

Rompere il business model della Disinformazione

M. Franchini

marcelino@franchini.email

M. Montanari

montanari.marg@gmail.com

P. Pedullà

popedulla@gmail.com

Abstract

I principali incentivi della disinformazione sono il profitto economico e l'influenza politica:

- La disinformazione for profit diffonde notizie false, basate su fatti e dati inesistenti e messi in circolazione da fonti non attendibili che dalla loro diffusione ricavano un guadagno, per combatterla è necessario convincere gli Ad Networks a tagliare la fonte di finanziamento, ovvero, la pubblicità display dove si diffonde disinformazione e facilitare la cooperazione con il regolatore per individuare i creatori di contenuti, domini e pagine, con l'obiettivo di bloccarli.
- La disinformazione con un fine politico è creata per manipolare l'opinione pubblica, la principale strategia di diffusione sono gli ads targettizzati per raggiungere le persone più propense a rilanciare questi contenuti per ottenere una diffusione esponenziale, è necessario abolire la pubblicità politica mirata, un messaggio politico non si dovrebbe poter comprare, si dovrebbe diffondere organicamente attraverso il consenso.

Introduzione

Sebbene la disinformazione sia un fenomeno di vecchia data, è negli ultimi anni che la circolazione di notizie false sta diventando motivo di preoccupazione per i governi di tutto il mondo. Internet ha reso possibile l'aumento del flusso di informazioni. Dati, fotografie, pensieri, opinioni, annunci pubblicitari vengono condivisi a una velocità fino a 20 anni fa inimmaginabile ed entrano in contatto con una platea molto estesa. Certamente, il web non è l'unico veicolo responsabile della circolazione di notizie false. Tuttavia, questa oggi riesce principalmente online. Basti pensare che le indagini di Eurobarometro evidenziano che il 72% dei cittadini europei che utilizzano Internet ogni giorno si imbatte in notizie false più volte al mese (Eurobarometro, 2020). E questo provoca danni su più fronti alle fondamenta su cui si basa il vivere democratico. Da una parte, riduce la fiducia nelle istituzioni e nei mezzi di comunicazione più tradizionali (la stampa si trova spesso ad essere target di azioni che cercano di screditarla); dall'altra, danneggia le nostre democrazie perché rende difficile ai cittadini prendere decisioni informate e in alcuni casi rischia di influenzarne l'opinione o di alimentare la radicalità delle idee.

In particolare, è sui social media che le informazioni vengono diffuse con più frequenza in maniera poco accurata, inesatta o intenzionalmente errata. Proprio le intenzioni che muovono la diffusione di contenuti falsi ci permettono di distinguere tra disinformazione, misinformazione e fake news. Per inquadrare meglio il tema di questo paper, è d'obbligo aprire una piccola parentesi semantica. Con il termine disinformazione si intende la diffusione intenzionale di notizie o informazioni inesatte o distorte allo scopo di influenzare le azioni e le scelte di qualcuno^[1]. La misinformazione ha di diverso che ad essere diffusa non è una falsità, bensì una notizia vera, ma condivisa per creare un danno. Il termine fake news, forse il più sdoganato nel linguaggio corrente, si riferisce invece a notizie false ma verosimili: di queste è difficile dimostrare la falsità, ma al tempo stesso non esiste una corrispondenza di dati che le avvalorano.

Nel presente policy paper affronteremo il tema della disinformazione online – e quindi diffusa attraverso le piattaforme –, distinguendo quella a scopo di lucro, e perciò nell'interesse di alcuni attori intenzionati a guadagnare dalla diffusione di contenuti falsi, da quella per scopo politico, orchestrata con l'intento di favorire o svantaggiare una personalità o un governo.

Focalizzandoci sul fenomeno della monetizzazione della disinformazione, cercheremo di spiegare come disarticolare il complesso business model che si nasconde anche dietro un semplice clic a un link apparentemente innocente o una pubblicità online. Un'azione sugli ad networks, a nostro avviso, è il primo passo per disincentivare *malicious actors* dal diffondere

contenuti falsi per trarne profitto. Sul fronte della disinformazione politica, invece, passeremo brevemente in rassegna gli attori coinvolti, le tecniche utilizzate, gli interessi in gioco, il problema delle interferenze esterne. Per poi indicare quali sono stati, fino ad ora, gli strumenti regolativi studiati da alcuni stati – ma soprattutto dall’Unione Europea – per rispondere a una minaccia che si è già dimostrata spaventosamente efficace nell’influenzare i comportamenti elettorali. Passeremo infine a spiegare in tutto sette proposte di policy per rompere il business model della disinformazione: tre per contrastare e disincentivare la disinformazione economica, quattro per la disinformazione politica.

Il terreno della lotta alla disinformazione online è un terreno scivoloso, dove i regolatori cercano un equilibrio tra la necessità di garantire la libertà di espressione e dei mezzi di comunicazione – internet è per eccellenza il luogo del libero pensiero – e l’obbligo di tutelare la sicurezza dei cittadini e preservare la tenuta degli organi democratici. In un ambiente ricco di informazioni, poi, l’azione del regolatore contro i contenuti falsi si scontra necessariamente con una questione di budget. Tra le quattro proposte che abbiamo studiato anche una modalità per finanziare l’operazione di vigilanza sui contenuti.

^[1] Definizione Treccani

Regolamentare: con quali modalità?

Di recente – l'interesse è stato accelerato dalla vicenda di Cambridge Analytica, risalente al 2018 – le autorità hanno studiato diverse possibili strade nella lotta alla disinformazione. Tra le pratiche che gli Stati possono mettere in campo per contrastare e disincentivare la disinformazione, vanno considerate infatti anche quelle di natura regolamentare. Innanzitutto, occorre definire la responsabilità che hanno le compagnie che gestiscono social media per i contenuti condivisi sulle loro piattaforme. Per far sì che la responsabilità sia effettivamente assunta da parte di questi attori, gli Stati hanno cominciato a definire diversi meccanismi legislativi che ritengono appropriati. In questo capitolo cercheremo di fare una panoramica dei tre principali livelli di regolamentazione su cui i governi si sono misurati o hanno provato a muoversi.

Autoregolazione

Il primo approccio affida un ruolo centrale nella lotta alla disinformazione alle grandi compagnie che gestiscono le piattaforme. Queste si mettono a disposizione per implementare volontariamente azioni volte a contrastare il proliferare di notizie false o campagne di disinformazione sui rispettivi canali. Lo stato, o l'ente, si limita a incentivare certi comportamenti virtuosi ma non ha la facoltà di imporre la propria volontà. L'esempio forse più strutturato di questa strategia viene dall'Unione Europea. L'UE è ora il più grande regolatore tecnologico a livello internazionale. Utilizza il suo potere normativo per rispondere alla disinformazione in un ambiente di controlli ricco di informazioni e di bassa qualità. Il contesto creato dal Piano di azione contro la disinformazione, adottato dall'Unione Europea nel dicembre del 2018, prevede un metodo di autoregolazione (*self-regulatory approach*). In poche parole, l'UE responsabilizza le piattaforme online – da Google a Facebook a Mozilla – dando a loro l'incarico di supervisionare sui contenuti e contrastare sui rispettivi canali la disinformazione. I big tech, firmando il codice di buone pratiche, hanno accettato di inviare report mensili in vista delle elezioni del Parlamento europeo (maggio 2019), in modo tale che potesse esserci una verifica sulle azioni portate avanti dalle compagnie per evitare campagne online che avrebbero potuto influenzare il voto. Questa regolamentazione proattiva delega quindi l'incarico di fact-checking e di intervento alle compagnie, dando all'UE un ruolo di supervisore. Il problema è che, nonostante i piccoli progressi compiuti, il meccanismo di autoregolamentazione non sta avendo l'effetto desiderato di combattere la diffusione di disinformazione online. Tanto che la Commissione stessa ha criticato la scarsa trasparenza nell'enunciare l'algoritmo che regola la selezione dei contenuti e l'insufficiente impegno dei Big Tech nell'intervenire quando si presentano fake news. Ad esempio, a febbraio 2020 la

Commissione europea ha chiesto uno sforzo in più sul fronte della disinformazione. Il problema di tale metodo “soft” è che dipende dalla volontà politica delle parti coinvolte. Compagnie private, con interessi economici precisi. Non fa il gioco di Facebook, per fare un esempio, intervenire per bloccare alcuni contenuti e utenti. È chiaro che la paralisi dei Big Tech è dovuta alla mancanza di volontà politica. Lo ha reso evidente l'emergenza legata alla pandemia da Covid-19. Nei mesi in cui il Coronavirus cominciava a diffondersi, le piattaforme hanno dimostrato un impegno senza precedenti contro la diffusione di contenuti falsi, rendendo evidente che gli strumenti per agire nella direzione di una maggiore trasparenza e controllo della disinformazione online ci sono, ma non sono declinati in un impegno costante, su base regolare e in circostanze ordinarie.

Assetto normativo

Una seconda linea di pensiero preferisce l'*hard law* alla regolamentazione flessibile. Alcuni stati, infatti, hanno legiferato con l'intento di obbligare i social media ad assumersi la responsabilità per i contenuti che diffondono sulle rispettive piattaforme. Il quadro giuridico esistente in Germania agisce in questo modo. Il Network Enforcement Act impone multe fino a 50 milioni di euro alle società che gestiscono piattaforme qualora non riuscissero a rimuovere contenuti indubbiamente illegali entro 24 ore dalla segnalazione di un reclamo. Non sempre, però, il confine è così netto. Quando a circolare sui social media è materiale denigrante, a quel punto la legge tedesca allarga le maglie del tempo, obbligando le aziende a bloccare il contenuto entro sette giorni. La sanzione è la stessa.

Al tempo stesso, però, il tentativo di regolamentare eccessivamente i contenuti pubblicati su uno spazio libero rischia di sbilanciare il delicato equilibrio tra libertà di espressione e necessità di combattere la disinformazione. In alcuni paesi, come per esempio l'Australia, al fine di contrastare il fenomeno, si è pensato di equiparare il trattamento dei social media a quello per i mezzi di stampa, in modo tale da garantire la libertà di espressione nei limiti espressi dalla legge in materia.

Co-regolazione

Esiste però una terza via, che sta nel mezzo: è la co-regolamentazione. In questo modo, si va a manovrare il processo di distribuzione dei contenuti e non a mettere mano ai singoli contenuti. Le compagnie che gestiscono le piattaforme online diventano in questo caso responsabili delle proprie azioni di fronte a un organo esecutivo deputato alla supervisione e alla regolamentazione. Una revisione normativa su questo stampo prevede che il Parlamento nazionale rediga un codice di condotta che obblighi per tutte le piattaforme che operano

nello spazio di pertinenza statale ad adeguarsi a determinati standard tecnici su cui il governo deve poi supervisionare. Verrebbe meno il principio di adesione volontaria, ma prevarrebbe una logica di enforcement, reso possibile da una serie di regole valide erga omnes.

Věra Jourová, Vicepresidente della Commissione Europea in carica, si occupa di valori e trasparenza. Durante la regolare supervisione da parte della Commissione dell'implementazione del Codice di buone pratiche contro la disinformazione, ha da poco dichiarato quanto segue: "Il codice ha dimostrato che le piattaforme online e il settore pubblicitario possono fare molto per contrastare la disinformazione, quando sottoposti al controllo pubblico. Ma è necessario un aumento della responsabilizzazione e della responsabilità delle piattaforme, che devono diventare più trasparenti. È giunto il momento di andare oltre le misure di autoregolamentazione". Con tali affermazioni viene messa in luce come migliore alternativa la modalità della co-regolazione o, comunque, è chiara l'intenzione della Commissione a procedere verso forme di monitoraggio meno flessibili. Ma sempre meno "robuste" rispetto a forme di regolazione in senso stretto.

Nella prassi, esiste già un caso di co-regolazione contro la disinformazione in Danimarca sul fronte delle intromissioni esterne con fine politico. Al fine di tutelare gli organi democratici da possibili minacce di influenza straniera, il sistema danese prevede il coinvolgimento del governo, del ministero degli Affari esteri e dell'intelligence. Un'azione trasversale e coordinata, inserita in un piano d'azione varato nel 2018 che prevede 11 contromisure con l'intento di fronteggiare campagne di disinformazione (European Parliament Report, marzo 2019).

L'approccio dell'Italia, in fatto di lotta alla disinformazione, al momento, non prevede alcuna regolamentazione o patto di corresponsabilità con componenti della società civile. È di giugno l'approvazione alla Camera dei Deputati di una commissione d'inchiesta sulla diffusione di fake news, composta da 20 deputati e senatori, che in 18 mesi di tempo dovranno inquadrare il fenomeno e proporre l'adozione di nuove leggi o misure amministrative, nell'ottica di ostacolare la disinformazione. Questo sarebbe un primo passo verso la definizione di un quadro giuridico utile anche per limitare la circolazione di contenuti falsi sui social network e sul web.

Piattaforme Social e gli altri fattori tecnologici che aiutano la diffusione della disinformazione online

Questo capitolo si focalizza sui fattori che favoriscono la diffusione della disinformazione e sugli attori delle campagne di disinformazione online. Negli ultimi anni, gli strumenti e i metodi utilizzati per diffondere la propaganda e la disinformazione online sono diventati sempre più sofisticati.

Nello specifico, il capitolo si focalizzerà sul ruolo delle piattaforme social e dei motori di ricerca web come veicoli per la creazione, pubblicazione di contenuti di disinformazione online e il conseguente ricavo dalla loro messa in circolazione.

Inoltre, le strategie di amplificazione della disinformazione online sono esaminate, tra cui i messaggi pubblicitari e i bot e le ragioni per cui alcuni utenti in buona fede condividono contenuti di disinformazione. Il capitolo si conclude con un focus sulle dinamiche emergenti in base alle quali l'intelligenza artificiale e il *deep learning* vengono impiegati per la diffusione di contenuti di disinformazione molto credibili.

I social media (ad esempio Facebook, Twitter, etc.) e i motori di ricerca (ad esempio Google) si sono affermati come importanti fonti di notizie e informazioni online per un numero crescente di cittadini in tutto il mondo.

Conseguentemente, miliardi di utenti in tutto il mondo sono diventati bersaglio di disinformazione online e campagne di propaganda attraverso queste piattaforme social.

Inoltre, il proliferare di disinformazione e misinformazione online sta influenzando anche i risultati della ricerca sul web, il che ostacola ulteriormente gli utenti nella ricerca e nella lettura di informazioni online affidabili.

Negli ultimi due anni Google ha investito in modo significativo per evitare che contenuti falsi apparissero nei suoi algoritmi di ricerca attraverso l'utilizzo di indicatori di verifica dei fatti e di fiducia (i.e. ClaimReview Markup), oltre a promuovere contenuti di editori più autorevoli e strumenti di feedback diretto. Sta inoltre investendo massicciamente nella Google Digital News Initiative, che ha finanziato numerosi progetti volti a sviluppare nuove tecnologie e strumenti per la verifica dei contenuti e l'individuazione di fenomeni di disinformazione. Queste misure, pur rappresentando passi nella giusta direzione, sono ben lungi dall'essere sufficienti, dal momento che le campagne di disinformazione online e i loro orchestratori sono in continua evoluzione in termini di sofisticazione.

I meccanismi più comuni di diffusione per la creazione, promozione e diffusione di contenuti di disinformazione online

- Profili e Gruppi Fake

- I profili e gruppi fake presenti nei social media hanno l'obiettivo di pubblicare contenuti di disinformazione e amplificarne la diffusione e l'importanza percepita. Gli account fake possono essere creati e acquistati per agire come *followers* fittizi e quindi gonfiare artificialmente l'importanza e la popolarità degli account di persone reali. Allo stesso tempo, gli account fake seguono account famosi reali con il duplice obiettivo di acquisire credibilità sulle piattaforme social e di passare inosservati.

I gruppi, in particolare, svolgono un ruolo sempre più importante nelle campagne di disinformazione. A titolo esemplificativo, alcuni gruppi sono creati per "diffondere contenuti sensazionalistici o fortemente di parte", spesso distorcendo i fatti per adattarli a una narrazione" (Weedon, Nulan & Stamos, 2017). L'individuazione dei gruppi fittizi, p.es. di Facebook, può essere agevolata dall'utilizzo dei segni di spunta di verifica di autenticità del gruppo (Silverman, 2017).

- Pubblicità online e Clickbait

- La pubblicità online è stata ampiamente utilizzata per realizzare profitti per i siti di notizie online falsi, in quanto ricevono pagamenti quando gli annunci vengono mostrati insieme al contenuto falso. Il traffico è guidato verso questi siti attraverso annunci sulle piattaforme social costruiti per indurre l'utente a cliccarci sopra (*clickbait posts* e *promoted posts*).

Il *clickbait* si avvale di titoli accattivanti e sensazionalisti che incitano l'utente a cliccare, facendo leva sull'aspetto emozionale di chi vi accede. Il suo obiettivo è quello di attirare chi apre questi link per incoraggiarli a condividerne il contenuto sulle reti sociali, aumentandone quindi in maniera esponenziale i proventi pubblicitari.

I *promoted posts* sono contrassegnati come contenuti pubblicitari sia su Facebook sia su Twitter e rispetto a questi l'utente può eseguire le stesse azioni come a qualsiasi altro post normale (ricondivisione, etc.). Gli inserzionisti sono

fatturati dalla piattaforma social sulla base del numero di interazioni generate con il post. In molti casi gli inserzionisti possono scegliere quali utenti visualizzeranno i *promoted posts*, sulla base di informazioni personali come la posizione geografica, il sesso, gli interessi, il tipo di dispositivo o altre caratteristiche specifiche. Quando gli annunci sono rivolti ad un insieme molto ristretto di utenti, con profili molto specifici, la pratica si chiama micro-targeting.

In generale, è importante rilevare come le piattaforme cerchino di controllare ciò che viene diffuso, vietando la pubblicità di beni e servizi illegali, comprese l'uso di sostanze stupefacenti. Il sistema dei controlli dei contenuti diffusi sulle piattaforme è sia algoritmico sia sulla base di un sistema di feedback degli utenti.

Tuttavia, la mancanza di trasparenza e di regolamentazione della pubblicità online all'interno di piattaforme quali Google, Facebook e Twitter ha causato preoccupazioni significative sia per gli attori governativi sia per il regolatore. Per far fronte a questo, il gruppo di esperti dell'Unione Europea sulle notizie false e la disinformazione online ha sviluppato la sua raccomandazione di policy in merito alla distruzione degli incentivi economici sottesi alla diffusione di contenuti di disinformazione.

Iniziative di carattere tecnologico per combattere la disinformazione

Questo capitolo si concentra sugli algoritmi e gli strumenti tecnologici per il rilevamento delle campagne di disinformazione.

C'è un'urgente necessità di affrontare queste grandi sfide e sviluppare una nuova generazione di strumenti di verifica dei contenuti e di verifica della disinformazione online.

Per contrastare la soggettività, la politica post-verità, la disinformazione e la propaganda, molte organizzazioni mediatiche e istituzioni apartitiche in tutto il mondo hanno avviato iniziative di verifica dei fatti, come elemento chiave di un approccio multidimensionale di successo per limitare la diffusione e l'influenza della disinformazione online.

La sfida più grande degli strumenti di fact checking è rappresentata dal portare all'attenzione dei cittadini i risultati dell'applicazione dei *fact - checks*.

A titolo esemplificativo, FactStream è un'applicazione per telefoni cellulari, sviluppata da Duke Reporters' Lab, che avvisa gli utenti americani degli ultimi factcheck, svolti dalle tre maggiori organizzazioni statunitensi di fact-check - il Washington Post, PolitiFact e FactCheck.org.

Ogni notizia online viene mostrata insieme alla sua fonte, al livello di veridicità della notizia e all'organizzazione di fact-checking che l'ha verificata. Il fact-checking viene anche utilizzato dalle piattaforme social al fine di contribuire alla limitazione della diffusione virale di contenuti falsi.

Facebook, ad esempio, permette agli utenti di segnalare contenuti pubblicati di provenienza sospetta, che vengono poi verificate da una rete di organizzazioni di fact-checking. La verifica dei contenuti è complementare al controllo dei fatti e si occupa di verificare se un'immagine, un video o un meme promuovono informazioni false.

Se da un lato la tecnologia esistente per la verifica dei contenuti e il controllo dei fatti può aiutare i giornalisti, i media e le piattaforme dei social media a segnalare o filtrare più efficacemente i contenuti falsi, dall'altro non riesce ad affrontare un'altra sfida fondamentale: quella di analizzare i meccanismi e la portata delle campagne di disinformazione e propaganda online, le loro fonti e i loro obiettivi, nonché i loro effetti sui cittadini, sulle imprese e sulla società.

- **Amplificazione e account falsi**

- L'importanza della ricerca sull'amplificazione è stata portata all'attenzione dell'opinione pubblica e della politica, come risultato di studi che cercano di stabilire il ruolo dei *bot* e dei *sockpuppets* nelle elezioni presidenziali americane del 2016, nel referendum sulla permanenza del Regno Unito nell'UE e nelle elezioni presidenziali francesi del 2017.

Le ricerche all'avanguardia sui metodi di rilevamento dei bot (Varol & al, 2017; Woolley & Howard 2016; Cresci & al. 2016) utilizzano prevalentemente caratteristiche di comportamento sociale (ad es. frequenza dei tweet, uso di hashtag, etc.) Esistono anche approcci che rilevano i bot sulla base delle elevate correlazioni nelle attività tra di loro.

L'abilitatore chiave dietro a tutto questo lavoro sono gli insiemi di dati di account di bot e sock puppet collaudati e tutti i loro dati sui social media (ad es. post, profilo sociale, condivisioni e simili). Questa base di dati ("*gold data*") sono necessari per l'addestramento degli algoritmi di machine learning per il rilevamento di bot e sock puppet.

- **Rilevamento di campagne di disinformazione e misinformazione**

- Giornalisti e ricercatori si trovano sempre più spesso ad affrontare la sfida di verificare non solo l'autenticità dei contenuti online (ad esempio, una notizia virale o un video), ma anche di scavare più a fondo, per trovare la fonte originaria della disinformazione virale, tracciare la sua diffusione nelle comunità e nelle reti online e stabilire il probabile impatto sui cittadini. È necessario studiare la disinformazione online e le campagne di propaganda nella loro interezza. A tal fine, sosteniamo l'adozione del modello concettuale AMI (Agent-Message-Interpreter) (definito nella sezione 1.2). Seguendo questo modello, è necessario esaminare tutti e tre gli elementi chiave delle campagne di disinformazione e propaganda (Wardle&Derakhshan, 2017): l'affidabilità degli agenti (fonte originale, amplificatori, e le connessioni tra di loro), la credibilità e l'accuratezza del messaggio, e gli effetti sulle credenze degli interpreti (cioè il pubblico target).

Un aspetto chiave di questo modello di analisi è l'affidabilità e la credibilità di questi agenti di disinformazione. La ricerca giornalistica (Lacy & Rosenteil, 2015) ha proposto diverse metriche per valutare la qualità delle notizie e dei

media online, come la parzialità di parte, la parzialità strutturale, l'attualità e la trasparenza delle fonti. Tuttavia, attualmente non esistono metodi automatizzati per calcolare questi indicatori di attendibilità delle fonti basati sui contenuti.

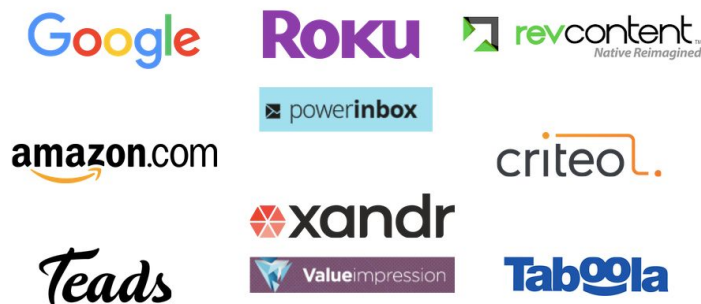
Inoltre, modelli di rete per la modellizzazione della fiducia sono stati proposti, ad esempio per le reti di utenti di Twitter (Wang, 2010), ma questi sono ancora molto sperimentali.

Una nuova iniziativa promettente è il Global Disinformation Index, che sta costruendo un sistema di valutazione in tempo reale dell'affidabilità delle fonti di notizie in tutto il mondo. L'obiettivo è quello di avere una fonte neutrale, indipendente e trasparente di rating del rischio di probabilità (simile a quella delle obbligazioni nella finanza), basata su algoritmi di IA che identificano automaticamente i domini spazzatura (Melford et al, 2018).

Gli agenti di disinformazione spesso non agiscono in modo indipendente, anche se a volte può essere difficile da stabilire. Per dare l'impressione che un gran numero di fonti indipendenti stiano rilanciando contenuti in modi diversi sugli stessi "fatti", alcuni siti di disinformazione e/o sock puppet account riutilizzano e ripubblicano contenuti, in una pratica nota come riciclaggio di informazioni (Starbird, 2017). Attualmente i giornalisti non dispongono di strumenti facili da usare che mostrino quali siti mediatici o account di social network hanno riutilizzato contenuti di un altro. Questo è importante, dal momento che i sock puppet stanno riconfezionando e/o ripubblicando i contenuti nel tentativo di dar loro credibilità e con l'obiettivo di essere accettati dagli utenti come fonti di notizie affidabili.

Proposte di policy principali

- Rompere il business model della disinformazione
 - La disinformazione for profit diffonde notizie false, basate su fatti e dati inesistenti e messi in circolazione da fonti non attendibili che dalla loro diffusione ricavano un guadagno, per combatterla è necessario convincere gli Ad Networks a tagliare la fonte di finanziamento, ovvero, la pubblicità display dove si diffonde disinformazione e facilitare la cooperazione con il regolatore per individuare i creatori contenuti, domini e pagine, con l'obiettivo di bloccarli, questo onere si può finanziare imponendo una tassa del 5% sul costo per l'inserzionista negli Ad Networks per finanziare il regolatore affinché vigili sulla validità dei contenuti dove si mostrano gli annunci.



- Abolire la pubblicità politica mirata
 - La disinformazione con un fine politico è creata per manipolare l'opinione pubblica, la principale strategia di diffusione sono gli ads targettizzati per raggiungere le persone più propense a rilanciare questi contenuti per ottenere una diffusione esponenziale, è necessario abolire la pubblicità politica mirata, un messaggio politico non si dovrebbe poter comprare, si dovrebbe diffondere organicamente attraverso il consenso.



Proposte di policy secondarie

- Creare un quadro giuridico transnazionale

- E' necessario evitare una frammentazione nel framework legale, un esempio è il "Grand Committee on Disinformation" composto da parlamentari di diversi paesi, inoltre si dovrebbero usare le istituzioni multilaterali come l'OCSE e il G7 per affrontare questi temi.
- Responsabilizzare le piattaforme senza compromettere la libertà di espressione, la privacy e la trasparenza, un esempio di linee guida sono i principi di Santa Clara [2]
 - Le piattaforme dovrebbero pubblicare il numero di post rimossi e account eliminati in modo permanente o temporaneamente sospesi a causa di violazioni delle linee guida sui contenuti.
 - Le piattaforme dovrebbero fornire un avviso a ogni utente il cui contenuto viene rimosso o sospeso fornendo il motivo della rimozione o della sospensione.
 - Le piattaforme dovrebbero fornire agli utenti la possibilità di fare appello se un loro contenuto è stato rimosso o sospeso.
- E' necessario rafforzare la fiducia nelle istituzioni pubbliche e nel discorso politico online. La responsabilità sui contenuti diffusi online è particolarmente sensibile quando si tratta di discorsi politici e campagne politiche. La pubblicità politica utilizza il micro targeting basato su dati personali e comportamentali al fine di raggiungere gli elettori indecisi, il che se fatto poco eticamente può minacciare la tenuta stessa di una democrazia.
- Quando si prendono in considerazione le risposte normative alla disinformazione, vi è un grosso rischio di mettere in pericolo la libertà di espressione cercando di definire quello che deve essere vero o falso.

- Definire gli standard delle campagne politiche e mediatiche

- Supportare e promuovere giornalismo di alta qualità e standard minimi nelle campagne politiche per evitare quello che successe nel referendum sull'uscita del Regno Unito dalla Unione Europea: diversi tabloid britannici pubblicarono

informazioni imprecise su migrazione, terrorismo e controllo delle frontiere, questo poi fu amplificato dai politici e sui social media, manipolando così le percezioni degli elettori. Al fine di salvaguardare la fiducia del pubblico nei media, si dovrebbero definire e adottare standard di stampa più rigorosi.

- Promuovere sforzi di fact checking con iniziative nazionali di verifica dei fatti, questo dovrebbe essere fatto come collaborazione tra diverse testate giornalistiche per garantire il pluralismo. Inoltre le figure politiche dovrebbero impegnarsi a non minare la reputazione dei media e dei giornalisti.
- Il fact checking non è sufficiente per combattere la disinformazione, perché è un processo molto dispendioso in termini di tempo, che non può essere completamente automatizzato a causa degli attuali limiti degli algoritmi d'intelligenza artificiale. Non essendo possibile verificare l'informazione in tempo questa viene quindi ampiamente condivisa attraverso i social network, dove può essere amplificata o presentata fuori dal suo contesto originale dai media online, il cui modello di business si basa principalmente sul traffico click through e sulla pubblicità display creando così un conflitto d'interesse tra il profitto e la diffusione di notizie non verificate che portano traffico.

- **Studiare approcci interdisciplinari**

- Quando si tratta di combattere la disinformazione online, la tecnologia può essere di grande aiuto, ma non bisogna dimenticare che questo fenomeno sfrutta le debolezze umane. L'analisi dell'impatto dei social media sulla società e sulla democrazia coinvolge una varietà di attori con background diversi, inclusi data scientist, ricercatori d'intelligenza artificiale, scienziati politici e sociali, nonché giornalisti. Queste figure dovrebbero poter cooperare a livello europeo e nazionale su piattaforme indipendenti che facilitano la ricerca nell'individuazione e nella prevenzione della disinformazione online. Le piattaforme potrebbero anche fungere da incubatori d'iniziative. Potrebbero essere concepiti come organi consultivi indipendenti, simili al Consiglio nazionale digitale in Francia, in combinazione con istituti di ricerca nazionali o europei che riuniscono scienziati di più organizzazioni e discipline.
- Consentire alla società civile di moltiplicare gli sforzi, Come già discusso, le società di social media hanno riconosciuto i vantaggi di lavorare a fianco di giornalisti e scienziati, al fine di ottenere una migliore comprensione della disinformazione e beneficiare degli ultimi progressi nella ricerca

sull'intelligenza artificiale. La società civile può e deve anche svolgere il ruolo di controbilanciamento e di stakeholder indipendente, lavorando a fianco e in cooperazione con società e piattaforme private per segnalare la disinformazione. Dovrebbe essere incoraggiata la condivisione delle conoscenze, sviluppando così competenze trasversali ed evitando progetti duplicati da parte di organizzazioni diverse. Poiché le campagne di disinformazione possono anche oltrepassare i confini linguistici, culturali e nazionali, vi è anche un forte argomento per incoraggiare iniziative della società civile a livello europeo.

- Superare i pregiudizi cognitivi, l'educazione all'alfabetizzazione mediatica viene ancora erogata su scala molto ridotta e localizzata e ci vorrà del tempo per aumentarla a livello europeo. È meglio consegnato attraverso iniziative indipendenti dallo stato, in quanto potrebbe essere frainteso come un'altra forma di propaganda governativa. Ancora più importante, i cittadini dovrebbero essere in grado di comprendere la logica alla base dell'educazione all'alfabetizzazione mediatica e la fonte dei materiali educativi. Ove possibile, la progettazione e l'erogazione dell'educazione all'alfabetizzazione mediatica dovrebbe essere allineata con le conoscenze e le convinzioni (sia politiche che non) del suo target demografico. Ciò è simile alle sfide affrontate dai verificatori di fatti e dalle piattaforme sul modo migliore di presentare le informazioni, che potrebbero rafforzare, piuttosto che indebolire, la fiducia nella disinformazione, specialmente nelle comunità altamente polarizzate.

- **Trasparenza e responsabilità nella moderazione dei contenuti**

- Le piattaforme dovrebbero pubblicare il numero di post rimossi e gli account sospesi in modo permanente o temporaneo a causa di violazioni delle linee guida sui contenuti. Come minimo, queste informazioni dovrebbero essere suddivise lungo ciascuna di queste dimensioni:
 - Numero totale di post distinti e account contrassegnati.
 - Numero totale di post distinti rimossi e account sospesi.
 - Numero di post distinti e account contrassegnati e numero di post distinti rimossi e account sospesi, per categoria di regola violata.

- Numero di post distinti e account contrassegnati e numero di post distinti rimossi e account sospesi, in base al formato del contenuto in questione (ad es. Testo, audio, immagine, video, live streaming).
 - Numero di post distinti e account contrassegnati e numero di post distinti rimossi e account sospesi, in base alla fonte del contrassegno (ad esempio, governi, segnalatori attendibili, utenti, diversi tipi di rilevamento automatico).
 - Numero di post distinti e account contrassegnati e numero di post distinti rimossi e account sospesi, in base alle posizioni dei segnalatori e degli utenti interessati (se evidente).
- Le piattaforme dovrebbero fornire un avviso a ciascun utente il cui contenuto è stato rimosso o l'account è stato sospeso sul motivo della rimozione o della sospensione. In generale, le aziende dovrebbero fornire indicazioni dettagliate alla comunità su quali contenuti sono vietati, inclusi esempi di contenuti consentiti e non consentiti e le linee guida utilizzate dai revisori. Le aziende dovrebbero anche fornire una spiegazione di come viene utilizzato il rilevamento automatico in ciascuna categoria di contenuto. Quando si fornisce a un utente un avviso sul motivo per cui il suo post è stato rimosso o un account è stato sospeso, un livello minimo di dettaglio per un avviso adeguato include:
- URL, estratto di contenuto o altre informazioni sufficienti per consentire l'identificazione del contenuto rimosso.
 - La clausola specifica delle linee guida che il contenuto è stato ritenuto violare.
 - Come il contenuto è stato rilevato e rimosso (contrassegnato da altri utenti, governi, fact checkers, rilevamento automatico o reclamo legale esterno o di altro tipo). L'identità dei singoli segnalatori in genere non dovrebbe essere rivelata, tuttavia, i contenuti segnalati dal governo dovrebbero essere identificati come tali, a meno che non sia proibito dalla legge.
 - Spiegazione del processo attraverso il quale l'utente può presentare ricorso contro la decisione.

- Le piattaforme dovrebbero fornire un'opportunità significativa per ricorrere tempestivamente a qualsiasi rimozione di contenuti o sospensione dell'account. Gli standard minimi per un appello significativo includono:
 - Revisione umana da parte di una persona o di un gruppo di persone che non è stata coinvolta nella decisione iniziale.
 - Un'opportunità per presentare ulteriori informazioni che saranno prese in considerazione nella revisione.
 - Notifica dei risultati della revisione e una dichiarazione del ragionamento sufficiente per consentire all'utente di comprendere la decisione.

- **Alfabetizzazione mediatica**

- Promuovere l'alfabetizzazione mediatica e il pensiero critico dei cittadini. L'educazione all'alfabetizzazione mediatica viene ancora erogata su scala molto ridotta e localizzata e ci vorrà del tempo per aumentarla a livello europeo. I cittadini dovrebbero essere in grado di comprendere la logica alla base dell'educazione all'alfabetizzazione mediatica e la fonte dei materiali educativi. Ove possibile, la progettazione e l'erogazione dell'educazione all'alfabetizzazione mediatica dovrebbe essere allineata con le conoscenze e le convinzioni del suo target demografico, questo è simile alle sfide affrontate dai fact checkers e dalle piattaforme sul modo migliore di presentare le informazioni, che potrebbero rafforzare, piuttosto che indebolire, la fiducia nella disinformazione, specialmente nelle comunità e contesti altamente polarizzati.

- **Innovazione in ambito tecnologico**

- E' necessario preservare i contenuti sui social media per studi futuri, i dati sono essenziali per addestrare l'apprendimento automatico e gli algoritmi di intelligenza artificiale. In un ambiente basato sui dati, la domanda principale non è solo se i dati vengono raccolti, ma se sono accessibili, come e quando. Poiché le tecniche di disinformazione si stanno evolvendo molto rapidamente, i ricercatori devono avere accesso ininterrotto a grandi quantità di dati, al fine di valutare, addestrare e riprogettare le metodologie.

- Promuovere la definizione di politiche per la ricerca etica e di tutela della privacy e l'analisi dei dati sui dati dei social media e anche per garantire l'archiviazione e la conservazione dei contenuti dei social media di valore storico fondamentale. Ciò potrebbe, ad esempio, riguardare tutte le pubblicazioni e gli annunci di partiti politici e candidati ufficiali in un deposito pubblico accessibile a ricercatori e terze parti fidate. Un altro esempio potrebbe essere un sistema di valutazione in crowdsourcing che consenta ai contenuti verificati di prevalere sulle notizie non verificate.
- Le tecnologie di intelligenza artificiale consentono di generare "deepfake" ancora più difficili da identificare e verificare, soprattutto perché immagini e video sono generalmente percepiti come più credibili. Giornalisti, ricercatori, piattaforme e tecnologi devono costruire una comprensione comune di queste minacce. È necessario un approccio condiviso per affrontare le sfide relative ai diritti umani, giornalistiche e tecnologiche sollevate da questo fenomeno. La ricerca collaborativa consentirebbe la creazione di strumenti come la ricerca video inversa e si preparerebbe ad affrontare rischi e scenari imminenti.
- Misurare l'efficacia delle soluzioni tecnologiche implementate dalle piattaforme social volti a rilevare bot e account falsi, identificare discorsi di incitamento all'odio e segnalare contenuti contestati da organizzazioni indipendenti di controllo dei fatti. Finora, tuttavia, non esiste alcun meccanismo in atto per verificare in modo indipendente l'efficacia di questi approcci contro la disinformazione.
- Implicazioni etiche delle soluzioni tecnologiche: Queste nuove tecnologie sono possibili solo all'interno di un solido quadro giuridico. Inoltre, la prospettiva di un monitoraggio completamente automatizzato della disinformazione solleva legittimi timori sulla censura e sulla violazione della libertà di espressione. Pertanto, la trasparenza dell'algoritmo è una condizione necessaria per garantire che le scelte algoritmiche siano verificabili e prive di pregiudizi. Ad esempio, sospendere erroneamente un account è una forma di censura e devono essere messi in atto meccanismi di riparazione rapidi e di revisione umana indipendenti. Allo stesso tempo, la piena trasparenza algoritmica corre il pericolo di indebolirne l'efficacia, in quanto consentirebbe agli attori che pubblicano e diffondono disinformazione di ideare strategie nuove.

Trends in Online Influence Efforts

https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2020/09/Trends_in_Online_Politics_Influence_Efforts_V2.0.pdf

Principi di Santa Clara

<https://santaclaraprinciples.org/>

Contrastare la disinformazione online: un approccio europeo

<https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>

EU Disinfo Lab, The few faces of disinformation; 11 Maggio 2020

[https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU\(2019\)624278_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU(2019)624278_EN.pdf)

Andrea Renda, The legal framework to address “fake news”: possible policy actions at the EU level; Ceps, giugno 2018