How to Combat Fake News in Social Media: A Technopolitical Analysis of Regulation Methods

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Kane Lee

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Sean M. Ferguson, Department of Engineering and Society

Introduction

In its early days, the public opinion of social media was optimistic, with many believing that it could serve to bring democracies to the world through globalization of accessible information. During Arab Spring, the series of anti-government protests that spread in Islamic countries such as Iran, Egypt, Tunisia, and Bahrain in the early 2010s, social media platforms were hailed for its contribution in galvanizing people and encouraging them to take active parts in revolts through globalization of local issues. This optimism did not last long into the decade, however, as it became apparent that social media platforms were not exempt from the dangers of misinformation and propaganda (Schiffrin, 2017).

With their rapid growth, social media platforms have become more influential as news media networks. In present day, over 60% of adults now get news primarily from social media sources, and that number is seemingly continuing to grow (Kim, Moravec, & Dennis, 2019). As a news network, social media has shown numerous advantages over pre existing traditional platforms such as newspaper and TV, including faster and more convenient access, the widespreadness of the information less limited by the geographies, ability to personalize the news by individual interest, and the sheer number of contents. However, these advantages have shown to contribute to social media's own disadvantages, as content can be replayed among users with no significant third party filtering, fact-checking, or editorial judgment (Allcott & Gentzkow 2017). These disadvantages have also become apparent, as the issue of fake news across these platforms have become more prevalent.

The two main actors that are most directly able to combat fake news through social media are the government and the corporations, through regulatory policies and changes to the design

of the platforms. Many scholars, such as Schiffrin, Marsden, Meyer, and Brown assert that neither government nor the corporations should be solely responsible in their endeavor to combat fake news, and that a co-regulation is needed to form an effective solution. Although this is a widespread idea, there seems to be a lack of sufficient empirical evidence for this claim. Thus, this paper seeks to analyze the efforts of both the US government and the large social media companies through comparative study on economical, political, and psychological factors, and provide insight to the roles and responsibilities of these actors in regulation of misinformation through social media.

Literature Review

One well-known incident that raised the awareness of fake news is the 2016 US presidential election, where many have expressed concern about the effect of circulation of these fake news in social media platforms. Research by Philip Howard, who studied bot activities and disinformation during the 2016 Election, showed that there were overwhelming levels of news from Russian outlets, Wikileaks, and junk news sources flooding Twitter just before the election. Furthermore, other researchers found alarming evidence that the most popular fake news stories were more widely shared on Facebook than the most popular mainstream news stories, and that many people who see fake news stories report that they believe them (Allcott & Gentzkow 2017). Although research by Allcott and Schiffrin argues that the evidence on influence of exposure to fake news in the election have been much smaller than Trump's margin of victory, the election has demonstrated the potential threat to democracy to the public, and has raised social awareness that changes are necessary. However, it has not been clear on what needs to be

done by whom in order to combat this issue, other than that the two main actors with responsibilities for intervention are the government and the social media corporations.

For government regulation agencies, control over social media contents is conflicted by the First Amendment rights, which includes the right to free speech online. The U.S. Supreme Court has ruled that the First Amendment applies in full measure to speech on the Internet, and attempts to regulate online speech based on content have been declared unconstitutional (Park, 2016). In case of fake news that is targeted to an individual, defamation claims provide a cause of action, as the First Amendment does not generally treat defamatory statements as protected speech (Walters, 2018). However, defamation claim is still limited to intentional or knowingly false statements, and with the usage of bot and anonymous accounts, it is limited in its abilities to combat fake news (Park & Youm, 2019).

As online platforms were being more commonly used as means to bypass regulations that affect their physical counterparts in recent years, commentators and legislators began questioning the reputations of social media platforms as digital public forum. Although for other kinds of media, political candidates had to declare their sponsorship and file copies with the FEC, there were no such policies prior to the election that mandated the public record of political advertisement in social media platforms. Furthermore, after the 2016 election, Facebook has continued to refuse to provide information to researchers about the political advertisements it displayed and who saw them, leaving the influences of these fake news on the election ambiguous (Schiffrin, 2017). This lack of government intervention has been argued to give too much power to the corporations, who themselves are capable of unfairly banning and restricting access to potentially valuable speech.

Currently, social media corporations are self-regulated through internal policies and design changes in each platform. Internal platform user agreements, such as terms of services, lays out how the corporations determine moderation standards and are used to moderate the user contents. These terms of services generally state that the companies have a right to remove user content, and have users agree not to upload contents that falls outside its policies. Although these internal policies are common across social media platforms, each platform is independent in their policies, which means that their implementation could differ in methods or efficacy (Grygiel & Brown 2018). Although these policies can be used to moderate contents that are universally agreed to be harmful, such as revenge porn and terrorism, they also give the corporations power over how they present the contents to the users as well.

Another way social media platforms are able to regulate contents is through development of machine learning and AI algorithms, which use automated techniques to detect and filter through contents for fact-checking. Advocates for these automated fact-checking notes their convenience and cost-effectiveness in their ability to go through a high number of contents generated in online platforms (Ozbay & Alatas 2020). However, critics such as Hildebrandt in Marsden, Meyer, and Brown's paper (2019) warns the dangers of automated solutions, as they are limited in their accuracy, especially for expression where cultural and contextual cues are necessary, and that we must avoid the machine learning version of the Thomas self-fulfilling prophecy theorem, that "if a machine interprets a situation as real, its consequences become real".

In their paper, Marsden, Meyer, and Brown (2019) further discusses the limitations to both corporate self-regulation and the government regulation. In corporate self-regulation,

termed non-audited self-regulation, the substantial failures to the regulatory ecosystem for the media in the 2016 election has already proven its own inadequacy. In government regulation, termed statutory regulation, authors argue how it is unclear what it could achieve without invoking direct censorship of non-comforming organisations. Similarly, Schffrin (2017) argues that both the laws and changes by platform companies are necessary in combating fake news, as relying on laws could give ways for corporate and government censorship, but platform companies should not be able to neglect action by hiding under the idea of free speech.

This ongoing evaluation of government and corporate policies on social media platforms and its ability to delegate power through control over contents ties closely with the ideas that artifacts can have politics, in that technologies can shape our social relations. In his paper, Winner (1980) talks about technopolitics, and how the technical development and the arrangement are critical in determining the politics of artifacts. Framing the social media system with Winner's idea of technopolitics will serve to highlight how the regulation and control over these news content will have to focus on the power, roles, and responsibilities of the government and the corporations.

Method

Analysis by Pavleska, Školkay, Zankova, Ribeiro, and Bechmann (2018) discusses methodology for performance analysis of fact-checking organizations, and establishes multiple aspects in which scholars can critically approach the fact-checking efforts. The three major aspects that they discuss are economical, political, and psychological. Economical criteria focuses on the cost of the implemented solution relative to efficacy of efforts, risk assessments

of potential failures, and contribution to the media development. Political criteria describes the political developments resulting from the implemented solution, such as how transparent and sustainable the method is, what are the implication of shifts in power and responsibilities from their functioning, and how does it affect the fundamental rights through its success or failures. Psychological criteria focuses on the social behaviors of the consumers resulting from the implemented solution, such as how it affects the human bias phenomena and dissuades users from information disorder. Using these aspects as evaluation criterias, this paper assesses how the proposed corporation, government, and hybrid solutions are fulfilling or lacking in each category, and use them as basis of comparison to provide analytical data.

My data includes case study of actual policies and solutions proposed by both social media corporations and US government, all of which comes after the events of 2016 US presidential election which marks the point of increased concerns for the misinformation through social media news networks. In order to focus on the different approaches made with the least amount of variables in policies and designs, data is chosen to focus on multiple solution proposals from a single platform. The platform chosen is for analysis is Facebook, which is the largest social media platform by monthly active users, and one of the main recipients of the criticism following the 2016 election. This paper uses comparative studies of this data by applying the proposed evaluation criteria in order to provide analytical evidence to the argument made by scholars that hybrid co-regulated solutions are more effective than the strictly corporate or government solutions.

Non-audited self-regulation

In December 2016, shortly after the US presidential election, Vice President of Facebook Adam Mosseri released an announcement (2016) of four new features in the platform that Facebook is working on to address the issue of fake news. These features are 1)Easier Reporting, 2)Flagging Stories as Disputed, and 3)Disrupting Financial Incentives for Spammers. Easier Reporting feature allows the users to report to Facebook if they believe that a post is a purposefully fake or deceitful news. Flagging Stories as Disputed connects with third-party fact checking organizations that are signatories of International Fact Checking Code of Principles by Poynter, a non-profit institute that works to promote higher levels of journalism. Using reports from the users and through other signals, the contents are verified and flagged by these organizations. Finally, Disrupting Financial Incentives for Spammers works to combat spammers that masquerade as well-known news organizations and post hoaxes to gain views and revenue through advertisements. To do this, Facebook has eliminated the ability to spoof domains, which reduces the prevalence of sites that pretend to be real publications.

Economical

Primary economic benefits to these three features is that the risk of potential failures are low. In both Easier Reporting and Flagging Stories feature, Facebook minimizes the risk by using multiple stages and parties for validity checks. Easier Reporting is first checked by the users, then by Facebook, and Flagging Stories goes through third-party fact checkers who document their process for fact-checking and link it to the flagged post. Disrupting Financial Incentives are also directly contributing to media development, as it is working to discourage fake sites that reduces the user trust and competes with real publication sites for views. The

ability to have multiple features tackle this issue through various angles are also beneficial to its effectiveness.

Political

Some of the proposed solutions are surprising in that they are preventing these features from leading to more power and responsibilities to the corporations. Easier Reporting delegates detection process of fake posts to the users, while Flagging Stories delegates verification process to the third-party organizations. This shift in role also leads to lowered risk of corporate and government censorship. However, the description of Facebook's methods of determining and regulating the contents are lacking in transparency.

Psychological

Psychologically, the features are highlighting the issue of fake news to help users be more critical of the news on social media. With Easier Reporting and Flagging Stories, the posts must be verified by multiple sources, thus users could become more aware that the news contents on social media platforms can be unreliable.

Statutory regulation

In looking at the government efforts to combat fake news, I have chosen three bills that are proposed by the congress that seeks to regulate social media corporations. The first bill is Biased Algorithm Deterrence Act of 2019, which states that the owner of social media service will be treated as a publisher or speaker of user-generated content and be held liable if it is manipulating display of contents that may lead to intended bias by the platform owner. If the service or algorithm displays user-generated content in any order that is not chronological, delays

the display of such content relative to other content, or hinders the display of such content for reasons other than to carry out the user's direction or to restrict material that the provider policy or user considers inappropriate, then it will be considered biased. The second bill, Voter Privacy Act of 2019, seeks to ensure privacy with respect to voter information by giving voters right to access and delete personal information obtained by social media, prohibit transfer of their data, and prohibit targeting based on personal information. The last bill is Bot Disclosure and Accountability Act of 2019, which seeks to require social media providers to establish and implement policies for public disclosure of automated software programs intended to impersonate or replicate human activity, and prohibit their usage for online political advertising.

Economical

The cost of implementation and risk of failure are minimal in these policies, because they create clear guidelines that social media companies must follow, and the failure or violation is enforced by fines or imprisonment. In Bot Disclosure and Accountability Act, enforcing restriction of automated programs on social media platforms will decrease the amount of spam bots that are used in promotion of fake news, but the other bills provide indirect solutions by creating transparency to the black box of social media content display algorithms, which means that their effectiveness are uncertain.

Political

Politically, the proposed bills seek to shift power and control that social media corporations had over the contents and data by enforcing them to be publicly disclosed. Under the Voter Privacy Act, the social media corporations would have to release personal user data and how they are being used, as well as give users the power to remove their own data or prevent

their transfers. Furthermore, success of the Voter Privacy Act will lead to fundamental rights of the social media users to privacy of their personal information. Policy regulations also seek to impose tangible responsibilities to social media corporations in combating fake news by holding them accountable for user contents if it fails to meet requirements, such as displaying contents without bias in the Biased Algorithm Deterrent Act.

Psychological

The Biased Algorithm Deterrent Act deters corporation's bias by preventing manipulation through ordering or hinderance of the contents, but it doesn't affect much about the user confirmation bias. Other acts do not attempt to affect the human bias phenomena or dissuade users from information disorder.

Co-regulation

Since there are no formal approaches made in the US for co-regulation at the time of writing this paper, I have used the co-regulation example described in the paper by Marsden, Meyer, and Brown (2019). Their regulation makes use of artificial intelligence and machine learning for swift detection of fake news, but also utilizes human regulators to check the detected contents. Without human intervention, accounts cannot be suspended, and if the contents are deleted, the users will have access to the appeal process. The government policies will also lay out the general principles that apply to AI regulators, and require public disclosure of details of the scheme design. It seeks to combine legitimacy of parliamentary approval for regulatory systems with general principles of good regulation, such as independence from regulatees, appeal

processes, and governance principles, while devoting responsibility to independent corporations which gives agility and flexibility to the regulation.

Economical

The success of this regulation method will lead to rapid and efficient countermeasures to fake news contents. Checking and appeal processes through human regulators also act as safety measures for failure of AI, which means that the potential risk will be low. However, the cost to implementing both AI regulation and human regulators that has to constantly supervise these softwares will be high.

Political

The government regulation policies act to ensure that the regulation process will be transparent to the public through disclosure of details of the AI scheme design, and make sure that the corporations will not have black box algorithm to take control over the detection process. Also, the human regulator will either be employed by government agencies or third-party organizations to be independent from corporations, which also shifts the power away from the corporations.

Psychological

There is no attempt to affect the human bias phenomena or dissuade users from information disorder.

Discussion

Overall, the self-regulation case was effective in its economic efficacy and their ability to most directly and swiftly tackle the fake and biased contents. Also, Facebook's use of user and

third-party for regulation was surprisingly effective in political aspect, as it prevented these solutions from leading to even more power and responsibilities for the corporations. Facebook's features were also the only regulation case that considered psychological aspects, although it did not go much in detail. On the other hand, the statutory regulation was most effective in its political aspects. The policies were aimed to create transparency in social media platforms and their functionalities and shifting the power away from the corporations to the users. Finally, the co-regulation method case was able to combine the economic efficacy and performance efficiency of the self-regulation with the political transparency and healthy power dynamic of the statutory regulation, but also failed to fulfill the psychological criteria.

The lack of psychological consideration may stem from the fact that many of these solutions are focused on removing the potential sources of bias rather than affect the social behaviors of the consumers. However, even if the solutions are effective and efficient in removing posts, they can't prevent the users from accessing the posts before they are detected. Study on human bias phenomena shows that once a belief has been established, people will constantly reiterate it for themselves and contrary opinions will be unable to influence their confidence. Once this belief is established, even a refutation by an authority has little impact on changing people's minds, which means that flags by third party or removal of posts may not revert the belief that the fake post has brought to the users (Borges & Gambarato 2019). Furthermore, multi-level check processes from self and co-regulation means that verification process would be more accurate, but also slower than fully AI implementation.

Outside of the psychological aspect, however, the analysis suggests that it agrees with the idea that co-regulation is more effective than self or statutory regulation because of their innate

limitations. Although Facebook's solution recognized the self-regulation's limitation in lacking formal and transparent guidelines for regulation and implemented user and third-party checking processes, it would be against their own interest for corporations to give up the power they hold over contents and data. On the other hand, the government is lacking in their ability to implement direct prevention or removal of fake news, and they lack the rapid and efficient micro-level content control that the corporations have access to. In terms of technopolitics, the co-regulation case was shown to delegate power and responsibilities of government and corporations based on their strengths and abilities, and use them to cover for other's limitations. Therefore, the overall analysis suggests that co-regulation was a more effective solution than the self and statutory regulations.

Bibliography

Allcott, H., & Gentzkow, M.. (2017). Social Media and Fake News in the 2016 Election. Retrieved from https://search.lib.virginia.edu/articles/article?id=bth%3A122833485.

Biased Algorithm Deterrence Act of 2019, H.R.492, 116th Congress (2019-2020).

Borges, P., & Gambarato, R.. (2019). The Role of Beliefs and Behavior on Facebook: A semiotic Approach to Algorithms, Fake News, and Transmedia Journalism. Retrieved from https://search.lib.virginia.edu/articles/article?id=ufh%3A139171717.

Bot Disclosure and Accountability Act of 2019, S.2125, 116th Congress (2019-2020).

Brown, E.. (2019). "Fake News" and Conceptual Ethics. Retrieved from https://search.lib.virginia.edu/articles/article?id=hlh%3A139889478.

Davies, H.. (2018). Redefining Filter Bubbles as (Escapable) Socio-Technical Recursion. Retrieved from https://collab.its.virginia.edu/access/content/group/d1e9a10e-31f0-4cbc-b357-2882a6e8647f/Truth%20Post%20Truth%20Fake%20News/Davies%20-%202018%20-%20Redefining%20Filter%20Bubbles%20as%20(Escapable)%20Socio-Tec.pdf.

Ending Support for Internet Censorship Act, S.1914, 116th Congress (2019-2020).

Gilchrist, Alan.. (2018). Post-Truth: An Outline Review of the Issues and What Is Being Done to Combat It. Retrieved from https://search.lib.virginia.edu/articles/article?id=zbh%3A132282535.

Grygiel, J., & Brown, N.. (2018). Are social media companies motivated to be good corporate citizens? Examination of the connection between corporate social responsibility and social media safety. Retrieved from https://www-sciencedirect-com.proxy01.its.virginia.edu/science/article/pii/S0308596118304178.

Honest Ads Act, S.1989, 115th Congress (2017-18).

Kim, A., Moravec, P., Dennis, A.. (2019). Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings. Retrieved from https://search.lib.virginia.edu/articles/article?id=bth%3A137887539.

Marsden, C., Meyer, T., Brown, I.. (2019). Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation? Retrieved from

https://reader.elsevier.com/reader/sd/pii/S026736491930384X?token=94E6A368CDD8ABB7622A2DCBBC70A4D2BA2691CEE3581748F99FFD56CD33E9FCC14D60716C2C54538C2F0871AEF33EB5

Mosseri, A.. (2016). Addressing Hoaxes and Fake News. Retrieved from https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/.

Ozbay, F., Alatas, B.. (2019), Fake News Detection Within Online SOcial Media Using Supervised Artificial Intelligence Algorithms. Retrieved from https://search.lib.virginia.edu/articles/article?id=edselp%3AS0378437119317546

Park, A., Youm, K.. (2019) Fake News From a Legal Perspective: The United States and South Korea Compared. Retrieved from https://www.swlaw.edu/sites/default/files/2019-04/7.%20Ahran%20Park%3B%20Kyu%20Ho%20Youm%2C%20Fake%20News%20from%20a%20Legal%20Perspective%20-%20The%20United%20States%20and%20South%20Korea%20Compared.pdf

Park, C.. (2016). Online Speech and Democratic Culture: A Comparison of Freedom of Online Speech Between South Korea and the United States. Retrieved from https://search.lib.virginia.edu/articles/article?id=edselp%3AS0736585315300617

Pavleska, T., Školkay, A., Zankova, B., Ribeiro, N., Bechmann, A.. (2018). Performance Analysis of Fact-checking Organizations and Initiatives in Europe: a Critical Overview of Online Platforms Fighting Fake News. Retrieved from http://compact-media.eu/wp-content/uploads/2018/04/Performance-assessment-of-fact-checking-organizations_A-critical-overiview-Full-Research-1-1.pdf

Schiffrin, A.. (2017). Disinformation and Democracy: The Internet Transformed Protest But Did Not Improve Democracy. Retrieved from https://search.lib.virginia.edu/articles/article?id=edsjsr%3Aedsjsr.26494367.

Winner, L.. (1980). Do Artifacts Have Politics?. Retrieved from https://www.jstor.org/stable/20024652?origin=JSTOR-pdf&seq=1#metadata_info_tab_contents

Walters, R.. (2018). How to Tell a Fake: Fighting Back Against Fake News on the Front Lines of Social Media. Retrieved from https://search.lib.virginia.edu/articles/article?id=a9h%3A134763624