



American ideals. Universal values.

Fake News, Free Speech, And Foreign Influence: The Smart Way the United States Can Combat Disinformation

On January 23, 2018, through the generous support of the Henry M. Jackson Foundation, Human Rights First hosted a public conference entitled, “Fake News, Free Speech, and Foreign Influence: The Smart Way the United States Can Combat Disinformation.” Experts discussed conditions enabling exploitation of the digital public square, the effectiveness of proposed solutions, and the potential impact of various interventions on universal human rights.

Human Rights First gratefully acknowledges the participation of the following experts, and notes that all conclusions and recommendations presented in this paper reflect the sole views of Human Rights First.

- Jamie Fly, Senior Fellow and Director, Future of Geopolitics, Asia Program, Alliance for Securing Democracy, German Marshall Fund
- Nina Jankowicz, George F. Kennan Fellow, Wilson Center
- Amy MacKinnon, Development Editor at Coda Story
- Tiffany Li, Resident Fellow, Yale Law School, Information Society Project and Head of the Wikimedia/Yale Law School Initiative on Intermediaries and Information
- Emma Llansó, Director, Free Expression Project, Center for Democracy & Technology
- Jason Pielemeier, Policy Director, Global Network Initiative
- Shanti Kalanthil, Director, International Forum for Democratic Studies, National Endowment for Democracy

The latest in our series of expert-led discussions on Russian influence in Europe and potential tech-driven solutions, the event brought together leading thinkers on disinformation, the media, free speech, and technology at the National Press Club in Washington D.C. The goal of the program was to assess the contemporary state-of-play and share ideas on the emerging responses and responsibilities of governments, the technology industry, and civil society.

The following report captures the event’s main conclusions, ideas, and recommendations, building upon analysis and research conducted in recent years by Human Rights First.

Introduction

In January 2017, the U.S. intelligence community publicly released its assessment that Russian state and non-state actors led an influence campaign to impact the 2016 U.S. presidential election and undermine public faith in the democratic process.¹ Since that time, new findings, including recent indictments issued by Special Counsel Robert Mueller, continue to shed light on past and ongoing manipulations of public perception and behavior, particularly through the spread of false and inflammatory content on social media.²

Various governmental, quasi-governmental, and non-governmental actors have weaponized the “attention economy”—the marketing and monetization of online users’ time and attention—by shifting public opinions and behavior through disinformation, hate speech, harassment,

propaganda, and other harmful content spread at a scale and efficiency not possible before the rise of social media. From trolls harassing journalists in Hungary,³ to officials and ultranationalists inciting ethnic violence through Facebook in Burma,⁴ the proliferation of harmful speech online is a global, multifaceted threat to both democracy and security.

These challenges are spurring public and regulatory pressure on major technology companies—including, most prominently, Google, Facebook, and Twitter—who are, in turn, increasingly governing digital expression on their platforms. These industry leaders have updated their content moderation and advertising policies, strengthened enforcement mechanisms, removed content and users (not without controversy and error), and made promises to continue improving safety.

Yet, on both sides of the Atlantic, many policymakers remain unconvinced that companies are up to the task. In Germany, lawmakers have already intervened—imposing hefty fines on platforms that fail to expeditiously remove illegal third-party content reported by users or complaint bodies.⁵ In the United States, bipartisan groups of senators and House members introduced the Honest Ads Act, legislation that would increase disclosure requirements for online political ads and require reasonable effort from platforms to prevent foreign political interference.⁶

Meanwhile, free speech advocates fear the consequences of increasingly zealous content regulation by private companies, entities with no legal obligation to protect users' right to free expression or to enforce policies fairly and transparently. Laws imposing liability on information intermediaries for third party content, as in Germany, are of particular concern. Critics argue that these policies incentivize overbroad enforcement by inherently risk-averse companies, and enable governments to outsource content

regulation – that sometimes includes censorship -- to private entities.⁷

A Crisis in Faith: Disillusionment, Distrust, and Disinformation

Western democracies are facing a significant decline in public trust of institutions, leaders, and the media. These trends are made worse by disinformation. A recent 2017 Gallup/Knight Foundation study reported that only 32% of Americans had a “great deal or fair amount of trust in the media.”⁸ Meanwhile, the 2018 Edelman Trust Barometer has reported that the U.S. general population’s trust of business, media, NGOs, and government institutions declined 37% in the last year.⁹ President Trump’s condemnation of legacy media as “fake news,” and the administration’s legitimization of fringe outlets, such as Breitbart and, previously, Gateway Pundit, only further degrades public trust in the media.

In this context, policymakers should avoid treating disinformation as a solely Russian threat. They should instead consider it a systemic vulnerability, exploitable by a variety of actors and enabled by larger trends that require long-term solutions. Ad hoc attempts to combat disinformation through censorship—recent bans of Russian social media sites and other media in Ukraine, for example—are often ineffective and counterproductive, potentially marginalizing groups and decreasing trust in the government.¹⁰ In a similar vein, U.S. policymakers may have erred in designating RT and Sputnik as foreign agents under the Foreign Agents Registration Act (FARA). Given the limited reach of these outlets, this largely symbolic gesture provides cover for Russian authorities, and autocrats elsewhere, to further suppress legitimate journalism by arguing that their legislation or policy mirrors that of the U.S.

Recommendations:

■ Invest in Social and Educational Programs:

Countering the decline in public trust of the media and other democratic institutions requires a broad range of solutions. As new rounds of elections approach in the United States and elsewhere, authorities must raise public awareness about the potential threat of disinformation. The cure should not, however, be worse than the disease—exaggerating the scope of the threat may further erode faith in democratic process and media. Long term solutions will require expanding media literacy, critical thinking, and civics education, beginning at an early age. Public education should empower citizens with the tools to consume news critically, foster faith in our national ideals, and encourage active citizenship. Tech companies should nudge users towards improved digital citizenship and media literacy through innovative platform design and investment.¹¹

- ### ■ Support High-Quality Journalism and Civil Society:
- Governments should increase public funding for high-quality, local journalism. Media consolidation disconnects citizens from political life in their own communities and makes it difficult to assess local impacts of national policies. Social media platforms should support organizations engaged in countering harmful speech, fact-checking, or high-quality local journalism.¹² Google's Ad Grants, which gives qualified nonprofits \$10,000 of free text-ads on Google Search, is a positive step in this direction.

A Case Study: Understanding and Repelling Russian Online Disinformation Operations

Russia's online influence operations in democratic countries, particularly in the lead up to recent

elections in the United States, France, and Germany, have been characterized by a dynamic toolkit. Primary tactics include use of email phishing and hacking, strategic leaking of sensitive information, troll factories and "botnets," state-sponsored media, and micro-targeted political advertising. The overarching aim of this work is to stoke anger, apathy, and distrust toward democratic processes, institutions, and leaders. This "sharp power" strategy, which consists of both on- and off-line components, focuses on undermining the moral authority, prestige, and unity of adversaries, rather than on improving Russia's own attractiveness and influence.¹³

Observation of Russian online influence operations through tools such as the German Marshall Fund's Hamilton 68 and Artikel 38 dashboards—tools monitoring content promoted by a network of approximately six hundred Russia-linked Twitter accounts in English and German—reveal tactics more sophisticated than simply advancing overtly pro-Kremlin positions. Instead, these networks seize upon highly politicized or divisive topics as an opportunity to pit citizens against one another and shift the window of acceptable public discourse further to the margins.

For example, following the February 14 mass shooting at Marjory Stoneman Douglas High School in Parkland, Florida, Russian trolls and bots (automated accounts) inundated Twitter with anti-gun control messaging and disinformation tying the shooter to the far-left Antifa movement.¹⁴ Notably, these accounts often promote opposing sides of wedge issues, as is the case during the ongoing NFL national anthem controversy.¹⁵ As detailed in Mueller's recent indictment related to Russian interference in the 2016 election, fanning flames on both sides of contentious issues was the *modus operandi* of the Internet Research Agency (IRA), a troll factory in St. Petersburg, Russia. With a monthly budget of more than \$1.25

million, the organization created Facebook pages and accounts that falsely represented themselves as belonging to American users and groups, purchased user-targeted advertisements, and instigated protests from half-way around the world on causes as divergent as immigration policy and Black Lives Matter.¹⁶

Not all communities are equally resistant to the spread of Russian disinformation.¹⁷ In general, social media users tend to strictly limit their news sources,¹⁸ and to group closer to one another on the basis of perceived similarities—a sociological phenomenon known as “homophily.”¹⁹ This dynamic may result in polarized information bubbles—a trend particularly salient among communities that are traditionally skeptical of mainstream media.²⁰ Academic research increasingly demonstrates that the strength of political homophily on social media varies between groups with different political orientations, with recent research indicating more ideologically conservative individuals exhibiting higher rates than more centrist or left-leaning individuals.²¹ How different communities network and share information is an evolving area of inquiry and subject to change, but policymakers should not ignore the disproportionate impacts these differences may cause.

Recommendations:

- **Neutralize Malicious Bots:** As part of a disinformation campaign, automated accounts artificially amplify narratives and create the illusion of broad support. Social media companies should endeavor to identify bots through content and similarity mapping methods—with a particular focus on bots engaged in political speech—and closely review whether the accounts should be either conspicuously tagged as bots or removed from the platform. These enforcement mechanisms should be designed in close consultation with

civil liberties groups to avoid unnecessary harm to legitimate speech and privacy concerns of users, to ensure a fair review process, and to anticipate potential risks.

- **Improved Monitoring and Information Sharing between Allies:** Treaty bodies, including NATO, are an ideal venue for sharing knowledge. Policymakers should improve their ability to detect “canaries in the coal mine” by listening to the concerns of allies and continuing investment in artificial-intelligence-driven technologies to monitor and detect online disinformation and hate campaigns before they metastasize.²² A crucial area of future inquiry is the extent to which China will adopt a similar playbook as Russia; a trend already observed in Taiwan.

Gaming Democracy: How Social Media Exacerbates Disinformation

Social media has dramatically altered the creation, distribution, and consumption of information, weakening the grip of traditional information gatekeepers and providing users with an unprecedented surplus of information. The democratic dividends of this change are significant—propelling grassroots movements such as those involved in the 2011 Arab Spring, Black Lives Matter, and the 2017 Women’s March. Yet connecting users and individuals with information at such scale has had unforeseen consequences.

The design of major platforms contributes to polarization and ideological echo-chambers. Platforms in the “attention economy” draw likeminded users together (homophily) and algorithmically amplify content intended to appeal to individual users, resulting in so-called “filter bubbles.” These trends fuel partisanship and enable narratives, including false narratives, to incubate outside the scrutiny of an open

marketplace of ideas. As seen in the case of the Internet Research Agency, dedicated actors can embed themselves into these isolated groups, and use the confined nature of the conversation to insert and amplify narratives that advance their own objectives. These insular ecosystems, which allow for the propagation of ideas without competition from other facts and ideologies, advantage actors seeking to sow division and distrust, enabling those who seek to enflame existing grievances, doubts, and fears.

Platform algorithms tend to reward popularity—measured through “likes,” shares, views, comments, or other types of engagement—with greater visibility.²³ This model incentivizes demagoguery, virality, and sensationalism, while further eroding reliance on experts and nuanced coverage.²⁴ Furthermore, popularity-based systems are manipulated through coordinated efforts, often with the assistance of bots and professional trolls. Such artificial boosting legitimizes deceptive narratives through the illusion of broad support, while drowning out dissenting voices.²⁵

The advent of social media has precipitated a litany of challenges to an increasingly digital public square: reduced civility due to a real or imagined feeling of online anonymity; the duopolistic control of digital advertising dollars, a main source of revenue for publishers, by Facebook and Google; a business model that incentivizes publishers to prioritize rapid and bulk production of video content, instead of high-quality, long-form journalism; and social atomization and manipulation through highly-targeted advertising and other means.

An increasingly common response to these threats is imposing liability on social media platforms for illegal content posted by users. This is a dangerous path for free speech. A consequence of imposing liability for third-party content is overbroad regulation of legitimate, but

controversial, speech because technology companies and their corporate boards have a natural aversion toward taking unnecessary, costly risks.²⁶ Compelling private companies to interpret and enforce the law outsources censorship to an unaccountable entity that does not provide citizens the same procedural safeguards and transparency as a formal state action. Admittedly, the scale and speed of illegal content online makes it impracticable for each instance to be resolved in a court of law, but under such constraints, democratic societies should opt for speech-permissive, rather than speech-restrictive systems, and pursue enforcement against the worst offenders.

This points to another troubling extra-judicial trend: authorities establishing bodies, such as Europol’s Internet Referral Unit, to identify and refer content to Internet and social media companies for removal due to violation of a company’s terms of service.²⁷ This allows law enforcement to effectively censor content outside of a legal process, circumventing adequate access to remedy, accountability, and transparency.²⁸ Such programs can be effective at removing extremist content, but terms of service-based referrals require safeguards with respect to transparency and access to effective remedy from both governments and companies.²⁹

Recommendations:

■ Empower Researchers, Share Data

Responsibly: Significant knowledge gaps remain with respect to the reach and impact of disinformation campaigns, the real-world relationship between online hate and violence, the effectiveness of various counter-speech and fact-checking programs, and other areas. Researchers are frustrated by lack of access to privately held data and proprietary algorithms. Facebook’s semi-closed application programming interface (API), and its various

private pages and groups, have made it more difficult for researchers to investigate the world's largest social media network.³⁰ Data from Instagram and WhatsApp are almost completely inaccessible to researchers, even though these services are used heavily for political messaging and manipulation. Wherever possible, companies should strive to make their algorithms open access, and offer access to proprietary data to validated third-parties. Increased collaboration with outside researchers in these areas will help drive public and private policy in the right direction.

- **Increased Commitment to Multi-Stakeholder Initiatives and Improved Oversight:** The dynamism of the threat—and the potential to inflict harm in the course of responding—necessitates a holistic and multi-stakeholder approach for both public and private regulators. Tech companies should increase their commitment to the principles, implementation guidelines, and accountability mechanism established by multi-stakeholder bodies such as the Global Network Initiative (GNI). Whether voluntarily or through the establishment of an official oversight body, technology companies ought to be more transparent and inclusive with respect to their internal ethical review processes.
- **Urge Tech Companies to Recognize their Role as Instruments of Democracy:** Elected officials and the public must maintain pressure on tech companies to affirmatively view their role, in addition to seeking profit, as one of working to safeguard democratic institutions. The industry must recognize that whether unwittingly or negligently, their platforms are used to attack democratic values, ideals, and institutions; and they will have to take actions to combat and reverse this trend. These interventions should be non-censorial to the greatest extent possible; awareness of societal

obligations by big tech must be built around respect for the fundamental rights of users and democratic, not paternalistic, principles.

Conclusion

The challenge disinformation poses to democracy requires a multidimensional response, engaging governmental, civil society, and private sector solutions. Crucially, it also requires a clear-eyed assessment that these malicious campaigns exploit preexisting cleavages, particularly today's hyper-partisanship political environment, as well as recent, major disruptions to the media landscape and information diet. While broad-based declines in societal trust of the media, leaders, and institutions constitute the real disease, actors in government, the technology sector, and civil society can, and should, take immediate action to mitigate future attacks and increase long-term resiliency while protecting fundamental freedoms.

Endnotes

- ¹ Assessing Russian activities and Intentions in Recent US Elections, Office of the Director of National Intelligence, National Intelligence Council, January 6, 2017, https://www.dni.gov/files/documents/ICA_2017_01.pdf.
- ² United States v. Internet Research Agency, LLC, No. 1:18-cr-00032-DLF (D.D.C. filed Feb. 16, 2018).
- ³ See generally, Tamas Tofalvy, [Online Harassment of Journalists in Hungary: Forms, Coping Mechanisms and Consequences for Press Freedom](#), (Scott Griffen ed. & Javier Luque ed., International Press Institute, April 2017), https://ipi.media/wp-content/uploads/2017/03/006_Hungary_Report_Online_2017.pdf.
- ⁴ Megan Specia & Paul Mozur, [A War of Words Puts Facebook at the Center of Myanmar's Rohingya Crisis](#), THE NEW YORK TIMES, (Oct. 27, 2017), <https://www.nytimes.com/2017/10/27/world/asia/myanmar-government-facebook-rohingya.html> (describing the utilization of Facebook by prominent nationalists and even verified government or military accounts in Burma to spread misinformation and hate speech against the Rohingya minority); see also Casey Hynes, [Internet Use Is on the Rise in Myanmar, But Better Options are Needed](#), FORBES, (Sept. 22, 2017), <https://www.forbes.com/sites/chynes/2017/09/22/internet-use-is-on-the-rise-in-myanmar-but-better-options-are-needed/#58927bcc448e> (drawing attention to low media and digital literacy in Burma due to longstanding government restrictions and Facebook's position as the *de facto* Internet for most Burmese due to cheap data-plans that only access Facebook and poor accommodation of Burmese languages on other platforms).
- ⁵ Beschluempfehlung und Bericht [Resolution and Report], Deutscher Bundestag: Drucksache [BT] 18/13013, <http://dip21.bundestag.de/dip21/btd/18/130/1813013.pdf> (for an English-language version of the statute, see https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf).
- ⁶ Honest Ads Act, S. 1989, 115th Cong. (2017).
- ⁷ [Germany is Silencing 'Hate Speech' but Cannot Define It](#), The Economist, (Jan.13, 2018), <https://www.economist.com/news/europe/21734410-new-social-media-law-causing-disquiet-germany-silencing-hate-speech-cannot-define-it>. See also Adam Nagy, [German Social Media Regulation Inspires Repressive Regimes](#), Human Rights First, (Aug. 3, 2017), <https://www.humanrightsfirst.org/blog/german-social-media-regulation-inspires-repressive-regimes>.
- ⁸ Gallup/Knight Foundation, [American Views: Trust, Media, and Democracy](#), (Jan. 16, 2018) <https://knightfoundation.org/reports/american-views-trust-media-and-democracy>.
- ⁹ Edelman, [The 2018 Edelman Trust Barometer](#), (Jan. 21, 2018) <https://www.edelman.com/trust-barometer>.
- ¹⁰ See Committee to Protect Journalists, [Ukraine Bans Russian Media Outlets, Websites](#), (May 17, 2017), <https://cpj.org/2017/05/ukraine-bans-russian-media-outlets-websites.php> (statement of CPJ Europe and Central Asia Program Coordinator Nina Ognianova) (“Attempts to ban Russian media in Ukraine are antidemocratic, are likely to be ineffective, and could easily backfire by making the government appear afraid of allowing citizens to make up their own minds...”).
- ¹¹ Researchers and companies have tweaked platforms and created mechanisms to, among other things, combat misinformation on Reddit, advance gender equality on Twitter, and encourage civility and discourage bullying, hate speech, and harassment on one of the most popular online computer games, League of Legends. See J. Nathan Matias, [Persuading Algorithms with an AI Nudge](#), CivilServant, (Feb. 1 2017), https://civilservant.io/persuading_ais_preserving_liberties_r_worldnews.html; J. Nathan Matias, Sarah Szalavitz, & Ethan Zuckerman, [FollowBias: Supporting Behavior Change Toward Gender Equality by Networked Gatekeepers on Social Media](#), Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work & Social Computing, 1082-1095, (Feb. 2017); Brendan Maher, [Can a Video Game Company Tame Toxic Behavior?](#), 531 Nature 568-571, (Mar. 31, 2016), <https://www.nature.com/news/can-a-video-game-company-tame-toxic-behaviour-1.19647>.
- ¹² Nina Jankowicz, [Russian Trolls Are Only Part of the Problem](#), The New York Times, (Jan. 25, 2018) <https://www.nytimes.com/2018/01/25/opinion/russian-trolls-fake-news.html>.
- ¹³ Christopher Walker & Jessica Ludwig, [The Meaning of Sharp Power](#), Foreign Affairs, (Nov. 16, 2017), <https://www.foreignaffairs.com/articles/china/2017-11-16/meaning-sharp-power>. See also Joseph S. Nye Jr., [How Sharp Power Threatens Soft Power](#), Foreign Affairs, (Jan. 24, 2018), <https://www.foreignaffairs.com/articles/china/2018-01-24/how-sharp-power-threatens-soft-power>.
- ¹⁴ Erin Griffith, [Pro-Gun Russian Bots Flood Twitter After Parkland Shooting](#), Wired, (Feb. 15, 2018), <https://www.wired.com/story/pro-gun-russian-bots-flood-twitter-after-parkland-shooting/>.
- ¹⁵ Devlin Barrett, [Lawmaker: Russian Trolls Trying to Sow Discord in NFL Kneeling Debate](#), Washington Post, (Sept. 27, 2017), https://www.washingtonpost.com/world/national-security/lawmaker-russian-trolls-trying-to-sow-discord-in-nfl-kneeling-debate/2017/09/27/5f46dce0-a3b0-11e7-ade1-76d061d56efa_story.html?utm_term=.ccd8aa0c85a8.
- ¹⁶ *Id.* at 7, 14.
- ¹⁷ See [The Impact of Russian Interference on Germany's 2017 Elections: Testimony Before the U.S. Senate Select Committee on Intelligence](#), 115th Cong. (2017) (testimony of Constanze Stelenmüller, Brookings Institute). (describing inherent strengths of Germany in countering Russian meddling: “Its politics are far less polarized than, say those of the United States, or the United Kingdom; income and education inequality is far less drastic than in the Anglo-Saxon world. It has healthy institutions, a functioning representative democracy, and genuine political pluralism. Its economy is strong, its public education good. It possesses a large number of independent, quality media organizations which are still trusted by consumers, and social media are less relevant for public debate than elsewhere... It also helps that Germany is not the first country to face this issue in an election...”). For more information on differences in social media consumption between Germany, France, and the US see Nic Newman et al., [Reuters Digital News Report 2017](#), Reuters Institute for the Study of Journalism, (May 2017), <http://www.digitalnewsreport.org/> (according to this report 41% of online users in the US share news electronically weekly, in comparison only 32% of French users and 18% of German users – additionally, sharing and commenting has risen in the US over the last three years and declined in Germany).
- ¹⁸ See Ana Lucía Schmidt et al. [Anatomy of News Consumption on Facebook](#), 114: 12, Proceedings of the National Academy of Sciences 3035–3039 (2017).
- ¹⁹ See generally Itai Himelboim et al., [Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter](#), 18: 2 Journal of Computer-Mediated Communication 40–60 (2013); Jae Kook Lee et al., [Social Media Network Heterogeneity, and Opinion Polarization](#), 64: 4 Journal of Communication 702–22 (2014); Miller McPherson et al.,

Birds of a Feather: Homophily in Social Networks, 27 Annual Review of Sociology 415 – 44 (2001) (discussing homophily in a variety of offline social networks).

²⁰ Robert Faris et. al., Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election, Berkman Klein Center Research Publication, (2017), https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3028120_code727672.pdf?abstractid=3019414&mirid=1 ; See also Andrew Guess et. al., Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 U.S. Presidential Campaign (2018), <https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf>.

²¹ See generally Andrei Boutyline & Robb Willer, The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks, 38: 3 Political Psychology, 551–569 (2017); Pablo Barberá et. al., Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?, 26: 10 PSYCHOLOGICAL SCIENCE, 1531–1542 (2015).

²² Defense Advanced Research Projects Agency, Social Media in Strategic Communication (SMISC) (Archived), <https://www.darpa.mil/program/social-media-in-strategic-communication> (“The general goal of the Social Media in Strategic Communication (SMISC) program is to develop a new science of social networks built on an emerging technology base. Through the program, DARPA seeks to develop tools to help identify misinformation or deception campaigns and counter them with truthful information, reducing adversaries' ability to manipulate events.”) a list of projects and papers funded by this program, none of which were classified, is available at the DARPA Open Catalog see DARPA, Social Media in Strategic Communication, (Nov. 13, 2015), <https://opencatalog.darpa.mil/SMISC.html>.

²³ Antonio García Martínez, How Trump Conquered Facebook—Without Russian Ads, *Wired*, (Feb. 23, 2018) <https://www.wired.com/story/how-trump-conquered-facebook-without-russian-ads/> (examining features of Facebook advertising infrastructure that give favor to more engaging – whether through a click, like, comment etc. – ads and posts, even)

²⁴ Bharat N. Anand, The U.S. Media's Problems are Much Bigger Than Fake News and Filter Bubbles, *Harvard Business Review*, (Jan. 05, 2017) <https://hbr.org/2017/01/the-u-s-medias-problems-are-much-bigger-than-fake-news-and-filter-bubbles> (discussing structural conditions on the supply and demand side of the contemporary media market that incentivize sensational and polarizing content).

²⁵ Gary King, Jennifer Pan, & Margeret E. Roberts, How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument, 111, 3, *American Political Science Review* 484—501 (2017) (estimating that the Chinese government is responsible for approximately 448 million social media comments a year and providing evidence that the goal of this program is not to engage in debate over controversial topics but to distract users and change subjects.)

²⁶ Daphne Keller, Empirical Evidence of “Over-Removal” By Internet Companies Under Intermediary Liability Laws, The Center for Internet and Society at Stanford Law School, (Jun. 15, 2017) <http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws> (a collection of academic studies documenting the over-removal of legal content by intermediaries as a consequence of notice and takedown systems in various jurisdictions); see also Jack Balkin, Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation, 51: 3 *U.C. Davis L. Rev.* 1149 (2018); Seth Kreimer, Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link, 155 *U. PENN. L. REV.* 11 (2006).

²⁷ Europol, Counter-Terrorism Specialists Team Up to Take Down Online Terrorist Propaganda, (Sept. 05, 2016) <https://www.europol.europa.eu/newsroom/news/counter-terrorism-specialists-team-to-take-down-online-terrorist-propaganda>.

²⁸ Lucie Krahulcova, Europol's Internet Referral Unit Risks Harming Rights and Feeding Extremism, *Access Now*, (Jun. 17, 2016), <https://www.accessnow.org/europols-internet-referral-unit-risks-harming-rights-isolating-extremists/>.

²⁹ Global Network Initiative, Extremist Content and the ICT Sector, (Nov. 2016) <https://globalnetworkinitiative.org/sites/default/files/Extremist-Content-and-the-ICT-Sector.pdf> (outlining recommendations to improve transparency and offer effective remedy to users with respect to government referral of alleged terms of service violations to companies).

³⁰ Jonas Kaiser, Loud, Influential, and Hyped: How we Overestimate Twitter, *Zurich Institute of Public Affairs Research*, (Dec. 22, 2017) <https://zipar.org/commentary/loud-influential-and-hyped-how-we-overestimate-twitter/>.