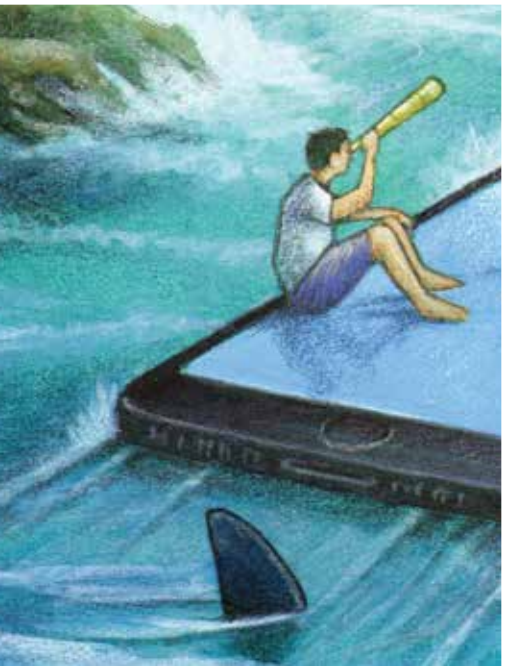
One in a Series of Working Papers from the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression



U.S. Initiatives to Counter Harmful Speech and Disinformation on Social Media

Adrian Shahbaz
Freedom House

June 11, 2019

TRANSATLANTIC WORKING GROUP

The Transatlantic Working Group Papers Series

Co-Chairs Reports

Co-Chairs Reports from TWG's Three Sessions: Ditchley Park, Santa Monica, and Bellagio.

Freedom of Expression and Intermediary Liability

Freedom of Expression: A Comparative Summary of United States and European Law
B. Heller & J. van Hoboken, May 3, 2019.

Design Principles for Intermediary Liability Laws
J. van Hoboken & D. Keller, October 8, 2019.

Existing Legislative Initiatives

An Analysis of Germany's NetzDG Law
H. Tworek & P. Leerssen, April 15, 2019.

The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications
J. van Hoboken, May 3, 2019.

Combating Terrorist-Related Content Through AI and Information Sharing
B. Heller, April 26, 2019.

The European Commission's Code of Conduct for Countering Illegal Hate Speech Online: An Analysis of Freedom of Expression Implications
B. Bukovská, May 7, 2019.

The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem
P.H. Chase, August 29, 2019.

A Cycle of Censorship: The UK White Paper on Online Harms and the Dangers of Regulating Disinformation
P. Pomerantsev, October 1, 2019.

U.S. Initiatives to Counter Harmful Speech and Disinformation on Social Media
A. Shahbaz, June 11, 2019.

ABC Framework to Address Disinformation

Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses
C. François, September 20, 2019.

Transparency and Accountability Solutions

Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry
M. MacCarthy, February 12, 2020.

Dispute Resolution and Content Moderation: Fair, Accountable, Independent, Transparent, and Effective
H. Tworek, R. Ó Fathaigh, L. Bruggeman & C. Tenove, January 14, 2020.

Algorithms and Artificial Intelligence

An Examination of the Algorithmic Accountability Act of 2019
M. MacCarthy, October 24, 2019.

Artificial Intelligence, Content Moderation, and Freedom of Expression
E. Llansó, J. van Hoboken, P. Leerssen & J. Harambam, February 26, 2020.

www.annenbergpublicpolicycenter.org/twg

A project of the Annenberg Public Policy Center of the University of Pennsylvania in partnership with
The Annenberg Foundation Trust at Sunnylands and the Institute for Information Law of the University of Amsterdam,
with support from the Embassy of the Kingdom of the Netherlands in Washington, D.C.

TRANSATLANTIC WORKING GROUP

RESEARCH BRIEF

U.S. Initiatives to Counter Harmful Speech and Disinformation on Social Media†

Adrian Shahbaz, Freedom House

June 11, 2019

Harmful speech

There are limited, if any, legislative efforts in the United States that directly target hate speech. Attempts to combat hate speech through legislation are restricted by (1) its broad definition, (2) the First Amendment, and (3) likely applications against minority groups.

Despite the lack of criminal legislation around hate speech specifically, there are a range of other legal tools available to target similar inflammatory and dangerous speech online. Many of these laws are problematic in that they criminalize behavior with often disproportionate penalties, yet do not take a preventative or structural approach to issues of inflammatory and hateful speech.

- Cyberbullying: A number of states have addressed cyberbullying. For example, a 2017 law in Texas, which received backlash, criminalized bullying of someone under the age of 18 online or via text messages. Another bill in Nebraska would provide materials to school districts to prevent and respond to instances of cyberbullying.

- Cyber-harassment: The federal government does not criminalize cyber-harassment, although some of the behavior could be targeted under other laws such as cyberstalking. Some states target harassment online; California's penal code criminalizes the use of electronic communication equipment to repeatedly contact someone with the intent to harass or annoy.

- Cyberstalking: States generally have anti-stalking laws that can apply to the online sphere, and the federal government has a cyberstalking statute (Title 18 U.S. Code § 2261A). Stalking differs, in part, from harassment due to the repeated nature of the communications.

- Other laws that could address online hate speech include the Violence Against Women Act, hate crime statutes, and other statutes in the U.S Criminal Code. Victims of inflammatory speech can also sue under civil law, although this remains expensive and time-intensive.

Amid fears that tech companies are not effectively monitoring their platforms, recent discussion in Congress has centered on modifying Section 230 of the Communications Decency Act of 1996. This

† A research brief prepared by Adrian Shahbaz, research director for technology & democracy at Freedom House, a nonprofit, independent watchdog organization, for the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. Read about the TWG here: https://www.ivir.nl/twg/.

provision generally shields intermediaries (such as social media companies and website owners) from legal liability for the activities of their users, although there are exceptions for criminal and state law (e.g., on harassment, stalking, protecting children), intellectual property law, and sex trafficking law. The latter is a recent development.

- In March 2018, Congress passed the Stop Enabling Sex Traffickers Act and Allow States and Victims to Fight Online Sex Trafficking Act of 2017, or SESTA/FOSTA, intended to address sex trafficking facilitated online. It represents one of the few legislative changes to intermediary liability in recent years. However, the law has had the unintended consequence of pushing companies to preemptively remove legitimate content, and sex workers and community advocates argue that it threatens their safety since targeted platforms – such as Backpage.com and sections of Craigslist – made it possible for sex workers to flee exploitive situations, communicate with one another, and build protective communities. The only two Senators to vote against SESTA were Ron Wyden (D-Ore.) and Rand Paul (R-Ky.).

- Sen. Wyden, a coauthor and staunch supporter of Section 230 as a fundamental pillar of the internet, argues that companies have embraced only the part of the law that provides protection against liability and have not actively used their censorial discretion to remove unwanted or illegal content.

- Wanting to modify the law, Sen. Mark Warner (D-Va.) published a 2018 white paper suggesting that platforms should be ordered to remove content after a court deems it defamatory or invading privacy, among other material.

- In a more aggressive approach against Section 230, Sen. Ted Cruz (R-Texas) argues that the provision requires that platforms be "neutral public forums," which could disqualify companies like Facebook from receiving special immunity if they act as "political speakers" when, as he contends, they operate with anti-conservative bias.

- Similarly, Rep. Devin Nunes (R-Calif.), who filed a lawsuit against Twitter, argues that Section 230 should not apply to the platform because it is a content creator and is politically biased against him. Likewise, Sen. Josh Hawley (R-Mo.) has also raised "viewpoint discrimination" in efforts to change Section 230.

When interpreting laws that relate to online speech, courts have generally upheld First Amendment protections. For example, in the 2014 case *Elonis v. United States*, the Supreme Court reversed the conviction of Anthony Elonis for threatening to kill his ex-wife. The conviction hinged on Elonis' threatening Facebook comments about his ex-wife, colleagues, a kindergarten class, local police, and an FBI agent. Specifically, the Court reversed the standard that allowed for criminal liability if a "reasonable person" would understand the accused's words as a threat, but ruled narrowly to only address principles of the accused's intent and not questions around whether there are "true threats" of violence.

Disinformation and foreign propaganda

A number of legislative efforts at both the federal and state levels have targeted foreign disinformation by promoting greater transparency around online advertising and foreign news, as well as by promoting digital media literacy.

Senators Amy Klobuchar (D-Minn.) and Warner, with endorsement from Sen. John McCain (R-Ariz.), introduced the Honest Ads Act in October 2017, which would require those who purchase and publish online political advertisements to disclose information about the ads to the public. In April 2018, Twitter announced its support of the act. The specifics of the bill include:

- Incorporating paid internet and digital advertisements in the definition of electioneering communication under the Bipartisan Campaign Reform Act of 2002
- Forcing platforms and websites with over 50 million unique visitors each month to publicly document people or groups spending more than $500 on political ads
- Mandating platforms to make "all reasonable efforts" to not allow foreign individuals and groups to advertise online

There have been a number of legislative efforts targeting disinformation, or viral deception, originating from foreign actors. For example, the FY 2017 National Defense Authorization Act (NDAA) incorporated the Countering Disinformation and Propaganda Act. The text created the Global Engagement Center, an interagency body housed in the Department of State that coordinates counter-propaganda efforts across the government, and also provided grant opportunities for civil society groups to work on related issues. The FY 2018 NDAA, building off its 2017 version, again included components aimed at countering foreign propaganda and disinformation. The FY 2018 omnibus appropriations bill included $250 million for a new "Countering Russian Influence and Aggression Fund." The FY 2019 omnibus increased this amount to $275 million.

At least 24 states have introduced bills establishing a council or committee focused on comprehensive media literacy education. For example, in September 2018, California passed a law encouraging media literacy in schools by forcing the state's Department of Education to provide online resources on best practices to analyze and evaluate the news. Similarly, a 2017 Connecticut law created a council in their Department of Education addressing digital citizenship, internet safety, and media literacy. In another example, a bill in Florida would require public schools to teach fifth and sixth graders how to responsibly use social media.

There have also been renewed efforts to enforce or update the 1938 Foreign Agents Registration Act (FARA) in a bid to increase transparency around the foreign funding of media outlets. Al Jazeera, RT and Sputnik, China Daily, Korean broadcaster KBS America, and Japanese broadcaster NHK Cosmomedia have registered under the law. Some civil society organizations have criticized the use of FARA against the media, noting that it could lead to politicized targeting of outlets.

Civil society initiatives

The private sector and civil society have been more actively engaged in tackling these issues. A joint Stanford-Oxford report contains a helpful primer on "What Facebook Has Done" on content moderation and News Feed controls, and WhatsApp (owned by Facebook) took measures aimed at combating the viral spread of false information through limiting users' abilities to bulk forward messages. Google also announced a series of actions to increase the integrity of news displayed on its platform. Civil society action can be categorized into new fact-checking initiatives and partnerships, increased investment and training on how reporters can verify user-generated content, and programs to increase digital media literacy among the population. There are also well-funded initiatives like the Credibility Coalition and First Draft that aim to establish standards for online content, provide educational resources, and conduct empirical research on best practices for combating misinformation.