TRANSATLANTIC WORKING GROUP

Co-Chairs Report No. 2: The Santa Monica Session

Susan Ness, Annenberg Public Policy Center
Nico van Eijk, Institute for Information Law, University of Amsterdam

July 22, 2019

Introduction

The Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression (TWG) convened its second session from May 9-12, 2019, at the Annenberg Community Beach House in Santa Monica, California.

Our first session, held February 27-March 3 at Ditchley Park in the United Kingdom, focused primarily on analyzing U.S. and European approaches to freedom of expression, and how these approaches could inform the ongoing initiatives to address hate speech and terrorism online. In particular, it examined the experience of four key initiatives to address online speech: Germany's Network Enforcement Act, or "NetzDG"; the EU's proposed Terrorism Content Regulation; the EC's Code of Conduct on Countering Illegal Hate Speech Online; and the Global Internet Forum to Combat Terrorism's "Hash-Sharing" Database. Our report on conclusions drawn from that discussion can be found here.

In Santa Monica, we reviewed recent developments in each of these areas, as well as the implications of the fallout from the tragic events in Christchurch, New Zealand. Among other things, we noted that:

- Increasingly, countries are moving toward statutory regulation of content moderation by online intermediaries, rather than improving the existing self- and co-regulatory mechanisms;
- Companies have tended toward over-removal (both based on their terms of service and in response to the increase in legally mandated short-removal times). They also lack independent oversight mechanisms for their content removal policies and practices under their terms of service and lack redress for such practices; and
- Current indicators of "success" for moderation policies, which tend to focus on the overall volume of content removed, are deficient. Other outcomes such as demonetizing content or reducing its visibility as well as the availability of redress should also be measured.

We then turned to the main theme of our second session: efforts to address "disinformation" or "viral deception," the term coined by Professor Kathleen Hall Jamieson to capture both intent to deceive *and* to disseminate. In contrast to illegal hate speech and incitement to violence, deceptive speech is not necessarily illegal. Arguably, it is protected in our transatlantic societies by freedom of expression and/or the First Amendment. That said, politicians, policymakers and the public increasingly see disinformation as causing serious societal harms, even when the content is not false but intentionally misleading. The rapid and broad (viral) dissemination may have been boosted artificially by "bots"

and fake accounts, by commercial actors (and sometimes even government officials), often with a malicious intent to weaponize social divisions, distrust in institutions, and other societal ills.

The group considered a number of specific initiatives that have either been adopted or are being considered in the United States and Europe to address disinformation:

- The EC Code of Practice on Disinformation;
- The United Kingdom's White Paper on Online Harms; and
- The Algorithmic Accountability Act, recently introduced in the Senate by Senators Ron Wyden and Cory Booker.

The TWG also discussed viral deception caused by or on behalf of a foreign government as part of an information operation. Government responses to information ops have a different set of tools available, ranging from diplomacy to sanctions to internet service denial.

Finally, the group began its review of intermediary liability in light of efforts underway in both Europe and the United States to condition or restrict the present forms of safe harbor for online platforms.

The background papers prepared for this session will be revised in light of the Santa Monica discussions and posted on the TWG website.

Key findings and recommendations

As co-chairs of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, we offer the following preliminary observations and recommendations, culled from the discussions in Santa Monica. There is no attribution, as the discussion proceeded under the Chatham House Rule. Members of the Working Group have reviewed our report and their comments generally are reflected in our conclusions below.

Adopt "Freedom of Expression by Design" as a guiding principle

First articulated at our Ditchley Park Session, this concept has even greater currency in the context of disinformation. Freedom of expression is a fundamental right underpinning our democracies, and is essential for holding governments accountable. Where speech – online or offline – clearly is illegal, it should be addressed according to applicable law.

When speech is not clearly illegal, governments must exercise extreme caution and refrain from requiring deletion either directly or indirectly (by essentially deputizing companies to take down offending speech). Governments and internet companies should consider positive measures instead, such as increasing both government and platform transparency regarding takedown requests, raising public awareness, investing in media literacy, and encouraging funding for high-quality news reporting and fact-checking. To avoid an actual or perceived conflict of interest, governments should refrain from directly funding news outlets or fact-checking organizations.

Focus on Actors and Behavior, rather than on Content that is odious but legal

The "A-B-C" analytical framework, which was presented in one of the papers and discussed during the meeting, helps policymakers to focus not on *Content* but rather on bad *Actors* and deceptive

Behavior. As content, "fake news" and "disinformation" are part of our democratic landscape, and are not *per se* illegal. Governments should not trigger the removal of undesirable, but legal, content, as such action is inconsistent with freedom of expression.

Governments may choose to address harmful behavior on the internet, where content is artificially propagated by bad actors – "bots, "astroturfers," or "troll farms" – as such conduct no longer reflects an authentic dialogue among citizens. This distortive behavior is akin to "spam," which companies have the technical expertise to address. But a cautionary note: some fake identities might be legitimate and should be protected, such as whistle-blowers calling attention to government corruption.

Strengthen enforcement of rules on foreign government interference

Foreign governments, too, have a right to have their voices heard in policy debates; that is diplomacy. But such engagement must be open and transparent. The United States as well as many European countries restrict interference from foreign governments in domestic political debates. Jurisdictions often prohibit foreign governments from making financial or in-kind contributions to political campaigns, and require foreign governments to register and report on their lobbying activities.

Covert foreign government manipulation of public opinion through artificial amplification and disinformation, or "information operations," is often deployed through multiple online channels and coordinated with real-world actions. These foreign governments may strive to deepen societal fissures by supporting both sides of contentious social issues. Such activities well may be illegal and better addressed through government channels, where additional tools are available, such as diplomatic pressure, sanctions, and disruption of internet service. Governments should determine whether and how to respond to such campaigns, bearing in mind that concerns about "information warfare" can be repurposed by authoritarian regimes to justify actions to impose "information sovereignty" within their borders.

Relevant European and American government agencies should strengthen collaboration against these "hybrid" tactics through NATO and other organizations, and should work cooperatively with companies and civil society to identify and derail such attacks.

Strengthen transparency and accountability

Companies should ensure that their terms of service and community standards are clear and accessible. Users whose content is deemed unacceptable and then removed or downgraded should be notified and provided a pathway for prompt redress. Both platforms and governments should disclose as much information as possible about enforcement actions taken.

In enforcing terms of service violations involving content, platforms should consider a variety of actions that lessen the impact on freedom of expression, including reducing content visibility through deceleration and demonetization, as well as deletion.

In an interim report in May, the French government suggested creation of a new regulatory regime to oversee both the transparency and accountability of platform content-moderation systems, rather than ruling on the content itself, to protect freedom of expression. It is an intriguing concept that deserves wider consideration.

During an election season, special attention should be given to both candidate and social issue advertising, as such communications are integral to the electoral process. If narrowly drafted,

governments could require specific disclosures for microtargeted candidate and social issue ads that state why the ad is being seen, the screening criteria, who paid for the ad, and the amount spent.

Transparency should require the logging and archiving of relevant data, to be made available for legitimate research purposes while guarding user privacy. Some platforms specifically block researchers from examining how their terms of service are enforced. Such restrictions should be lifted.

Consider an online court system or other independent body to adjudicate content moderation decisions

One proposal to resolve the sometimes conflicting roles of users and intermediaries is to create a system of specialized online courts that could quickly hear and adjudicate these disputes based on the digital record. These "e-courts" could be fast, simple and cheap; they would operate entirely online with no physical presence of complainant or defendant and no right of appeal (but still leave open the choice to file the case in the regular court system in lieu of the internet court). They would focus on whether content removal violated freedom of expression (based on the law of the complainant's jurisdiction); use specially trained magistrates; and, over time, build a public record of published decisions to serve as guideposts. Such a system could reduce the number of inappropriate removals, and could also protect platforms against undue government pressure to remove content that is troublesome but not illegal. The TWG will further develop this concept at its third session in November.

Separately, an independent body could be empanelled to review and redress cases of content removal or inappropriate termination of accounts and to provide guidance for platforms in novel situations such as the Christchurch attack. The selection of members, scope of authority, and scalability of social media councils are among the factors that the TWG should flesh out in the months ahead.

Be cautious if considering changing intermediary liability laws

Both in Europe under the e-Commerce Directive and in the United States under CDA Section 230, internet intermediaries have been protected to some degree against liability for content posted by users, in part to protect freedom of expression, but also to promote innovation and economic growth. These "safe harbor" protections are being revisited in Europe and North America, as legislatures and the public press for conditioning protection on proactive removal of troubling content. They want intermediaries to assume greater responsibility – a "duty of care" or even liability – for the actors, behavior and content on their platforms. The largest platforms often are viewed as controlling the public square.

The elimination of liability protections would likely result either in over-removal of lawful content, thus limiting freedom of expression, or passive posting of user content without moderation, thus elevating the amount of hate speech and viral deception online.

More nuanced approaches may offer alternatives to reducing intermediary liability protections. The TWG discussed an initial briefing paper on intermediary liability, which will be revised to participate in the public debate.

Promote media literacy, quality journalism, and fact checking

Viral deception is most effective when citizens are unaware of malicious attempts to influence their behavior. That impact can be reduced if the public knows how to identify stories that are false or misleading and promoted for malevolent ends. Governments have a duty to provide digital literacy education, not just for children but also for adults.

One tool in the fight against "disinformation" is serious fact-checking, although its scalability and effectiveness are limited. Major social media companies are investing in quality journalism and in respected fact-checking organizations. Platforms should be transparent about these efforts and protect the independence of these organizations. While elevation of trustworthy news sources is appropriate, there is a significant risk that lesser-known yet quality sources will be down-ranked, presenting risks to freedom of expression.

Governments should support and promote efforts to strengthen fact-checking organizations and journalism, provided that they maintain an arms-length relationship to preserve the independence of these entities.

Good Corporate Governance Encompasses Good Corporate Citizenry

Today's economy depends on a vibrant, global internet. Most internet companies, large and small, are legitimate and beneficial private-sector actors in our economies. To the extent that they give voice to the public by uploading their content, they contribute to democratic discourse and freedom of expression. But the internet has also spawned bad actors that take advantage of the openness of the network to rip apart the fabric of society.

As good corporate citizens, platforms should work proactively with policymakers and stakeholders to find scalable solutions to make the internet as safe and beneficial as possible while respecting freedom of expression. Solutions should take into account the size and variety of companies involved.

Governments should also strengthen consumer protection rules to ensure that platforms engage in appropriate behavior toward their users and other ecosystem companies.

Next Steps

Our final Transatlantic Working Group Session in November will examine:

- best practices in the use of artificial intelligence to address harmful content including algorithmic accountability;
- platform and government transparency;
- policy recommendations on intermediary liability; and
- policy recommendations for internet courts and social media standards councils.

During the third quarter, we will hold roundtables with stakeholders for additional feedback and engagement.

Our final report will be released at the end of the March 2020.