Institutional Counter-disinformation Strategies in a Networked Democracy

Jonathan Stray
Columbia Journalism School, New York, NY, USA, jms2361@columbia.edu

ABSTRACT

How should an organized response to disinformation proceed in a 21st century democratic society? At the highest level, what strategies are available? This paper attempts to answer these questions by looking at what three contemporary counter-disinformation organizations are actually doing, then analyzing their tactics. The EU East StratCom Task Force is a contemporary government counter-propaganda agency. Facebook has made numerous changes to its operations to try to combat disinformation, and is a good example of what platforms can do. The Chinese information regime is a marvel of networked information control, and provokes questions about what a democracy should and should not do. The tactics used by these organizations can be grouped into six high level strategies: refutation, exposure of inauthenticity, alternative narratives, algorithmic filter manipulation, speech laws, and censorship. I discuss the effectiveness and political legitimacy of these approaches when used within a democracy with an open Internet and a free press.

CCS CONCEPTS

• Security and privacy~Social aspects of security and privacy • Social and professional topics~Censorship • Information systems~Content ranking

KEYWORDS

Disinformation, policy, democracy, refutation, content ranking, censorship

1 Institutional Responses to Disinformation

There has been a great deal written about responses to online disinformation within the computer science, social science, and political science communities in the last few years. Such work includes technical approaches such as automated fact checking [1], psychological experiments concerning belief and persuasion [2], documentation of active disinformation campaigns [3], and proposed responses to specific threats [4].

There is also a need for a big-picture view. A robust societal response to disinformation will require long-lived institutions executing politically acceptable strategies. Institutionalized disinformation response within democracies peaked during the Cold War. At that time the U.S. had multiple agencies and programs tasked with countering disinformation and getting out an American narrative [5], the most significant of which was the United States Information Agency (1953-1999) which once had a staff of thousands and a budget of billions [6].

While there are good descriptions of Cold War counter-disinformation institutions, there is much less documentation of current organized efforts which must deal with the networked world. There are a few theory-driven analyses (such as Hellman and Wagnsson's typology of strategies [7]), but little that synthesizes current practice. This paper attempts to fill this gap by outlining the work of three contemporary counter-disinformation organizations, then categorizing their tactics into higher-level strategies.

Having identified these strategies, this paper considers which are effective and legitimate in a pluralistic democracy. However, I make no attempt to define "disinformation." Adjudicating which speech is harmful is a profound problem with millennia of history. Instead, the emphasis here is on potential responses and their alignment with the values of an open society, *given that* a particular narrative is held to be damaging and worthy of countering.

Section 2 relates brief case studies of three organizations. Section 3 groups and discusses the types of tactics used across these organizations. Each strategy has certain advantages and disadvantages in terms of efficacy (how well it works) and legitimacy (alignment with the values of an open society.) Section 4 concludes.

2 Institutional Case Studies

2.1 Case study: the East StratCom Task Force

The EU East StratCom Task Force was created in 2015 at the direction of the European Council to respond to "Russia's ongoing disinformation campaigns" [8]. It is a modern example of a government counter-propaganda organization, and focusses on tracking and rebutting state propaganda.

It currently has just 14 full time staff [8] and funding of €1.1 million per year [9]. This small team of coordinates with a volunteer network of some 500 journalists, diplomats, NGO staff, and interested amateurs who assist in media monitoring and analysis in many countries [10]. The unit publishes under the "EU vs. Disinfo" brand, and produces the weekly "Disinformation Review" along with rebuttals of specific items of disinformation. The content is majority English with substantial Russian.

The unit reports that it debunked 1310 "disinformation cases" in 2017 [11] and over 4,700 since founding [12]. It's encouraging that individual false messages are being systematically tracked, and this database of cases is online and searchable.

One of the most striking things about the East StratCom Task Force is how small it is, both in comparison to its main adversary and relative to a previous generation of counter-disinformation efforts. For comparison, the Russian Internet Research Agency -- just one arm of the Russian propaganda effort -- employs on the order of a thousand people [13] [14]. Many EU politicians and experts are concerned that the East StratCom Take Force effort is greatly underpowered [15].

2.2 Case study: Facebook

American Social media platforms' disinformation practices have been under intense scrutiny since the 2016 election [16]. Facebook is particularly interesting because of its enormous scale, which both increases its impact and makes it a frequent target of organized disinformation campaigns.

Facebook says that most of the politically-themed disinformation or "fake news" on the platform is not politically motivated propaganda, but financially motivated clickbait [17], produced to drive users offsite and generate ad revenue [18]. Regardless of motivation, Facebook has responded to disinformation in three general ways: identifying false content, identifying untrustworthy sources, and identifying "inauthentic" accounts.

Since December of 2016, Facebook users have been able to report an article as "fake news." [19] Facebook also maintains machine learning models that try to identify posts as disinformation, including links, photos, and videos [20]. However identified, these items are then sent to third party fact checkers for verification. These fact checkers are selected from those certified under Poynter's International Fact Checking Code of Principles [21] and include well-known organizations such as Politifact. When a fact checker rates an article as false, it is demoted in the news feed [22]. Regardless of rating, Facebook displays the resulting fact check as a "related article" whenever the original link is shared [23]

Most of what is publicly known about this program comes from interviews with the fact checking organizations involved [24]. In the U.S., Facebook currently works with six fact-checking partners who receive candidate disinformation URLs through a web "dashboard." At any one time there are perhaps a few thousand posts on the dashboard. Properly checking a claim is a meticulous and often slow research process, and one organization reported that it checked 90 stories in one year [24]. It is unclear if the overall scale of the effort has any appreciable impact on user exposure to disinformation.

Perhaps because of this potential scale mismatch, Facebook has also started tackling disinformation at the level of sources, not just individual posts. A 2018 survey asked users “how much do you trust each of these domains?” and the resulting scores are used to “inform ranking” [25]. Again, there is no public information on the results of this program. It’s also not clear how well user trust ratings align with the truth or falsity of sources, though perhaps surprisingly there may not be much partisan difference in American users’ source rankings [26].

Meanwhile, there is an active effort at Facebook to identify and remove "inauthentic" accounts and pages, which involve a false identity or deceptive activities, and not necessarily false content. The efforts of the Russian Internet Research Agency (IRA) during the 2016 election are the best documented example of this sort of “inauthenticity.” [27] The IRA created approximately 470 Russian-controlled pages on divisive topics, with titles such as "Blacktivist", "Secured Borders," and "LGBT United" [28]. The true operators of these pages were hidden in a variety of ways, including fake accounts used to administer the page.

2.3 Case study: the Chinese information regime

21st century China provides an in-depth study of the possibilities of state information control in the networked era. Clearly the Chinese Communist Party has a particular view of what constitutes "disinformation," but consider that some kinds of Chinese disinformation are easily recognizable to Westerners. There are commercial scams of various sorts [29] and also rumors, like the viral misconception that iodized salt could prevent radiation poisoning from Fukushima fallout, which caused a buying panic [30].

Chinese control of online information dissemination has tightened considerably since President Xi Jinping took power in 2013. Creemers has written a thorough review [29]. For a few years peaking at about 2012, Chinese "microblogging" services such as Weibo provided a relatively uncontrolled public conversation. Many celebrities and other high profile people had millions, in some cases tens of millions of followers. But then a number of widely shared viral messages painted the regime in unflattering ways. Sometimes the claims were true, as when citizens collected and posted evidence of corrupt officials, and sometimes they were false, as with the claim that after the 2011 Wenzhou train crash the government "had paid more than 200 million Yuan in compensation for a foreign passenger who died." The original poster of this falsehood was sentenced to three years' imprisonment [29] [31].

The Xi government decided they had a new information battleground, and made a number of moves over the next few years. Online sites and platforms must be licensed by the government. Platforms are legally responsible for their users' posts, in stark contrast to the U.S. legal framework [32]. A Real Names law ensures that platforms and apps can link users to offline identities. If a post found to be "unlawful" is shared 500 times the poster faces up to three years in prison. A 2014 report claims two million people are involved in online content monitoring, across 800 organizations [33]. And the "fifty cent party" of paid pro-government accounts posts an estimated 450 million comments per year. [34]

The government has long created and disseminated its own narratives through state media organs. The so-called "50c party" (meaning "fifty cents per post") is a different approach: a network of pro-party social media accounts which are not publicly connected to the government. Yet a trove of leaked 50c worker reports shows that most posters are in fact local government employees, who often have unrelated day jobs while posting on the side [34]. But the posts do not directly engage disinformation. King, Pan, and Roberts analyze the leaked reports to conclude,

> ...almost none of the Chinese government’s 50c party posts engage in debate or argument of any kind. They do not step up to defend the government, its leaders, and their policies from criticism, no matter how vitriolic; indeed, they seem to avoid controversial issues entirely. Instead, most 50c posts are about cheerleading and positive discussions of valence issues.

We also detect a high level of coordination in the timing and content in these posts. A theory consistent with these patterns is that the strategic objective of the regime is to distract and redirect public attention from discussions or events with collective action potential.

3 A Taxonomy of Tactics

Each of the institutions in these case studies has adopted specific tactics for fighting disinformation. I've collected and classified these to produce the taxonomy in Table 1. Each of these strategies will be discussed in turn, with particular attention to which are effective or desirable within a democracy. Centralized media control is not usually possible with a free press, and heavy handed strategies may conflict with democratic values such as free speech. Again, there is no attempt here to say which sorts of narratives are worthy of countering. No matter what is decreed to be disinformation, there is still the ethical question of which means of countering it best align with democratic values.

Table 1: Counter-disinformation strategies used by the three institutions in this paper, and their effectiveness and legitimacy in a democratic society.

Strategy	Used by	Effectiveness	Legitimacy
Refutation	EU Stratcom Facebook via fact-checkers	Works if consistent, but not all disinfo is about facts.	Generally legitimate to speak the truth, though people will disagree on what truth is.
Expose inauthenticity	EU Stratcom Facebook	Discredits the source, provides justification for further measures.	Content-neutrality is appealing. Important to preserve legitimate anonymity.
Alternative narratives	EU Stratcom China	Helps displace disinfo, inoculates against it if seen first.	Can itself be disinfo or distraction.
Algorithmic filter manipulation	Facebook China via 50c party	Media algorithms have huge effect on information exposure.	Platforms may abuse this power, users may game it.
Speech laws	Facebook enforces such laws China	Can be effective at targeting narrow categories of speech.	Broad laws against untruth are draconian.
Censorship	China	Effective when centralized media control is possible.	Generally conflicts with free speech.

3.1 Refutation

Refutation, rebuttal, or debunking might be the most obvious counter-strategy. It's also well within the bounds of democracy, as it's simply "more speech."

It may not be the most effective, because reasoned counter-arguments don't necessarily change someone's belief [4] [35] [36]. While some experiments have shown a "backfire" effect where attempted corrections entrench false beliefs [37], more recent research suggests that backfire is less common than previously thought [38] [2] and that refutation does work to correct factual knowledge ("oh, Clinton was wrong, truck

drivers do not pay more taxes than hedge fund managers...") while being unlikely to change political attitudes ("...but I'm still voting for her.") It's most effective if it's done consistently over the long term [4] and in any case it's practiced by most counter-disinformation organizations.

The East StratCom Task Force states that it operates primarily by refuting "stories contradicting publicly available facts." [39] This same "non-factual" claim is the moral heart of journalistic fact-checking, and Facebook's practice of demoting articles rated false. Alas, the real world is more complex than true or false. East StratCom's database of disinformation cases [12] includes many items that are factually correct, but omit information in order to suggest a different meaning (Wardle classifies this as "misleading content" [40]). Such messaging is more difficult to refute as simply "false."

Interestingly, China does not engage with government critics, at least not on social media. Non-response is also a classic PR strategy. This suggests that public refutation is not a necessary tactic, and can even be harmful.

3.2 Expose inauthenticity

One of the oldest and best-recognized forms of disinformation is pretending to be someone you are not. Bot networks, "astroturfing," and undisclosed agendas or conflicts of interest could all be considered inauthentic communication. The obvious response is to discredit the source by exposing it.

China requires all apps and services to register users under their full legal names, which could be considered a pre-emptive defense against misattribution. Yet there are democratically important uses for anonymity. Insisting on traceable identities can be a severe security problem for human rights workers, journalists, and activists. Facebook says that this is why it does not require page administrators to publicly disclose their identity.

However, Facebook still knows who administers a page, and they use this information to help detect "inauthenticity." According to the company, the divisive pages created by the Russian IRA were not shut down not because of what they said -- after all, they often repeated memes and quotes from people who were genuinely involved in various political issues. Rather, they were removed because the administrators were hiding their true identities as Russian agents. Similarly, Facebook removed 30,000 French "sock puppet" or "bot" accounts which they identified "by analyzing the inauthenticity of the account and its behaviors, and not the content the accounts are publishing" [27] and continues to remove accounts involved in "coordinated inauthentic behavior" in many countries [41].

The authenticity approach is morally and politically appealing because it rests on a widely shared communicative norm: pretending to be someone you are not is unethical. This avoids the difficult question of deciding what someone can be allowed to say. This idea of "content neutrality" is a key concept in U.S. First Amendment law, referenced in several Supreme Court decisions [42]. Critics charge that a "content neutrality" doctrine is fundamentally incoherent because speech policies are never truly neutral, and can always be shaped to target particular content indirectly. Indeed, in crafting policies and directing enforcement, Facebook is exercising considerable discretion in what sort of speech to allow.

3.3 Alternative narratives

A long line of experimentation suggests that merely saying that something is false is less effective than providing an alternative narrative [4] [43] [37], and the non-platforms in this paper combat disinformation in part by promoting their own narrative.

One of the East StratCom Task Force's primary missions is to represent EU actions and values to its target audiences [44]. This provides a ready-made positive narrative which may "inoculate" citizens against disinformation, or even displace it. The Chinese 50c party also promotes alternative narratives, albeit while pretending to be regular citizens as opposed to government workers.

At the nation-state level, alternative narratives could easy wander into suspicious territory. For this reason, much of the US government is prohibited by law from trying to influence domestic opinion [45]. Public messaging is an essential part of a government's role, but it should not itself be disinformation.

3.4 Algorithmic filter manipulation

The rise of platforms creates a truly new way of countering disinformation: demote it by decreasing its ranking in search results and algorithmically generated feeds. Conversely, it is possible to promote alternative narratives by increasing their ranking.

On Facebook, items judged as false "appear lower in News Feed" [19] and “typically lose 80 percent of their traffic." [46] Users ratings of source trustworthiness similarly “inform ranking” [25].

While platform operators have the power to directly punish disinformation, there are other strategies for manipulating search rankings. High volume alternative narratives, as in the case of the 50c party, may be interpreted by algorithmic ranking systems as “popular” or “trending” and thereby displace other content from the resulting recommendations. This can be thought of as a technical means of distraction.

Because of their vast scale, platforms must rely on users to help moderate content by flagging spam, pornography, incitement to violence, and other undesirable posts. Facebook's disinformation detection system includes user reports [19]. Such reporting systems can be influenced by organized campaigns, which have sometimes been able to demote content by directing an allied audience to report it, down-vote it, etc., a form of "user-generated censorship" [47]. But ignoring audience ratings entirely is not a good answer. A distributed, bottom-up campaign to bring algorithmic attention to some issue may be a perfectly legitimate and democratic tactic. This puts platforms in the difficult position of deciding which types of collective action to reward and which to punish.

3.5 Speech laws

While there are U.S. based fake news factories targeting domestic audiences [48] [49] I do not know of a case where a contemporary disinformation creator was prosecuted. This may not even be possible because the Supreme Court has held that the First Amendment generally protects lying; the major exceptions concern defamation and fraud [50]. In Europe, the recent report of the High Level Expert Group on Fake News and Online Disinformation recommended against attempting to regulate disinformation [51].

China has a different approach, where people who have posted widely-shared false information on social media have been fined or imprisoned. China has also imposed licensing requirements on online platforms which publish "news," and makes platforms liable for the posts of their users. These policies create a credible threat which forces platforms to fight falsehoods (and politically unacceptable stories.) [29]

By contrast, platforms in the U.S. operate under the legal framework of section 230 of the Communications Decency Act and are not generally responsible for the speech of their users. A number of experts and scholars feel that this is part of the reason there has been so much American innovation in social media [32]. But in most democracies platforms are still legally liable for hosting certain types of content. For example, Germany requires platforms to remove Nazi-related material within 24 hours or face fines [52]. These sorts of targeted regulations can be successful in the sense of largely removing certain narrow categories of posts from platforms.

3.6 Censorship

One way of combatting disinformation is simply to remove it from public view. In the 20th century, censorship was sometimes possible through control over broadcast media. This is difficult with a free press, and it is even harder to eliminate information from a networked ecosystem. Yet platforms do have the power to remove content entirely and often do, both for their own reasons and as required by law.

Censorship is generally considered suspect in a democracy. The EU's High Level Expert Group on Fake News and Online Disinformation holds that "any form of censorship either public or private should clearly be avoided." [51]

There is no such aversion in China, which employs a variety of interlocking censorship measures. The Great Firewall prevents domestic access to offending foreign sources. Domestic platforms and publishers are licensed and held responsible for the content they distribute, which creates a self-censorship regime. Indeed, every Chinese platform has an army of people whose job it is to monitor user content for falsehoods and politically sensitive speech [53] [29]. WeChat has an ever-changing list of words that cannot be used in a group chat [54].

American platforms also employ huge numbers of moderators tasked with removing things like pornography, incitements to violence, copyright violations, and certain other material as required by law, but not typically falsehoods or political material [55]. Outside of such specific categories, most Western platforms attempt to be "content neutral" and do not remove user material. This does not mean they do not address disinformation. Facebook does not remove content marked as false by outside fact checking organizations, but it does present that marking to its users. This seems more in line with democratic ideals.

4 Conclusion

Despite their differences, there are many common patterns between the East StratCom Task Force, Facebook, and the Chinese government. I've grouped these into six broad categories of contemporary counter-disinformation tactics, summarized in Table 1. Each has certain advantages and disadvantages in terms of efficacy and legitimacy -- that is, alignment with the values of an open society.

A cross-sector response -- both distributed and coordinated -- is perhaps the biggest challenge. In societies with a free press there is no one with the power to direct all media outlets and platforms to refute or ignore or publish particular items, and it seems unlikely that people across different sectors of society would even agree on what is disinformation and what is not. In the U.S. the State Department [56], the Defense Department [57], academics [58], journalists [21], technologists [59] and others have all launched their own more-or-less independent counter-disinformation efforts. Distributed work is essential, but coordination is an operational advantage. In many countries, a coordinated response will require coming to terms with a deeply divided population. Citizens will require strong assurances that the strategies employed to counter disinformation are both effective and aligned with democratic values.

REFERENCES

[1] B. Adair, C. Li, J. Yang and C. Yu, "Progress Toward "the Holy Grail": The Continued Quest to Automate Fact-Checking," in *Computation + Journalism Symposium*, Evanston, 2017.

[2] E. Porter and T. Wood, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior,* January 2018.

[3] S. C. Woolley and P. N. Howard, *Computational Propaganda Worldwide: Executive Summary,* Oxford Internet Institute, 2017.

[4] C. Paul and M. Matthews, "The Russian "Firehose of Falsehood" Propaganda Model Why It Might Work and Options to Counter It," RAND Perspectives, 2016. [Online]. Available: https://www.rand.org/pubs/perspectives/PE198.html.

[5] H. Romerstein, "Counterpropaganda: We can't win without it," in *Strategic Influence: Public Diplomacy, Counterpropaganda, and Political Warfare*, J. M. Waller, Ed., Institute of World Politics Press, 2009, pp. 137-180.

[6] N. J. Cull, The Cold War and the United States Information Agency: American Propaganda and Public Diplomacy, 1945-1989, Cambridge University Press, 2009.

[7] M. Hellman and C. Wagnsson, "How can European states respond to Russian information warfare? An analytical framework," *European Security,* vol. 26, no. 2, 2017.

[8] "Questions and Answers about the East StratCom Task Force," European External Action Service, 12 May 2018. [Online]. Available: https://eeas.europa.eu/headquarters/headquarters-homepage_en/2116/%20Questions%20and%20Answers%20about%20the%20East%20StratCom%20Task%20Force.

[9] J. Rankin, "https://www.theguardian.com/world/2017/nov/25/eu-anti-propaganda-unit-gets-1m-a-year-to-counter-russian-fake-news," *The Guardian,* 17 November 2017.

[10] C. McDonald-Gibson, "The E.U. Agency Fighting Russia's Wildfire of Fake News with a Hosepipe," *Time,* 11 September 2017.

[11] EU Vs. Disinfo, "Figure of the Year: 1310," 20 December 2017. [Online]. Available: https://euvsdisinfo.eu/figure-of-the-year-1310/.

[12] EU Vs. Disinfo, "Disinformation Cases," [Online]. Available: https://euvsdisinfo.eu/disinformation-cases/. [Accessed 13 1 2019].

[13] T. Lister, J. Sciutto and M. Ilyushina, "Putin's 'chef,' the man behind the troll factory," 17 October 2017. [Online]. Available: https://www.cnn.com/2017/10/17/politics/russian-oligarch-putin-chef-troll-factory/index.html.

[14] N. MacFarquhar, "Inside the Russian Troll Factory: Zombies and a Breakneck Pace," 18 February 2018. [Online]. Available: https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html.

[15] M. Maksimovic, "EU officials warn of 'underestimating' Russian propaganda in Balkans," *Deutsche Welle,* 14 November 2017.

[16] M. Isaac and D. Wakabayashi, "Russian Influence Reached 126 Million Through Facebook Alone," *New York Times,* 30 October 2017.

[17] E. Schrage, "Hard Questions: Russian Ads Delivered to Congress," 2 October 2017. [Online]. Available: https://newsroom.fb.com/news/2017/10/hard-questions-russian-ads-delivered-to-congress/.

[18] L. Moses, "'The underbelly of the internet': How content ad networks fund fake news," 28 November 2016. [Online]. Available: https://digiday.com/media/underbelly-internet-fake-news-gets-funded/.

[19] A. Mosseri, "News Feed FYI: Addressing Hoaxes and Fake News," Facebook, 15 December 2016. [Online]. Available: https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/.

[20] A. Woodford, "The Hunt for False News," Facebook, 19 October 2018. [Online]. Available: https://newsroom.fb.com/news/2018/10/inside-feed-hunt-false-news-october-2018/.

[21] Poynter Institute, "The International Fact-Checking Network," 2017 September. [Online]. Available: https://www.poynter.org/channels/fact-checking.

[22] D. Zigmond, "Machine Learning, Fact-Checkers and the Fight Against False News," Facebook, 8 April 2018. [Online]. Available: https://newsroom.fb.com/news/2018/04/inside-feed-misinformation-zigmond/.

[23] J. Smith, "Designing Against Misinformation," 20 December 2017. [Online]. Available: https://medium.com/facebook-design/designing-against-misinformation-e5846b3aa1e2.

[24] M. Ananny, "The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation," Tow Center for Digital Journalism, 4 April 2018. [Online]. Available: https://www.cjr.org/tow_center_reports/partnership-press-facebook-news-outlets-team-fight-misinformation.php.

[25] A. Mosseri, "Helping Ensure News on Facebook Is From Trusted Sources," Facebook, 19 January 2018. [Online]. Available: https://newsroom.fb.com/news/2018/01/trusted-sources/.

[26] G. Pennycook and D. G. Rand, "Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality," *Proceedings of the National Academy of Sciences,* vol. 116, no. 7, 2019.

[27] J. Weedon, W. Nuland and A. Stamos, *Information Operations and Facebook,* Facebook, 2017.

[28] C. Timberg, "Russian propaganda may have been shared hundreds of millions of times, new research says," The Washington Post, 5 October 2017. [Online]. Available: https://www.washingtonpost.com/news/the-switch/wp/2017/10/05/russian-propaganda-may-have-been-shared-hundreds-of-millions-of-times-new-research-says.

[29] R. Creemers, "Cyber China: Upgrading Propaganda, Public Opinion Work and Social Management for the Twenty-First Century," *Journal of Contemporary China,* vol. 26, no. 103, 2017.

[30] L. Burkitt, "Fearing Radiation, Chinese Rush to Buy…Table Salt?," *The Wall Street Journal,* 17 March 2011.

[31] C. Yim, "Rumors Confession," 12 April 2014. [Online]. Available: https://www.pressreader.com/china/china-daily/20140412/281530814003426.

[32] The Electronic Frontier Foundation, "CDA 230: The most important law affecting Internet speech," [Online]. Available: https://www.eff.org/issues/cda230. [Accessed 31 12 2017].

[33] M. Fong, "China Monitors the Internet and the Public Pays the Bill," *Global Voices Advox,* 20 July 2014.

[34] G. King, J. P. Pan and M. E. Roberts, "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument," *American Political Science Review,* vol. 111, no. 3, 2017.

[35] S. Lewandowsky, U. K. H. Ecker, C. M. Seifert, N. Schwarz and J. Cook, "Misinformation and Its Correction Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest,* vol. 13, no. 3, 2012.

[36] M. Chessen, "Understanding the Psychology Behind Computational Propaganda," Advisory Commission on Public Diplomacy, May 2017. [Online]. Available: https://uscpublicdiplomacy.org/story/can-public-diplomacy-survive-internet.

[37] C. Silverman, "The Backfire Effect," *Columbia Journalism Review,* 17 June 2011.

[38] E. Porter, "Facts matter, and people care: An empirical perspective," Advisory Commission on Public Diplomacy, May 2017. [Online]. Available: https://uscpublicdiplomacy.org/story/can-public-diplomacy-survive-internet.

[39] EU vs. Disinfo, "COMMENTARY: Means, goals and consequences of the pro-Kremlin disinformation campaign," 19 January 2017. [Online]. Available: https://euvsdisinfo.eu/commentary-means-goals-and-consequences-of-the-pro-kremlin-disinformation-campaign.

[40] C. Wardle, "Fake news. It's complicated.," *First Draft News,* 16 February 2017.

[41] N. Gleicher, "Coordinated Inauthentic Behavior Explained," Facebook, 6 December 2018. [Online]. Available: https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/.

[42] E. Chemerinsky, "Content Neutrality as a Central Problem of Freedom of Speech: Problems in the Supreme Court's Application," *Southern California Law Review,* vol. 74, 2001.

[43] J. Henick and R. Walsh, "U.S. 2016 Elections: A case study in "inoculating" public opinion against disinformation," Advisory Commission on Public Diplomacy, May 2017. [Online]. Available: https://uscpublicdiplomacy.org/story/can-public-diplomacy-survive-internet.

[44] European External Action Service, "Questions and Answers about the East StratCom Task Force," 8 November 2017. [Online]. Available: https://eeas.europa.eu/headquarters/headquarters-homepage_en/2116/%20Questions%20and%20Answers%20about%20the%20East%20StratCom%20Task%20Force.

[45] W. R. Sager, "Apple Pie Propaganda: The Smith-Mundt Act before and after the Repeal of the Domestic Dissemination Ban," *Northwestern University Law Review,* vol. 109, no. 2, 2014.

[46] T. Lyons, "News Feed FYI: Replacing Disputed Flags with Related Articles," Facebook, 20 December 2017. [Online]. Available: https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/.

[47] C. Peterson, "User-generated censorship : manipulating the maps of social media," June 2013. [Online]. Available: http://hdl.handle.net/1721.1/81132.

[48] L. Sydell, "We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned," *NPR,* 2016 November 23.

[49] T. McCoy, "Inside a Long Beach Web operation that makes up stories about Trump and Clinton: What they do for clicks and cash," 16 November 2016. [Online]. Available: http://www.latimes.com/business/technology/la-fi-tn-fake-news-20161122-story.html.

[50] E. Cherminsky, "The First Amendment and the Right to Lie," 12 September 2012. [Online]. Available: http://www.abajournal.com/news/article/the_first_amendment_and_the_right_to_lie.

[51] "Final report of the High Level Expert Group on Fake News and Online Disinformation," 12 March 2018. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation.

[52] M. Eddy and M. Scott, "Delete Hate Speech or Pay Up, Germany Tells Social Media Companies," 30 June 2017. [Online]. Available: https://www.nytimes.com/2017/06/30/business/germany-facebook-google-twitter.html.

[53] "'It's seen as a cool place to work' – how China's censorship machine is becoming a growth industry," 29 September 2017. [Online]. Available: http://www.scmp.com/news/china/policies-politics/article/2113377/its-seen-cool-place-work-how-chinas-censorship-machine.

[54] L. Ruan, J. Knockel, J. Q. Ng and M. Crete-Nishihata, "One App, Two Systems: How WeChat uses one censorship policy in China and another internationally," CitizenLab, 30 November 2016. [Online]. Available: https://citizenlab.ca/2016/11/wechat-china-censorship-one-app-two-systems/.

[55] A. Chen, "The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed," *Wired,* 23 October 2014.

[56] U.S. Department of State, "Global Engagement Center," [Online]. Available: https://www.state.gov/r/gec/. [Accessed 1 1 2018].

[57] U.S. Department of Defense, "Strategy for Operations in the Information Environment," June 2016. [Online]. Available: https://www.defense.gov/Portals/1/Documents/pubs/DoD-Strategy-for-Operations-in-the-IE-Signed-20160613.pdf.

[58] A. Marwick and R. Lewis, "Media Manipulation and Disinformation Online," [Online]. Available: https://datasociety.net/output/media-manipulation-and-disinfo-online/.

[59] L. H. Owen, "The Trust Project brings news orgs and tech giants together to tag and surface high-quality news," 16 November 2017. [Online]. Available: http://www.niemanlab.org/2017/11/the-trust-project-brings-news-orgs-and-tech-giants-together-to-tag-and-surface-high-quality-news/.