



FOR IMMEDIATE RELEASE

June 17, 2019

Contact: Michael Rozansky | michael.rozansky@appc.upenn.edu | 215.746.0202

How governments and platforms have fallen short in trying to moderate content online

Efforts to curb hate speech, terrorism, and deception while protecting free speech have been flawed

PHILADELPHIA and AMSTERDAM – The Transatlantic Working Group, created to identify best practices in content moderation on both sides of the Atlantic while protecting online freedom of expression, has released its first working papers that explain why some of the most prominent efforts to date have failed to achieve these goals.

In a series of papers, members of the group find that current efforts by governments and platforms are flawed and have fallen short of the goals of adequately addressing the problems of hate speech, viral deception, and terrorist extremism online while protecting free speech rights.

The Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, or TWG, also released a report of the co-chairs with interim recommendations for government and platforms coming out of its initial meeting earlier this year in Ditchley Park, U.K. The co-chairs are former Federal Communications Commission member **Susan Ness**, a distinguished fellow of the Annenberg Public Policy Center of the University of Pennsylvania, and **Nico van Eijk**, professor of law and director of the Institute for Information Law (IViR) at the University of Amsterdam.

In their first report ([download here](#)), the TWG co-chairs recommended that:

- Specific harms to be addressed by content moderation must be clearly defined and based on evidence and not conjecture;
- Transparency must be built in – both by governments and platforms – so that the public and other stakeholders can more accurately judge the impact of content moderation;
- Due process safeguards must be provided so that authors of material taken down have clear and timely recourse for appeal;
- Policy makers and platforms alike must understand the risks of overreliance on artificial intelligence, especially for context-specific issues like hate speech or disinformation, and should include an adequate number of human reviewers to correct for machine error.



The set of working papers includes:

- **Freedom of Expression:** A pillar of liberal society and an essential component of a healthy democracy, freedom of expression is established in U.S., European, and international law. This paper looks at the sources of law, similarities and differences between the U.S. and Europe, and how laws on both sides of the Atlantic treat hate speech, violent extremism, and disinformation. [Download](#)
 - **Brittan Heller**, The Carr Center for Human Rights Policy, Harvard University
 - **Joris van Hoboken**, Vrije Universiteit Brussels and University of Amsterdam
- **An Analysis of Germany's NetzDG Law:** Arguably the most ambitious attempt by a Western state to hold social media platforms responsible for online speech that is deemed illegal under a domestic law, German's Network Enforcement Act took effect January 1, 2018. While instituting some accountability and transparency by large social media platforms, its provisions are likely to have a chilling effect on freedom of expression. [Download](#)
 - **Heidi Tworek**, University of British Columbia
 - **Paddy Leerssen**, Institute for Information Law (IViR), University of Amsterdam
- **The Proposed EU Terrorism Content Regulation:** The European Commission proposed the Terrorism Content Regulation (TERREG) in September 2018, and it is currently in the final stages of negotiation between the European Parliament and the Council. As drafted, the TERREG proposal raises issues of censorship by proxy and presents a clear threat to freedom of expression. [Download](#)
 - **Joris van Hoboken**, Vrije Universiteit Brussels and University of Amsterdam
- **Combating Terrorist-Related Content Through AI and Information Sharing:** Through the Global Internet Forum to Counter Terrorism, the tech industry uses machine learning and a private hash-sharing database to flag and take down extremist information. The analysis of this private sector initiative raises transparency and due process issues, and offers insight into why AI failed to promptly take down the Christchurch, New Zealand, shooting videos. [Download](#)
 - **Brittan Heller**, The Carr Center for Human Rights Policy, Harvard University
- **The European Commission's Code of Conduct for Countering Illegal Hate Speech Online:** The Code, developed by the European Commission in collaboration with major tech companies, was introduced in 2016. As a voluntary code, it was seen as less intrusive than statutory regulation. But the code is problematic: It delegates enforcement actions from the state to platforms, lacks due process guarantees, and risks excessive interference with the right to freedom of expression. [Download](#)
 - **Barbora Bukovská**, ARTICLE 19

The full set of papers and the co-chairs report may be [downloaded as a single PDF](#).

The Transatlantic Working Group consists of more than two dozen representatives of government, legislatures, the tech industry, academia, journalism, and civil society organizations in search of common ground and best practices to reduce online hate speech, terrorist extremism and viral deception without harming freedom of expression. The group is a project of the [Annenberg Public Policy Center](#) (APPC) of the University of Pennsylvania in partnership with the [Institute for Information Law](#) (IVIIR) at the University of Amsterdam and [The Annenberg Foundation Trust at Sunnylands](#). Additional support has been provided by the Embassy of the Kingdom of the Netherlands.

For a list of TWG members, [click here](#).

Additional reading about these issues and the TWG:

- [How \(Not\) to Regulate the Internet](#) (Peter Pomerantsev, The American Interest, June 10, 2019)
- [Regulating the Net is Regulating Us](#) (Jeff Jarvis, Medium, May 31, 2019)
- [A Lesson From 1930s Germany: Beware State Control of Social Media](#) (Heidi Tworek, The Atlantic, May 26, 2019)
- [What Should Policymakers Do To Encourage Better Platform Content Moderation?](#) (Mark MacCarthy, Forbes, May 14, 2019)
- [Proposals for Reasonable Technology Regulation and an Internet Court](#) (Jeff Jarvis, April 1, 2019)
- [Protect Freedom of Speech When Addressing Online Disinformation, Transatlantic Group Says](#) (APPC, March 6, 2019)
- [Transatlantic Working Group Seeks To Address Harmful Content Online](#) (APPC, Feb. 26, 2019)
- [Wake Up Call](#) (Susan Ness, Medium, Oct. 12, 2018)

The [Annenberg Public Policy Center](#) (APPC) was established in 1993 to educate the public and policy makers about the media's role in advancing public understanding of political, health, and science issues at the local, state, and federal levels. Follow us on [Facebook](#) and Twitter [@APPCPenn](#).

For information on the Privacy Policy of the University of Pennsylvania, please [click here](#).

If you would like to stop receiving news about the Transatlantic Working Group from the Annenberg Public Policy Center of the University of Pennsylvania, [click here](#) to send an email to be removed from this list.