

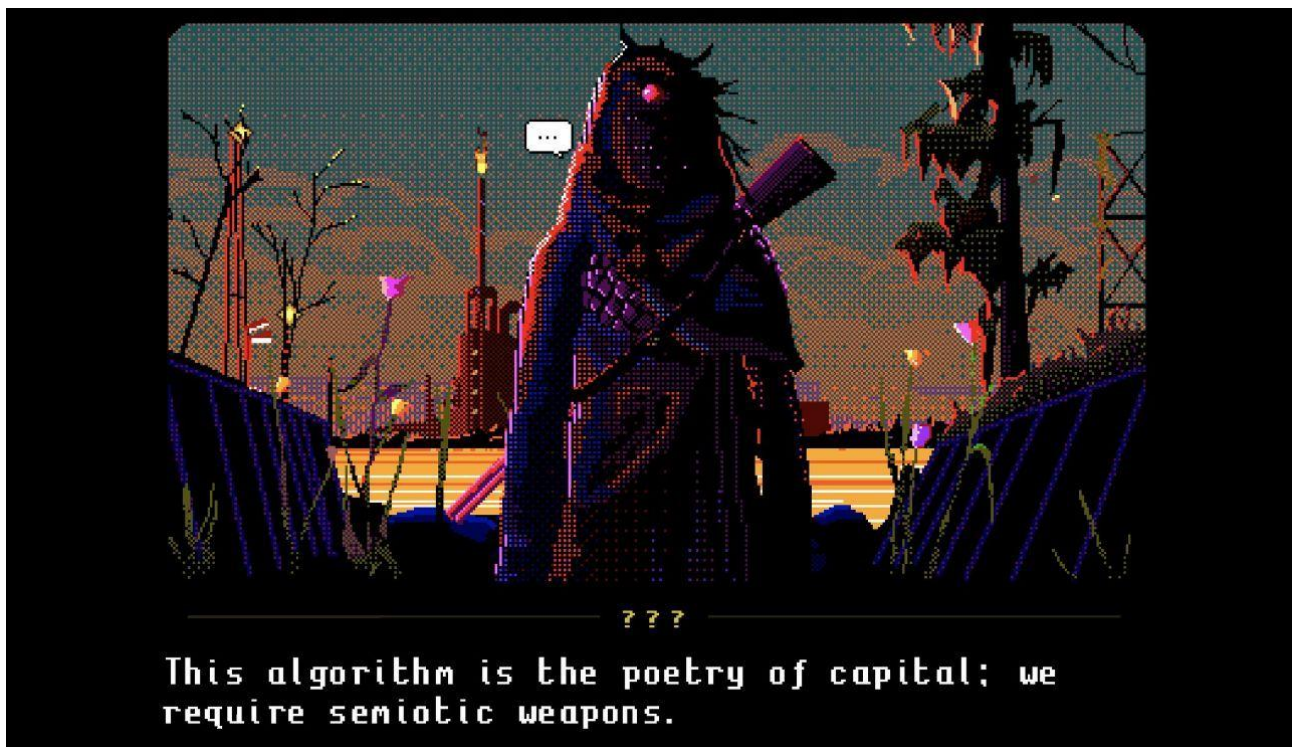
# TRex discorso – versione 0.5 – the ships they're all on fire

Daniele, Claudio, Giulia per Tracking Exposed

11 gennaio 2021

## Indice

Introduzione . . . . .	2
Internet e Big Data . . . . .	3
Algoritmica . . . . .	4
Influenza politica . . . . .	5
<b>Nascita di Tracking Exposed</b>	<b>5</b>
Etica dell'informazione . . . . .	6
Diritti . . . . .	7
Sulle spalle dei giganti . . . . .	7
<b>Gli Algoritmi di personalizzazione</b>	<b>9</b>
<b>TRex: Tracking Exposed</b>	<b>10</b>
<b>Metodo di Tracking Exposed</b>	<b>10</b>
Il proprio Algoritmo . . . . .	12
Scoperte esperienziali . . . . .	13
<b>Proposta</b>	<b>13</b>
<b>Note conclusive</b>	<b>14</b>



## Introduzione

Vorremmo introdurre con parole semplici, o meglio: con parole naturali, ciò di cui andremo a parlare tra poco. Del fatto che Internet possa condizionare il nostro comportamento e l'orientamento politico. E che lo possa fare con intenzione, semplicemente decidendo a quali informazioni esporci. Di come avvenga la scelta delle informazioni e naturalmente di come sopravvivere.

Non è corretto parlare di Internet in senso generale, il protocollo TCP/IP non è l'oggetto del nostro discorso. Parliamo dell'internet dei Social media, che si manifestano come piattaforme oramai indistinguibili, data la loro preponderanza, dalla grande rete.

I social non sono più l'agorà, la piazza di discussione libera che si ipotizzava negli anni 90, ma sono un *luogo* di esposizione a enorme quantità di informazione. Non informazione causale, ma con un obiettivo. Potremmo dire che l'esposizione ai media sociali polarizza il discorso e radicalizza verso opinioni non conciliabili tra loro. In una frase semplice: divide la popolazione.

Avete visto anche voi il tizio con l'elmo e le corna nel Campidoglio americano che si faceva i selfie? Bene, questo non sarebbe avvenuto senza i social media e di questo evento recente siete stati testimoni.

Quando parliamo di Social non parliamo di un concetto astratto di socialità in rete, ma di concrete piattaforme, di alcune note aziende. Quando parliamo di interazione non parliamo di azione creativa e consensuale, ma di una azione guidata. Quella che ognuno di noi chiama: "la rete" è in realtà un circoscrit-

to giardinetto che qui chiameremo “la bolla”. Le informazioni che circolano nella bolla sono decise da algoritmi costruiti dall’azienda che ha fabbricato la piattaforma.

Come funzionano questi algoritmi? Sono efficaci? Per cosa? Non si sa, perché sono segreti come la ricetta della Coca-cola.

Nel discorso che seguirà, proveremo a introdurre l’idea che l’attuale dimensione del distanziamento sociale aumenta di frequenza e di importanza la comunicazione tramite Internet, e di conseguenza anche il suo impatto. La cosa aggiunge urgenza a questo discorso. Presenteremo il progetto TRex, Tracking Exposed, che si occupa di cercare evidenze di come gli algoritmi usati dalle piattaforme social, tracciano e utilizzano i dati che ricavano dai loro utenti.

Se a quanto pare oggi non si può collettivamente sfuggire allo: “stare sui social”, ma pare anche che non si è immuni alla propaganda, proveremo a raccontare cosa possiamo fare.

Il nostro oggetto di studio sono gli algoritmi usati dalle grandi piattaforme, quali Facebook o Youtube. Una volta capito e accettato che queste piattaforme hanno il potere di influenzare il comportamento, bisognerà pur porsi la domanda con che criterio lo facciano e di come evitare che lo facciano malamente.

Nel testo che segue si tenterà di spiegare perché gli algoritmi siano rilevanti per la società, perché non debbano rimanere segreti, perché le grandi Imprese che li usano non possano regolamentarsi da sole e qual è l’importanza del lavoro di Tracking Exposed, progetto che attraverso il metodo empirico ricerca e ottiene dati di fatto sul funzionamento degli algoritmi.

La modalità di TRex è studiare il funzionamento degli algoritmi per sottoporli a scrutinio pubblico.

È importante ricordare come l’algoritmo che decide, ad esempio quale informazione appare più frequentemente nella navigazione su Facebook, è un segreto industriale. Non è libero né aperto, ma è un algoritmo prigioniero in una gabbia. Non è possibile conoscere la sua composizione, dunque possiamo solo assistere al suo funzionamento e alle conseguenze che provoca. Raccogliere e analizzare i dati per studiarli e risalire alla sua natura. Esporre il tracciamento.

*Un algoritmo cattivo ti fa male. / A captive algorithm it’s bad for ya.*

## Internet e Big Data

La connessione ininterrotta a internet che sperimentiamo in questi anni è una novità, la rete è un formidabile strumento di raccolta dati; scollegarsi non è un’opzione praticabile, a meno di rinunciare a quasi ogni servizio possibile, dalla relazioni con le proprie cerchie, al noleggio di un’automobile o la gestione di un

conto in banca. Inoltre, la pandemia ha imposto di recente l'introduzione del distanziamento sociale, aumentando di frequenza e importanza le pratiche di connessione remota interattiva come la teledidattica e il telelavoro.

Con l'aumento della comunicazione mediata, vengono utilizzati come tramite degli insiemi tecnologici per sopperire al contatto e alla presenza fisica. Questo particolare uso di un supporto tecnologico complesso è particolarmente visibile, in quanto viene tentativamente strutturato e regolamentato, nel telelavoro, infatti chiamato: *Lavoro agile* e nella teledidattica, chiamata: *D.A.D. Didattica a Distanza*, ma aumentano anche lo scambio e la vendita a distanza; naturalmente le relazioni nel senso più vasto ne sono coinvolte. La tecnologia è più importante di prima. Il grande mediatore tecnologico del nostro tempo è Internet.

A ogni connessione gli individui lasciano dietro di sé un grande numero di dati e metadati, che vengono utilizzati come merce dalle ditte private, che li raccolgono tramite servizi "gratuiti", per esempio i social network. Questi dati, sempre aggiornati e aggregati in grandi quantità, vengono chiamati "Big Data" e sono commerciabili in quanto di grande utilità per disegnare il carattere e le abitudini degli individui; vengono utilizzati non solo per distinguere e catalogare gusti e opinioni, ma anche per stimolarli. Possono infatti indirizzare l'opinione del singolo individuo a scopo pubblicitario, come anche influenzare il comportamento sessuale e creare consenso per la governance.

La capacità di raccogliere e processare grandi quantità di dati è privilegio delle entità che possiedono i mezzi di produzione adatti a farlo. Queste non corrispondono alle istituzioni locali o statali, ma sono ditte private come per esempio Google, Amazon e Facebook. Tali aziende quali accumulano i dati, li trattano, li usano e a volte li rivendono in un'ottica di profitto privato che non per forza corrisponde a un bene comune. Ci troviamo di fronte a una asimmetria informativa e dunque di potere. Le leggi sulla Privacy, anche se in costante aggiornamento, possono per definizione solo seguire e non certamente precedere l'innovazione scientifica e sono dunque inadeguate a rapportarsi con il potere dei colossi privati, spesso extra nazionali, che hanno a disposizione i dati dei cittadini.

## Algoritmica

Il modo in cui i Big Data vengono raccolti e trattati, nonché le modalità di analisi dei dati, chiamano in causa la parola Algoritmo. Gli algoritmi sono calcoli finiti che hanno lo scopo di trarre conclusioni, come una ricetta. Sappiamo che la Ricetta non è neutrale, come non lo è la Statistica e non lo è l'Algoritmo. Il punto d'osservazione, le modalità di acquisizione, la capacità e la volontà di cercare o dimostrare qualcosa influiranno sul risultato finale, che verrà poi

fornito, formattato, per ulteriore interpretazione. Il modo in cui viene presentata un'informazione contiene già un giudizio, un punto di vista e un obiettivo.

Gli algoritmi che vengono utilizzati per raccogliere e utilizzare i dati dalle ditte che offrono servizi online non sono pubblici, il loro funzionamento è un segreto aziendale. Non sappiamo con quale criterio dopo un determinato video su Youtube ce ne viene proposto un altro, possiamo intravedere lo scopo pubblicitario, ma non quello politico. Non sappiamo ancora come funziona, ma ci siamo accorti di come queste informazioni siano utilizzate per determinare il consenso e per regolare le comunità al suo interno.

## Influenza politica

Il caso Cambridge Analytica è significativo. Nel 2018, grazie a un'inchiesta giornalistica del canale Tv inglese Channel 4, viene rivelato che la Società UK Cambridge Analytica ha usato a scopo politico i dati degli utenti, fornendo loro informazioni costruite e mirate. Un'azione questa, che nel loro gergo tecnico viene detta *segmentazione*, allo scopo di influenzare le scelte elettorali. La ditta ha ricevuto i dati da Facebook, ufficialmente senza pagarli direttamente, ma figurando come ricercatori. Negli scorsi anni ha influenzato le elezioni di Argentina, Nigeria, Italia e Stati Uniti. Per stessa ammissione del responsabile Alexander Nix, *registrata in video*, le attività di gestione del consenso addirittura non si limitavano all'uso dei dati, ma all'occasione potevano comprendere attività più palesemente illecite, come l'utilizzo di sex workers per ricattare i candidati politici scomodi. Lo scandalo ha travolto Cambridge Analytica, costringendola a chiudere, ma ha anche coinvolto Facebook, sollevando domande sul suo potere e responsabilità. Quando il suo fondatore Mark Zuckerberg sarà interrogato al Senato USA, alla domanda su quale sia la natura dei suoi affari risponderà: *“Senatore, noi facciamo pubblicità / Senator, we run ads”*. È importante notare che fino a quel momento Facebook non si era mai riferita a sé stessa come un'agenzia pubblicitaria, ma come un servizio gratuito e utile a mettere in collegamento le persone nel mondo. Il suo potere è da allora noto al pubblico: Facebook ha la possibilità di influenzare la politica, attraverso la manipolazione del consenso.

## Nascita di Tracking Exposed

Tracking Exposed nasce nel 2016, per analizzare dall'esterno il funzionamento di un algoritmo.

Per andare verso un'etica dell'informazione, dobbiamo essere in grado di interpretare il funzionamento degli algoritmi che trattano le informazioni che ci riguardano. Non avendo accesso al codice sorgente degli algoritmi, possiamo

applicare il metodo scientifico creando una situazione mirata, raccogliendo i dati che ne scaturiscono e confrontandoli. Sì, stiamo tirando palline da ping pong verso un muro invisibile e cerchiamo di capire la natura del muro studiando le palline che ci ritornano.

Una delle funzioni degli algoritmi è di tracciare i comportamenti personali e fornire risposte diverse a persone diverse. Per esempio in quale diverso ordine vengono proposti gli articoli sulla timeline di Facebook. Per poter de-costruire questa funzione, abbiamo creato un componente aggiuntivo per il navigatore in modo che, su base volontaria, le persone possano fornire i dati che Facebook crea durante l'esperienza di utilizzo e in seguito analizzare i dati ricavati per rilevare le diversità di trattamento tramite comparazione.

Lo scopo di Tracking Exposed è rivelare la natura e il funzionamento delle tecnologie di tracciamento, dato che le conseguenze sul mondo reale, come abbiamo visto, sono importanti. È un compito non risolvibile con un approccio puramente legislativo e certamente è vano chiedere alla stessa Società che trae profitto dall'uso dell'algoritmo -che ha essa stessa progettato come segreto industriale- di spiegare e rendere pubblico il suo funzionamento. Ancor più illusorio è credere che l'Azienda stessa possa e voglia porre un rimedio per mitigare le conseguenze nocive dei suoi algoritmi, per evidente contrasto con il proprio interesse privato. Le richieste periodicamente provenienti dalle istituzioni nazionali nei confronti delle grandi piattaforme internazionali, mancando della capacità esecutiva, finiscono per somigliare a lamenti. Queste aziende private, grazie ai dati che raccolgono, conoscono meglio i cittadini delle istituzioni che li governano.

## **Etica dell'informazione**

Ogni tentativo di interazione di tipo legislativo si è rivelato sino a oggi insufficiente, ma di fronte all'obsolescenza delle leggi - il legislatore segue l'innovazione tecnologica, non può precederla - l'approccio non può essere puramente di tipo legislativo. L'aumentare dell'automazione nella burocrazia di ambito giuridico, l'uso delle intelligenze artificiali in ambito normativo e la tendenza a costruire regole basandosi sui dati, la cosiddetta: *data driven policy*, rende urgente avere la possibilità di conoscere e poter de-costruire gli elementi che partecipano alla formazione dei Big data. Occorre incamminarsi verso un tipo di etica dell'informazione che consideri l'individuo come un agente libero e responsabile, portatore di un pacchetto di informazioni di cui va salvaguardata l'integrità e non solo la proprietà, e che sia questo approccio etico a guidarci nelle scelte future. Occorre che le istituzioni di monitoraggio etico-politico, dette *watchguard*, siano in grado di effettuare le analisi in proprio e in maniera indipendente, sia avendo a disposizione gli strumenti tecnici necessari sia il metodo di analisi. In

modo che quest'ultime siano in grado di collaborare tra loro e con la comunità scientifica.

Sappiamo che la tecnologia non è neutrale, che l'intelligenza artificiale fallisce nel comprendere contesto e intento e che l'algoritmo è oggi strumento di governance politica. Un discorso di de-costruzione del potere ha bisogno, per farsi comprendere, di un approccio interdisciplinare. Per comprendere e farci comprendere chiederemo aiuto alla sociologia, alla semiotica e alla psicologia ma prima di tutto crediamo che servano dei fatti: evidenze tecnologiche ricavate empiricamente. Così è nata la ricerca di Tracking Exposed.

## Diritti

Gli algoritmi, filtrando le informazioni, organizzano il tempo e il consenso. Non esiste una soluzione tecnica, la cui credenza chiamiamo: *tecno-soluzionismo*. Perché il problema non è solo tecnico, ma sociale, giuridico-politico e naturalmente *anche* tecnico. La rete non è uno spazio staccato dalla realtà, le conseguenze della repressione algoritmica varia a seconda della situazione geopolitica dove si esprime, ma ricade sempre e con grande violenza sul corpo, cioè sulla vita reale delle persone. Per esempio un post “sbagliato” su Facebook in Italia può causare il licenziamento, ma lo stesso “errore” in un Paese in guerra o con diritti civili insufficienti, può costare la vita.

Crediamo che ai *Big Data* si debba rispondere con *Big Rights*. È una questione politica e certamente non solo economica. Per avere dei diritti bisogna esercitarli e dunque conoscerli. I dati possono essere utilizzati, in un contesto democratico, nell'interesse collettivo e per il bene comune, ma in attesa che questo sia possibile, vogliamo contribuire a tenere aggiornata la pubblica opinione e la comunità scientifica sulle dinamiche dello sfruttamento dati e stimolare un dibattito a questo proposito. Lo scopo di Tracking Exposed è anche di evidenziare e rendere esercitabili i diritti legati allo sfruttamento dei dati, fare informazione e sviluppare critica. Allo stato attuale non possiamo ambire a promuovere soluzioni, ma possiamo far conoscere il funzionamento degli algoritmi proprietari, evidenziando come sia un requisito necessario a ogni forma di democrazia moderna, perché Internet può anche essere un mezzo di oppressione. Le libertà civili nel secolo 21 sono inestricabilmente connesse alla resistenza alla sorveglianza elettronica.

## Sulle spalle dei giganti

Abbiamo trovato aiuto, nel tentativo di farci comprendere, dalle parole scritte da grandi persone del passato. Scoprendo che il discorso della manipolazione delle informazioni non è nato ieri. Pensiamo sia il caso di lasciar loro la parola.

*“Non si può mangiare un confetto pretendendo di sentire – solo perché si ha una vasta cultura e un forte controllo delle proprie sensazioni – sapore di sale. La chimica non sbaglia mai. Siccome esiste anche una chimica delle emozioni, e uno dei composti che per antica tradizione suscitano emozioni è un intreccio ben congegnato, se un intreccio è ben congegnato suscita le emozioni che si era prefisso quale effetto. Potremo poi, après coup, criticarci per averle provate, o criticarle come emozioni repellenti, o criticare le intenzioni con cui è stata congegnata la macchina che le ha provocate. Ma questa è un’altra storia. Un intreccio ben temperato produce gioia, terrore, pietà, riso o pianto.”*

Umberto Eco, *Il superuomo di massa*, 1976

*“Quando la signora Thatcher è stata eletta per la seconda volta, aveva ingaggiato la Saatchi & Saatchi, una grossa Compagnia pubblicitaria, per la sua campagna. E i pubblicitari hanno utilizzato tutti i possibili trucchi, dai giri di frase calcolati per suscitare facili emozioni, ai colori dei suoi vestiti o delle tende davanti alle quali si metteva, sino al calcolo preciso di quando comparire e scomparire dai media. E intanto la sua nobile opposizione socialista disprezzava quei trucchi e i media. Abbiamo potuto osservare bene l’attenta regia della campagna della signora Thatcher, seguendo un geniale programma televisivo. Quando dico ”noi” intendo quella minoranza del paese che l’ha seguito, mentre ritengo che guardarlo avrebbe dovuto essere obbligatorio. Siamo arrivati a un punto in cui un leader politico non solo usa abilmente i vecchi trucchi teatrali per fomentare la folla, come faceva Giulio Cesare nel dramma di Shakespeare, ma ingaggia gli esperti che rendono tutti quei trucchi più efficaci. L’antidoto però consiste nel fatto che in una società aperta possiamo anche esaminare quei trucchi mentre vengono usati con noi. Naturalmente sempre che scegliamo di esaminarli, sempre che non cambiamo canale per vedere Dallas o qualsiasi altra cosa.”*

Doris Lessing, *le Prigioni che abbiamo dentro*, 1987

*“Mentre le merci sono separate da noi e possono essere portate a casa per essere analizzate prima di essere consumate, l’informazione entra a far parte di noi non appena viene acquistata, e provoca in noi cambiamenti determinanti, se considerati nella prospettiva dell’intellettualismo etico, per la quale l’ignoranza e la disinformazione sono la differenza decisiva fra il male e il bene.”*

Platone, Protagora (314b), circa 388 A.C.

*“Internet non è più uno spazio libero e indipendente, ma è commercialmente controllato e personalizzato, Google e Facebook ciberanno gli utenti di ciò che vogliono fargli vedere. Il computer è diventato uno specchio che riflette i tuoi interessi e rinforza i tuoi pregiudizi.”*

Eli Pariser, *Filter bubble*, 2011



*“Conoscere gli individui meglio di quanto conoscano sé stessi è un formidabile mezzo di controllo.”*

Edward Herman e Noam Chomsky, *Manufacturing consent*, 1988

*“Libertà non sta nello scegliere tra bianco e nero, ma nel sottrarsi a questa scelta prescritta.”*

Theodor Adorno

*“L’individuo in rete è rappresentato dall’insieme dei suoi dati.”*

Stefano Rodotà

## Gli Algoritmi di personalizzazione

Gli algoritmi di personalizzazione sono nati perché le Compagnie volevano rendere l’esperienza online meno noiosa, più efficiente e soddisfacente e sono serviti a evitare lo spam. Oggi invece regolano il discorso pubblico all’interno dei social media e di conseguenza sulla governance, sino all’individuo. Disincentivano il pensiero critico: se la radio e la tv possono essere strumenti di manipolazione di massa, capaci di amalgamare l’immaginario e unificatori linguistici, almeno altrettanto si può dire dell’internet della raccolta dati. Che sia per limitare la *information overload*, cioè l’eccesso di esposizione, o al contrario per mantenere l’utente attaccato allo schermo e vendergli più pubblicità, l’algoritmo svolge una funzione di *gatekeeping*, cioè di filtro automatizzato tra le piattaforme e le persone, ma senza che ne siano accessibili i presupposti ideologizzati, o le finalità. In altre parole: siamo consapevoli che la comunicazione non è mai neutrale ma ha sempre un obiettivo, per esempio pensiamo che sia fondamentale poter riconoscere in un’informazione la differenza tra notizia e pubblicità.

Gli algoritmi non sono comprensibili al pubblico a causa della loro complessità. Inoltre sono secretati. Le Compagnie non cedono i dati e tanto meno le formule. È irrealistico anche chiedere alle Compagnie di autoregolarsi. Perciò, per verificare che le imprese non mettano in atto pratiche scorrette nei confronti del consumatore, è necessaria un’analisi terza e indipendente sugli algoritmi. Inoltre, non si può discutere di vessatorietà di clausole contrattuali senza considerare come effettivamente i dati del consumatore vengono sfruttati per discriminare, non solo sulla base di nazionalità e luogo di residenza, ma anche sulla base di dati molto più sensibili quali orientamento sessuale, reddito e disponibilità economica, stato civile e livello di istruzione. Le dichiarazioni delle Società non sono sufficienti. È necessaria la possibilità di documentare e analizzare prove circa l’aggressività e la pervasività di pratiche commerciali ritmate da tempi e modi personalizzati. Gli algoritmi devono essere conosciuti al pubblico attraverso un’analisi esterna e indipendente.

# TRex: Tracking Exposed

Tracking Exposed è un progetto no-profit che usa software libero per analizzare prove di personalizzazione algoritmica. Permettiamo cioè agli utenti dei social media di setacciare e raccogliere i dati che li riguardano e di analizzarli per comparazione, offrendo loro interfacce rispettose della privacy. Lo scopo è di rivelare quanto sia aggressivo e manipolatorio il moderno panorama di Internet e le sue conseguenze.

TRex investiga gli algoritmi; riportiamo il potere di analisi e controllo dei dati agli utenti che producono essi stessi i dati e che dovrebbero dunque possedere. Forniamo alle persone gli strumenti per *controllare il controllore*. Crediamo che per attenuare la repressione algoritmica, da parte di chi dispone del potere di raccogliere dati e utilizzarli, sia necessario che il funzionamento degli algoritmi sia trasparente al pubblico. Ci proponiamo di agire sulla realtà, rivelando il funzionamento dei sistemi di profilazione personalizzati e producendo prove del loro funzionamento. Lo scopo è quello di fornire strumenti, di abilitare le persone, ovvero quelle che vengono chiamate: “utenti” nel loro rapporto con la tecnologia di consumo, a mantenere e gestire la parte di potere esistente nei loro dati.

Per poter fare tecnologia, per programmare strumenti di analisi aggiornati, per fare educazione ed essere in grado di divulgare gli strumenti e i metodi di lavoro, sia per altri analisti che in ambito di ricerca, abbiamo bisogno di una copertura economica che ci permetta di non basarci esclusivamente sul volontariato. Soprattutto considerando la statura materiale delle entità commerciali con cui ci confrontiamo. Crediamo che il momento storico sia paragonabile a un “medioevo”, termine che qui usiamo nella imprecisa accezione di epoca oscura e superstiziosa, per quanto riguarda gli strumenti digitali. L’uso della rete e il suo impatto sulla società è in divenire e crediamo che il nostro lavoro sia fondamentale per evitare o almeno limitare le conseguenze della repressione algoritmica. Pensiamo che l’importanza di questo lavoro acquisterà un’importanza sempre più visibile negli anni a venire.

*“La libertà non può essere gratis.” Seneca.*

## Metodo di Tracking Exposed

Tracking Exposed utilizza il metodo scientifico. Fa uso di tecnologie aperte e verificabili, rende disponibili e così utilizzabili dal pubblico e dalla comunità scientifica i risultati delle analisi e gli strumenti utilizzati, la metodologia, il contesto e i dati raccolti nella loro interezza. Perché siano sottoposti a ulteriore verifica. Accompagniamo le analisi a una descrizione trasparente per permettere replica e confutazione. Costruiamo in proprio gli strumenti di analisi quando

questi non sono già esistenti e li rendiamo disponibili, scaricabili e utilizzabili liberamente, senza alcun bisogno di autorizzazione da parte nostra. Siamo indipendenti nella raccolta dei dati, nella scelta di quale obiettivi perseguire e di quale posizione osservazione adottare. Tendiamo a un approccio interdisciplinare per tener conto della complessità dell'osservazione e per mantenere la capacità di comunicare i risultati, ma aderiamo al metodo sperimentale di raccolta prove e analisi comparativa.

Utilizziamo lo schema scientifico di riproducibilità sull'apprendimento automatico, versione 2,7 del aprile 2020. Reproducibility checklist: Per ogni modello e algoritmo presentato, viene inclusa una descrizione chiara degli assunti matematici e dei modelli considerati, analisi della complessità che comprende: tempo, spazio e quantità di dati per ogni algoritmo. Ogni ipotesi teorica comprende la prova in completezza e la sua descrizione. Ogni banca dati utilizzata è accompagnata dalla statistica di rilievo, come numero di esempi, andamento, validazione e diversità dei test effettuati; quali dati sono stati esclusi e perché, assieme ai passi preliminari. Pubblichiamo la versione scaricabile dei dati e dell'ambiente di simulazione, ogni dato raccolto è accompagnato da una descrizione completa del processo di collezione, incluse le istruzioni e i metodi per la verifica. I codici condivisi comprendono le specifiche delle dipendenze, l'evoluzione del codice, preparazione, modelli e documento: "LEGGIMI" con i comandi da eseguire per riprodurre i risultati.

La base del nostro metodo consiste nel cercare evidenze nel diverso trattamento automatizzato che ricevono due utenti simili. Per ricavare i dati di utenti simili mentre accedono allo stesso servizio, abbiamo adottato la tecnica di creare utenti ad hoc, e anche quella di chiedere i dati su base volontaria. La comparazione tra i dati raccolti da un browser pulito, senza tracce di navigazione precedente e uno già utilizzato, dunque contenente tracce personalizzate, offre per la nostra analisi i risultati più interessanti.

Riconosciamo tre categorie di attori sociali tra: utenti, produttori di news e piattaforme. Dividiamo tra dieta informativa e diversità informativa all'interno delle analisi di quella che chiamiamo: la "Filter bubble". Descriviamo in modo accurato il metodo con il quale effettuiamo il test, sia per mettere a disposizione degli interessati gli strumenti necessari per poterlo replicare, che per avere la possibilità di mettere in discussione le inevitabili mancanze che ogni test contiene. Per raccogliere i dati del trattamento che ricevono gli utenti facendone una copia, abbiamo costruito un'estensione del browser per collezionare quello che appare all'utente.

Rilasciamo i dati dei nostri esperimenti per permettere ad altri gruppi di ricerca di verificare o confutare le nostre analisi, e perché possano usare lo strumento. L'algoritmo continua a cambiare, il nostro software è appositamente rilasciato sotto licenza libera perché sia utilizzabile da tutte le persone. Pensiamo che solo

collettivamente ci si possa interrogare su quali sistemi debbano regolare i nostri discorsi online, che condividere la capacità di analisi sia un modo possibile per evitare una deriva tecnocratica.

Abbiamo infine sperimentato l'opposizione delle piattaforme, per esempio Facebook, alle analisi che le riguardano. Non meramente sul piano legale, ma soprattutto su quello più materiale dell'opacità e dell'assenza delle informazioni, va ricordato a questo proposito che il rilascio dei dati è sempre incompleto. Infine anche un'opposizione sul piano propriamente tecnocratico del rendere le informazioni esplicitamente inaccessibili. L'analisi della Interfaccia di Programmazione di una Applicazione non è sufficiente e inoltre le Compagnie aggiornano e chiudono le API quando gli pare opportuno, godendo della evidente asimmetria di potere. Anche per questi motivi stiamo transitando da una metodologia di analisi quantitativa verso un'analisi qualitativa e di inserimento di ulteriori elementi contestuali.

## **Il proprio Algoritmo**

I dati appartengono e sono generati da persone. Il nostro scopo non è solo fornire un metodo utilizzabile in ambito di ricerca o di studio, ma permettere all'utilizzatore finale di compiere le sue proprie ricerche e analisi. Pensiamo che la capacità di analisi dei dati sia da riportare alla persona cui i dati stessi appartengono. Le persone usando la rete dovrebbero avere persino la possibilità di comporre il loro, proprio algoritmo. Abbiamo scoperto che le nostre analisi hanno l'utilità di de-costruire l'azione delle piattaforme e in alcuni casi di permettere l'individuazione di campagne di disinformazione.

Tracking Exposed è un punto d'osservazione indipendente e si propone di rendere visibile il tracciamento, cioè il modo in cui i social network studiano le persone, raccolgono e processano i loro dati secondo attività e interazioni per arrivare a proporre contenuti mirati. La sorveglianza non è visibile per le persone, perché gli strumenti di analisi, gli algoritmi, sono oscuri. Questo produce discriminazione algoritmica. Nell'intento di rendere le nostre analisi comprensibili e utilizzabili nel dibattito contemporaneo, tentiamo di non produrre solo dati percentuali, ma anche analisi che partendo da dati ricavati con metodo scientifico esperienziale o sperimentale, producano elementi di critica e di utilità sociale.

Oggi non sappiamo ancora giudicare appieno gli algoritmi e il loro potere. Analisti come noi stanno inventando tassonomie e le mettono a disposizione per la verifica, ma quando la società civile pone delle richieste alle piattaforme, queste sono sempre volte a risolvere un problema di tipo politico, come il rimuovere la disinformazione o i messaggi d'odio, o evitare il bullismo online. Pensiamo che da una parte non sia accettabile delegare una responsabilità censoria a un'azienda privata, e inoltre che non vada accettata questa tendenza a delegare

lo spirito critico. Le dinamiche di potere delle reti vanno affrontate decostruendo e scorporando questi poteri. Dobbiamo poter usare i dati in modo costruttivo.

## **Scoperte esperienziali**

Abbiamo scoperto che l'acquisizione dei dati è il momento più sensibile.

Nella ricerca di nuove metriche per capire e misurare l'algoritmo, abbiamo iniziato a usare l'espressione: Dieta informativa, per dare un nome al regime cui Facebook sottopone i suoi utenti. A prescindere dai valori utilizzati da Facebook, lo scopo è evidenziare ed eventualmente trasformare l'arbitrarietà con la quale l'algoritmo decide per nostro conto.

Abbiamo incontrato, durante questo esperimento, anche il concetto di diversità informativa. Ossia la probabilità con la quale l'algoritmo propone contenuti che l'utente non hai mai visto prima. Possiamo dedurre che l'esposizione a contenuti diversi o la riproposizione di contenuti già visti impatti direttamente sulla diversità di opinioni alla quale l'utente viene esposto e di conseguenza sulla percezione della pluralità. Abbiamo chiamato questa variabile di proposta di contenuti sempre diversi: Diversità informativa.

Osservando come i giornali online vengono trattati da Facebook abbiamo constatato che l'algoritmo non considera solamente le preferenze, cioè i Like. Come neppure considera solo le risorse investite dalle attività commerciali, cosa che avrebbe causato il ripetersi delle stesse proporzioni di proposte agli utenti. Sembra che il giornale più avvantaggiato sia quello centrista, come se l'algoritmo penalizzasse le posizioni più radicali, e nel verificare che le pagine del giornale più seguito sono proposte più spesso, si potrebbe dedurre che l'algoritmo tenda a preservare lo status quo, ma purtroppo non abbiamo ancora una chiara determinazione; la metodologia serve a verificare una ipotesi che impatta sulle nostre diete informative individuali.

È opportuno ricordare che Facebook ostacola attivamente le analisi indipendenti. Non attraverso pressioni legali, come invece è il caso di Spotify. Facebook usa il suo strapotere tecnologico. Rendendo difficile la creazione di utenti sperimentali, individuando le attività che si discostano dal comune, ma anche attraverso il boicottaggio: inserendo nel loro HTML parti di codice con l'unico scopo di offuscare i dati che ci servono per effettuare un raffronto. La protezione dell'algoritmo proprietario non avviene con meccanismi legali, ma tecnologici.

## **Proposta**

Tracking exposed offre strumenti, esperienze, metodo e discorso sull'analisi degli algoritmi. Dall'analisi passiva delle esperienze personalizzate, sino alla

declinazione degli effetti sugli esperimenti effettuati.

È fondamentale per noi fare chiarezza che su qualunque risposta al funzionamento di un algoritmo, non si possa avere una risposta soddisfacente da parte della stessa entità che questo algoritmo ha inventato e utilizza, per evidente conflitto di interessi. Non è verosimile che la stessa ditta che produce l'algoritmo sia chiamata ad autoregolarsi. Inoltre l'algoritmo è così personalizzato, geolocalizzato e regionalizzato per cui una versione che viene applicata in una nazione, non vale per un'altra.

Le competenze e le esperienze pratiche di TRex sono condivisibili: utilizzabili da ricercatori come metodo di analisi o da un'entità produttrice di contenuti che voglia capire le logiche di valutazione e di un partner che voglia supportare l'esperienza e il lavoro di TRex.

Offriamo test affidabili, in un panorama dove anche i test inaffidabili vengono comunque ripresi e utilizzati e i racconti aneddotici rischiano di essere il riferimento più ricorrente.

## Note conclusive

Il testo che avete letto è un lavoro in fase di sviluppo, alcune porzioni, come la storia del progetto o le metodologie di analisi, non sono ancora state descritte. Aggiornamenti possono essere trovati a <https://tracking.exposed>.