

Lotta contro i discorsi d'odio, preservando la libertà d'espressione

8 Settembre 2020

Nuove tecniche sono pronte all'uso per generare contro-narrazioni sulle molte dimensioni dell'hate speech: estremismo, islamofobia, razzismo, odio online contro la comunità LGBTI+, ecc. Intervista a Marco Guerini

GS: Esiste un modo per regolamentare e prevenire l'hate speech online? E in particolare: cos'è l'hate speech e cosa non è sui social media? Come si può contenere l'incitamento all'odio senza influire sulla libertà d'espressione?

MG: In realtà ci sono molte definizioni di incitamento all'odio, infatti è difficile definirlo dato l'ampiezza e le sfumature delle culture e delle lingue. Di solito preferisco attenermi a una definizione ampia quale: l'incitamento all'odio si riferisce a “espressioni che attaccano o sminuiscono, che incitano alla violenza o all'odio contro gruppi o singoli obiettivi, sulla base di caratteristiche specifiche quali l'aspetto fisico, la religione, la discendenza, l'origine etnica, l'orientamento sessuale, l'identità di genere o altro. “

Sotto questo aspetto il mezzo non è così rilevante: sia online che nel mondo reale, l'incitamento all'odio è sempre lo stesso. In effetti, dobbiamo chiarire che le piattaforme social non sono generatori di odio in quanto tali; sono invece un mezzo attraverso il quale espressioni di odio assumono nuova linfa vitale e nuova forza. Grazie alla combinazione tra il modo in cui funzionano gli algoritmi di selezione dei contenuti (ovvero quegli algoritmi che decidono cosa visualizzare sul feed delle notizie dell'utente) e il modo di ragionare dell'uomo, l'odio ha la possibilità di diffondersi come non mai. In effetti, le persone tendono a reagire più rapidamente e con maggiore forza ai contenuti che generano emozioni negative come la paura e l'odio, mentre gli algoritmi tendono a promuovere contenuti che generano un maggiore coinvolgimento. In questa configurazione, i commenti d'incitamento all'odio hanno quindi trovato terreno fertile per diffondersi rapidamente e pervasivamente.

Bisogna riconoscere che le piattaforme online, le autorità e le organizzazioni non governative (ONG) hanno fatto uno sforzo senza precedenti per prevenire la diffusione dell'incitamento all'odio. Tuttavia, gli approcci standard basati sulla moderazione dei contenuti quali la sospensione dell'account, la cancellazione dei contenuti, l'interdizione[la cosiddetta shadow-banning] possono essere accusati di censura, overblocking e, infine, di ostacolare la libertà d'espressione.

Un'altra possibile soluzione sarebbe quella di intervenire sul funzionamento degli algoritmi di selezione ma ciò presuppone in qualche modo di aver già identificato l'odio, quindi scivoliamo nuovamente nella situazione precedente.

Un ulteriore problema è che queste strategie di identificazione ed eliminazione non possono essere facilmente applicate a discorsi pericolosi, ad esempio contenuti che suscitano odio e divisione ma non rientrano in una delle definizioni formali d'incitamento all'odio.

“Da quando ha iniziato ad arrivare nel nostro paese c'è stato un picco di reati di violenza”. che si basa su una correlazione vera, ma spuria;

GS: I giganti dei social media devono sviluppare un progetto speciale dedicato alla lotta contro l'odio. Come la vederesti? Quali altri modi ci sono per combattere l'incitamento all'odio sui social media?

MG: Mi risulta che si stia attivamente cercando di trovare nuovi modi per combattere l'hate speech online. Sono convinto, però, che dobbiamo cambiare completamente la nostra prospettiva: dobbiamo superare le strategie di identificazione ed eliminazione reattive. Dobbiamo concentrarci su una strategia alternativa utilizzata da alcune ONG. Queste ONG stanno addestrando gli operatori a intervenire direttamente nelle chat contenenti incitamento all'odio scrivendo risposte di testo gentili, chiamate contro-narrative, che hanno lo scopo di contrastare il contenuto d'odio con prove credibili e di impedire che esso si diffonda ulteriormente.

GS: In che modo le persone sui social media possono aiutare nella battaglia contro l'incitamento all'odio?

MG: La gente sta già aiutando le piattaforme social con le loro attività di segnalazione, ma ci sono due problemi principali:

1. Capita spesso che vengano segnalati contenuti che piacciono e che fanno arrabbiare ma che non costituiscono necessariamente incitamento all'odio.
2. alcune persone usano talvolta viene usata questa possibilità di segnalazione per mettere a tacere gli avversari. Per questo motivo le piattaforme multimediali richiedono molto intervento manuale da parte dei moderatori per controllare ogni singola segnalazione degli utenti.

L'idea alternativa di scrivere manualmente le risposte a tutto l'odio online richiede molta esperienza e tempo e alla fine non è un'attività scalabile, è una vera e propria fatica di Sisifo. A questo proposito crediamo che in futuro sarà cruciale unire i vantaggi dell'IA e dell'apprendimento profondo all'esperienza umana per superare la proliferazione e la rapida diffusione del discorso d'odio online. Non vedo l'ora che arrivi una nuova Era dell'Illuminismo in cui la ragione sostituisca l'odio e la censura, dove le differenze possano arricchire la nostra società attraverso il dialogo.

GS: Cosa stanno trovando ultimamente i ricercatori che monitorano l'incitamento all'odio in rete? Quanto è alto il rischio che i discorsi d'odio su Internet diventino veri e propri reati?

MG: Il rischio che l'incitamento all'odio diventi reato d'odio è sempre elevato: può portare a depressione o suicidio dei bersagli, promuovere l'uso della violenza, incoraggiare la discriminazione e aumentare le divisioni sociali. È successo in passato e sfortunatamente accadrà ancora.

Un aspetto chiave dell'odio in rete è che consente alle persone di vedere gli altri usare violenza verbale rendendola normale. Le norme sociali hanno il potere di influenzare i comportamenti delle persone. Se vedi un flusso d'insulti ti viene da pensare che queste cose siano accettabili e in qualche modo "normali". Ecco perché crediamo che dobbiamo entrare nel flusso dell'odio e cambiare le narrazioni violente con narrazioni positive e inclusive. Forse non saremo in grado di far cambiare idea agli hater, ma di sicuro impediremo agli astanti di cadere nella trappola dell'odio.

GS: Il discorso d'odio online ha molte dimensioni: estremismo di destra, islamofobia, razzismo, odio in rete contro la comunità LGBTI + ... Quali dovrebbero essere le risposte diverse da presentare loro?

MG: **Parlando di contro-narrative, mentre ci sono principi generali alla radice di queste risposte (per esempio, è sempre l'individuo che è responsabile di qualcosa, mai il gruppo), ci sono anche "argomenti" che sono specifici per ogni bersaglio – ad es. sfatare tutte le canard sugli ebrei dai "Protocolli degli Anziani di Sion". Questo è il motivo per cui abbiamo bisogno anche di esperti e di progetti educativi per vincere questa lotta.**

LINK

<https://magazine.fb�.eu/it/news/lotta-contro-i-discorsi-dodio-preservando-la-liberta-despressione/>

TAG

- #artificial intelligence
- #counter narratives
- #diversity
- #education
- #freedom
- #hatespeech
- #internet
- #social media

AUTORI

- Giancarlo Sciascia