IDEAs

Inclusive Dialogues for Equal Actions

Contronarrative e moderazione degli HS

CPS CULTURE POLITICA SOCIETÀ

di.unito.it
DIPARTIMENTO DI INFORMATICA

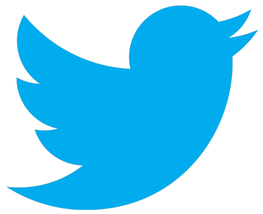
Università di Torino
Dipartimento di Giurisprude

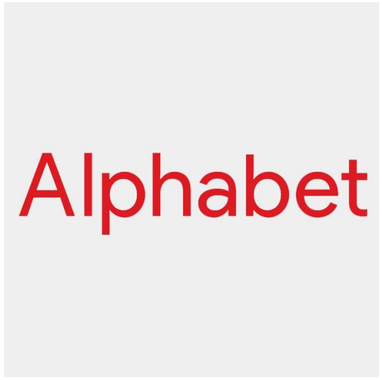
Studi um

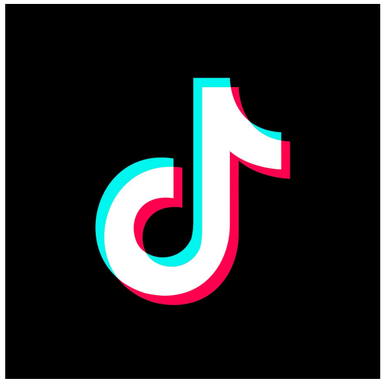
Fisica Torino
Dipartimento di Eccellenza

Il Codice di Condotta per il contrasto allo Hate Speech illegale

Un strumento giuridico non vincolante, che rappresenta un impegno per l'Unione Europea e aziende ICT per contrastare il fenomeno dello Hate Speech online







Gli obiettivi del Codice di Condotta

1. Migliorare le procedure di rimozione di Hate Speech dai social media
2. Sensibilizzare gli utenti di queste piattaforme sul tema

1. Rimozione dei contenuti

Rimozione dei contenuti

Le aziende informatiche **predispongono procedure chiare ed efficaci** per esaminare le segnalazioni riguardanti forme illegali di incitamento all’odio nei servizi da loro offerti, in modo da poter rimuovere tali contenuti o disabilitarne l'accesso. Le aziende informatiche **predispongono regole o orientamenti per la comunità degli utenti** volte a precisare che sono vietate la promozione dell'istigazione alla violenza e a comportamenti improntati all'odio

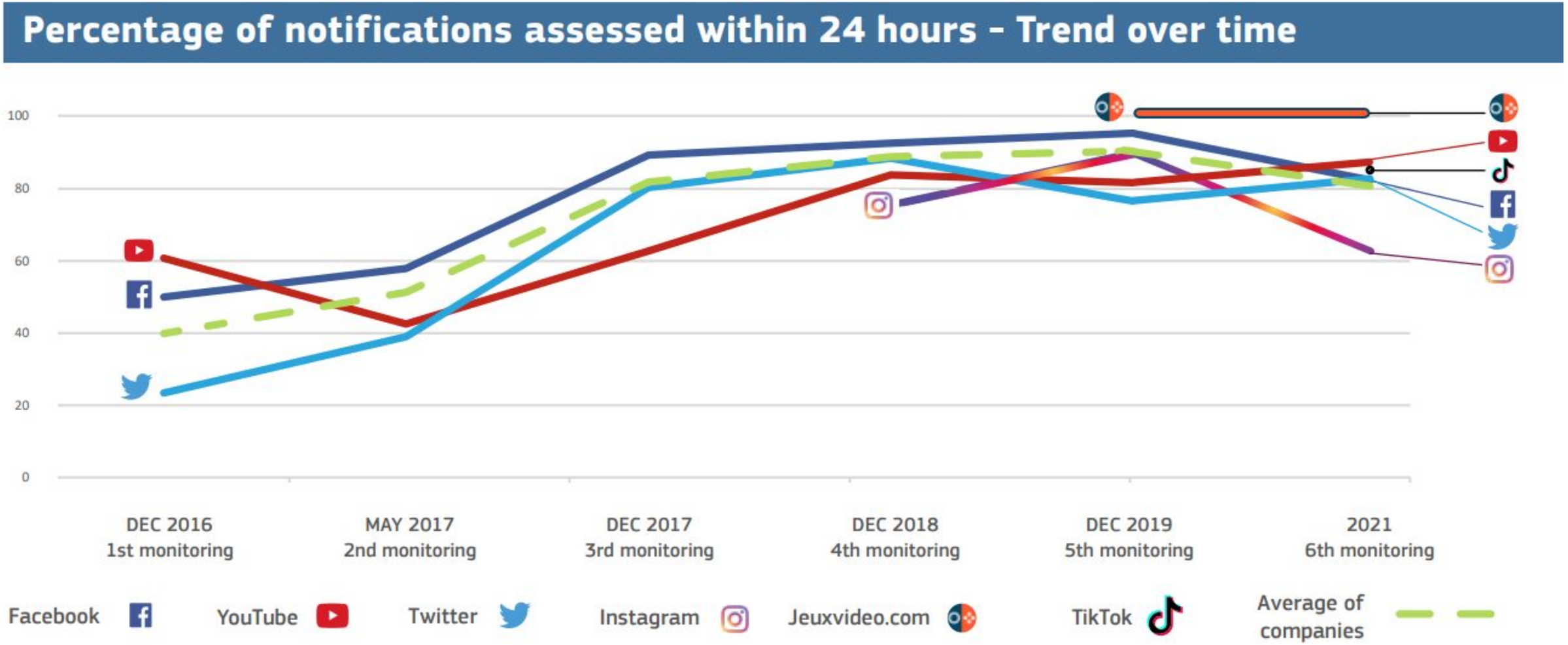
Rimozione dei contenuti

Al ricevimento di una segnalazione valida mirante alla rimozione di forme illegali di incitamento all'odio, **le aziende informatiche la esaminano** alla luce delle regole e degli orientamenti da esse predisposti per la comunità degli utenti

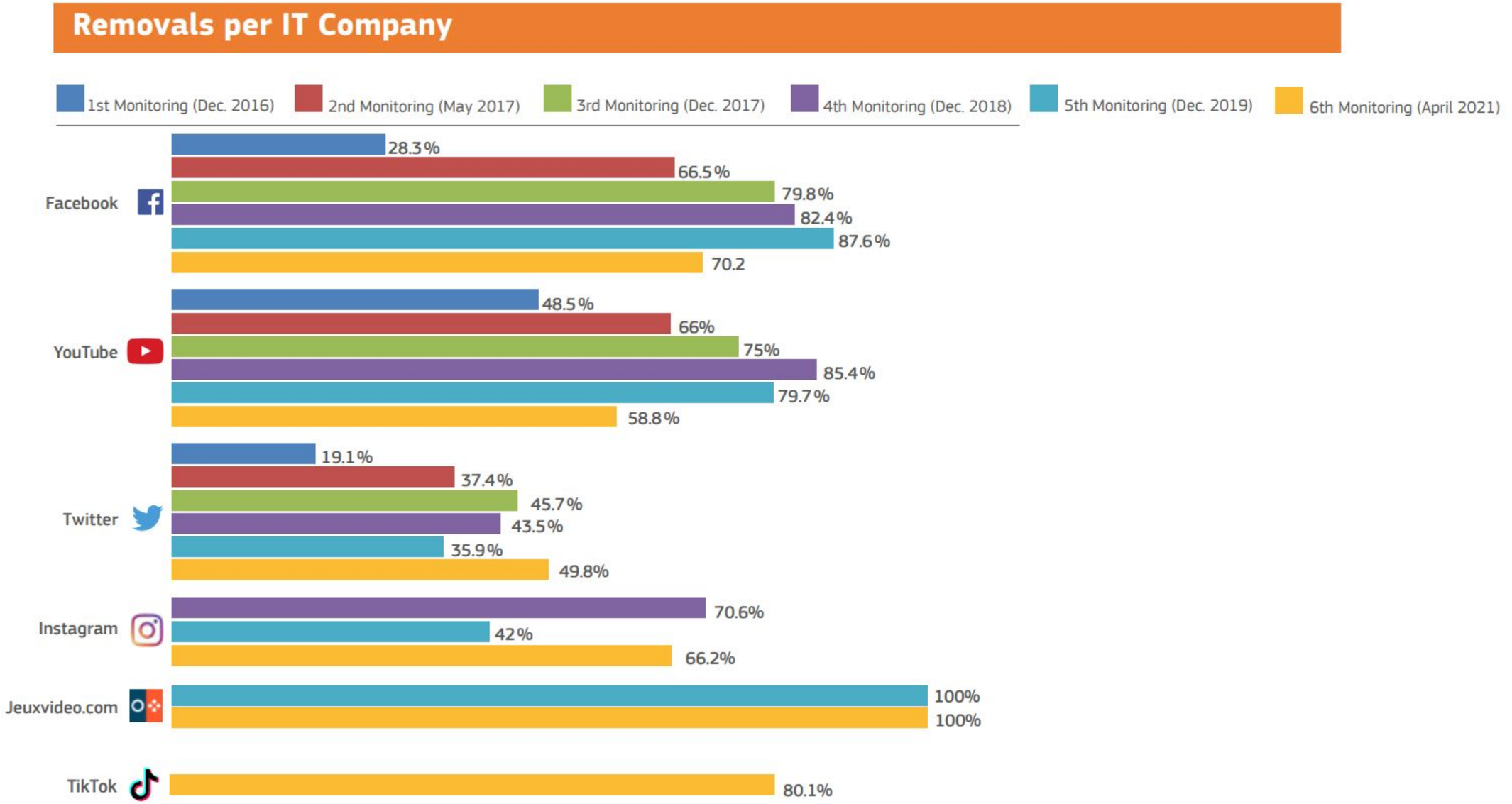
Rimozione dei contenuti

Le aziende informatiche **esaminano in meno di 24 ore la maggior parte delle segnalazioni valide** miranti alla rimozione di forme illegali di incitamento all’odio e, se necessario, rimuovono tali contenuti

Rimozione dei contenuti



Rimozione dei contenuti



Rimozione dei contenuti

Nel primo trimestre del 2021, dei circa 62 milioni di video che abbiamo rimosso a livello globale, il 2% è stato rimosso per aver violato le nostre norme sui comportamenti di odio. Abbiamo rimosso in modo proattivo il 67% dei video che incitano all'odio prima ancora che ci venissero segnalati e l'84% è stato rimosso entro 24 ore dalla pubblicazione.

Rimozione dei contenuti

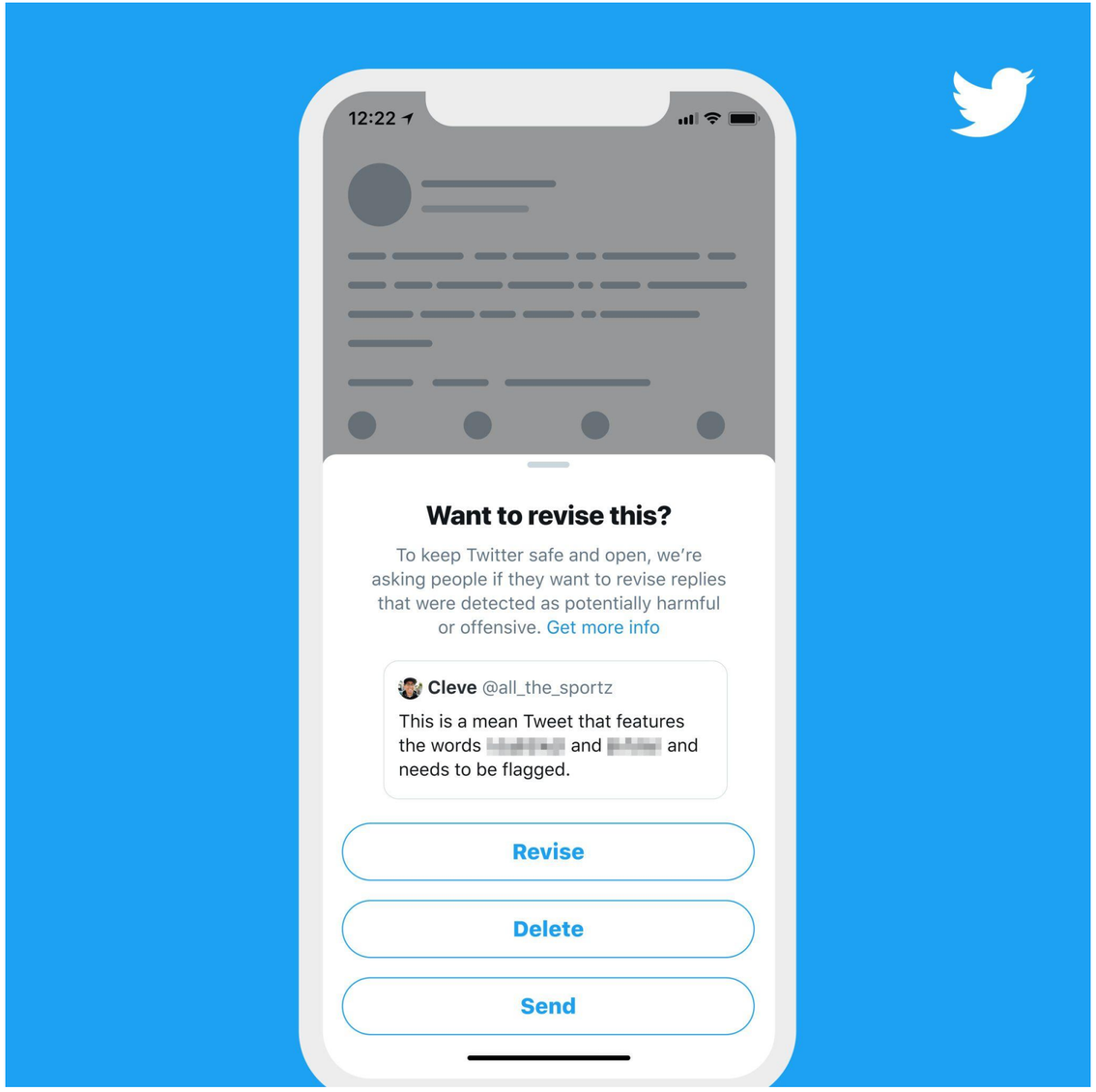
- Abbiamo preso provvedimenti su 9,8 milioni di contenuti di incitamento all'odio su Instagram, rispetto ai 6,3 milioni del primo trimestre del 2021. Di questi contenuti, il 95,1% è stato rilevato in modo proattivo prima di essere segnalato;
- Non abbiamo pubblicato i dati sui contenuti sottoposti a ricorso e ripristinati su Instagram nell'ultimo CSER perché, a causa di una riduzione temporanea della nostra capacità di revisione a causa del COVID-19, non potevamo sempre offrire alle persone la possibilità di presentare ricorso su Instagram. Abbiamo comunque concesso alle persone la possibilità di dirci che non erano d'accordo con la nostra decisione, il che ci ha aiutato a migliorare la nostra precisione.

2. Sensibilizzazione degli Utenti

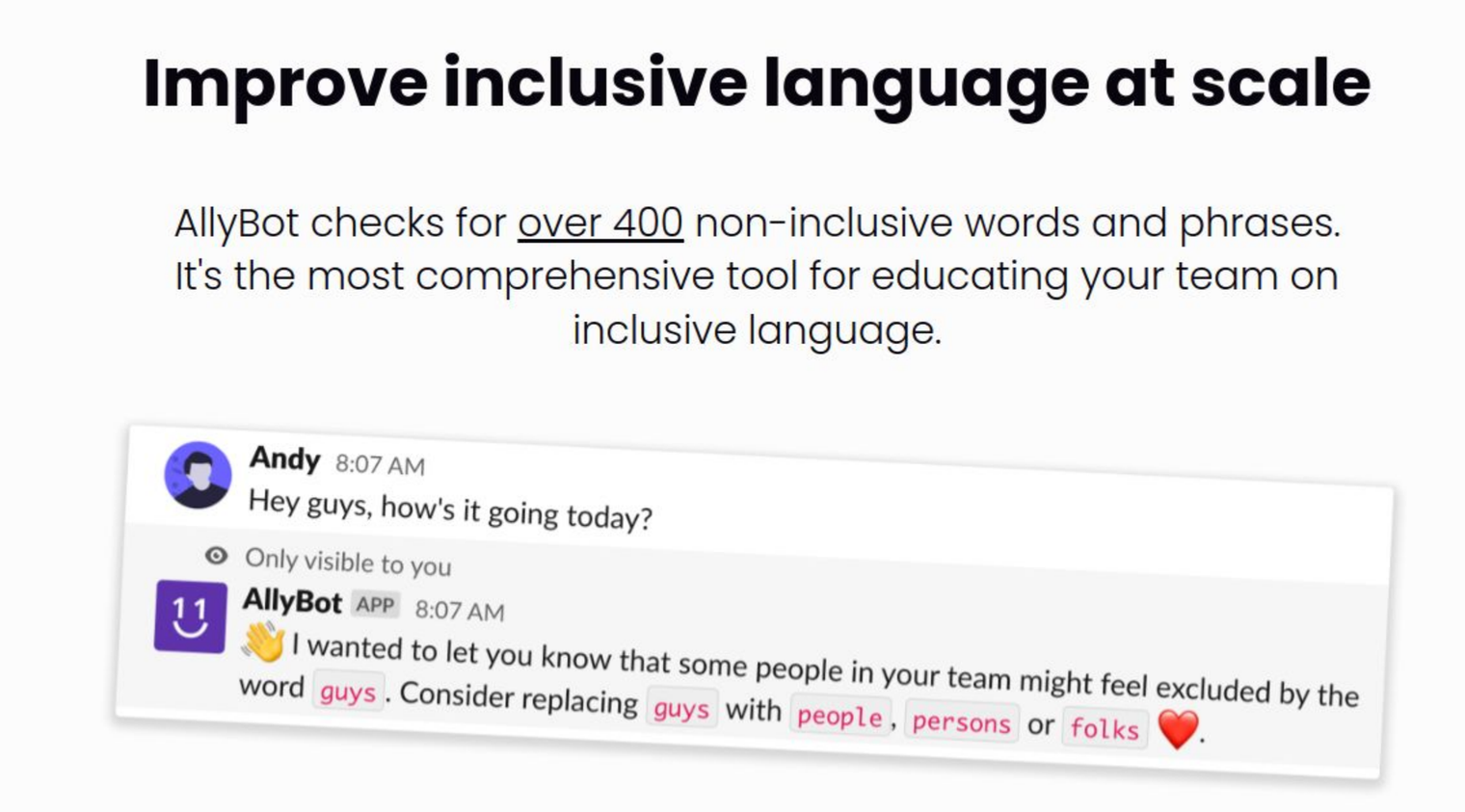
Sensibilizzazione degli utenti

Inoltre, le aziende informatiche **svolgono presso i loro utenti un'opera di educazione e di sensibilizzazione** sulle tipologie di contenuti non autorizzate in base alle regole e agli orientamenti da esse predisposti per la comunità degli utenti.

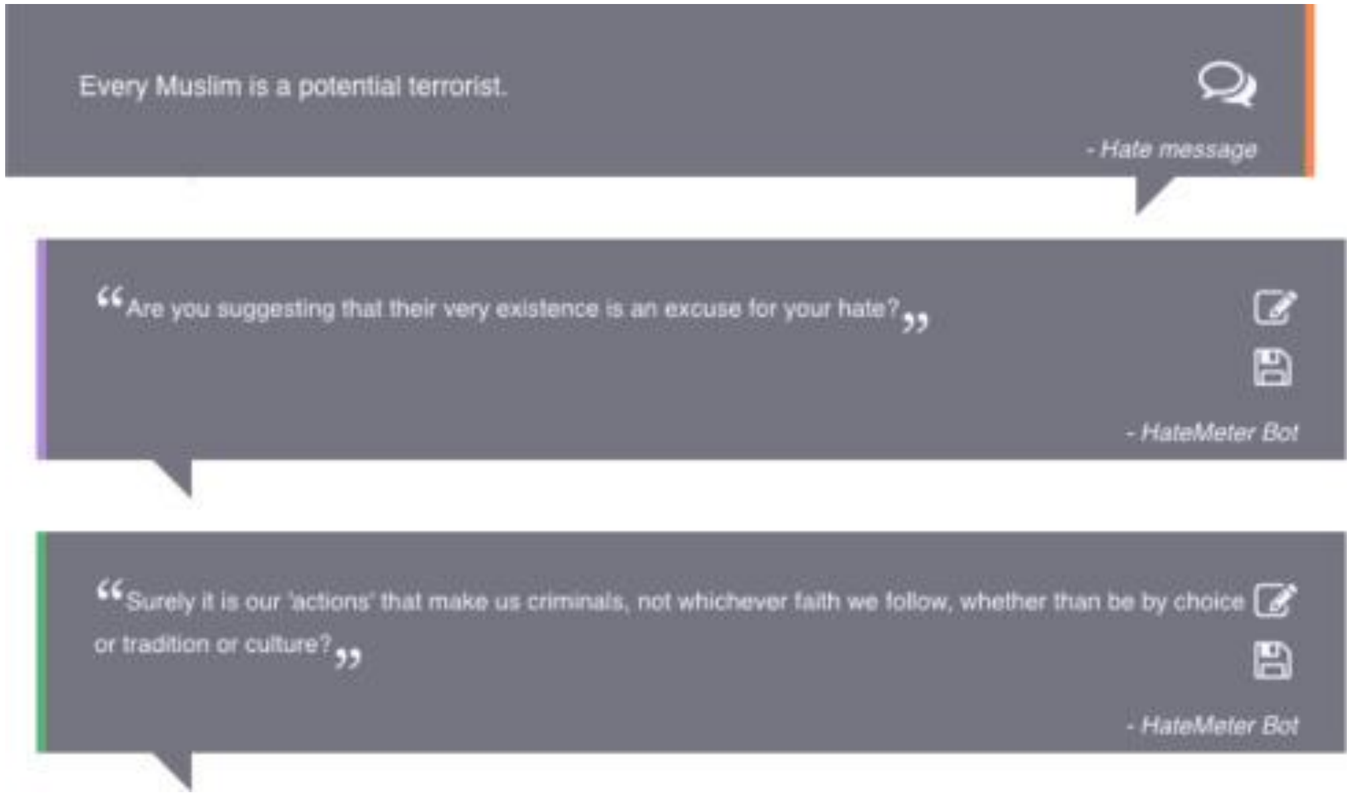
Sensibilizzazione degli utenti



Sensibilizzazione degli utenti



Sensibilizzazione degli utenti



Tre approcci contro l’Hate Speech

1. Rimuovere
2. Rispondere
3. Riscrivere

Esercizio

Dividetevi in 4 gruppi e, dopo aver letto il regolamento della piattaforma e i 15 messaggi che vi abbiamo assegnato:

1. Scegliete uno dei 15 messaggi da **rimuovere** e spiegate perché
2. **Rispondete** a uno dei 15 messaggi
3. **Riscrivete** uno dei 15 messaggi

Esercizio

Linee guida di Youtube: https://bit.ly/3u6Jnbh e https://bit.ly/3veCVOX
Linee guida di META (Facebook e Instagram): https://bit.ly/3NPYb5I
Linee guida di Twitter: https://bit.ly/3DFR6A5