TRANSATLANTIC WORKING GROUP

FOR IMMEDIATE RELEASE
November 12, 2019
Contact: Michael Rozansky | michael.rozansky@appc.upenn.edu | 215.746.0202

A 'Disinformation ABC' & other strategies to address malevolent speech online

Policy makers should focus on Actors and Behavior rather than Content, paper says

PHILADELPHIA and AMSTERDAM – In addressing disinformation online, members of the Transatlantic Working Group encourage policy makers to focus on bad actors and deceptive behavior, and avoid basing decisions on content, because much of what is characterized as "fake news" or "viral deception" may be odious but protected speech in a democratic society.

"Governments should refrain from triggering the removal of undesirable, but legal, content, as such action is inconsistent with freedom of expression," said Susan Ness, the group's co-chair and founder, in releasing a report and a set of working papers concerning U.S. and EU efforts to control disinformation.

In their second set of working papers, members of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression (TWG) examine codes of online content regulation including the EU Code of Practice on Disinformation, the UK White Paper on Online Harms, and the proposed U.S. Algorithmic Accountability Act of 2019, as well as the intermediary liability safe harbors under Section 230 of the (U.S.) Communications Decency Act and the EU's e-Commerce Directive. Refined following the TWG's second meeting, from May 9-12, 2019, in Santa Monica, Calif., the papers represent the opinions of the authors, informed by the discussions. The TWG session was co-sponsored by the Annenberg Public Policy Center of the University of Pennsylvania and the Annenberg Foundation Trust at Sunnylands.

The group will hold its third and final session this week, hosted by the Rockefeller Foundation Bellagio Center in Como, Italy.

The new report (download here), highlights best practices from the discussions:

- Tech companies and governments should adopt "freedom of expression by design" as a guiding principle;
- Companies should strengthen transparency and accountability and ensure that their terms of service and community standards are clear, accessible, and consistently enforced, with mechanisms in place for users seeking redress from adverse decisions;

IVIR

THE ANNENBERG PUBLIC POLICY CENTER OF THE UNIVERSITY OF PENNSYLVANIA

- Wider consideration should be given to a French government proposal for a new regulatory regime that would oversee transparency and accountability of platform content-moderation systems, rather than ruling on the content itself;

- An online court system or other independent body should be created to swiftly adjudicate content moderation decisions;

- Policy makers should use caution in modifying liability safe harbors for online services that display user-generated content to avoid over-deletion of legal content or under-removal of hateful, violent, or malevolently deceptive content;

- Governments should strengthen enforcement of rules on foreign government interference – and consider the use of tools such as diplomatic pressure, sanctions, and disruption of internet service, if necessary, to deter illegal foreign activity;

- Platforms should act as good corporate citizens and work proactively with policy makers and other stakeholders, including independent researchers, to find scalable solutions to make the internet as safe and beneficial as possible while respecting freedom of expression.

Working papers: Online harms, disinformation, and artificial intelligence

The set of working papers and authors includes:

- **The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem**: Announced by the European Commission in 2018, the code was heralded as the first such self-regulatory initiative in the world. While it is clearly making a difference in the behavior of the largest digital platforms, according to the author, the "structurally incoherent" code will not solve the ill-defined problem of addressing harmful but not illegal content. Digital platforms are only one part of a broader internet ecosystem that must be enlisted in this effort. Appendices include data from Google, Facebook and Twitter. Download

    - **Peter H. Chase**, The German Marshall Fund of the United States

- **Actors, Behavior, Content: A Disinformation ABC**: "A" is for manipulative actors, "B" is for deceptive behavior, and "C" is for harmful content. This ABC framework lays out three key vectors characteristic of viral deception in order to guide regulatory and industry responses while minimizing the impact on freedom of expression. The author notes that while the public debate in the U.S. is largely concerned with actors – who is a Russian troll online? – the technology industry has invested in identifying and addressing malevolent, coordinated behavior, while governments have focused on content. Download

    - **Camille François**, Graphika and Berkman Klein Center, Harvard University

- **A Cycle of Censorship: The UK White Paper on Online Harms and the Dangers of Regulating Disinformation:** This two-part paper offers an analysis of the Online Harms White Paper published in March 2019, which proposes a regulatory model for mitigating what

it calls "harms" to society. Critics have cited the White Paper's hazy definition of "harm" and stressed its chilling effect on free speech. The second part of the paper recommends shifting the focus to transparency, and away from content, to empower people with more information; provide news organizations with more insight; and serve as a unifying principle for regulating the internet and promoting democratic values. Download

  - **Peter Pomerantsev**, Institute of Global Affairs, London School of Economics

- **Design Principles for Intermediary Liability Laws:**

  Recognizing the fundamental role that intermediary liability safe harbors have played in the growth and innovation on the internet and acknowledging the changed public perception due to the spread of online harms, this paper provides a tool kit for policy makers considering redesigning the present regime. It outlines the "dials and knobs" by which changes could be made and the impact such changes would have on freedom of expression and innovation. This analysis looks at the provisions of intermediary liability laws and the safe harbors offered by the U.S. Communications Decency Act of 1996, Section 230; and the EU's e-Commerce Directive. Download

  - **Joris van Hoboken**, Vrije Universiteit Brussels and University of Amsterdam
  - **Daphne Keller**, Stanford Center for Internet and Society

- **An Examination of the Algorithmic Accountability Act of 2019**: While unlikely to receive congressional approval as a stand-alone bill, the proposal could become part of a national privacy law being discussed in congressional committees. The act would require companies to assess automated decision-making systems for risks to privacy and risks of biased or discriminatory decisions, which would apply to the AI systems platforms use to detect and counter hate speech, terrorist extremism, and disinformation. This analysis notes that AI-guided content moderation decisions that contain built-in biases can have serious consequences for individuals in health care, credit, housing and education. Download

  - **Mark MacCarthy**, Georgetown University

- **U.S. Initiatives to Counter Harmful Speech and Disinformation on Social Media**: Though there are limited, if any, legislative efforts in the U.S. that directly target hate speech, there are other legal tools available to address inflammatory and dangerous speech online, including measures on cyberbullying, cyber-harassment, cyberstalking, and hate crime statutes. In addition, the Honest Ads Act would compel disclosure of information about political ads. Download

  - **Adrian Shahbaz**, Freedom House

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 current or former officials from government, legislatures, the tech industry, academia, journalism, and civil society organizations in North America and Europe. They search for 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 (IViR) 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.

Learn more about these issues on "The New Censorship," a BBC radio documentary from TWG member Peter Pomerantsev.

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.