Centre for International Governance Innovation

This policy brief is part of CIGI's project on freedom of thought: Legitimate Influence or Unlawful Manipulation?

Find out more at: www.cigionline.org/fot

Policy Brief No. 4 — January 2024

Protecting Freedom of Thought: Mitigating Technological Enablers of Disinformation

Alexa Raad

Key Points

- The internet and ad-supported business models facilitate the targeting, dissemination and amplification of false information online.
- Algorithmic designs, cognitive biases and platform structures enable false information to spread faster than truth on the internet and are synergistically exploited in influence operations, compromising freedom of thought.
- Conceptual frameworks like DISARM and RICHDATA provide insights into disinformation campaigns' stages and tactics, emphasize the importance of countering harmful content amplification, and recommend strategic measures to increase cost and complexity for influence operators.
- The United States needs to enact national data collection laws, data brokerage industry regulation, social media company oversight, regulation of third-party players facilitating disinformation, and other measures to enhance media literacy and protect the information ecosystem.

Introduction

The advent of the internet and ad-supported business models that capitalize on capturing attention has introduced powerful technological enablers that greatly enhance the targeting, dissemination and amplification of false information (such as misinformation, disinformation and malinformation) to an unprecedented degree.

Influence operations leverage a synergistic blend of factors, including algorithmic designs that prioritize engagement, human cognitive biases and platform structures that incentivize content sharing, to adeptly manipulate and compromise our freedom of thought. Freedom of thought is the right to keep our thoughts and opinions private, the right not to have our thoughts and opinions manipulated, and the right not to be penalized for our thoughts and opinions alone (Alegre 2022).

To understand the technology enablers of disinformation, new conceptual frameworks have emerged that lay out the specific stages and the associated tactics, techniques and procedures (sometimes abbreviated as TTPs) used by malicious influence operators. The most applicable framework, called DISARM, also provides specific countermeasures for each tactic.

Effective remediation efforts need to prevent amplification of harmful content, and increase the risk, cost and complexity to the influence operator. In the United States, regulation is needed in at least four

About the Author

Alexa Raad is recognized as a thought leader in the cybersecurity, DNS (Domain Name System), cloud, internet infrastructure and Internet of Things industries. Alexa authored a patent in cybersecurity and served in C-level roles and as a strategic adviser to CEOs, executive teams, boards and private equity teams. In addition to running her consulting practice, she is a Qualified Risk Director of the Directors and Chief Risk Officers Organization. She is a member of the Verified Voting Foundation Advisory Board, a co-founder and co-host of the *TechSequences* podcast (techsequences.org) and part of a panel of experts sought by Pew Research for commenting on internet and technology. Alexa is also a member of the International Academy of Digital Arts and Sciences and has served, by appointment, as a Webby's judge since 2008.

areas: establishment of a national data collection and privacy law; regulation of the data brokerage industry; regulation of social media companies and big tech; and regulation of other third-party players that provide services or tools to facilitate disinformation campaigns. In addition, social media and big tech can be incentivized to implement other measures such as throttling of viral content, establishing data provenance as a means of user education, and employing leading-edge technology such as artificial intelligence (AI) to identify and root out deepfakes. None of these recommendations are novel; however, when paired with effective public-private collaboration (including academia and civil society) and increased media literacy, they would help mitigate the pollution of our information ecosystem and protect our freedom of thought.

To understand technological enablers for influence operations, first we need to examine how content is targeted, prioritized and fed to users. Second, we need to understand why false information has a far greater potential for engagement and viral spread than the plain and simple truth. Lastly, we would need to use a framework for understanding the stages of disinformation campaigns to learn where influence campaigns can be stymied, throttled or mitigated with some measure of efficacy.

Ad-Supported Business Models and Algorithms

Historically, the traditional media channels such as television, radio, magazines and newspapers accounted for much of the advertising spend. However, with the rise of the internet, the advertising landscape shifted. Currently, the big tech platforms — which include search engines, social media companies and a smattering of big retail aggregators — are the major recipients of internet advertising revenue. In 2021, Google, Meta (formerly Facebook) and Amazon accounted for more than seven out of every 10 dollars, or 74 percent of global digital ad spending, which is 47 percent of all money spent on advertising that year (Joseph 2022). Consequently, the traditional news outlets (for example, local and national newspapers and broadcast journalism) are severely underfunded by comparison. Global digital advertising is forecasted to grow 85 percent

from $407 billion in 2022 to $753 billion in 2026 (Woodford 2022).[1] The ad-supported model provides "free" services to users who would otherwise have to pay for those same services. A 2020 survey found that Americans place a value of more than $1,400 on the free services, mobile apps and digital content they receive that are otherwise funded by advertising (Digital Advertising Alliance 2020). However, there is still a cost.

While the ad-supported model has fuelled internet adoption and innovation, it has also led to "surveillance capitalism," defined by Shoshana Zuboff as a "new economic order" and "an expropriation of critical human rights that is best understood as a coup from above" (Zuboff 2019, n.p.). Foundational to internet advertising is the ability to collect massive troves of personal data, the digital breadcrumbs left behind by our online presence. This data is collected and fed into sophisticated algorithms designed for user engagement. The algorithms employed by social media platforms are designed to optimize advertising revenue by capturing user attention and engagement. Consequently, these algorithms tend to prioritize emotionally charged and divisive content that sustains user engagement within their feed. That content is more than just convenience or entertainment; it also tends to be "news-oriented." Just under half of US adults now rely on social media sites for news "often" or "sometimes," even as these sites harbour misinformation and disinformation on their platforms (Walker and Matsa 2021).

An internal Facebook presentation to executives in 2018 made clear that the company knew its recommendation engine actively promoted extreme and polarizing content (Horowitz and Seetharaman 2021). A particular slide from the presentation stated, "Our algorithms exploit the human brain's attraction to divisiveness" (ibid.). Another internal report, this one from 2016, found "64% of all extremist group joins are due to our recommendation tools," with most users joining at the suggestion of Facebook's "Discover" and "Groups You Should Join" algorithms. "Our recommendation systems grow the problem," the presentation stated (ibid.). However, the executives took no action, since doing so would negate their efforts to grow revenues. Facebook (Meta) is not alone. Mozilla researchers found that YouTube not only hosts but actively recommends videos that violate its own policies concerning inappropriate content, hate speech, and medical and political misinformation (McCrosky and Geurkink 2021). Another social media giant, Twitter (rebranded as X in July 2023), changed its ranking algorithm in 2016 from prioritization based on chronological order to prioritization based on a "relevance model" score (Koumchatzky and Andryeyev 2017). Relevancy, in this case, meant the platform's deep-learning algorithm was changed to place greater weight on popularity and on engagement with similar types of tweets or authors, that is, retweets or likes. In all three of these cases, the needs of the business model dictated the design of the algorithms.

What Drives the Viral Spread of False Information?

It has been said that a lie can travel halfway around the world while the truth is putting on its shoes. The internet and social media have simply supercharged the spread. The question is why.

The term *cognitive bias*, introduced by Amos Tversky and Daniel Kahneman (1974), describes a systematic yet potentially flawed pattern of response by individuals to judgment and decision-making challenges. Cognitive biases arise from individuals' reliance on quickly applied, yet fallible, cognitive strategies referred to as *heuristics*. No less than 180 cognitive biases have been identified (Nortje 2020).

Cognitive hacking, which focuses on exploiting vulnerabilities in human cognition and perception, targets individuals' psychological or cognitive processes to manipulate their thoughts, decision making or behaviour. Cognitive hacking attacks often leverage psychological biases, social engineering or manipulation tactics to deceive or influence individuals.

Cognitive hacking also cleverly exploits human beings' instinctive group affiliations, triggering a subconscious "us" versus "them" mentality. By tapping into the automatic threat-detection mechanisms of our brain's limbic system, which

1 All currency in US dollars.

is concerned especially with emotion and motivation, it manipulates us into perceiving those who are different or outside our "in-groups" as potential threats to our survival. This insidious manipulation intensifies divisions and fosters distrust within communities. Moreover, cognitive hacking renders us more vulnerable to believing and accepting propaganda (Heslen 2020).

But we cannot blame viral spread of disinformation on these cognitive biases alone. There is a symbiotic and synergistic relationship between our cognitive biases and the design of social media platforms and sharing features that supercharges the spread of harmful content and false information. A recent study found that harmful and false content not only has higher engagement rates but also is much more likely to be reshared than factual content (Allen 2022). The study found that the biggest factor in its virality was the structure of rewarding users for habitually sharing information (ibid.). Another study by the University of Southern California found that "due to the reward-based learning systems on social media, users form habits of sharing information that attracts others' attention. Once habits form, information sharing is automatically activated by cues on the platform without users considering response outcomes such as spreading misinformation. As a result of user habits, 30 to 40% of the false news shared in our research was due to the 15% most habitual news sharers" (Ceylan, Anderson and Wood 2023). This finding is also supported by the results of a study from the Massachusetts Institute of Technology on the spread of false information on Twitter/X. The researchers found, as expected, that false news spreads far more rapidly than the truth. But they also found that the spread of false information is due to people retweeting the false news, not just to bots programmed to amplify the disinformation (Vosoughi, Roy and Aral 2018). If false information gets user engagement faster than the truth, it implies that fact-checking efforts will not be as effective, because the content has already garnered most of its engagement by the time the fact-checkers have completed their work.

Influence operations leverage a synergistic blend of factors, including algorithmic designs that prioritize engagement, human cognitive biases and platform structures that incentivize content sharing. Through these mechanisms, they adeptly manipulate and compromise our freedom of thought.

The Anatomy of a Disinformation Campaign

To understand technological enablers of disinformation, it is valuable to analyze organized disinformation campaigns using a structured framework that encompasses both the stages and the associated tactics, techniques and procedures employed by influence operators to manipulate thoughts.

One approach is to consider disinformation campaigns within the context of a digital marketing campaign, albeit a malicious one. Companies frequently employ digital marketing campaigns to encourage interaction, to enhance awareness or to generate revenue (or to do all three) using diverse tactics, content and messages in increasingly narrow and targeted stages. The Center for Security and Emerging Technology at Georgetown University devised the RICHDATA framework, which takes a digital marketing campaign as a conceptual model and lays out its seven stages as a disinformation "kill chain."[2] The seven stages correspond to the framework's acronym: **reconnaissance**, **infrastructure**, **content**, **deployment**, **amplification**, sustained engagement through **trolling**, and **actualization** (Sedova et al. 2021).

Another approach is to view the problem through the lens of cybersecurity and information security frameworks by analyzing the tactics, techniques and procedures used by adversaries during disinformation campaigns. The original model was proposed in 2013 by the MITRE Corporation, in a project called ATT&CK (standing for adversarial tactics, techniques & common knowledge) initially focused on APTs, or advanced persistent threats. Over time, this framework was expanded to cover a broader range of cyberthreats (Strom et al. 2018). Although it was not specifically designed to be applicable to disinformation campaigns, the framework was later adapted and extended. The latest and perhaps most applicable adaptation is the DISARM series of frameworks, which

2 According to Wikipedia, the term *kill chain* is a military and information security concept that identifies the structure of an attack. The chain consists of identifying the target, dispatching forces to the target, initiating an attack on the target and destroying the target. "'Breaking' an opponent's kill chain is a method of defense or preemptive action." See https://en.wikipedia.org/wiki/Kill_chain.

envisions digital disinformation campaigns as cognitive attack campaigns.[3] In the information security context, a digital cognitive attack is "a computer or information system attack that 'modifies certain user behaviors in a way that violates the integrity of the entire user information system'" (Cybenko, Giani and Thompson 2002). As the MITRE ATT&CK framework is used for combatting cyberattacks, the DISARM framework is used for identifying and mitigating *cognitive information security attacks*.[4] DISARM is open source and offers two versions of the framework: red and blue.[5] The red team conducts simulated attacks to test the system's resilience and uncover potential weaknesses. And the blue team takes on a defensive role, implementing and managing security measures, monitoring network activity and responding to threats.

Both the RICHDATA model and the DISARM set of frameworks document the planning, preparation, deployment and amplification stages. The harmful consequences of disinformation campaigns occur when the content is disseminated and amplified, leading to viral spread within the target audience.

Effective remediation efforts need to prevent amplification of harmful content, and increase the risk, cost and complexity to the influence operator. Amplification of harmful content, especially within echo chambers, creates repetition. As researchers Aumyo Hassan and Sarah J. Barber (2021, 38) write, "Repeated information is often perceived as more truthful than new information. This finding is known as the illusory truth effect, and it is typically thought to occur because repetition increases processing fluency. Because fluency and truth are frequently correlated in the real world, people learn to use processing fluency as a marker for truthfulness." This is another reason why fact-checking is not as effective in combatting disinformation. Alongside the time delay it entails, its effectiveness is undermined by the requirement for repeated exposure to false information to correct it.

DISARM's blue cognitive security framework lists comprehensive and excellent countermeasures for each technique used by adversaries in a particular phase of a disinformation campaign. Some countermeasures are tactical and short-term and others more strategic and long-term. The next section lays out the most strategic and critical remediation steps. Some are covered in DISARM's blue framework, while others are not.

Potential Regulatory, Policy and Technical Mitigation

Assuming advertising-based business models are here to stay, at least for now, we need to address the problematic elements associated with these models that are currently leveraged by influence operators. One key area is data collection and privacy. Although the EU data privacy protection measures (General Data Protection Regulation) went into effect in 2018,[6] presently there is no federal law in the United States that discourages the excessive gathering of crucial data or personally identifiable information. Instead, there is a patchwork of state-level privacy laws, which has resulted in a fragmented and complex landscape that still fails to address fundamental concerns. A federal privacy law should establish rules and regulations regarding collection, use and disclosure of personal information. It should also explicitly define the types of critical data that can be collected, stored or accessed within the United States and among its allies.

The lack of a comprehensive baseline US privacy law has also allowed the proliferation of data

3 See https://github.com/DISARMFoundation/DISARMframeworks.

4 See Gray and Terp (2019). Although the terms *disinformation campaign* and *cognitive information security attack* are sometimes used interchangeably, they are a bit different. The latter typically refers to an attack on an individual's or an organization's information systems with a specific focus on exploiting vulnerabilities in human cognition and perception. This term is also used interchangeably with *cognitive hacking*. The goal here is to manipulate thoughts, decision-making processes or behaviour. The term *disinformation campaign* is broader and often involves a deliberate attempt to spread false or misleading information with the aim of influencing public opinion, perceptions or behaviour. Please see the policy brief by Wesley Wark (2024) for this project, which focuses on online disinformation campaigns used for foreign interference in electoral and other democractic processes.

5 See "DISARM Blue framework: Latest Framework" at https://github.com/DISARMFoundation/DISARMframeworks/blob/main/generated_pages/disarm_blue_framework.md.

6 The General Data Protection Regulation protects the personal data of individuals located in the European Economic Area, which includes the European Union, plus Iceland, Liechtenstein, Norway and the United Kingdom.

brokers. There are thousands of data brokers worldwide and the market is expected to reach $462 billion by 2031 (Transparency Market Research 2022). Data brokers are for-profit companies that collect and aggregate personal and company data from public and private sources, then resell or license such information to third parties. As far back as 2014, the Federal Trade Commission reported that just one of these data brokers already had "3000 data segments for nearly every U.S. consumer" (Ramirez et al. 2014, iv); another had "information on 1.4 billion consumer transactions and over 700 billion aggregated data elements" (ibid.). In addition, data brokers can compromise national security as they are unfettered by restrictions on sale of data to foreign nation states (Leong and Yi-Ling 2020). Comprehensive data privacy regulation and oversight of data brokers will hamper micro-targeting of individuals and communities and thus hinder the effectiveness of disinformation campaigns.

In the United States, where most social media platforms and big tech are based, Section 230 of the Communications Decency Act has afforded the big tech platforms shelter from liability.[7] This must change. Regulatory mechanisms should be established to ensure accountability, transparency and the availability of redress mechanisms. These measures are essential for holding social media companies responsible for their actions, promoting transparency in their operations, and providing avenues for addressing grievances or concerns. The following are a few potential requirements:

- **Transparency on the algorithm goals, design, data set used for training, and output generated.** For example, implementing a regulatory framework based on an algorithmic impact assessment (AIA) would require the system's developers to evaluate the potential societal harm *prior* to implementation. An AIA assessment would involve documentation of such impacts, providing a means for accountability as well as future policy development.
- **Adjustment of algorithms to redefine relevancy in favour of authenticity rather than popularity.** This adjustment would elevate trustworthy information and reduce the visibility of misinformation.
- **Independent third-party auditors and researchers to assess the efficacy of disinformation mitigation efforts.** Having independent auditors and researchers can enhance credibility and provide objective evaluations. Sharing the findings of such audits can demonstrate the company's commitment to transparency and accountability.
- **Public-private partnerships and collaboration, including with academia and civil society.** Companies can be required to share timely insights and information in disinformation patterns and trends, thus fostering a better understanding of the information ecosystem.
- **Provision of user education and media literacy tools to help users evaluate and identify reliable sources of information.** Fortunately, there are examples of successful media literacy education programs such as those used in Finland. This Scandinavian country has integrated a comprehensive media literacy program into core educational classes. As a result, Finland has secured the top spot in the Media Literacy Index (Open Society Institute Sofia 2023, 7) for the sixth year in a row.
- **Fact-checking and content moderation teams and operations.** Technologies such as AI can enable faster flagging of harmful content, such as identifying and flagging deepfakes.
- **Clear and consistent enforcement of policies against hate speech, disinformation and posting of harmful content.** These same policies should have simple and transparent redress mechanisms.
- **Takedown of botnet servers and inauthentic accounts.** These accounts are instrumental in amplifying disinformation. At the very least, accounts suspected of being operated by bots should be labelled as such to provide transparency for human users.

7 See www.govinfo.gov/app/details/USCODE-2021-title47/USCODE-2021-title47-chap5-subchapII-partI-sec230.

- **Tighten regulation on political campaign advertisements on social media platforms.** While in the United States political campaigns and organizations are required to report their expenditures to the federal government, there is no reporting requirement for consultancies and agencies that spend money on their behalf, making social media company platforms ideal targets for dubious ads.

In July 2023, US senators Elizabeth Warren and Lindsey Graham introduced a bipartisan proposal to create a Digital Consumer Protection Commission empowered to oversee digital platforms on issues such as competition, transparency, privacy and national security. Additionally, the draft Digital Consumer Protection Act[8] mandates that major platforms secure a licence to conduct their operations, providing the commission the ability to revoke the licence in instances of persistent, severe and unlawful misconduct resulting in substantial harm to consumers. Although the act does not cover many of the above recommendations, it does address, for example, targeted advertising based on users' personal data and offers users transparency on how their data is used. However, as of this writing in December 2023, the act is still in the first stage of the legislative process, and may not come to fruition.

Similar to the proliferation of data brokers, there are many other third-party players that aid and abet in the creation and distribution of disinformation. As an example, as of 2018 there were more than 65 firms offering digital influence operations or "computational propaganda" as a service (Bradshaw, Bailey and Howard 2020, i). To prevent further industrialization of disinformation, governments should hold private firms that provide influence operations as a service accountable.

Beyond being regulated, social media platforms and big tech can be encouraged or incentivized to adopt additional measures. Incentives can include government or private sector grants, advertiser and government tax incentives, or reduction of previously levied fines. These additional measures could include the following activities:

- **Implementing data provenance.** In other words, identify and label the *chain of custody*[9] or the information regarding the origin of content (for example, memes, videos, audio, pictures and so forth) that have potential to go viral.
- **Throttling content sharing.** Throttling can be used especially for content that has been shared many times. It could include adding additional steps (friction) for users who want to share the content to their followers. A relatively small way to throttle sharing is to clearly label content that has been shared many times. This is currently done by WhatsApp: when a message in WhatsApp has been forwarded five times or more away from its original sender, it is labelled with a double arrow icon. A significant share of the disinformation and misinformation online is not publicly shared and instead is passed through end-to-end encryption. This is a challenge for platform operators. However, throttling measures such as those referenced above with WhatsApp can be used to at least curb some of the spread.
- **Setting a higher bar for social media users with a significantly large following.** This higher bar is necessary due to the disproportionate influence these users hold compared to the average user. A prime example is the finding by the Center for Countering Digital Hate (2021, 6) that only 12 individuals accounted for 65 percent of the disinformation and outright lies about COVID-19 vaccines that proliferated on Facebook, Instagram and Twitter during a six-week period.

Conclusion

While regulation, policies and even technical solutions can play an important role in combatting disinformation, there are also challenges to their successful implementation, such as inadequate enforcement mechanisms,

8 See a title-by-title summary at www.warren.senate.gov/imo/media/doc/DCPC%20Section-By-Section.pdf.

9 In the realm of disinformation content, the chain of custody pertains to the chronological documentation of the origin, creation, transmission and dissemination of a piece of content. It traces the journey of the information, identifying the individuals or entities responsible at each stage, offering insights into its creation, distribution and potential manipulation. Understanding the chain of custody is crucial for assessing the credibility and authenticity of content in the complex landscape of misinformation and disinformation.

free speech concerns, and the rapidly evolving disinformation and technology landscape. Enforcement can be difficult, not only in terms of proving culpability, but also in terms of the complexities of establishing proper jurisdiction. Free speech concerns make it difficult to strike the right balance between protecting free speech and curbing disinformation. Lastly, the threat landscape for disinformation is rapidly evolving, due to technological advances and geopolitical events. As a result, remediation steps tend to lag. Despite these formidable challenges, disinformation is an existential threat worthy of tackling, as discerning truth from fiction is fundamental to every consequential decision we make as a society.

Works Cited

Alegre, Susie. 2022. *Freedom to Think: Protecting a Fundamental Human Right in the Digital Age*. London, UK: Atlantic Books.

Allen, Jeff. 2022. "Misinformation Amplification Analysis and Tracking Dashboard." Integrity Institute, October 13. https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard.

Bradshaw, Samantha, Hannah Bailey and Philip N. Howard. 2020. *Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*. February. Oxford, UK: Oxford Internet Institute. www.oii.ox.ac.uk/news-events/reports/industrialized-disinformation-2020-global-inventory-of-organized-social-media-manipulation/.

Center for Countering Digital Hate. 2021. *The Disinformation Dozen: Why Platforms Must Act on Twelve Leading Online Anti-vaxxers*. Washington, DC: Center for Countering Digital Hate. https://counterhate.com/research/the-disinformation-dozen/.

Ceylan, Gizem, Ian A. Anderson and Wendy Wood. 2023. "Sharing of misinformation is habitual, not just lazy or biased." *Proceedings of the National Academy of Sciences of the United States of America* 120 (4): e2216614120. https://doi.org/10.1073/pnas.2216614120.

Cybenko, George, Annarita Giani and Paul Thompson. 2002. "Cognitive Hacking: A Battle for the Mind." *Computer* 35 (8): 50–56. https://doi.ieeecomputersociety.org/10.1109/MC.2002.1023788.

Digital Advertising Alliance. 2020. "Americans Value Free Ad-Supported Online Services at $1,400/Year; Annual Value Jumps More Than $200 Since 2016." Press release, September 28. Washington, DC: Digital Advertising Alliance. https://digitaladvertisingalliance.org/daa-news/press-releases#year2020.

Gray, John F. and Sara-Jayne Terp. 2019. "Misinformation: We're Four Steps Behind its Creators." Cambridge, MA: Berkman Klein Center for Internet & Society at Harvard University. https://cyber.harvard.edu/sites/default/files/2019-11/Comparative%20Approaches%20to%20Disinformation%20-%20John%20Gray%20Abstract.pdf.

Hassan, Aumyo and Sarah J. Barber. 2021. "The effects of repetition frequency on the illusory truth effect." *Cognitive Research: Principles and Implications* 6 (1): 38. www.ncbi.nlm.nih.gov/pmc/articles/PMC8116821/.

Heslen, John J. 2020. "Neurocognitive hacking: A new capability in cyber conflict?" *Politics and the Life Sciences* 39 (1): 87–100. https://doi.org/10.1017/pls.2020.3.

Horowitz, Jeff and Deepa Seetharaman. 2021. "Facebook Executives Shut Down Efforts to Make the Site Less Divisive." *The Wall Street Journal*, May 26. www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499.

Joseph, Seb. 2022. "The Rundown: Google, Meta and Amazon are on track to absorb more than 50% of all ad money in 2022." Digiday Media, February 4. https://digiday.com/marketing/the-rundown-google-meta-and-amazon-are-on-track-to-absorb-more-than-50-of-all-ad-money-in-2022/.

Koumchatzky, Nicolas and Anton Andryeyev. 2017. "Using Deep Learning at Scale in Twitter's Timelines." *Insights* (blog), May 2017. https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.

Leong, Dymples amd Teo Yi-Ling. 2020. "Data Brokers: A Weak Link in National Security." *The Diplomat* (blog), August 21. https://thediplomat.com/2020/08/data-brokers-a-weak-link-in-national-security/.

McCrosky, Jesse and Brandi Geurkink. 2021. *YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm*. Mountain View, CA: Mozilla Foundation. July. https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf.

Nortje, Alicia. 2020. "What Is Cognitive Bias? 7 Examples & Resources (Incl. Codex)." PositivePsychology, August 5. https://positivepsychology.com/cognitive-biases/.

Open Society Institute Sofia. 2023. "'Bye, bye, birdie': Meeting the Challenges of Disinformation. The Media Literacy Index 2023." Policy brief, June. Sofia, BG: Open Society Institute. https://osis.bg/wp-content/uploads/2023/06/MLI-report-in-English-22.06.pdf.

Ramirez, Edith, Julie Brill, Maureen K. Ohlhausen, Joshua D. Wright and Terrell McSweeny. 2014. *Data Brokers: A Call for Transparency and Accountability*. May. Washington, DC: Federal Trade Commission. www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

Sedova, Katerina, Christine McNeill, Aurora Johnson, Aditi Joshi and Ido Wulkan. 2021. "AI and the Future of Disinformation Campaigns. Part 1: The RICHDATA Framework." Center for Security and Emerging Technology Policy Brief, December. https://cset.georgetown.edu/wp-content/uploads/CSET-AI-and-the-Future-of-Disinformation-Campaigns.pdf.

Strom, Blake E., Andy Applebaum, Doug P. Miller, Kathryn C. Nickels, Adam G. Pennington and Cody B. Thomas. 2018. *MITRE ATT&CK®: Design and Philosophy*. Project Number No. 10AOH08A-JC, July. McLean, VA: MITRE Corporation. https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf.

Transparency Market Research. 2022. *Data Brokers Market Outlook 2031*. TMRGL130237, July. Wilmington, DE: Transparency Market Research. www.transparencymarketresearch.com/data-brokers-market.html.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31. www.jstor.org/stable/1738360.

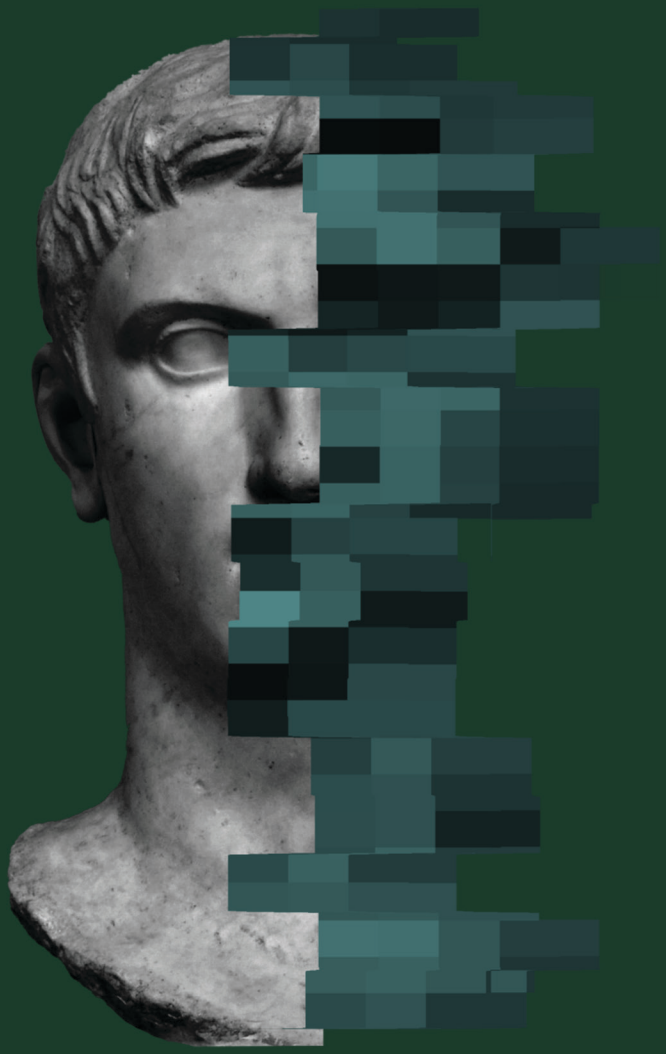
Vosoughi, Soroush, Deb Roy and Sinan Aral. 2018. "The Spread of True and False News Online." Research Brief. Cambridge, MA: MIT Initiative on the Digital Economy. https://ide.mit.edu/wp-content/uploads/2018/12/2017-IDE-Research-Brief-False-News.pdf.

Walker, Mason and Katerina Eva Matsa. 2021. "News Consumption Across Social Media in 2021." Pew Research Center, September 20. www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/.

Wark, Wesley. 2024. *Foreign Interference Online: Where Disinformation Infringes on Freedom of Thought*. Legitimate Influence or Unlawful Manipulation? Policy Brief No. 3. Waterloo, ON: CIGI. www.cigionline.org/publications/foreign-interference-online-where-disinformation-infringes-on-freedom-of-thought/.

Woodford, Scarlett. 2022. *Digital Advertising: Market Forecasts, Emerging Trends & Key Opportunities 2022–2026*. Basingstoke, UK: Juniper Research. www.juniperresearch.com/researchstore/innovation-disruption/digital-advertising-data-research-report.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London, UK: Profile Books.



About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Managing Director and General Counsel Aaron Shull
CIGI Senior Fellow and Project Co-Leader Susie Alegre
Director, Program Management Dianna English
Program Manager Jenny Thiel
Publications Editor Lynn Schellenberg
Senior Publications Editor Jennifer Goyder
Graphic Designer Abhilasha Dewan

This policy brief was made possible thanks to the financial support of the Konrad-Adenauer-Stiftung (KAS) Canada.

For publications enquiries, please contact publications@cigionline.org.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org