogni termine è proporzionale alla frequenza di utilizzo.

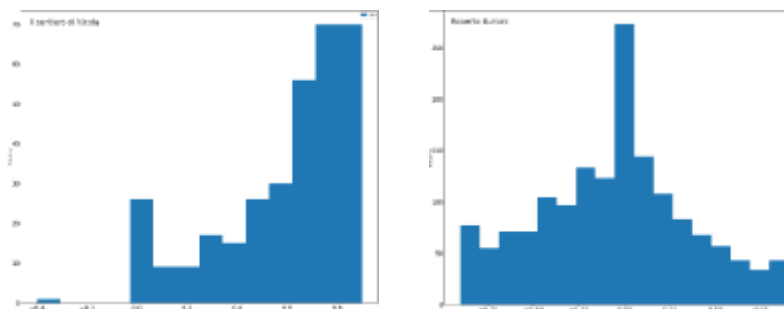


Da questa considerazione, si può innanzitutto concludere che, come già osservato dall'analisi del dataset di Twitter, la polarizzazione delle comunità non passa attraverso la forma comunicativa che si struttura all'interno delle comunità stesse. Su Twitter, si è invece osservato che la polarizzazione si sostanzia in una differente modalità di interazione

verso i contenuti pubblicati.

Infine, è stata effettuata una *sentiment analysis* sui commenti sotto il post più commentato di ogni pagina. La distribuzione finale del valore dell'indice è stata considerata un segnale per studiare la condivisione del contenuto pubblicato all'interno di una pagina. In Figura 11, è raffigurata la distribuzione dell'indice ottenuto dalla *sentiment analysis* effettuata tramite la piattaforma di analisi del linguaggio di Google Cloud del post più commentato di due pagine prese a modello dell'analisi effettuata: "Roberto Burioni" (pagina appartenente al gruppo pro-vax e riconducibile ad un noto immunologo italiano) e "Il sentiero di Nicola" (pagina appartenente al gruppo no-vax). Come si evince dalle due distribuzioni, sotto i commenti delle due pagine, si osservano differenti *patterns* d'interazione. La comunità di utenti attorno alla pagina de "Il sentiero di Nicola" è di certo coerente con il post pubblicato dalla pagina nel mostrare un'uniformità di reazioni, come dimostra il picco positivo dell'indice. Al contrario, la distribuzione dei commenti sotto il post della pagina "Roberto Burioni" presenta un picco centrato a zero, e code che si estendono sia nell'asse positivo che in quello negativo, sintomo di una mancanza di uniformità nei commenti riportati sotto il post. Tale andamento è sintomo di una polarizzazione nei commenti della pagina verso posizioni opposte e dunque di una trasversalità della diffusione di tale post all'interno delle due comunità. Quest'ultimo andamento si può riscontrare in una netta minoranza di pagine all'interno sia dei gruppi pro-vax che di quelli no-vax e sono sintomo di una parziale sovrapposizione delle due comunità.

Fig. 11. Distribuzione dell'indice finale ottenuto da una *sentiment analysis* effettuata sul testo dei commenti riportati sotto il post più commentato di due pagine prese a modello



6. CONCLUSIONI E FUTURI SVILUPPI

L'utilizzo degli OSN nella dieta informativa di frammenti sempre più consistenti della società è un fenomeno in rapida espansione e che ha una rilevanza mediatica sempre maggiore. Negli ultimi decenni, gli OSN si sono dimostrati determinanti in fenomeni collettivi, quali il sostegno alla partecipazione politica attiva ed ai movimenti sociali. Tuttavia, l'orizzontalità nella comunicazione all'interno degli OSN ed il superamento di un modello top-down tipico dei mezzi di comunicazione tradizionali vengono negate dalla presenza di confini virtuali imposti alla sfera pubblica digitale, ossia all'universo virtuale di discussione in cui si forma l'opinione pubblica. Dietro la presenza di tali confini, esistono processi algoritmici alla base della diffusione dei contenuti all'interno degli OSN. Tali processi sono alimentati dai dati dell'utente che, interagendo con i contenuti della *News Feed*, ne alterano il funzionamento. L'algoritmo è dunque l'artefice principale della formazione di confini virtuali, attraverso un processo di personalizzazione e selezione di contenuti che, per questo motivo, non può essere considerato neutrale.

Purtroppo, non essendo accessibile la conoscenza dei processi interni del loro funzionamento, l'articolo si è concentrato nell'osservazione degli effetti di tali algoritmi sulla sfera pubblica digitale, in particolare sulla formazione di gruppi di utenti maggiormente connessi tra loro. Pur consapevoli di un dibattito in corso sull'esistenza di tali gruppi, meglio noti come *echo chamber*, e sulla difficoltà nella stima della reale influenza che gli algoritmi esercitano nella loro formazione, questo articolo ha mostrato, attraverso il *case study* del dibattito italiano attorno all'obbligo vaccinale, come si può studiare la loro effettiva esistenza. L'articolo si situa all'interno del dibattito relativo alla presenza delle *echo chamber* all'interno degli OSN attraverso un'analisi effettuata all'interno di due OSN, quali Twitter e Facebook, che ha portato ad una loro osservazione indiretta. A partire da esso, si può concludere che le *echo chamber* si sostanziano in queste due piattaforme attraverso comunità di utenti e di pagine che mostrano una tendenza ad interagire in maniera privilegiata tra loro attorno ad un tema specifico come quello del dibattito italiano sull'approvazione del DL Lorenzin e dell'efficacia dei vaccini. All'interno di tali comunità, che si strutturano anche attraverso la presenza di specifiche parole d'ordine maggiormente ricorrenti in una singola comunità, la diffusione di notizie avviene tramite pagine o profili, chiamati *hubs*, che sono responsabili della quasi totalità della diffusione di notizie all'interno di

tali gruppi. Tale conclusione è certamente più marcata nell'analisi dei dati Twitter in cui sono disponibili le reti di utenti ed in misura minore nell'analisi dei dati Facebook dove invece il nostro studio si sofferma sull'analisi semantica dei post e dei commenti.

D'altra parte, si è consapevoli di come entrambi gli OSN in esame hanno limiti intrinseci che portano ad interrogarci sull'effettiva portata delle nostre conclusioni. Oltre ad essere limitati dalla specificità del dibattito italiano sull'efficacia dei vaccini, vi sono considerazioni legate all'utilizzo e al funzionamento di entrambi gli OSN. Innanzitutto, pur essendo noto che la diffusione di contenuti all'interno di Twitter avviene attraverso il retweet, a quest'ultimo si aggiungono ulteriori forme di interazioni tra utenti, come le "mention". Tali interazioni potrebbero essere prese in esame in un futuro sviluppo delle analisi riportate in questo articolo e svelare discussioni più ampie che non sono deducibili dalle reti precedentemente analizzate. Se invece si sposta l'attenzione sull'analisi effettuata sui dati Facebook, si vuole porre l'attenzione sul fatto che le interazioni prese in esame sono limitate fortemente dal numero di utenti che seguono le pagine considerate nello studio. Per ampliare l'analisi e verificare l'ipotesi di formazione di *echo chamber* su un campione più ampio di utenti, in un futuro sviluppo della nostra ricerca si potrebbero studiare i post ed i commenti di altri gruppi di pagine, ad esempio riconducibili a media tradizionali (quotidiani nazionali, emittenti televisive, etc.) ed a giornalisti.

RIFERIMENTI BIBLIOGRAFICI

- BARISONE, M., MICHAÏLOU, A., AIROLDI, M. (2017). Understanding a digital movement of opinion: the case of #RefugeesWelcome. *Information, Communication & Society*, [online Dec 08 2017], 1-20.
- BASTIAN, M., HEYMAN, S., JACOMY, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating network. In *Proceedings of the Third International ICWSM Conference* (pp. 361-362). San José: AAAI Press.
- BASTOS, M. T., MERCEA, D., CHARPENTIER, A. (2015). Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media. *Journal of Communication*, 65(2), 320-350.
- BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R., LEFEBVRE, E. (2008). Fast Unfolding of Communities in Large network. *Journal of Statistical Mechanics: Theory and Experiment*, 10, 10008.
- BOCCI, M. (2017). La battaglia dei vaccini su Twitter. Così la rete si è
-

- schierata sul decreto del governo. *R.it*, 15 giugno.
- BRUNS A. (2017a). Echo chamber? What echo chamber? Reviewing the evidence. *6th Biennial Future of Journalism Conference (FOJ17)*, 14-15 September 2017, Cardiff (UK.) (Unpublished).
- , HIGHFIELD T. (2017b). Is Habermas on Twitter? Social Media and the Public Sphere. In A. Bruns, G. Enli, E. Skogerbø, A. O. Larsson, C. Christensen (eds.). *The Routledge Companion to Social Media and Politics*. (pp. 56-73) New York: Routledge.
- BUCHER, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 14(7), 1164-1180.
- CASTELLS, M. (2009). *Potere e Comunicazione*. Milano: EGEA.
- COLTELLI, M. (2015). Autismo per colpa di nove vaccini?. *Butac*, 27 luglio.
- (2017). Barnard e lo studio di Goldman e Miller del 2011. *Butac*, 26 maggio.
- DAVIS, C. A., CIAMPAGLIA, G. L., AIELLO, L. M., CHUNG, K., CONOVER, M. D., FERRARA, E., MENCZER, F. *et al.* (2016). OSoMe: the IUNI observatory on social media. *PeerJ Computer Science*, 2(e87).
- DOTTI, G. (2017). Conferenza anti-vaccini alla Camera, in realtà non esiste un'altra verità. *Wired.it*, 13 aprile.
- FACEBOOK (2015). Using the graph API. Disponibile su: <https://developers.facebook.com/docs/graph-api/using-graph-api>
- (2018). A Platform Update. Disponibile su: <https://newsroom.fb.com/news/2018/07/a-platform-update/>
- GILLESPIE, T. (2012). Can algorithm be wrong?. In L. Irani, C. M. Kelty, N. Seaver N. (eds.). *Crowds and Clouds*. New York-Los Angeles: Limn.
- HABERMAS, J. (2005). *Storia e critica dell'opinione pubblica*. Roma-Bari: Laterza.
- HARGREAVES, E., AGOSTI, C., MENASCHÉ, D., NEGLIA, G., REIFFERS-MASSON, A., ALTMAN, E. (2018). Biases in the Facebook News Feed: a Case Study on the Italian Elections. *arXiv.org*, 1807.08346.
- HOWELL, L. (2013). *Global Risks 2013. Digital Wildfires in a Hyper-connected World*. Ginevra: World Economic Forum.
- JACOMY, M., VENTURINI, T., HEYMANN, S., BASTIAN, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, 9(6), e98679.
- KINCAID, J. (2010). EdgeRank: The secret sauce that makes Facebook's news feed tick. *TechCrunch*, April 22.
- KRACKHARDT, D., STERN, R. N. (1988). Informal network and
-

- Organizational Crises: An Experimental Simulation. *Social Psychology Quarterly*, 51(2), 123-140.
- MERCEA, D., FUNK, A. (2016). The social media overture of the pan-European Stop-ACTA protest: An empirical examination of participatory coordination in connective action. *Convergence*, 22(3), 287-312.
- NIKOLOV D., OLIVEIRA D.F.M., FLAMMINI A., MENCZER F. (2015). Measuring online social bubbles. *PeerJ Computer Science*, 1 (e38).
- PARETO V. (2013). Corso d'economia politica. Torino: UTET.
- PARISER, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York: Penguin Press.
- SCHMIDT, A. L., ZOLLO, F., DEL VICARIO, M., BESSI, A., SCALA, A., CALDARELLI, G., QUATTROCIOCHI, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*. 114(12), pp. 3035-3039.
- SHEARER, E., GOTTFRIED, J. (2017). News use across social media platforms 2017. *Journalism & Media*, September 7.
- VACCARI, C. (2018). *How Prevalent are Filter Bubbles and Echo chamber on Social Media? Not as Much as Conventional Wisdom Has It*. Loughborough: Centre for Research in Communication and Culture – Loughborough University.
- VECCHI, B. (2017). *Il capitalismo delle piattaforme*. Roma: ManifestoLibri.

APPENDICE

1. Elenco delle pagine Facebook contro la vaccinazione obbligatoria

<https://www.facebook.com/vacciniinforma/>
<https://www.facebook.com/autismovaccini/>
<https://www.facebook.com/COMILVA/>
<https://www.facebook.com/ASSOCIAZIONE-VACCINARE-INFORMATI-135706221964/>
<https://www.facebook.com/laleva.org/>
<https://www.facebook.com/MedNat.org/>
<https://www.facebook.com/Medicinenon/>
<https://www.facebook.com/Informasalus/>
<https://www.facebook.com/vaccinibasta/>
<https://www.facebook.com/vacciniliberascelta.it/>
<https://www.facebook.com/vacciniscltaconsapevole/>
<https://www.facebook.com/riformacostituzionale/>
<https://www.facebook.com/vaccinistop/>
<https://www.facebook.com/vaccinipuliti/>
<https://www.facebook.com/VitalMicroscopio/>
<https://www.facebook.com/noiavacciniobbligatori/>
<https://www.facebook.com/sicurezzaavaccini/>

<https://www.facebook.com/vaccinifacoltativi/>
<https://www.facebook.com/Libertà-di-vaccini-736935763155189/>
<https://www.facebook.com/vacciniebugiardini/>
<https://www.facebook.com/infovax.it/>
<https://www.facebook.com/Vaccini-Sicuri-Analizziamoli-575396842640250/>
<https://www.facebook.com/il.sentiero.di.Nicola/>
<https://www.facebook.com/Testimonianze-sui-danneggiati-dai-vaccini-1966023793617016/>
<https://www.facebook.com/topcarlo68/>
<https://www.facebook.com/NoVacciniForzosi/>
<https://www.facebook.com/Giorgio-Tremante-Danni-Dei-Vaccini-158473024170284/>
<https://www.facebook.com/Vaccini-Informati-per-Sopravvivere-123458221136843/>
<https://www.facebook.com/noaivaccini.it/>
<https://www.facebook.com/laveritasuivaccini/>
<https://www.facebook.com/Vaccini-perché-No-1358710034161275/>
<https://www.facebook.com/farmacovaccino/>
<https://www.facebook.com/NoObbligoVaccini/>
<https://www.facebook.com/gherard.inwind.it/>
<https://www.facebook.com/Quando-i-vaccini-uccidono-o-creano-malattie-104049992977922/>
<https://www.facebook.com/vaccinineonatianni/>
<https://www.facebook.com/reteinfovaccini/>
<https://www.facebook.com/Vaccini-Per-Una-Scelta-Consapevole-188802734594918/>
<https://www.facebook.com/Insieme-per-lautismo-insieme-per-dire-no-ai-vaccini-474077772647396/>
<https://www.facebook.com/vaccinarsiconsapevolmente/>
<https://www.facebook.com/consticazzichemivax/>
<https://www.facebook.com/corolaita/>
<https://www.facebook.com/IononVaccino/>
<https://www.facebook.com/Vacci-NO-il-rovescio-della-medaglia-1485684698399390/>
<https://www.facebook.com/CoordinamentoLiberaSclta/>
<https://www.facebook.com/liberasclta.org/>
<https://www.facebook.com/Libertà-di-sclta-Vaccinale-Italia-1203413723121751/>
<https://www.facebook.com/Vaccisano/>
<https://www.facebook.com/liberascltaitalia/>
<https://www.facebook.com/Liberascltastampa/>
<https://www.facebook.com/Libertà-di-Sclta-Proposta-di-Legge-di-Iniziativa-Popolare-158974298109784/>
<https://www.facebook.com/vaccinipediatrici/>
<https://www.facebook.com/Libertà-di-sclta-Vaccinale-1833374846913905/>
<https://www.facebook.com/novacciness/>
<https://www.facebook.com/Clivatoscana/>
<https://www.facebook.com/azioni.contro.obbligo.vaccinale/>
<https://www.facebook.com/StefanoRecheparlacongliscientisti/>
<https://www.facebook.com/SiAmoUnMovimento/>
<https://www.facebook.com/SiamoLiguria/>
<https://www.facebook.com/AURET-Autismo-Ricerca-e-Terapie-1686952251548554/>
<https://www.facebook.com/I-vaccini-fanno-male-118241631615692/>
<https://www.facebook.com/vaccinismo/>

2. Elenco delle pagine Facebook a favore della vaccinazione obbligatoria

<https://www.facebook.com/ivacciniealtricomplottelegendari/>

<https://www.facebook.com/reteinformazionevaccini/>
<https://www.facebook.com/iovaccino/>
<https://www.facebook.com/AlleanzaItalianaStrategieVaccinali/>
<https://www.facebook.com/Vaccini-AslTeramo-Numero-Verde-472244736443665/>
<https://www.facebook.com/antivaccinisti/>
<https://www.facebook.com/ivaccinisullamiapelle/>
<https://www.facebook.com/VacciniPediatri-e-Asl-Verità-provax-senza-bugie-390079854744987/>
<https://www.facebook.com/I-vaccini-tomano-a-scuola-1637425496370734/>
<https://www.facebook.com/robertoburioniMD/>
<https://www.facebook.com/BeatriceLorenzin/>
<https://www.facebook.com/Vacciniamo/>
<https://www.facebook.com/vaccinarsi/>
<https://www.facebook.com/VaccinarSI.Vaccinfo/>
<https://www.facebook.com/PLopalcoPublic/>
<https://www.facebook.com/MedBunker/>
<https://www.facebook.com/VaccinarSi-253290528189135/>
<https://www.facebook.com/vaccinarsinveneto/>
<https://www.facebook.com/pagatidabighpharma/>
<https://www.facebook.com/iononmiinchinoevaccino/>
<https://www.facebook.com/dannidanonvaccino/>
<https://www.facebook.com/todiiomivaccino/>
<https://www.facebook.com/perchevaccino/>
<https://www.facebook.com/VaccinarSi-511573789002990/>
<https://www.facebook.com/VaccinarSiInSicilia/>
<https://www.facebook.com/vaccinarsilazio/>
<https://www.facebook.com/Epidemiologiavaccinale/>
<https://www.facebook.com/NoAllePseudoscienze/>
<https://www.facebook.com/iomivaccino/>
<https://www.facebook.com/Vaccinazioni-Toscana-sud-est-629197280549855/>
<https://www.facebook.com/Vaccinazione-Asp-Caltanissetta-571029169766980/>

Mario Tirino, Antonio Tramontana

I riflessi di «Black Mirror». Glossario su immaginari, culture e media della società digitale

Roma, Rogas Edizioni, 2018, 280 pp.

di *Irene Psaroudakis**

In un futuro (presente) prossimo saranno gli algoritmi e l'intelligenza artificiale a orientare le nostre azioni; ma se codici alfanumerici e flussi di dati sono già adesso degli agenti in grado di guidare il nostro quotidiano – si pensi, tra i vari esempi, al marketing, alle reti sociali online, all'imperativo dell'immagine, alle tecniche di comunicazione virale, etc. –, che genere di impatto avranno sul nostro essere “corpo relazionale”, ovvero sulla nostra condizione di individui dotati di senso e capaci di attribuire significati all'esperienza, sia essa vissuta o immaginata?

Tutto quello che stiamo vivendo – una “frattura epocale, nella contemporaneità, determinata dalla pervasività con cui le tecnologie e i media digitali ridisegnano la vita quotidiana di milioni di persone” (172) – delinea non soltanto un vero e proprio cambio di paradigma, ma soprattutto una sfida da comprendere e quindi da agire. Assumendo questa prospettiva, e attingendo a suggestioni mitologiche, si può affermare che l'interazione veicolata dai social media, il controllo



* IRENE PSAROUDAKIS è dottore di ricerca in “Storia e sociologia della modernità” e ricercatrice post-doc presso il Dipartimento di Scienze Politiche dell'Università di Pisa, dove collabora alle attività di ricerca e didattica, e contribuisce allo sviluppo della comunità degli Interazionisti Simbolici pisani.

Email: irene.psaroudakis@sp.unipi.it

e la *manipolazione* spazio-temporale permessa dalla tecnologia stiano assumendo la forma simbolica di tre vertici di un labirinto asfissiante ma eternamente aperto, indefinito, e in certo qual modo perturbante, rappresentato dal post-umanesimo. La tecnologia digitale diventa una sorta di novello Minotauro la cui ragione indiscussa, valore assoluto weberiano, è data dall'orientamento al futuro (tecnologico) in un presente post-umano, la cui "mostruosità" non è più solo esplicita, manifesta, ma latente nella sua serialità e in una riproducibilità de-individualizzante, il cui esito è la perdita del Sé in un'alterità tecnica e mediata.

La complessità della questione coglie perciò un ampio spettro di criticità, che possono essere sintetizzate in alcuni quesiti. Quali limiti alla realtà (il rapporto individuo-tecnica)? Quali categorie per le emozioni, che tipo di aggettivi per i nostri sensi (il mutamento delle peculiarità umane e quindi la questione del Self)? Quali effetti imprevedibili di questo essere "gettati" in questo particolare tempo (l'esperienza mediatica e mediale), in cui realtà digitale e mondo dei media si fanno biografia quotidiana, dissolvendosi in una virtualità ormai più che reale, più che radicata, oltre la "cultura della virtualità reale" (Castells, *The Rise of the Network Society*, 1996/2000)?

L'uomo postmoderno, dopo aver perduto le certezze del secolo scorso, esperisce un *everyday life* il cui significato è stravolto dall'incedere preponderante della tecnologia e dalla sovra esposizione dei/nei social media, e si fa testimone di un cambiamento di paradigma sociologico, antropologico e fisico. Il medesimo tipo di esperienza è raccontato dai protagonisti di *Black Mirror*, la serie pensata al futuro ma ispirata dalla contemporaneità creata e prodotta da Charlie Brooker per Endemol, che si muovono disorientati a seguito di piccoli o grandi eventi scatenati dalla tecnologia o dall'imperante ruolo dei mass media. Le atmosfere angoscienti narrate dalla fiction britannica costituiscono il perfetto "riflesso" disturbante, uno "specchio sul mondo o via di fuga da esso" (pag. 204), di cui Mario Tirino e Antonio Tramontana si servono per orchestrare un vocabolario distopico in grado di descrivere e interpretare questa nuova realtà non più aliena ma in sé alienante. Nel bel volume *I riflessi di Black Mirror*, da loro curato e uscito nel 2018 per i tipi di Rogas Edizioni, offrono al lettore un glossario multidisciplinare tra immaginari, culture e media dell'era digitale. Ed il *riflesso* è proprio il campo di azione di questo framework "altro", ipertecnologico, del nero dietro lo schermo ma anche del retro di ogni specchio, oltre la prolifica simbologia sociologica e non solo (da Anselm Strauss a Jean

Baudrillard, ma il riferimento è anche a Jack Lacan ed al mito di Narciso). Come ricordato nella prefazione del libro, il riflesso rappresenta un *limen* tra concreto e visionario, tra reale e illusorio, e diventa il milieu per eccellenza in (per) cui discutere della maestosità del dramma identitario dell'uomo post contemporaneo, non più a una dimensione ma sperso in mille e più pixel audiovisivi, tra innumerevoli relazioni online e pratiche comunicative mediate, nomade e monade, mutevole e frammentato.

Tirino e Tramontana riferiscono alle vicende di Black Mirror per presentare un individuo che non è più soggetto né robot, ma un *cyborg* prodotto e produttore di una condizione post/transumana in cui la dimensione tecnologica si è fatta antropologia. In questo ragionamento, la tecnologia non è più solo un "additivo biologico" ma anche principio di alterità (il non-umano), dispositivo regolatore di interazioni e relazioni, agente di significato e, perciò, struttura di costruzione identitaria. Da questo punto di vista, la serie può essere intesa come una perfetta sintesi del dramma dell'attore contemporaneo: "un'opera, non ancora conclusa, capace di tenere in sé le tensioni di un'epoca. È la rappresentazione più tangibile di una serie di angosce che accompagnano quelle azioni sempre più frequenti e, ormai, sempre più tangibili riassumibili nell'uso della tecnologia" (20-21).

Ma Black Mirror, oltre ad essere l'*oggetto* della riflessione, diventa anche il *soggetto* di pratiche e rituali contemporanei. La fiction stessa è effetto della propria serialità (una *post-serialità*), che si propaga confondendosi in una convergenza mediale (Jenkins, *Convergence Culture: Where Old and New Media Collide*, 2006), e replicandosi nelle reazioni dei suoi spettatori, perfetti interpreti della realtà che rappresenta: si dipana riproducendosi nei social networks, o "riflettendosi" negli schermi in cui viene visualizzata come in un gioco di bambole russe in cui l'identità reale, rappresentata, mediata (dello spettatore e dei protagonisti) si moltiplica nella varietà di scenari (riflessi) possibili.

Tutto il volume ruota attorno ad un elemento critico: il rapporto ambivalente tra *humanitas* e *techne*, a partire dal riconoscimento che, come argomentato da Gehlen nei suoi studi di antropologia filosofica, la relazione con la tecnologia è fondamento della realtà antropologica umana, le cui limitatezze fisiche e biologiche – e, di conseguenza, il concetto stesso di sviluppo – necessitano dell'apporto di un organo artificiale che è, appunto, la tecnica. Ciò che viene messo in discussione non è pertanto la tecnologia (digitale) in sé, ma il *come* l'elemento "protesico" della tecnologia cambi il modo in cui gli attori

sociali fanno esperienza del mondo. Poiché l'uomo esperisce e comunica attraverso un linguaggio universale fatto di miti, rappresentazioni, segni, archetipi etc., egli finisce per attribuire anche agli oggetti artificiali un universo simbolico, un *immaginario*; pertanto, l'esperienza quotidiana postmoderna rappresenta un continuo riorganizzarsi dell'immaginario attorno ad atti, dimensioni, esperienze le cui categorie sono in sé diverse, innovative perché nuove. Come scrivono gli autori, la triade meadiana (Mead, *Mind, Sense, and Society*, 1934) si può attualizzare nella formula "tecnologia, Sé e società", in cui la tecnologia è, parafrasando Simmel, quel legame armonizzante tra il singolo (l'immagine che ognuno ha di sé, ricostruita e ridefinita per mezzo del dispositivo digitale) e il complesso sociale (l'elemento interazionale e collettivo) in un quadro unitario ma potenzialmente univoco o, come afferma Calasso in *L'innominabile attuale* (2017), codificato "universalmente e onnilateralmente", senza che gli individui ne siano consapevoli.

Tirino e Tramontana cercano di rispondere alla complessità della questione interpretando il ruolo del digitale - da loro definito il re nascosto (Simmel, *Il conflitto della società moderna*, 1918/1999) della nostra epoca - attingendo parimenti dalla prospettiva sociologica dell'immaginario, dall'antropologia culturale e dalla *Media Archaeology*, formulando l'ipotesi per cui "una narrazione seriale come Black Mirror racconti proprio il momento in cui si incrinano gli immaginari sociali del Novecento, sulla spinta di tecnologie digitali che riconfigurano l'accordo intersoggettivo su cosa è reale e cosa non lo è" (172). Per dare ragione di un processo culturale che è ancora in pieno divenire lasciano spazio alla voce di diciassette autori, i cui punti di vista esprimono altrettante parole chiave; infatti, ogni capitolo approfondisce un concetto ed è un piccolo saggio a sé, come fosse uno dei tanti episodi in cui si leggono i mutamenti della società contemporanea. I capitoli sono presentati nel collettaneo secondo un ordine alfabetico, ma per il loro contenuto possono essere organizzati in tre aree che richiamano le tematiche principali sopra menzionate: 1) i capitoli Algoritmo, Corpo, Democrazia, Interazione, Tecnica, Zoo-tecnica descrivono i conflitti nell'immaginario dovuti ai cambiamenti nella relazionalità tra uomo e tecnica; 2) le conseguenze che tali contraddizioni agiscono sull'identità degli individui (i sensi e le facoltà umane) sono discussi nei capitoli Atmosfera, Esperienza, Illusione, Memoria, Morte, Paranoia, Pathos, Paura; 3) in Audience, Schermo e Serialità si pone invece l'esperienza mediale attuale in chiave fenomenologica.

Nei contributi si descrivono, tra gli altri, la funzione degli algoritmi nei rituali di *dating online*, l'autenticità delle interazioni agite anche a livello emozionale, la possibilità di ribellione al sistema e di esercitare il libero arbitrio e quindi il principio di umanesimo; tuttavia, se l'interazione è mediata, allo stesso modo la percezione del corpo e delle sue funzioni è mediata dalle sue rappresentazioni e autorappresentazioni, pur nella sporadica difesa dell'integrità individuale (umana). Un'altra questione è quella dello spazio contemporaneo, fosco, metropolitano, eppure scintillante come le vetrine dei *passages* di Benjamin: uno spazio alienante, in cui sono la tecnologia e i mass media ad imporsi come *tools* e ambiti relazionali, veicolando l'immaginario collettivo e facendosi territorio (fenomenologico) di esperienza. Ma l'affermarsi della rete come spazio di relazione evidenzia il processo di dematerializzazione del corpo prodotto dal regime dell'immagine e dalle sue dinamiche identitarie collettive fondate sui rituali digitali, che sostituiscono l'uniformità alla pluralità, modificano il significato della memoria e manifestano l'inadeguatezza del linguaggio verbale. Ne è un chiaro esempio l'affermarsi del populismo in politica, ma anche le nuove "pratiche di visione" che sono l'effetto di un radicale mutamento nei dispositivi mediatici e nelle loro strategie comunicative. In questo senso la bramosia bulimica di immagini ed il loro consumo, il rapporto tra mass media e spazio urbano (l'ambiente come schermo), ma anche il concetto di videosorveglianza post-foucaultiana per cui siamo "vigilati" da tanti agenti quanti sono gli individui della società stessa sono soltanto alcune delle riflessioni sviluppate nel volume. Ne conseguono una discussione sul ruolo degli oggetti/dispositivi tecnologici in relazione alla società e all'oggettivizzazione del Sé, ma anche sul relativo cambiamento del concetto di morte intesa come oblio, da contrapporre all'immortalità digitale. Molteplici contributi affrontano così la questione degli stati d'animo e dei sentimenti, e quindi del senso più profondo dell'individuo umano.

In definitiva, Tirino e Tramontana sono riusciti a rompere lo specchio e ad andare oltre il buio del riflesso, riconoscendo la perdita di centralità dell'uomo postmoderno a fronte di un processo di tecnologizzazione che è già in pieno atto, senza però demonizzarlo o guardare nostalgicamente al passato, ma dando ragione – anche criticamente – dell'evoluzione continua del sociale e quindi dei significati, degli immaginari, delle pratiche e delle identità dei suoi attori. Da cui deriva la considerazione che discutere in maniera dicotomica del binomio umanesimo-tecnologia risulta ormai essere obsoleto, moderno, disallineato rispetto allo spirito del tempo.

Se questo rappresenta la chiave interpretativa dell'intero volume, al termine della sua lettura resta tuttavia una (la) domanda: parafrasando Susca, "Qualcosa si è rotto, in effetti, oppure si è compiuto? Qual è la catastrofe di cui siamo tutti oggetti e soggetti – prima oggetti e poi soggetti? Resta, tra tanti frammenti di uno specchio rotto, qualcosa da ricostruire, da cui ripartire? [...] In un siffatto scenario, cosa rimane del nostro essere umani? Resta qualcosa da vivere? Che ne è della nostra esperienza?" (115). E il lasciare in certo qual modo aperta la risposta, esattamente come lo è il futuro, rappresenta il valore aggiunto del lavoro.

Daniele Gambetta

Datacrazia. Politica, cultura algoritmica e conflitti al tempo dei big data

Roma, D Editore, 2018, 360 pp.

di *Junio Aglioti Colombini**

Datacrazia è una raccolta antologica di saggi multidisciplinari che si interrogano e ci interrogano sui processi e sulle conseguenze che regolano una società in cui dati e algoritmi sono diventati dispositivo di potere e di controllo. Ricostruire gli eventi che portano al bisogno di voler stringere tra le mani un volume con l'immagine di Trump in copertina non è semplice, ma fondamentale.

La fine del XX secolo è stata caratterizzata da un improvviso sviluppo tecnologico che, a partire dagli Stati Uniti, ha investito i paesi industrializzati generando una vera e propria rivoluzione: è l'inizio di quella che l'economista Daniel Bell definisce come "società dell'informazione" (1973) riferendosi ad una struttura economica che vede nell'informazione un valore da adottare come risorsa capitalizzabile di scambio. L'avvento di internet e delle piattaforme digitali modifica ulteriormente questo panorama generando un'esponenziale crescita della quantità di dati disponibili (si comincia qui a parlare di *big data*) a causa del cambiamento radicale nelle modalità di produzione, consumo e conservazione dell'informazione.



* JUNIO AGLIOTI COLOMBINI è studente magistrale di "Informatica Umanistica" presso l'Università di Pisa.

Email: junio.aglioticolombini@gmail.com

Digitalizzazione e datificazione diventano parole chiave dei nuovi processi informativi e figure come quelle dei “*prosumer*” (Toffler, 1987) guadagnano centralità in un ambiente digitale sempre più disintermediato e accessibile. Il volume, in pratica, si configura come un lungo discorso sul destino moderno, come appunto questa dimensione appare attraverso la ricostruzione del pensiero habermasiano più maturo, quello che va dagli anni Ottanta ai giorni nostri, e del suo rapporto con tutto il panorama intellettuale di questa fase. Etica, sociologia, filosofia del linguaggio, filosofia del diritto e della politica, costituiscono gli ambiti di ricerca fondamentali che caratterizzano questa articolata e complessa analisi.

La rapida evoluzione tecnologico-strutturale è inoltre responsabile della creazione di uno spartiacque generazionale tra *Digital Immigrants*, una popolazione formatasi in un contesto storico pre-digitale, e *Digital Natives*, una fascia più giovane cresciuta in un ambiente digitale del quale padroneggia gli strumenti tecnici (Prensky, 2001); due mondi ai quali è richiesta la sfida di trovare un linguaggio comune per la trasmissione del patrimonio sociale e culturale.

È in un contesto complesso e a volte opaco che il contributo di *Datacrazia*, edito per D Editore, si inserisce e prova con disincanto a partire da scenari fantascientificamente distopici e paurosamente attuali per (ri)costruire in maniera collettiva una cultura degli algoritmi che sia strumento critico e attento per leggere i processi con cui le tecnologie sono entrate in simbiosi con le nostre vite. La scelta della forma antologica non ha dunque solo il merito di rendere stimolante la lettura ma risulta anche imprescindibile per l'ambizioso tentativo di restituire «l'overcomplessità di una tecnologia ai limiti della comprensione» [p.33] e consegnare a chi legge una visione globale e prismatica dei numerosi campi in cui *big data* e algoritmi sono entrati.

Il libro è composto da 23 saggi, suddiviso in cinque sezioni tematizzate: “Geopolitica e macroeconomia dei big data”, “La teoria è morta, evviva la teoria! – L'analisi dati nella ricerca”, “Macchine intelligenti”, “Vite datificate,” “Strategie dei corpi-macchina”. Apre il volume una prefazione di Raffaele Alberto Ventura, direttore della collana, che si interroga sulla rimediabilità di un'ormai troppo pervasivo rapporto con gli algoritmi; a chiudere un breve ma significativo glossario di supporto e approfondimento. Nonostante la divisione tematica delle sezioni, sarebbe riduttivo pensare di affrontare la lettura in maniera lineare perché i nuclei tematici fondamentali si inseguono e si intrecciano attraverso tutti i saggi offrendo un continuo cambio di fuoco che tal volta approfondisce e tal volte contestualizza temi e

questioni già trattate o da trattare. Ventiquattresimo saggio, fuori sezione, è il contributo del curatore – Daniele Gambetta – che ripercorre l'intera raccolta con un'elaborazione che raccoglie tutti i principali nuclei tematici affrontati nei vari saggi, restituendo in maniera chiara la necessità di un libro che sia strumento critico e collettivo per: «elaborare piattaforme collaborative e non estrattive, creare strumenti di indagine e inchiesta che svelino i meccanismi, spesso proprietari e oscuri, degli algoritmi che determinano le nostre vite, far emergere contraddizioni utili nel rivendicare il proprio ruolo di sfruttati diffusi rimettendo al centro la questione del reddito» (p. 31).

Datacrazia propone un percorso dichiaratamente politico che si prefigge lo smascheramento del “feticismo della neutralità della tecnologia” (Ciccotti, Cini, De Maria, Jona-Lasinio, 1976) dietro cui giacciono i «processi di controllo sociale, di valutazione come strumento disciplinare, di estrazione di valore dalla vita quotidiana» (p. 17) che trovano spazio nell'*online* ma sono mutuati da relazioni di potere e di classe ben radicate nell'*offline*, connessioni che è fondamentale individuare per mettere in campo soluzioni collettive a contraddizioni oramai sistemiche anche nei nuovi ecosistemi digitali.

Mettere in discussione il procedere della datificazione delle nostre vite passa anche attraverso un processo di comprensione e di risignificazione della parola “dato”; ne è consapevole Flavio Pintarelli il quale, nel suo contributo “Sull'etimologia della parola dato”, riparte proprio dalla contraddizione tra l'etimo latino *datum*: regalo, e i meccanismi di sfruttamento economico che invece questo “regalo” subisce. Molti sono i (pas)saggi che affrontano la capitalizzazione e la sussunzione del lavoro vivo con una lente marxista che ne osserva e analizza le caratteristiche contemporanee

In quest'ottica si sviluppa la riflessione di Andrea Fumagalli in “Per una teoria del valore-rete” che individua nei processi di elaborazione dei dati il prodotto di un “lavoro astratto” (Marx, 1975) che sfrutta il “lavoro concreto” di soggetti che, spesso inconsapevoli, introducono i propri dati di vita quotidiana in piattaforme digitali che con algoritmi di *data mining* e modelli di *business intelligence* trasformando un lavoro collettivo non salarizzato in capitale economico privato.

Fa eco Giorgio Grizziotti che, in “Big emotional data”, riprende il concetto marxista di capitalismo industriale per argomentare la tesi di un novello capitalismo cognitivo e biocognitivo – che definisce Neurocapitalismo (Grizziotti, 2016) – esercitato da grandi operatori di *platform capitalism* come Amazon o Facebook e che si muove all'interno di un ambiente – il Bioipermidia – in un cui il rapporto

sempre più simbiotico tra corpi incarnati e corpi macchinici è tale da creare nuovi dispositivi di controllo sociale.

Intrinsecamente connesso al nucleo tematico della datificazione delle vite e ad esso strettamente contiguo è dunque il tema dei nuovi meccanismi di sorveglianza e repressione politica e sociale che si assemblano nel nuovo ecosistema mediale. È il caso, come racconta Simone Picenni in “Cina e Big Data”, dei tentativi di Pechino di introdurre un sistema di “credito sociale”, un meccanismo che attraverso l'estrazione dei dati online parametrizza il grado di affidabilità delle cittadine e dei cittadini, regolandone le possibilità di accesso ai servizi di welfare e di prestito; *Nosedive*, la prima puntata della terza stagione della serie *Black Mirror*, è uno dei tanti esempi di fantascienza che trova riscontro in una reale distopia contemporanea.

Michel Foucault viene citato esplicitamente una sola volta in tutto il testo – a farlo è Raffaele Alberto Ventura nella sua prefazione, richiamando il concetto di *société panoptique* (Foucault, 1993) – ma è uno sguardo assolutamente foucaultiano quello che guida la lettura degli strumenti di potere e dei meccanismi di resistenza che possono essere proposti.

I dispositivi di controllo politico e sociale non sono infatti visti come singoli elementi che agiscono in maniera indipendente ma come: ««insieme assolutamente eterogeneo che implica discorsi, istituzioni, strutture architettoniche, decisioni regolative, leggi, misure amministrative, enunciati scientifici, proposizioni, morali e filantropiche, in breve tanto del detto che del non-detto» (Foucault, 2001).

E sono numerosi gli esempi dell'interazione di vincoli di diversa natura che, combinandosi, provocando effetti di oppressione: come leggi sul copyright che impediscono di utilizzare per il *machine learning* dei dataset contemporanei così che *Il grande Gatsby* o i drammi di *Shakespeare* diventano la risorsa su cui le intelligenze artificiali imparano, garantendo l'assimilazione di un canone letterario decisamente più bianco, più ricco e più occidentale (si veda il contributo di Signorelli, *La guerra dei bias*); o di programmi istituzionali come l'*All of Us* del *National Institutes of Health* che si prefiggono lo scopo di distribuire diecimila *FitBit* per raccogliere dati in relazione alle risposte farmacologiche, ma che nel farlo creano una banca dati di crescente valore che contiene dati biometrici strettamente individuali (si veda il contributo di Mauro Capocci, *I big data e il corpo*); o, infine, delle implementazioni di *user experience* promosse da Facebook e Google che mentre promuovono funzioni avanzate di interazione (condivisioni, *reactions*, ecc.), di fatto raffinano gli algoritmi di profilazione (Lelio

Simi, *Come i dati hanno cambiato (per sempre) il giornalismo*).

Ma è proprio dentro queste reti di potere che si situano i punti di resistenza (Foucault, 2004) che le autrici e gli autori dei saggi cercano di intercettare e valorizzare come esempi e come pratiche da cui ripartire. È il caso dello sguardo internazionale a progetti come la piattaforma *Decidim* (Manconi, *Nuovi codici della politica nel caos*) che rappresenta uno esempio virtuoso di strategia tecnopolitica il cui tentativo è quello di ribaltare le relazioni verticali di potere riappropriandosi della potenzialità degli strumenti digitali e incentivare movimenti *grassroot* che si basino su relazioni orizzontali e su principi di trasparenza delle infrastrutture informatiche e dove le sorgenti degli algoritmi non si nascondono in oligarchiche *black box*. Altro esempio di resistenza è la genesi del movimento #15M (Javier Tored Medina, *Uno sguardo tecnopolitico sui primi giorni del #15M*) che si innesta all'interno del processo di ibridazione tra mondo reale e mondo virtuale sfruttando le interconnessioni create nello spazio dei flussi (Castells, 2002) per dare corpo a mobilitazioni nello spazio dei luoghi, secondo il paradigma "*Think global, act local*" (Bauman, 2005) per cui le resistenze digitali diventano mobilitazioni locali.

Le mani che contribuiscono alla creazione di *Datacrazia* non sono quelle di chi semplicemente assiste al mutarsi di una società bensì quelle di soggetti coinvolti e travolti da una rivoluzione profonda e sentono la responsabilità di interrogarsi, come scienziate e scienziati, sul nuovo ruolo che hanno o possono avere in questa quarta rivoluzione industriale.

Opporsi ad una ricerca acriticamente *data driven* che si basa sulla dibattuta affermazione di Chris Anderson: "*correlation is enough*" (Anderson, 2008), diventa presto la conclusione di diversi saggi in cui si denuncia la fallacia di questo assunto che stabilisce che la correlazione implica anche la causalità (Eleonora Priori, *Is correlation enough? (spoiler: No!)*) e come questo postulato abbia esasperato l'ossessione di parametrizzare qualsiasi elemento, anche con risultati paradossali come la misurazione del sentimento religioso attraverso la lunghezza dei certi voti (Roberto Paura, *L'illusione della psicostoria*). È dunque un ruolo attivo di corpo intermedio che interpreta e crea correlazioni, quello che rivendica chi produce discorso scientifico; è decostruzione di una cultura algoritmica che si basa sull'assunto mendace che avere i dati è condizione sufficiente per avere garanzia di infallibilità; è denuncia di un sistema che, totalmente automatizzato, crea un'arendtiana "banalità dell'algoritmo" dagli esiti pericolosi (Massimo Airoidi, *L'output non calcolabile*).

Chi legge *Datacrazia* lo fa in maniera ciclotimica: si alternano rapidamente sensazioni di paranoica consapevolezza di essere parte di

un sistema panotticamente controllato a inaspettati sentimenti di orgoglio e appartenenza ad un genere umano che sviluppa tecnologie solidali e mette in campo strumenti per il progresso e per il benessere collettivo. È difficile accettare la pillola azzurra e ritrovarsi fuori da Matrix ma il pregio di questo libro è quello di raccogliere la sfida della complessità, accettandone anche le numerose ambivalenze, e leggerle con gli strumenti scientifici senza mai mettere in discussione il fondamentale assunto che, citando Najerotti in “Hapax Legomenon: la guerra delle parole”: «Da qui non si torna indietro: oggi produrre dati è un requisito imprescindibile per essere parte della società moderna; una società assolutamente interconnessa, in grado di abbattere i confini geografici e che si muove a velocità ben lontane dalla nostra umana comprensione».

RIFERIMENTI BIBLIOGRAFICI

- ANDERSON, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. Wired
- BAUMAN, Z. (2005). *Globalizzazione e glocalizzazione*. Roma: Armando.
- BELL, D. (1973). *The Coming of Post-Industrial Society: A Venture in Social Fore-casting*. New York: Basic Books.,
- CASTELLS, M. (2002). *La nascita della società in rete*. Milano: Università Bocconi Editore.
- CICCOTTI, G., CINI, M., DE MARIA, M., JONA-LASINIO, G. (1976). *L'Ape e l'architetto: paradigmi scientifici e materialismo storico*. Milano: Feltrinelli.
- FOUCAULT, M. (1993). *Sorvegliare e punire*. Torino: Einaudi.
- FOUCAULT, M. (2001). *Le jeu de Michel Foucault. Dits et Écrits II, 1976-1988, 299-300*. Parigi: Gallimard.
- FOUCAULT, M. (2004). *Storia della sessualità. Vol. 1: La volontà di sapere*. Milano: Feltrinelli.
- GRIZIOTTI, G. (2016). *Neurocapitalismo, mediazioni tecnologiche e linee di fuga*. Milano: Mimesis.
- MARX, K. (1975). *Il Capitale. Critica dell'economia politica*. Torino: Einaudi.
- PRENSKY, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon*, 9(5), 1-6.
- TOFFLER, A. (1987). *La terza ondata*. Sperling & Kupfer.
-

Safiya Umoja Noble

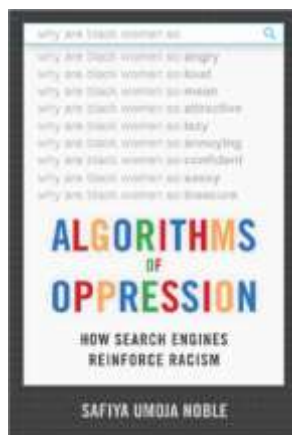
Algorithms of Oppression: How Search Engines Reinforce Racism

New York, New York University Press, 2018, 256 pp.

di Paola Imperatore*

Algorithm of Oppression. How Search Engines reinforce racism di Safiya Umoja Noble, studiosa nera presso la University of California (UCLA), è un testo che ci porta ad esplorare con vivo interesse le dinamiche sessiste e razziste all'interno del web.

Nonostante le promesse emancipatorie che il web sembrava poter realizzare, i limiti e le contraddizioni del “capitalismo digitale”, così come definito da Harvey (2005) e Fairclough (1995), stanno emergendo, e studiosi provenienti da diversi settori disciplinari hanno intrapreso un percorso di riflessione critica su tale modello. Non mancano infatti studi che prendano in esame da diverse angolazioni le implicazioni del web in termini politici, economici, sociali e culturali. Tra gli aspetti che stanno recentemente ricevendo attenzione nell'ambito accademico si può senz'altro annoverare la questione della rappresentazione delle identità subalterne nel web e dell'utilizzo talvolta strumentale delle stesse. Alcuni contributi hanno ad esempio evidenziato come piattaforme digitali quali Airbnb (Edelman, Luca, 2014), TaskRabbit (Schoenmaub, 2016), Uber (Jefte, 2015), o altre simili,



* PAOLA IMPERATORE è Dottoranda di Scienza Politica presso il Dipartimento di Scienze Politiche dell'Università di Pisa. Si interessa di movimenti sociali ambientalisti, urbani e femministi.

Email: paola.imperatore@sp.unipi.it

riproducano in modi differenti discriminazioni di genere e di razza. La richiesta di inserire dati personali su tali piattaforme accresce la salienza dell'identità, e così le donne, ritenute più dedite alla cura, saranno considerate più affidabili per chi usa piattaforme dove lo scambio avviene virtualmente, mentre le persone nere risulteranno maggiormente discriminate in termini di utenti interessati alla loro offerta. In altri casi, come quello del progetto di coworking CraftWorks, si è riscontrata una divisione del lavoro su base di genere associando le donne ad attività tradizionalmente femminili e gli uomini ad attività tradizionalmente maschiline (Schor *et al.*, 2016). Altri studi hanno tentato di inquadrare la discriminazione digitale da una prospettiva legale, mentre altri ancora hanno cercato di analizzare l'economia collaborativa/economia delle piattaforme da una prospettiva di genere (Fuster Morell, Imperatore, 2018). Quello che questi contributi hanno in comune è l'interesse ad analizzare il web attraverso una lente specifica focalizzata sulla questione delle identità storicamente oppresse.

Allo stesso tempo, l'ondata di mobilitazioni femministe che dal 2016 investe numerosi paesi, insieme all'affermarsi negli USA di movimenti per i diritti degli afroamericani, si pensi a Black Lives Matter, mostra una rinnovata volontà di mettere in discussione le relazioni di potere basate su genere e razza in tutte le sue declinazioni. In questo contesto, anche il web è finito nell'occhio del ciclone dopo una serie di casi che hanno sollevato numerose critiche sulla non-neutralità dei meccanismi di machine learning. Ne sono degli esempi software quali Amazon Rekognition, programma di riconoscimento facciale venduto alle forze dell'ordine di vari paesi e che, testato dalla American Civil Liberties Union, ha mostrato di associare con maggiore frequenza il volto di criminali con i volti di persone nere (Signorelli, 2019). O l'algoritmo COMPAS, molto utilizzato negli Stati Uniti per calcolare la probabilità di recidività di un criminale e che, come denunciato da *ProPublica*, indicherebbe i neri come più propensi a ricompiere crimini (Cossins, 2018). Il susseguirsi di questi episodi ha dunque condotto la società civile a denunciare la tendenza degli algoritmi di riprodurre le discriminazioni razziali, questione sollevata nel mese di gennaio anche da Alexandria Ocasio-Cortez, neodeputata democratica negli Stati Uniti (Kosoff, 2019).

In una fase storica caratterizzata da queste spinte, dentro e fuori l'accademia, il testo della Noble appare assolutamente necessario e utile non solo nel contesto statunitense, ma altresì in quello italiano e, più generalmente, europeo, dove la questione migratoria sta ponendo nuove problematiche relative alla rappresentazione stereotipata degli immigrati neri che nel web ha trovato ampio spazio. È in questo quadro che

Algorithm of Oppression necessita di essere collocato.

Il testo affronta in modo estremamente appassionante le implicazioni dell'intelligenza artificiale per le categorie che sono già sistematicamente marginalizzate e oppresse. Nello specifico, l'autrice utilizza una chiave di lettura "black feminist" per osservare empiricamente come razzismo e sessismo siano elementi strutturali nell'architettura del web, da cui la scelta di parlare di "algoritmo dell'oppressione". Nella prospettiva della Noble, adottare una lente intersezionale, capace di far comunicare oppressione di genere e di razza, permette di dare spazio ad una serie di interrogativi scientifici che derivano dall'esperienza diretta dei soggetti subalterni. Riprendendo l'approccio metodologico elaborato da Sandra Harding in "Feminism and Metodology" (Harding, 1987), l'autrice suggerisce di uscire dalla dominante prospettiva borghese, maschio-centrica e bianca, con cui si indagano le tecnologie, al fine di disporre di una ricerca capace di trattare il fenomeno del capitalismo digitale nella sua complessità e non solo da un punto di vista parziale. È necessario precisare che l'autrice si riferisce alle categorie di genere e razza non come concetti naturali ma come costrutti sociali, rifiutando un approccio essenzialista e binario uomo/donna, bianco/nero (Secondo la teoria essenzialista, la differenza tra i sessi è un dato naturale e imm modificabile. È la biologia dunque a dettare il comportamento dei maschi e delle femmine. Al contrario, la visione costruttivista, evidenzia come la dualità maschio-femmina, e i comportamenti associati a questi, siano prodotto di processi storici e culturali). Di conseguenza, l'utilizzo nel corso della sua analisi, della categoria "black woman", così come di altre categorie, non vuole legittimare l'esistenza della categoria di per sé, né attribuire ad essa delle specifiche caratteristiche, ma serve ad evidenziare la relazione di potere tra soggetti dominanti e soggetti subalterni.

Uno dei meriti che va riconosciuto ad *Algorithm of Oppression* è quello di riuscire a guardare alla questione di genere e di razza andando oltre all'aspetto dell'esclusione di alcuni soggetti dall'accesso alle tecnologie digitali. Nel testo, infatti, Noble identifica in primo luogo la discriminazione razzista e sessista nella presenza di un "digital divide" tra soggetti maschi e bianchi e gli "altri", ma sposta successivamente il focus sulla rappresentazione stereotipata delle identità oppresse. Il dibattito va a investire la questione dei contenuti associati nel web alle donne in generale, ai neri, alle donne nere, latine, asiatiche o alle minoranze religiose, mettendone in discussione la naturalità e la neutralità. L'analisi che l'autrice conduce si intreccia con la questione dei grandi monopoli tecnologici, come Google, considerati come una minaccia alla democrazia nella misura in cui detengono il monopolio dei flussi di informazione. È

proprio su Google, e nello specifico su Google Search, che lo studio si concentra, andando ad osservare come le discriminazioni razziali e sessiste vengono riprodotte all'interno di tale motore di ricerca. Noble tenta attraverso questo testo non solo di mostrare i processi discriminatori che si svolgono all'interno di questo spazio digitale, ma altresì di indicarne le cause e le possibili vie di uscita.

Algorithm of Oppression si struttura in sei capitoli, sebbene la riflessione centrale si sostanzia nei primi due dedicati rispettivamente al tema del controllo delle corporazioni sull'informazione e al modo in cui Google rinforza gli stereotipi maschilisti e razzisti per interesse economico. Già nelle pagine introduttive, Noble riporta un dato che emerge dalle sue prime ricerche condotte nel Settembre 2011, durante le quali, digitando su Google Search "black girls" trova nella prima pagina risultati legati a siti pornografici. La prima riflessione che il lettore si trova a sviluppare è la seguente: il collegamento tra "black girls" e siti pornografici non è l'esito delle ricerche prodotte dagli utenti stessi? Ed è esattamente a partire da questo interrogativo che il primo capitolo prende vita. Noble, pur riconoscendo il processo di ricerca come processo simbiotico che al contempo "*informs and is informed in part by users*", puntualizza che i meccanismi di ricerca, lungi dall'essere neutrali, sono frutto del "human thinking", del pensiero di soggetti portatori di specifici valori che vengono riprodotti nell'algoritmo attraverso il *deep machine learning*. Gli stessi pregiudizi che esistono nella società vengono "codificati" tramite l'algoritmo, per poi presentarsi ai nostri occhi come neutrali, frutto di un processo meccanico. In questo senso, la prima questione che Noble affronta è senz'altro orientata ad attribuire ai meccanismi di apprendimento e di ricerca nel web una natura politica piuttosto che tecnica e neutrale. I risultati delle ricerche che conduciamo non sono dunque esito di un processo oggettivo, specchio della collettività che utilizza il web, ma sono mediati da meccanismi funzionali alla ricerca del profitto. È in questo passaggio che l'autrice riprende Google e esplicita la sua natura di piattaforma pubblicitaria piuttosto che informativa. Sebbene una cospicua parte dei risultati delle ricerche condotte su Google sia esito dell'algoritmo di PageRank, ve ne sono molti altri che derivano dall'utilizzo AdWords, lo strumento di pubblicizzazione e ottimizzazione dei prodotti di Google che consente, a chi investe economicamente nelle inserzioni, di vedere i propri siti privilegiati nel sistema di ranking. Gli "advertisers" selezionano una serie di parole chiave che ritengono descrittive del prodotto pubblicizzato, gli utenti, d'altra parte, quando effettueranno una ricerca otterranno come primi risultati quelli che hanno ricevuto maggiore pubblicità. Google, come

mediatore tra l'inserzionista (Google customer) e l'utente (Google consumer), guadagnerà per ogni click ricevuto (cost per click, CPC). Questo meccanismo finisce col privilegiare gli interessi di quelle aziende che potranno investire maggiormente nelle inserzioni, a scapito di altri soggetti, come Noble dimostra efficacemente nel secondo capitolo, intitolato "Searching for Black Girls". In questa parte, l'autrice corre avanti e indietro nel tempo, riprendendo diversi esempi volti a dimostrare la costante presenza di un sistema oppressivo basato sul genere e la razza, che la moderna tecnologia riproduce e rinforza.

Dalle mostre imperiali di fine Ottocento, durante le quali gli stati colonialisti sfoggiavano il loro impero composto non solo di territori ma anche di umani messi in mostra a riprova di una differenza razziale e della superiorità di quelle bianche su quella nera, sino a eventi più recenti come il Sweden's World Art Day, tenutosi nel 2012, durante il quale una grottesca installazione è stata costruita al fine di denunciare la pratica della mutilazione genitale femminile, Noble evidenzia il permanere nel tempo di una visione stereotipata delle donne nere a uso e consumo di un pubblico bianco. Prendendo ad esempio i risultati correlati alla ricerca di "black girls", "asian girls" e "latin girls" su Google Search, Noble mostra la ricorrente associazione di queste parole con "hot", "sugar", "pussy", "sex", "porn star" ed altre che connotano in senso sessuale la ricerca. I primi siti ad apparire sono infatti legati a siti pornografici. L'autrice parla di "pornification" delle donne nere, così come di quelle latine o asiatiche, rappresentate nel web, coerentemente a quanto avviene in contesti offline, come oggetti sessuali. La Noble imputa questi risultati alla natura *profit-oriented* di Google, che preferisce dare maggiore visibilità ai siti pubblicitari dai quali ricava dei guadagni economici. Infatti, mentre sono le industrie pornografiche ad associare black/asian/latin women con termini sessualizzanti attraverso il processo di Search Engine Optimization (SEO) che permette di ottimizzare l'associazione tra termini, è Google che successivamente privilegia determinati siti, a scapito dei soggetti subalterni che vengono oggettificati e stereotipati anche nel web. La stessa ricerca viene presentata nel capitolo con altre parole, come "jews", "black teenagers", "white teenagers", "beauty", "doctor", "nurse", "professional hairstyle for work", "unprofessional hairstyle for work", producendo risultati che confermano la rappresentazione discriminante su base razziale. Commentando questi risultati, Noble riabilita il concetto di razzializzazione, processo nel quale si definisce una gerarchia delle razze che mantiene la bianchezza all'apice sociale, politico ed economico, e lo contestualizza nell'ambito digitale parlando di

technological racialization intesa come particolare forma di oppressione dell'algoritmo derivante dalle pratiche di ricerca per fini commerciali.

Nei capitoli seguenti il testo affronta, da un lato, ulteriori meccanismi che operano in senso discriminatorio verso le categorie marginalizzate, dai sistemi di classificazione delle informazioni alla detenzione temporalmente illimitata che Google ha dei dati degli utenti, dall'altro invece propone delle forme di fuoriuscita da tali meccanismi. Concentrandosi in particolare su quest'ultime, la Noble interpella a più riprese la comunità scientifica. Un primo ambito su cui agire è, per l'autrice, legato ai tradizionali corsi di studi dedicati al "technology design", nei quali sono assenti approfondimenti relativi all'analisi di genere, alla storia degli afroamericani o di altri soggetti subalterni. L'assenza di un expertise capace di garantire la dovuta attenzione alle rappresentazioni delle identità sul web rappresenta secondo la Noble un primo gap da colmare. Inoltre, è necessario sviluppare un filone di "black feminist technology studies" fondato su una chiave epistemologica capace di indagare le identità genderizzate e razzializzate nel web. Se da una parte, quindi, l'autrice suggerisce di investire nella formazione di una comunità accademica capace di approcciarsi alla discriminazione tecnologica con tutti gli strumenti necessari, dall'altra vi sono altri elementi che vanno messi in discussione affinché tale oppressione non si protragga. In particolare, la Noble ritiene centrale la questione dei monopoli tecnologici che, a suo avviso, devono essere infranti e regolamentati. Il monopolio dell'informazione è sempre più concentrato in soggetti privati/commerciali: questo conduce ad una progressiva erosione della qualità dell'informazione. In alternativa, l'autrice propone un modello di ricerca pubblica/non-commerciale libera dall'imperativo del profitto, nel quale l'informazione sia qualitativamente superiore e nel quale le rappresentazioni delle varie identità non siano discriminanti. In questo processo, le politiche pubbliche che disciplinano il web possono assumere un ruolo rilevante.

Il testo si chiude, acutamente, con il seguente interrogativo: "where are black girls now?". Dopo aver sollevato in precedenti articoli la problematica, Noble replica la ricerca "black girls" nel 2016, ottenendo risultati differenti da quelli del 2011: nessuno dei risultati è collegato a siti pornografici. Il miglior modo per dimostrare che il linguaggio e i meccanismi tecnologici non sono neutrali e che, al tempo stesso, possiamo agire su di essi, cambiarli e ridefinirli in modo tale da rendere il web uno spazio libero dalla discriminazione.

RIFERIMENTI BIBLIOGRAFICI

- COSSINS, D. (2018). Discriminating algorithm: 5 times AI showed prejudice. *NewScientist*, April, April 12.
- EDELMAN, B., LUCA, M. (2014). *Digital Discrimination: The Case of Airbnb.com*. Harvard Business School, NOM Unit, WP n. 14-054.
- FAIRCLOUGH, N. (1995). *Critical Discourse Analysis*. London: Longman
- FUSTER MORELL, M., IMPERATORE, P. (2018). Inclusion and discrimination at the platform economy: A gender focus. In M. Fuster Morell (ed.), *Sharing Cities. A worldwide cities overview on platform economy policies with a focus on Barcelona* (pp. 159-168). Barcelona: Editorial UOC.
- HARDING, S. (1987). *Feminism and Metodology*. Buckingham, UK: Open University Press.
- HARVEY, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.
- JEFFE, S. (2015). Uber isn't driving gender equality: Why its new hiring scheme is a road to nowhere. *Salon*, March 28.
- KOSOFF, M. (2019). Alexandria Ocasio-Cortez Says Algorithms Can Be Racist. Here's Why She's Right. *LiveScience*, January 29.
- SCHOENMAUB, N. (2016). Gender and the Sharing Economy. *Fordham Urban Law Journal*, 43(4), 1023-1070.
- SCHOR J.B. et al. (2016). Paradoxes of openness and distinction in sharing economy. *Poetics*, 54, 66-81.
- SIGNORELLI, D.A. (2019). Alexandria Ocasio-Cortez ha ragione: gli algoritmi sono razzisti. *Wired.it*, gennaio 26.
-

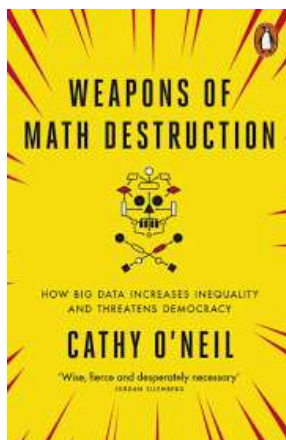
Cathy O'Neil

Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

New York, Broadway Books, 2016, 272 pp.

di *Davide Beraldo**

Scritto prima che lo scandalo Cambridge Analytica fosse esposto su vasta scala, questo libro affronta una tematica la cui importanza trova sempre maggiore riconoscimento in varie sfere della società: gli effetti indesiderati della crescente mediazione algoritmica dei processi sociali. Cathy O'Neil riporta numerosi e ben dettagliati esempi, con l'intento di denunciare un mondo che ha avuto modo di conoscere professionalmente in prima persona durante la sua carriera di *data scientist*. Facendo leva su un gioco di parole abbastanza efficace, il titolo mette in guardia dagli effetti potenzialmente devastanti di questi strumenti matematici che promettono di salvare il mondo.



I capitoli sono suddivisi tematicamente, ciascuno focalizzato su esempi tratti da uno specifico settore sociale. La lettrice, tramite uno stile di narrazione accessibile, dettagliato e persuasivo, è accompagnata attraverso una lunga carrellata di ingiustizie perpetrate a mezzo di big data e imperscrutabili modelli matematici. Accedere al college, ottenere o perdere un lavoro, essere assegnati a determinati turni, richiedere un mutuo o un prestito, interfacciarsi con il sistema giudiziario, decidere per

* DAVIDE BERALDO è ricercatore postdoc presso il Dipartimento di Media Studies - University of Amsterdam. Si interessa degli effetti dell'algorithmic mediation e di dataactivism dal punto di vista dei movimenti sociali.

Email: d.beraldo@uva.nl

chi votare: ciascun capitolo illustra come pressoché ogni sfera e fase dell'esperienza personale di un comune cittadino è al giorno d'oggi soggetta a delle decisioni (solo apparentemente) non imputabili a soggetti umani, con conseguenze potenzialmente disastrose.

Uno dei punti di forza di "Weapons of Math Destruction" è certamente quello di evitare una banale demonizzazione a 360 gradi di quelli che, l'autrice sottolinea in più passaggi, possono essere validissimi contributi in supporto delle più nobili cause - e che a volte vengono pensati come tali, ma finiscono col tradire le migliori intenzioni a causa di una serie di fattori sistematici. Dopotutto, modellare e, quindi, in una certa misura distorcere la realtà è inevitabile al fine di interfacciarsi con la complessità del mondo. Per qualificarsi come una *weapon of math destruction*, piuttosto, un modello matematico / algoritmo / procedura decisionale automatizzata deve esibire generalmente tre proprietà, come viene spiegato nel primo capitolo ed esemplificato nei successivi: opacità, scalabilità, iniquità.

Il primo rappresenta forse l'aspetto più importante nel qualificare un'arma di distruzione a mezzo di matematica come tale: la mancanza di trasparenza degli algoritmi che regolano la nostra quotidianità è ciò che più chiaramente rende lo scenario della loro applicazione capillare estremamente preoccupante per una società che si vorrebbe aperta e democratica. Questo vale sia a livello individuale che a livello aggregato. Perdere ingiustamente il proprio posto di lavoro per via di un punteggio generato sulla base di test la cui logica non può essere questionata, perché mantenuta nascosta o troppo complessa, è un sopruso ancora più odioso per via della difficoltà di mettere in questione una decisione presentata come oggettiva. A livello macro, avere a che fare con modelli protetti da segreto aziendale, o troppo complessi per essere compresi anche dai loro stessi creatori, significa evidentemente abdicare a qualsiasi velleità di sovranità democratica, specie quando questi modelli possono (potenzialmente) spostare un numero consistente di voti o condannare migliaia di persone alla miseria.

Meno scontato, ma egualmente pernicioso, il secondo aspetto riguarda la scala alla quale le *weapons of math destruction* agiscono. Prendere decisioni sulla base di modelli potenzialmente fallaci è parte del funzionamento fisiologico di un qualsiasi essere senziente. Tuttavia, per proseguire con la metafora militare, laddove le fallacie logiche umane hanno generalmente un raggio di distruzione limitato, le fallacie algoritmiche possono raggiungere una magnitudine tale da trasformarsi in ordigni nucleari.

Quanto all'iniquità, si tratta di una componente ovvia, ma importante

da sottolineare per ribadire che non è la società dei dati e degli algoritmi di per sé a essere messa sotto accusa dall'autrice. Bensi, sono l'uso contingente - deliberato o meno - che industria e politica spesso ne fanno, e determinate condizioni contestuali, a creare un doveroso allarme.

Uno degli aspetti più interessanti, sottolineato dall'autrice in più passaggi, è quello dei *feedback loop* innescati dall'interazione tra il processo di addestramento di questi modelli, le loro predizioni e la realtà sociale. Se uno strumento di *predictive policing* categorizza un certo quartiere come "ad alto rischio", più risorse di polizia verranno investite nel sorvegliare quel quartiere, andando così a gonfiare il tasso di criminalità registrato, finendo per confermare la predizione iniziale. Alcuni algoritmi, specie quelli di *machine learning* che hanno la pretesa di correggere gradualmente le proprie predizioni inizialmente inaccurate, "imparando dall'esperienza", sono soggetti a trasformarsi in "dispositivi tautologici", che contribuiscono alla definizione stessa delle condizioni alle quali il loro "regime di verità" opera. In altre parole: le *weapons of math destruction*, a differenza di modelli propriamente scientifici, non imparano dai propri errori. Laddove sulla carta un modello inizialmente fallace potrebbe essere calibrato o auto-calibrarsi per aumentare la propria accuratezza, molti di questi modelli non sono pensati per identificare i propri errori, ma per nasconderli dietro un manto di presunta scientificità. I *data set* con cui sono nutriti possono essere incompleti; le relazioni tra variabili che modellano possono essere spurie; i feedback che raccolgono per calibrare i propri parametri possono essere distorti. E le persone che licenziano, imprigionano, escludono, ecc. possono essere *outlier*, vittime di un giudice di cui non conoscono nemmeno il volto.

Le categorie sociali affette da queste armi, come è illustrato da diversi esempi, sono nella maggioranza dei casi categorie già marginali: lavoratori ricattabili, famiglie povere, minoranze etniche. L'ingiustizia algoritmica penalizza in misura sproporzionata le fasce più deboli della società, riproducendo e fortificando disuguaglianze strutturali già esistenti.

La missione degli algoritmi è presentata spesso come quella di eliminare pregiudizi umani che affliggono i processi decisionali. Si tratta di una missione nobile, considerando i molti e ben documentati bias sistematici a cui il nostro sistema cognitivo è soggetto. Tuttavia, questo bias, fa notare correttamente l'autrice, è spesso solo camuffato da un modello che si presenta oggettivo, ma che produce vinti e vincitori, magari come conseguenza di una fallacia statistica difficile da individuare. Se ad un giudice è costituzionalmente proibito discriminare un imputato sulla base della razza, come interpretare decisioni relative alla

custodia cautelare prese sulla base di modelli di recidività, che magari escludono (direttamente) la variabile razza nel calcolo del punteggio, ma includono una serie di variabili con essa altamente correlate? In casi come questo, un modello che si presenta come soluzione agli inevitabili pregiudizi umani finisce per contribuire a reintrodurre quegli stessi pregiudizi, mascherandoli in valutazioni oggettive attraverso strati imperscrutabili di elaborazioni statistiche basate su premesse potenzialmente errate – e, fattore cruciale, non trasparenti.

Quest'ultima considerazione punta a un aspetto importante: il tema dell'autorità ideologica esercitata da modelli statistici e formalismo matematico, il cui rigore scientifico può facilmente essere compromesso da *buzz words* abusate dal marketing, come *big data*, *machine learning* o intelligenza artificiale. Così come la frenologia prometteva in passato di stabilire i tratti caratteriali di una persona dalla misurazione del cranio, sostiene O'Neill, così la *data science* contemporanea viene spesso abusata per trarre conclusioni che non le pertengono, quali inferire la pericolosità sociale di una persona incrociando una mole sufficiente di dati di dubbia qualità.

Nelle conclusioni viene sottolineata una serie di problematiche ulteriori e accenna ai percorsi che potrebbero migliorare la situazione. In primo luogo, il problema è di tipo sistemico e non contingente: le *weapons of math destruction* fanno parte di un'ecologia, interagiscono tra loro e generano complessi schemi di retroazione (si potrebbe aggiungere: hanno una dimensione ideologica molto forte); per affrontarle, di conseguenza, non è possibile seguire un approccio selettivo, sopprimendone una per volta. Inoltre, non è possibile fare affidamento sulle proprietà salvifiche del libero mercato: non esistono nel *business* incentivi per la loro riconversione in strumenti di miglioramento della società; piuttosto si può dire con certezza come valga l'esatto contrario. Tra le strade da percorrere che individua l'autrice spicca quella di un "giuramento ipocratico" che i *data scientists* dovrebbero essere chiamati a sottoscrivere in virtù delle grandi responsabilità che il loro lavoro comporta. Ovviamente, un intervento legislativo e la possibilità di condurre forme di *algorithmic auditing* rappresentano condizioni di intervento imprescindibili, assieme alla condizione forse più difficile da accettare nel contesto economico-culturale contemporaneo: l'idea di sacrificare un po' di efficienza e (presunta) accuratezza in nome della giustizia sociale e dell'equità.

Da un lato, i maggiori punti di forza di questo testo sono certamente vastità e dettaglio degli esempi riportati, insieme all'accortezza di non cadere nella trappola di un tecno-determinismo di carattere apocalittico,

ponendo l'accento sulle condizioni alle quali certi processi di quantificazione dei fenomeni e automazione dei processi decisionali divengono armi nefaste. D'altro lato, alcuni spunti interessanti, sia pur richiamati più volte, vengono presentati in forma di semplice accenno piuttosto che di trattazione sistematica, quest'ultima la quale avrebbe potuto conferire un maggiore spessore teorico al testo.

Se modelli matematici e algoritmi possono essere armi, ci si potrebbe chiedere, chi le impiega e, soprattutto, perché? L'autrice richiama più volte il ruolo che il primato del profitto e dell'efficienza a tutti i costi giocano nella diffusione di questo fenomeno. Inoltre, molti degli esempi mostrano come l'effetto netto delle *weapons of math destruction* sia quello di rinforzare schemi di esclusione e strutture di disuguaglianza già esistenti, piuttosto che di generarne di nuovi. Tutto ciò dovrebbe suggerire che una critica alla società degli algoritmi non possa prescindere da una sua articolazione con una più ampia critica del capitalismo informazionale (e non).

Anello di congiunzione tra le due, anch'esso sottostimato nell'economia complessiva del testo, è infatti il momento di produzione ed estrazione di quei dati che vanno ad alimentare le *weapons of math destruction*. Per più che legittima scelta tematica, l'autrice si focalizza sul momento di elaborazione e valorizzazione dei cosiddetti *big data*; tuttavia, si potrebbe sostenere che una critica sistematica alla società degli algoritmi debba passare per forza di cose dalla tematica della produzione dei *data subjects* e della espropriazione dei dati ad opera del circuito del capitalismo della sorveglianza e delle piattaforme.

Weapons of Math Destruction, per concludere, è una ottima risorsa didattica e pedagogica per introdurre la questione dei problemi legati alla mediazione algoritmica della società, con il *plus* ulteriore di essere una critica proveniente da una insider nel mondo della *data science*, piuttosto che dal solito scienziato sociale. Ad un'audience accademica potrebbe mancare un coinvolgimento teorico più sistematico, specie per quanto riguarda i temi del contesto storico-sociale della società degli algoritmi. Tuttavia, un testo dettagliato, puntuale ed accessibile come questo rappresenta un grande contributo, dato che generare maggiore consapevolezza diffusa riguardo i rischi delle società dataficate è una precondizione imprescindibile per agire su di esse.

John Cheney-Lippold

We Are Data: Algorithms and The Making of Our Digital Selves

New York, New York University Press, 2017, 320 pp.

di *Letizia Chiappini**

John Cheney-Lippold in “We Are Data” descrive un mondo in cui informazioni e relazioni sociali vengono gestite, mediate e definite da algoritmi, capaci a loro volta di interpretare e influenzare i nostri comportamenti. Affronta la spinosa tematica della regolazione degli algoritmi e come prepararsi al cataclisma che si abatterà sulla nostra società, se non saremo in grado di proteggere i nostri dati. Per descrivere come il nostro *self* digitale sia mediato dalle tecnologie, l'autore propone un'attenta analisi osservando come gli algoritmi selezionano e classificano gli individui in base al genere, età, consumo culturale e così via. Le informazioni che divulghiamo volontariamente a fronte dell'utilizzo di una qualsiasi *app* vengono a loro volta mediate da *cookies* (piccoli file di testo che si trovano sul nostro computer e trasmettono informazioni sulla nostra attività di navigazione su reti pubblicitarie). Il fatto che ora ogni sito ci chieda se acconsentiamo all'utilizzo dei *cookies* non significa che siamo protetti da coloro che ne faranno uso finale per scopi commerciali.



Tale appropriazione surrettizia avviene regolarmente quando esprimiamo le nostre preferenze sui media, sia quando acquistiamo un

* LETIZIA CHIAPPINI è sociologa urbana, dottoranda all'Università degli Studi di Milano-Bicocca e Università di Amsterdam - doppia laurea. Ha una cattedra di ricerca e metodo all'Università di Science Applicate di Utrecht.

Email: L.Chiappini@uva.nl

prodotto su Amazon sia quando ci prepariamo al nostro appuntamento su OkCupid o Tinder. Nel primo caso si esprime una preferenza in base al nostro consumo di prodotti e *brand*, mentre nel secondo caso abbiamo condiviso dati sul nostro gusto personale nello scegliere un partner per un appuntamento. In questa rappresentazione digitale del nostro *self*, possiamo apparire come “bisessuale bianca” oppure “gay conservatore, o ancora “bianco hippie”. Anche nel caso dell’applicazioni per il ciclo mestruale si può intuire come questi dati vengano acquisiti e rivenduti a scopo di lucro. Il lavoro di queste *app* va ben oltre quello degli strumenti della cura del sé per i periodi di monitoraggio. Il monitoraggio dei cicli mestruali si è anche dimostrato un promettente modello di business per startup e aziende che lavorano con la logica del *datasucker*. Il nostro ciclo mestruale è prezioso, noi siamo dati, i dati sono profitto non per noi ma per le aziende che se ne impossessano.

Le pubblicità che ritroviamo nei nostri *feed* personali si attengono a classificazioni che gli algoritmi hanno prodotto di fronte alle nostre attività digitali. Non si tratta solo di scopi commerciali, talvolta si categorizzano aspetti ben più sensibili. Nel libro viene trattato il caso del governo degli Stati Uniti che identifica gli utenti di Internet come “cittadini” o “stranieri” basandosi sulla loro attività online, utilizzando i criteri conformi a quelli della NSA (National Security Agency). Secondo il calcolo algoritmico della NSA, comunicare con account esteri su Twitter, in una lingua diversa dall’inglese, e avere una lista di contatti email che include indirizzi di persone che vivono fuori dagli Stati Uniti aumenta il grado di “estraneità” dell’utente e dunque del cittadino.

L’autore chiarisce con precisione che non si tratta solo di “*profiling*” quando si parla di algoritmi, l’aspetto più critico è come questi ultimi mediano la nostra esperienza vissuta attraverso un processo chiamato “*modulation*”. In questo passaggio si riconosce lo sforzo di Cheney-Lippold di proporre una versione *soft* di un’analisi biopolitica alla Foucault: il *profiling* e la *modulation* governano la produzione e la selezione di informazioni oltre alla definizione di chi consuma tali contenuti. Pertanto, se un uomo bianco etero visita il sito di un giornale come il Guardian, il suo profilo aggregato di informazioni demografiche diviene ancora più bianco e più maschile. Non solo per l’effetto conosciuto come “*echo-chamber*”, in cui gli utenti di internet vengono sottoposti in modo esponenziale ai media e ai conoscenti che condividono le stesse notizie e dunque consolidano le loro opinioni, ma per ciò che Foucault definisce come “topografia discorsiva” della società. In breve, il processo che modula il discorso pubblico attraverso i media secondo prassi e retoriche dell’epoca attuale, e delimitando in tal modo ciò che può

essere detto e compreso.

La questione della *privacy* diviene cruciale nell'ultimo capitolo del libro. Se nei capitoli iniziali, l'autore ragiona a un livello più astratto di categorizzazione, controllo e soggettività, nell'ultima parte si nota lo sforzo di tradurre tali questioni in un risvolto pratico. Che fare dunque di fronte a due fenomeni come la mercificazione dei nostri dati personali e la costruzione di un nostro *self* digitale, un *avatar* che rappresenta in modo computazionale le nostre preferenze di consumo e il nostro stato di salute? L'anonimato e la "pratica dell'offuscamento" proposta da Helen Nissenbaum e ripresa da Cheney-Lippold sono due aspetti che ci fanno sperare in un futuro più roseo e meno tragico di quello descritto nella prima parte del libro. In particolare, si discute di estensioni dei *browser* (come ad esempio, TrackMeNot) e di *browser* come Tor, il quale, grazie a dei plugin e una rete di *proxy*, consente di oscurare il traffico internet impedendo il tracciamento e l'identificazione dell'utente.

Nelle conclusioni il lettore viene condotto verso dibattiti di tipo giuridico e legislativo che riguardano l'effettiva regolazione degli algoritmi e dei giganti delle piattaforme come Facebook e Google, ma anche i nuovi "unicorni" della *sharing economy*, come Airbnb e Uber. Sebbene alcuni autori protagonisti del dibattito attuale sulla politica delle piattaforme e sulla regolazione degli algoritmi, come la teorica Tiziana Terranova e lo studioso dei media Joseph Turow, siano citati nel libro, risultano poco sviluppati il dibattito e la connessione diretta con la struttura dell'opera. Il resoconto della legge statunitense e lo stato dell'arte delle misure UE per garantire la protezione dei dati e la *privacy* proposta dall'autore, appare incompleta al lettore e non dipana l'intricata questione della regolazione degli algoritmi.

Praticare una forma di resistenza verso gli stessi algoritmi che generano il nostro *self* digitale per proteggere non solo i nostri dati personali ma le nostre limitate risorse di attenzione, è ciò a cui dobbiamo ambire. Quello che "WeAreData" non suggerisce è una sorta di risposta collettiva o una via d'uscita dalle ineluttabili azioni operate dagli algoritmi. Tali azioni sono a loro volta determinate dall'operato di individui. La protezione della *privacy* descritta nel libro appare come una risposta individuale a un regime neoliberale esacerbato dalla foga e dalla rapidità dell'economia digitale, inserito in una congiuntura storica-economica capitalistica in cui le piattaforme digitali funzionano come infrastrutture sociali e facilitano la diffusione dei dati in scala globale. La risposta del libro si pone in una vena post-ideologica in cui la protezione della *privacy* diviene una sorta di illusione sociale individuale invece che una chiamata alle armi collettiva. "WeAreData" non suona come lo slogan di Anonymous: "Noi siamo Anonymous. Noi siamo legione. Noi

non perdoniamo. Noi non dimentichiamo. Aspettateci!”, nemmeno come “We Are Family” (Sister Sledge), neppure come “We Are the World” (Micheal Jackson), e neanche “We Are the Champions” (Queens). Il titolo ci ricorda che è un fatto tragico senza anima e senza colore.

Numero chiuso il 30 marzo 2019

2018/2 (aprile-giugno):

1. ILARIA IANNUZZI, L'ebraismo nella formazione dello spirito capitalistico. Un excursus tra le opere di Werner Sombart;
2. NICOLO' PENNUCCI, Gramsci e Bourdieu sul problema dello Stato. Dalla teoria della dominazione alla sociologia sto-rica;
3. ROSSELLA REGA, ROBERTA BRACCIALE, La self-personalization dei leader politici su Twitter. Tra professionalizzazione e intimizzazione;
4. STEFANO SACCHETTI, Il mondo allo specchio. La seconda modernità nel cinema di Gabriele Salvatores;
5. GIULIA PRATELLI, La musica come strumento per osservare il mutamento sociale. Dylan, Mozart, Mahler e Toscanini;
6. LUCA CORCHIA, Sugli inizi dell'interpretazione sociologica del rock. Alla ricerca di un nuovo canone estetico;
7. LETIZIA MATERASSI, Social media e comunicazione della salute, di Alessandro Lovari.

2018/3 (luglio-settembre):

1. RICARDO A. DELLO BUONO, Social Constructionism in Decline. A "Natural History" of a Paradigmatic Crisis;
2. MAURO LENCI, L'Occidente, l'altro e le società multiculturali;
3. ANDREA BORGHINI, Il progetto dei Poli universitari penitenziari tra filantropia e istituzionalizzazione;
4. EMILIANA MANGONE, Cultural Traumas. The Earthquake in Italy: A Case Study;
5. MARIA MATTURRO, MASSIMO SANTORO, Madre di cuore e non di pancia. Uno studio empirico sulle risonanze emotive della donna che si accinge al percorso adottivo;
6. PAULINA SABUGAL, Amore e identità. Il caso dell'immigrazione messicana in Italia;
7. FRANCESCO GIACOMANTONIO, Destino moderno. Jürgen Habermas. Il pensiero e la critica, di Antonio De Simone.
8. VINCENZO MELE, Critica della folla, di Sabina Curti.

2018/4 (ottobre-dicembre):

1. ENRICO CAMPO, ANTONIO MARTELLA, LUCA CICCARESE, Gli algoritmi come costruzione sociale. Neutralità, potere e opacità;
2. MASSIMO AIROLDI, DANIELE GAMBETTA, Sul mito della neutralità algoritmica;
3. CHIARA VISENTIN, Il potere razionale degli algoritmi tra burocrazia e nuovi idealtipi;
4. MATTIA GALEOTTI, Discriminazione e algoritmi;
5. BIAGIO ARAGONA, CRISTIANO FELACO, La costruzione socio-tecnica degli algoritmi;
6. ANIELLO LAMPO, MICHELE MANCARELLA, ANGELO PIGA, La (non) neutralità della scienza e degli algoritmi;
8. LUCA SERAFINI, Oltre le bolle dei filtri e le tribù online;
9. COSTANTINO CARUGNO, TOMMASO RADICIONI, Echo chambers e polarizzazione;
10. IRENE PSAROUDAKIS, Mario Tirino, Antonio Tramontana (2018), I riflessi di «Black Mirror»;
11. JUNIO AGLIOTI COLOMBINI, Daniele Gambetta (2018), Datacrasia;
12. PAOLA IMPERATORE, Safiya Umoja Noble (2018), Algorithms of Oppression;
13. DAVIDE BERALDO, Cathy O'Neil (2016), Weapons of Math Destruction;
14. LETIZIA CHIAPPINI, John Cheney-Lippold (2017), We Are Data.