MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®

CHARACTERIZATION OF CURRENT EFFORTS TO UNDERSTAND AND MITIGATE (D)EGREE AND (E)FFECT OF MIS/DISINFORMATION

July 2021

Purpose

In the summer of 2021, the Aspen Institute Commission on Information Disorder expressed interest in better understanding the state of the field for assessing "Degree" and "Effect" of mis/disinformation. Additionally, Aspen Institute Commissioners sought evidence-based recommendations for mitigating harms from mis/disinformation, including policy options.

This work was originally performed as part of MITRE's partnership with the Aspen Institute Commission on Information Disorder and was delivered in July 2021.

Overview

By default, social media platforms have become the first line of defense for assessing, preventing, and responding to online harms. For any online harm, including mis- and disinformation, civil society, academia, industry, and government stakeholders must partner to determine:

- How to appropriately define the problem to establish thresholds for action and cross-sector division of roles and responsibilities
- How to establish cross-platform and cross-sector standards to ensure consistency
- How best to coordinate data and information sharing to enable tactical response and effective Research, Development, Test, and Evaluation (RDT&E) on prevention, detection, and intervention options
- How to assess the effectiveness of various interventions, and to provide robust descriptions of their contextual dependencies
- How to effectively implement, maintain, and audit automated solutions
- Current and potential policy approaches to mitigate harms

Recommendations:

We believe that the Aspen Institute Commission on Information Disorder can create positive social impact by concentrating on approaches to:

- **Overcome data and information sharing barriers** to promote independent research on mis/disinformation impacts, and on the effectiveness of interventions on and across social media platforms.
- **Develop cross-platform standards and metrics** for characterizing the kind and degree of mis/disinformation.
- **Promote innovative research** on indicators of impact and offline effects of mis/disinformation.
- **Develop novel incentives** to develop and implement platform design features that diminish the spread and impact of mis/disinformation.
- **Develop, promote, and scale effective programs** to promote societal resilience to mis/disinformation.
- **Explore sensible policy options** that mitigate online harms while affirming values of openness, diversity, and freedom of expression.

Contents

Background on ABCDE Framework

Camille François, Chief Innovation Officer at Graphika, originated the "ABC" disinformation framework (A = Actors, B = Behaviors, C = Content) to guide industry and regulatory remedies to disinformation (François 2019). In April 2020, Alexandre Alaphilippe, Executive Director at EU Disinfo Lab, suggested adding the "D" (for "Distribution") (Alaphilippe 2020). In September 2020, James Pamment, then nonresident scholar in the Technology and International Affairs Program at the Carnegie Endowment for International Peace, altered the "D" to mean "Degree" and added the "E" ("Effect"), resulting in the "ABCDE" framework (Pamment 2020). In May 2021, Aspen Institute Commissioners expressed interest in better understanding the "D" and "E" components of this framework and potential related interventions.

What Are "D" and "E"?

Alaphilippe argued for expanding the ABC framework to include "D" for "Distribution" to refer to "[h]ow disinformation diffuses and spreads," which "owes largely to the digital architectures of online platforms" (Alaphilippe 2020). Alaphilippe pointed out that understanding distribution is hampered by lack of transparency: "We cannot understand how disinformation operates online, much less counter it effectively, if there is not clear and trustworthy data about how it is spreading and its impact" (Alaphilippe 2020). Alaphilippe's main recommendation is that [p]latforms … provide independent and reliable ways of accessing data on content audience and impact. As it stands, researchers are dependent on metrics determined and voluntarily released by the platforms themselves, with few ways to verify their veracity" (Alaphilippe 2020).

Pamment further developed François's and Alaphilippe's insights. In the ABCDE framework, "D" refers to "Degree," which Pamment defines as "information related to the distribution of the content in question and the audiences it reaches." In other words, "degree" relates to content distribution but also to an assessment of reach of materials. To evaluate (D)egree, Pamment suggests that researchers focus on determining these components: Audiences, Platforms, Virality, Targeting, and Scale (Pamment 2020). (E)ffect refers to the use of indicators of impact, including those from the first four components (A, B, C, D) to assess the overall effect, or "how much of a threat" a given case poses. Pamment offers critical questions for analyzing content to determine the nature of the threat (Pamment 2020).

The framework authors above do not provide a comprehensive list of indicators of impact, nor a method for assessing relative importance of different indicators. Assessing (D)egree components relies on social media platform data. (D)egree and (E)ffect of mis/disinformation are part of an overall assessment of *impact*.

(D)egree and (E)ffect: What Is Known, and How? Capabilities and Barriers

Most technical solutions in the mis/disinformation problem space can be grouped into three general types: attribution or validation capabilities, independent monitoring capabilities, or media literacy efforts.[i] Some capabilities provide information on actors, behaviors, or content that could form the basis of an overall assessment of impact. However, a robust, cross-sector set of indicators of impact that aligns available data sources and capabilities is not currently available. Additionally, around the world, many organizations and projects are focusing on analyzing and countering influence operations (see Smith 2020 for an overview and catalog of over 460 organizations).

Evaluating degree or effect of mis/disinformation campaigns is fundamentally connected to other attributes, such as actors, behaviors, and content. As such, many of the capabilities and current barriers outlined in this section pertain to the broader challenge of rapidly and accurately evaluating mis/disinformation campaigns.

Platforms' Assessments of (D)egree

Large social media platforms release threat reports or transparency reports (examples include Twitter Transparency Report on Influence Operations, which pertains specifically manipulation attempts by state-linked entities, Facebook Threat Report: The State of Influence Operations 2017-2020, and Reddit Transparency Report 2020). Such reports indicate trends and platform responses (including, for example, information on geography, types, and number of online influence campaigns, as well as take-down numbers) to a variety of harmful or illegal content or behaviors on their platforms, but do not present the methods used to identify and assess the degree of information campaigns. It is therefore not possible to validate platforms' assessments of trends, or to get a broader sense of the scope and scale of mis/disinformation on a platform or across platforms. Despite this limitation, individual platforms' transparency or trend reports can provide valuable information on their assessments of tactics, techniques, and trends.

Independent Assessment of (D)egree on and across Social Media Platforms

In this concise summary, we will not summarize the research findings that have resulted from independent study supported by reports and publicly available datasets such as those listed above.[ii] Instead, we offer overall observations that reflect a range of points that we believe are crucial:

- Independently assessing the degree of mis/disinformation on a single platform requires definitional clarity, attribution capabilities, and data that is rarely available.
- Although there has been success in particular topic areas (such as election misinformation and COVID-19 misinformation), robust cross-platform data sharing mechanisms and incentives to enable independent research and audit of platform findings do not exist.
- Academic research has moved ahead with available datasets, which primarily come from Twitter and Reddit. Findings may be limited by features of those platforms and may not be pertinent to activities on other platforms.
- Assessing cross-platform degree of mis/disinformation, as well as the spread of mis/disinformation among platforms, is an enduring problem.

One suggestion for overcoming this suite of issues has been to establish a multi-stakeholder collaboration model through federally funded research and development centers (FFRDCs), or through an approach modeled on FFRDCs (Shapiro et al. 2020).

(E)ffect: Bringing It All Together

To assess effect, one would need to overcome limitations of platform assessments and develop cross-platform awareness, both of which are challenges due to data access issues. Further, an assessment of effect would integrate actor, behavior, content, and degree indicators of impact. Given that mis/disinformation can manifest for a wide variety of topics, audiences, and genres of communication (beyond news, or propositional content for which simple true/false judgments are possible), assessments of effect must be calibrated to the specifics of the communicative event, and

thresholds established to appropriately diminish harms of specific kinds of actor, behavior, content, or degree attributes. On the issue of cross-platform variance in community standards and enforcement, see Bateman et al. 2021.

Data Sharing

Research to assess (D)egree and (E)ffect of mis/disinformation is dependent on data. For an overview of the complexity and significance of social media data and data sharing, see Shapiro et al. 2021. Development of standards and capabilities hinges on the ability of cross-sector partners to define requirements and overcome barriers for data sharing, to enable independent research while ensuring legal and privacy standards are maintained.

Pasquetto et al. (2020) provide 15 opinions from misinformation researchers detailing the research that "could hypothetically be conducted if social media data were more readily available." This commentary article presents perspectives on specific research projects that would be possible given different kinds of platform data, as well as a consolidated view of the current state of collaboration and data sharing with platforms, which has remained a strong area of emphasis within the research community over the past few years, and for which there has been ongoing experimentation (including efforts such as Social Science One that have been fraught with challenges and have not lived up to early expectations).

In addition to data sharing between social media platforms and independent researchers, data sharing among social media platforms and government stakeholders is another important dimension to a whole-of-society approach to diminishing online harms and ensuring rapid response to threats as well as improved long-term strategic advantage (including robust research to enable development of new capabilities for prevention, detection, and response). In government contexts, there are legal and privacy considerations that impact potential data sharing arrangements. Given the global reach of social media platforms (or "technology companies," as they may prefer to be known; see Napoli and Caplan 2017), complex governance issues that impact data sharing remain.

Interventions

This section provides an overview of a variety of online and offline interventions to prevent the spread of mis/disinformation, and to mitigate its effects. The subheadings below (on-platform interventions, off-platform interventions, and policy options) are rough, overlapping categories, which are intended to provide a high-level orientation to intervention approaches. On-platform interventions refer to those interventions that may be used by platforms. Off-platform interventions refer to those interventions, both online and offline, that other stakeholders may use. Policy options refer to potential regulatory measures, including those that impact or promote on- and off-platform interventions.

On-Platform Interventions

Partnership on AI (PAI) has created an Intervention Database, a public resource geared toward understanding the options that might be used by platforms to act on misinformation, as a response to the difficulty that partners in civil society, academia, and industry have expressed in determining what works (and does not work) and making comparisons across platforms (Saltz and Leibowicz 2021). Such interventions can be applied to content, accounts, or groups, and include labeling (credibility and context labels), ranking, and removal. Other interventions can be applied platform-wide, such as "'shadow banning'

of certain tags, keywords, or accounts," widespread labeling (of content or context) triggered by searches, or various digital literacy efforts (Saltz and Leibowicz 2021).[iv] Saltz and Leibowicz (2021) point out that there is a "lack of standardized goals and metrics for interventions" and that "while many platforms regularly release public statistics, these rarely include information about specific interventions other than high-level counts of actions such as 'posts removed.'"

In the United States, a small but influential group of industry participants and non-governmental organizations met in February 2018 and defined the Santa Clara Principles on Transparency and Accountability in Content Moderation (Santa Clara Principles 2018) for enhanced transparency and accountability among technology platforms. While some platforms have subscribed to the principles, participants at the time noted that the principles were "meant to serve as a starting point, outlining minimum levels of transparency and accountability [to] serve as the basis for a more in-depth dialogue in the future" (Santa Clara Principles 2018).

Off-Platform Interventions

While the primary focus of the Partnership on AI Intervention Database is on-platform interventions, there are also off-platform interventions worthy of note. Chief among those are efforts that promote resilience to mis/disinformation (including media and digital literacy efforts, some of which are carried out or promoted on platforms but others of which operate separately, with online and offline components). The Database of Informational Interventions (Consortium for Elections and Political Process Strengthening (CEPPS)) (funded by United States Agency for International Development (USAID)), for example, offers a snapshot of the tools and capabilities available internationally to create a healthy information space for political engagement and elections. Another notable suite of efforts to promote resilience, which comes from the Cybersecurity and Infrastructure Security Agency (CISA), offers materials that range from graphic novels to COVID-19 communication toolkits for state, local, tribal, and territorial officials for addressing mis/dis/malinformation. It is not possible in this paper to canvas the full range of educational initiatives promoted by academia, civil society, industry, and government. It is worth noting, however, that educational initiatives at the local, national, and international levels are advancing to educate children and young adults about recognizing and not spreading mis/disinformation.

Another bundle of interventions pertains to identifying and countering online narratives, including through fact-checking or debunking. Approaches vary. For example, the European Union (EU), through its EUvsDisinfo program, developed a group of 200 individuals across the EU who attempt to identify and share misinformation as it unfolds. EUvsDisinfo, in turn, applies additional searches and analytics to identify narratives of concern and to publish summaries weekly. In the United Kingdom, the government has taken a proactive role by establishing a Rapid Response Unit to promote accurate, fact-based news when mis/disinformation about a topic is deemed problematic (United Kingdom Government Communication Service 2018). Some nations are establishing special programs to deal with specific topics of concern.[v]

Other off-platform actions against actors include cease-and-desist letters and civil actions. Platforms can also escalate concerns to government partners, which enables further potential interventions, such as law enforcement actions or sanctions of foreign entities.

It is beyond the scope of this paper to evaluate the effectiveness of approaches, or to canvas the international landscape of approaches to misinformation.

Policy Options

This subsection provides a high-level view of a range of policy options for dealing with mis/disinformation currently under consideration in the US context. This paper does not provide an exhaustive list of proposed legislation at the US state or federal level, nor an assessment of the merits or drawbacks of policy options taken or under consideration by foreign partners. Given that the international internet governance policy context is complicated, this paper focuses specifically on high-level concerns in the US context.

Foreign Agents Registration Act (FARA)

FARA is a US law passed in 1938 that requires agents representing the interests of foreign powers in a political capacity to disclose their relationship with the foreign government or foreign principals, and to report information about related activities and finances. In recent years, policymakers have proposed reforms to FARA to improve enforcement, and to provide definitional clarity. FARA is relevant to the broader mis/disinformation discussion; it may be applied or extended to foreign actors' behaviors on social media, such as the placement of political ads or other funded content. For an overview of FARA, see Fattal 2019.

The Committee on Foreign Investment n the United States (CFIUS)

CFIUS is an interagency committee established in 1975 that reviews national security concerns related to foreign investments in US companies. Reviews by CFIUS, and proposals for its reform, are relevant to the broader mis/disinformation discussion because social media applications and the data they possess may pose national security risks. Foreign ownership of social media platforms may contribute to the mis/disinformation threat.[vii] For an overview of CFIUS and the Foreign Investment Risk Review Modernization Act enacted in 2018, see Tarbert 2020.

Social Media Platforms and Liability

The United States has traditionally exempted social media platforms from liability for either allowing the placement of mis/disinformation or removing it. This posture results from Section 230 of the Communications Decency Act. A substantial policy focus in the area of internet governance in the US context has been on potential amendments to Section 230 of the Communications Act of 1934 (47 U.S.C. §230), enacted as part of the Communications Decency Act of 1996, which broadly protects operators of "interactive computer services" from liability for publishing, removing, or restricting access to another's content" (Gallo and Cho 2021). For a detailed overview and history of Section 230, see Kosseff 2019. For an overview of proposed Section 230 and COVID-19 misinformation legislation, see Appendix B of Gallo and Cho 2021.

Other nations are adopting a range of approaches to intermediary liability. According to research conducted by the non-profit think tank Information Technology and Innovation Foundation, there are three common approaches to platform liability in democratic nations: the "actual knowledge" approach, in which platforms are responsible for removing false or intellectual property (IP) infringing content if they have actual knowledge that it has been posted; the "notice-and-takedown" approach, in which platforms are responsible for removing false or IP-infringing content once notified it has been posted; and the "mere conduit" approach, in which a technology platform that maintains a completely non-interventionist stance toward content has no liability for its truthfulness or non-truthfulness (Johnson and Castro 2021).

Recommendations

- Develop data sharing arrangements and mechanisms to enable independent research on the impacts of mis/disinformation.
  - Develop and fund innovative research on the offline impacts of online mis/disinformation and on the effectiveness of interventions.
- Encourage development of standards for assessing the impacts of mis/disinformation and the impact of interventions.
- Establish the conditions for cross-sector collaboration on standards to enable trend identification, cross-platform comparison, and intervention performance metrics.
  - Novel data sharing arrangements and shared community standards will:
    - Enable third-party audit and validation of platform reports and findings.
    - Assess and validate the effectiveness of various interventions.
- Develop incentives with industry partners to promote innovations (social, business, and technical) that diminish the spread and impact of mis/disinformation.
- Develop the conditions to enable effective cross-sector information sharing arrangements among partners with shared values and complementary expertise (industry, academia, non-profit, US Government, and foreign allies and partners).
  - Collaborate with researchers to develop and promote platform-specific and platform-independent indicators of impact (for mis/disinformation).
  - Based on a shared set of indicators of impact, develop and validate thresholds for intervention and appropriate response options.
  - Establish the US Government role in facilitating information sharing in a variety of circumstances.
- Collaborate with stakeholders to determine which policy options will improve the online ecosystem, and how policy impacts will be measured and evaluated.

Summary

This paper provides a snapshot of the issues pertinent to determining the degree and effect of mis/disinformation, the range of on- and off-platform interventions currently available to prevent or respond to online mis/disinformation, and a view of the policy options and considerations essential in the US context. The set of recommendations and resources provided in this paper is intended to orient Aspen Institute Commissioners to viable paths forward in establishing cross-sector priority areas for collaboration.

Assessing the impact of mis/disinformation and assessing the effectiveness of interventions to mitigate harmful impacts of mis/disinformation are challenging and interrelated problems of active interest to the research, practitioner, and policy communities involved in understanding, preventing, and countering mis/disinformation. Engagement metrics (impressions, views, likes, shares, etc.) can be a starting point for assessing the scope and degree of impact of online content, but such metrics do not tell the full story about what changes beliefs or real-world behaviors. Indeed, researchers such as Sinan Aral (2020) have pointed out that demonstrating behavior change attributable to an intervention (causal "lift") of online advertising campaigns is difficult. Counter-mis/disinformation researchers and practitioners consistently prioritize better understanding impact as one of the most important—and difficult—topics for additional research. Such research is fundamentally interdisciplinary, leveraging both qualitative and quantitative methods, and a variety of approaches to assess online and offline dimensions and their interconnections. Enabling independent research on the impacts of online content is a key area for the Commission to investigate and promote.

Data, standards, and partnerships will enable evaluation of the impacts of mis/disinformation, which in turn will enable new prevention and response measures. Beyond the impacts of any individual campaign, however, researchers consistently have pointed out that US societal vulnerabilities remain considerable. Supply-side measures, such as media and digital literacy interventions, are crucial dimensions of promoting resilience to mis/disinformation. Resilience programs can be better characterized, promoted, scaled, and evaluated.

The resonance and harms of mis/disinformation are a societal issue enabled by technology. While the focus in this paper has been on assessing the degree and effect of mis/disinformation in the US context, fundamental issues such as decline of trust in institutions (such as government and the media) and an increase in political polarization form preconditions that make the spread and impact of mis/disinformation possible. While social media platforms have made possible the low-cost amplification of potentially harmful content (Phillips 2019), they have also taken the lead on developing novel interventions to detect and prevent the spread of harmful content. Paired with resilience approaches focused on promoting critical thinking, on-platform interventions (including the redesign of algorithmic and user interface features) will make a difference. The Aspen Commission on Information Disorder can promote cross-sector collaboration on measurement of intervention effectiveness.

Works Cited

Alaphilippe, Alexandre. April 27, 2020, "Adding a 'D' to the ABC Disinformation Framework." Brookings. https://www.brookings.edu/techstream/adding-a-d-to-the-abc-disinformation-framework/, Accessed August 11, 2022.

Aral, Sinan. 2020. *The Hype Machine*. New York: Currency.

Bateman, Jon, Natalie Thompson, and Victoria Smith. April 1, 2021. "How Social Media Platforms' Community Standards Address Influence Operations." Carnegie Endowment for International Peace. https://carnegieendowment.org/2021/04/01/how-social-media-platforms-community-standards-address-influence-operations-pub-84201, Accessed June 25, 2021.

Consortium for Elections and Political Process Strengthening (CEPPS). "Database of Informational Interventions." https://counteringdisinformation.org/interventions, Accessed July 1, 2021.

Cybersecurity and Infrastructure Security Agency (CISA). "Mis, Dis, Malinformation." https://www.cisa.gov/mdm, Accessed July 1, 2021.

EUvsDisinfo. https://euvsdisinfo.eu, Accessed July 1, 2021.

Facebook. May 2021. "Threat Report: The State of Influence Operations 2017-2020." https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf, Accessed June 22, 2021.

Fattal, Joshua R. 2019. "FARA on Facebook: Modernizing the Foreign Agents Registration Act to Address Propagandists on Social Media." NYU Journal of Legislation & Public Policy 21:903-947. https://nyujlpp.org/wp-content/uploads/2019/10/Fattal-FARA-On-Facebook-21-NYU-JLPP-903.pdf, Accessed July 1, 2021.

François, Camille. September 20, 2019. "Actors, Behavior, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses." Working Paper of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. 1-10. https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf, Accessed June 14, 2021.

Funke, Daniel, and Daniela Flamini. 2019. "A Guide to Anti-Misinformation Actions around the World." Poynter. https://www.poynter.org/ifcn/anti-misinformation-actions/, Accessed August 11, 2022.

Gallo, Jason A., and Clare Y. Cho. January 27, 2021. "Social Media: Misinformation and Content Moderation Issues for Congress." R46662. Congressional Research Service. 1-32. https://crsreports.congress.gov/product/pdf/R/R46662, Accessed June 9, 2021.

Jankowicz, Nina, and Shannon Pierson. December 2020. "Freedom and Fakes: A Comparative Exploration of Countering Disinformation and Protecting Free Expression." Wilson Center. https://www.wilsoncenter.org/publication/freedom-and-fakes-comparative-exploration-countering-disinformation-and-protecting-free, Accessed June 24, 2021.

Johnson, Ashley, and Daniel Castro. February 22, 2021. "How Other Countries Have Dealt with Intermediary Liability." Information Technology & Innovation Foundation. https://itif.org/publications/2021/02/22/how-other-countries-have-dealt-intermediary-liability, Accessed June 16, 2021.

Kosseff, Jeff. 2019. *The Twenty-Six Words That Created the Internet*. Ithaca, NY: Cornell University Press.

Maddox, J.D. September 2019. "Lessons from the Information War: Applying Effective Technological Solutions to the Problems of Online Disinformation and Propaganda." GW Program on Extremism. 1-14. https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/Lessons%20from%20the%20Information%20War.pdf, Accessed June 9, 2021.

Napoli, Phil, and Robyn Caplan. 2017. "Why Media Companies Insist They're Not Media Companies, Why They're Wrong, and Why It Matters." First Monday.

Pamment, James. September 2020. "The EU's Role in Fighting Disinformation: Crafting a Disinformation Framework." Future Threats, Future Solutions #2. Carnegie Endowment for International Peace. https://carnegieendowment.org/files/Pamment_-_Crafting_Disinformation_1.pdf, Accessed June 14, 2021.

Pasquetto, Irene V., et al. December 9, 2020. "Tackling Misinformation: What Researchers Could Do with Social Media Data." Commentary. Harvard Kennedy School Misinformation Review. https://misinforeview.hks.harvard.edu/article/tackling-misinformation-what-researchers-could-do-with-social-media-data/, Accessed June 22, 2021.

Phillips, Whitney. 2019. "The Oxygen of Amplification: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online." Data & Society. https://datasociety.net/wp-content/uploads/2018/05/FULLREPORT_Oxygen_of_Amplification_DS.pdf, Accessed June 24, 2021.

Reddit. 2020. "Transparency Report 2020." https://www.redditinc.com/policies/transparency-report-2020-1, Accessed July 1, 2021.

Roudik, Peter, et al. April 2019. "Initiatives to Counter Fake News in Selected Countries." The Law Library of Congress, Global Legal Research Directorate. https://www.loc.gov/item/2019668145/, Accessed June 30, 2021.

Saltz, Emily, and Claire Leibowicz. June 14, 2021. "Fact-Checks, Info Hubs, and Shadow-Bans: A Landscape Review of Misinformation Interventions." Partnership on AI. https://www.partnershiponai.org/intervention-inventory/, Accessed August 11, 2022.

The Santa Clara Principles on Transparency and Accountability in Content Moderation. February 2018. https://santaclaraprinciples.org, Accessed June 28, 2021.

Shapiro, Elizabeth H., Michael Sugarman, Fernando Bermejo, and Ethan Zuckerman. February 2021. "New Approaches to Platform Data Research." NetGain Partnership. https://www.netgainpartnership.org/resources/2021/2/25/new-approaches-to-platform-data-research, Accessed August 11, 2022.

Shapiro, Jacob N., Natalie Thompson, and Alicia Wanless. December 2020. "Research Collaboration on Influence Operations Between Industry and Academia: A Way Forward." Carnegie Endowment for International Peace. https://carnegieendowment.org/files/Shapiro_Thompson_Wanless_Instantiating_Models_final.pdf, Accessed August 11, 2022.

Smith, Victoria. December 14, 2020. "Mapping Worldwide Initiatives to Counter Influence Operations." Carnegie Endowment for International Peace. https://carnegieendowment.org/2020/12/14/mapping-worldwide-initiatives-to-counter-influence-operations-pub-83435, Accessed August 11, 2022.

Tarbert, Heath P. 2020. "Modernizing CFIUS." George Washington Law Review 88(6):1477-1524. https://heinonline.org/HOL/PrintRequest?collection=journals&handle=hein.journals/gwlr88&id=1560&print=section&div=44&ext=.pdf&format=PDFsearchable&submit=Print%2FDownload, Accessed July 1, 2021.

United Kingdom Government Communication Service. July 19, 2018. "Alex Aiken Introduces the Rapid Response Unit." https://webarchive.nationalarchives.gov.uk/ukgwa/20200203104056/https:/gcs.civilservice.gov.uk/news/alex-aiken-introduces-the-rapid-response-unit/, Accessed August 11, 2022.

United States Senate Select Committee on Intelligence. December 17, 2018. "New Reports Shed Light on Internet Research Agency's Social Media Tactics." https://www.intelligence.senate.gov/press/new-reports-shed-light-internet-research-agency's-social-media-tactics, Accessed August 11, 2022.

Walshe, Taylor, and Tan, Shining. May 13, 2020. "TikTok on the Clock: A Summary of CFIUS's Investigation into ByteDance." Center for Strategic and International Studies. https://www.csis.org/blogs/trustee-china-hand/tiktok-clock-summary-cfiuss-investigation-bytedance, Accessed August 15, 2022.

---

[i] J.D. Maddox (2019), former Director of Analytics at the Global Engagement Center (Department of State), has identified types of technological solutions that pertain to the problems of online disinformation and propaganda. These include the following capability types:

- Content Validation (including use of distributed ledger technologies)
- Account Validation
- Website Ratings (independent organizations that provide rapid identification of mis/disinformation or harmful content to avoid amplification)
- Crowdsourced Information Verification
- Online Advertisement and Content Attribution
- "Yellow Alerts" to malign campaigns (by analytic teams using social media monitoring systems)
- Media and Digital Literacy efforts

[ii] Discussion of the two independent third-party reports commissioned by the Senate Select Committee on Intelligence (SSCI) in 2018 and based on data provided by SSCI is also out of scope. These reports focus on social media tactics used by Russia's Internet Research Agency (IRA). See: https://www.intelligence.senate.gov/press/new-reports-shed-light-internet-research-agency's-social-media-tactics

[iii] In certain cases, Twitter has released datasets of state-linked information campaigns. These are available in an archive, which enables independent researchers to study cases. Notably, in 2018, Twitter released a dataset with 3,841 accounts affiliated with the IRA and 770 accounts potentially originating in Iran, which in total included over 10 million tweets and 2 million images, with materials dating back to 2009. In years following, Twitter's archive has grown, and additional accounts associated with platform manipulation have been added, including in other countries, such as United Arab Emirates, Egypt, Spain, Ecuador, China, Hong Kong, and others.

[iv]In this context, "shadow banning" is a content moderation technique that refers to the practice of blocking content in a way that is not visible to the user.

[v]For example, Argentina may maintain a Commission for the Verification of Fake News within the National Election program. This topic-specific program would work to identify news "of doubtful credibility" during election periods (Roudik 2019).

[vi]See Funke and Flamini (2019) for an overview (with ongoing updates). See Roudik 2019 for initiatives to counter fake news in selected countries (Argentina, Brazil, Canada, China, Egypt, France, Germany, Israel, Japan, Kenya, Malaysia, Nicaragua, Russia, Sweden, United Kingdom). See Jankowicz and Pierson 2020 for counter-disinformation case studies focused on regulations in Germany, Brazil, Singapore, and Ukraine.

[vii]For example, consider the case of TikTok (Walshe and Tan 2020).