Journal of Systemics, Cybernetics and Informatics (2022) 20(5), 118-137
https://doi.org/10.54808/JSCI.20.05.118

An Investigation of the Effectiveness of Facebook and Twitter Algorithm and Policies on Misinformation and User Decision Making

Jordan Harner[1], Lydia Ray[2], Florence Wakoko-Studstill[3]

[1]TSYS School of Computer Science, Columbus State University
Columbus, GA 31906, US
[2]TSYS School of Computer Science, Columbus State University
Columbus, GA 31906, US
[3]Criminal Justice & Sociology Department, Columbus State University
Columbus, GA 31906, US

Abstract

Prominent social media sites such as Facebook and Twitter use content and filter algorithms that play a significant role in creating filter bubbles that may captivate many users. These bubbles can be defined as content that reinforces existing beliefs and exposes users to content they might have otherwise not seen. Filter bubbles are created when a social media website feeds user interactions into an algorithm that then exposes the user to more content similar to that which they have previously interacted. By continually exposing users to like-minded content, this can create what is called a feedback loop where the more the user interacts with certain types of content, the more they are algorithmically bombarded with similar viewpoints. This can expose users to dangerous or extremist content as seen with QAnon rhetoric, leading to the January 6, 2021 attack on the U.S. Capitol, and the unprecedented propaganda surrounding COVID-19 vaccinations. This paper hypothesizes that the secrecy around content algorithms and their ability to perpetuate filter bubbles creates an environment where dangerous false information is pervasive and not easily mitigated with the existing algorithms designed to provide false information warning messages. In our research, we focused on disinformation regarding the COVID-19 pandemic. Both Facebook and Twitter provide various forms of false information warning messages which sometimes include fact-checked research to provide a counter viewpoint to the information presented. Controversially, social media sites do not remove false information outright, in most cases, but instead promote these false information warning messages as a solution to extremist or false content. The results of a survey administered by the authors indicate that users would spend less time on Facebook or Twitter once they understood how their data is used to influence their behavior on the sites and the information that is fed to them via algorithmic recommendations. Further analysis revealed that only 23% of respondents who had

seen a Facebook or Twitter false information warning message changed their opinion "Always" or "Frequently" with 77% reporting the warning messages changed their opinion only "Sometimes" or "Never" suggesting the messages may not be effective. Similarly, users who did not conduct independent research to verify information were likely to accept false information as factual and less likely to be vaccinated against COVID-19. Conversely, our research indicates a possible correlation between having seen a false information warning message and COVID-19 vaccination status.

***Keywords**: Social Cybersecurity, Social Media, Filter Bubble, Disinformation Campaign, COVID-19, Facebook, Twitter*

1. Introduction

Social media usage is pervasive in the United States and most other countries. Particularly in the wake of the COVID-19 pandemic and the necessity of physical isolation, social media usage has increased significantly. Unfortunately, the policies and underlying algorithms designed to generate increased traffic and interactions on social media sites often create safety and security issues in both cyberspace and the real world especially when factoring in human psychology and behavior. These algorithms create filter bubbles and feedback loops which negatively impact human lives, society and country. One such harmful impact is observed in the role of social media in disseminating deliberate wrong information.

Social media sites are increasingly used by people as a source of information. According to a report from the Pew Research Center, about half of all American adults consume news from social media [8]. Another report suggests that US adults who consume news from social media tend to be more misinformed about the coronavirus and politics [6]. However, unlike prominent news media organizations, in social media, information can be posted without appropriate vetting by regular users. As a result, a significant amount of misinformation is disseminated widely in social media circles may or may not have originated elsewhere. While some misinformation is produced or shared without any malicious intent, there are disinformation campaigns that purposely target specific user bases in order to gain domestic or international political advantages. The impact of the disinformation campaign may not always stay within the limits of cyberspace. Often, our physical world is also severely impacted.

Prominent social media sites such as Facebook and Twitter use content and filter algorithms that are designed to grow user bases and extend the time spent on these

sites. While the intentions are seemingly innocuous and business-driven, the nature of these algorithms play a significant role in creating filter bubbles that may captivate many users. These bubbles reinforce existing beliefs and expose users to content they might have otherwise not seen. In addition, filter bubbles create feedback loops with dangerous or extremist content as seen in the QAnon rhetoric, leading to the January 6, 2021 attack on the U.S. Capitol, and the unprecedented propaganda about COVID-19 vaccinations.

The algorithms of social media sites have other underlying biases that also create havoc in people's lives directly or indirectly. In the social media space, Facebook still permits, to varying extents, ad targeting by race and other demographics which was found to be an issue when housing companies used it to target ads at only specific people. It is also possible for politicians and organizations to target advertisements towards, for instance, a white male between the ages of 18 and 30 that lives in a particular zip code with a certain political affiliation and interests [18].

While content algorithms are designed to either increase engagement for advertising purposes or to moderate harmful content, they are treated as intellectual property protected by applicable laws in the United States. Not much is understood about the intricacies of how content algorithms work. Due to their secretive nature, is it difficult for users to make informed decisions about the types of content they need to use and what information Facebook and Twitter are collecting from them which can lead to a perpetuating filter bubble. This paper hypothesizes that the lack of understanding of content algorithms and their ability to perpetuate filter bubbles creates an environment where dangerous false information is pervasive and not easily mitigated with the existing algorithms that are meant to provide warning messages against false information. In a sense, one set of algorithms is working to recommend content while another set of algorithms may be working to provide content warnings even to the recommended content. These algorithms are not widely understood on a detailed level and are extremely complex.

This research falls under the emerging interdisciplinary field of social cybersecurity which aims to understand how cyber-mediated changes in human behavior influence social and political consequences. Traditional definitions of cybersecurity have focused primarily on more technical aspects of the subject such as hacking, malware, physical security, and business continuity planning. Social cybersecurity is a newer, emerging space within the discipline that examines the ways in which people consume information, primarily on the Internet, and what role social media companies and government agencies play to ensure spaces remain safe. Kathleen M. Carley et al, in [19] refers to this space as the nexus between cybersecurity and

other disciplines such as sociology, marketing, forensics, information warfare, and social psychology. In a traditional cybersecurity sense, the goal is often to implement procedures to prevent cyber incidents from occurring and to have mitigation and recovery plans in place for when they do occur. In social cybersecurity, the goal is, in many ways, very similar. Social cybersecurity is also concerned with ensuring social media companies, ISPs, and other stakeholders in the social media space are acting in an ethical way that presents as little risk to users and society as possible. Social media policies and content algorithms are at the crux of the issue. In social cybersecurity research, these policies and algorithms are reviewed through the lens of other disciplines such as social sciences and psychology in an effort to understand the human component of cyber safety.

Through analysis of social media cybersecurity policies and content algorithms in conjunction with a survey, our research seeks to understand the relationship between these policies and algorithms and user behavior outside of cyberspace. In addition, we seek to gauge how cybersecurity policies contributed to the nature of people's experiences with the COVID-19 pandemic. With data collected from 97 Facebook and Twitter users, we tested the following hypotheses:

- Facebook and Twitter false information warning messages are effective in changing opinions.
- If Facebook or Twitter revealed more information of how they use artificial intelligence to influence the time users spent on the websites and the content that was pushed to them, users would choose to spend less time on the websites.
- Exposure to false information warning messages on Facebook and/or Twitter leads to users being more likely to have been vaccinated against COVID-19.
- Facebook and Twitter users that reported not independently researching content marked as "misleading" or "false" were more likely to have not received a COVID-19 vaccination.

Results indicate that users would spend less time on Facebook or Twitter once they understood how their data is used to influence their behavior on the sites. Further analysis revealed that a majority of the people who responded believed in the warning messages they read. Only a small proportion of the respondents indicated that they are not influenced by false information warning messages. Similarly, users who did not not conduct independent research outside of the social media sites were less likely to have been vaccinated against COVID-19.

In this paper, a brief review of the related literature is presented in section 2. In section 3, we provide a brief review of the policies and algorithms of Facebook and

Twitter. Section4 contains our research methodology, section 5 provides a description of our results and discussion, and we draw our conclusions in section 6.

2. Literature Review

Social media algorithms, filter bubbles and disinformation campaigns have been a focus of interdisciplinary research since the past few years. In this section, we briefly describe the research done in this interdisciplinary area and the results obtained by researchers in Computer Science, Political Science and Social Sciences. Peter M. Dahlgren in [1] states that social media-created personalization algorithms create filter bubbles that reinforce existing beliefs and limit exposure to differing opinions. The author presents critical arguments against the notion of filter bubbles arguing that people are generally exposed to differing opinions and that, in the context of political polarization, that politics composes a relatively small part of most people's lives. The author states that "the more confident an individual is in their belief, attitude, or behavior, the more exposure they have to challenging information."

Schelling et al. in [2] review AI algorithms for news consumption as they apply to social media outlets. They speak to how these algorithms can lead to polarization and echo chambers. This study specifically reviews Dutch news articles to define what they call "Ideology Spaces." It also proposes solutions using the very AI that creates the issue.

M. Geetha Yadav et al. in [3] look at ways in which AI and natural language processing might be solutions for social media companies to identify false news quickly and efficiently. The AI model presented has a goal of reviewing snippets of articles then provides a percentage estimate to automatically identify the likelihood that the article contains false information.

David Lauer in [4] argues that some social media companies and, in this case in particular, Facebook's business model thrives on creating filter bubbles. It states that Facebook has a financial incentive to create and maintain algorithms that promote groupthink and extreme content because this type of content typically receives the highest levels of engagement. By driving engagement and longer usage periods, Facebook is able to expose users to more ads which is its primary revenue source.

Stephen Neely et al. in [5] reviews the ways in which patients and the public in general consume health-related information particularly regarding social media and

the COVID-19 pandemic. A survey was conducted of 1003 US-based adults to gather how they use information from social media to stay up to date with COVID-19 information and the extent to which these adults fact-checked the things they read. The results showed that most adults surveyed used social media to better understand the evolving pandemic and that nearly two thirds did not fact-check with a healthcare professional. In terms of the real-world effect of social media disinformation relating to COVID-19, adults who were cited as consuming social media information from more credible sources were more likely to choose to be vaccinated. The study suggests that healthcare professionals need to evolve and adapt to the ways in which the public receives information.

Jianming Zhu et al. in [6] discusses the way in which positive reinforcement of certain ideas can make the effect of an echo chamber even stronger. Influence maximization is studied which is the theory that an initial number of users to first disseminate this information can have an impact on the information's ultimate overall influence. User activation occurs from echo chamber influence rather than peer-to-peer influence which is differentiated by an echo chamber's groupthink effect. Zhu et al. suggest that users that like the same topic are considered to be in the same social media group. Looking through other resources, it will be relevant to study if this statement can be further proven or disproven.

Fatimah Alzamzami and Abdulmotaleb El Saddik in [7] propose a real-time method of monitoring social media user actions in the U.S. and Canada during the pandemic. Using artificial intelligence (AI) to monitor information and connectivity trends among social media users is one way in which officials and healthcare professionals can begin to understand how users absorb information. The paper stresses the importance for officials and experts to monitor common social media information and trends to be able to understand positive and negative sentiments that spill over into the real world creating offline implications. It describes a method to monitor information on a domain-by-domain basis utilizing keywords common across all domains.

Felix Drinkall and Janet B. Pierrehumbert in [8] demonstrate ways in which caseloads of COVID-19 can be predicted across specific geographic subreddits (Reddit sub-domains) and other domains. The study found a strong correlation between studying these specific subreddits and short-term COVID-19 caseload predictions. Like the paper in [7] this paper uses specific keywords in reviewing several COVID-19 subreddits aimed at several US cities.
Christopher Whitfield et al in [9] review the four largest subreddits in North Carolina similar to the Reddit research conducted in [8]. They also utilized keywords in their research but, in this case, their goal is to study the public's uptake

of COVID-19 precautions such as wearing a mask, washing your hands, and social distancing. This study looks at some of the ways in which researchers can use natural language processing (NLP) to determine what is being discussed the most. Some limitations do exist with the modeling, namely its ability to accurately categorize certain posts and comments as positive, negative, and neutral sentiments. Joanne Chen Lyu et al. in [10] explore social media and, in particular, Twitter to examine the public sentiments of receiving a COVID-19 vaccination. The study attempts to divide tweets into 16 COVID-19 or vaccine-related topics. Vaccination accounts for 15% of the tweets with the majority being rated as "positive" especially when following major announcements such as the 90% efficacy of the Pfizer vaccine. One potential limitation of this study is to consider if those distrustful of the COVID-19 vaccinations are less likely to share their sentiments in an online space as compared to those more in favor.

Antonio F. Peralta et al. in [11] discuss the motivations of social media companies to maximize usage to drive revenue growth, oftentimes as the expense of promoting accurate information. Filter algorithms are commonly used by social media companies to further engage users by showing them related content to what they may have interacted with previously. This is called algorithmic bias. This paper proposes a new method of measuring information exchanges as they apply to algorithmic bias, filter bubbles, and echo chambers. It finds that under algorithmic bias, the opinions of a smaller part of an online group tend to be deprioritized and, thus, do not gain as much traction as opinions shared by most of the group.

Matthew Andreotta et al. in [12] propose a method of extracting and analyzing large datasets from social media and other websites with the purpose of performing a quantitative analysis. It also gives an example of this approach through looking at commentary in Australia on Twitter surrounding climate change.

Alexander Chkhartishvili and Ivan Kozitsin in [13] attempt to measure the effects of echo chambers in different social networks. This study differs from others in that the opinions of social network users are utilized to determine the extent of echo chambers. The paper argues that simply looking at a user's online interactions, for example, with a particular politician does not necessarily equate to them aligning with that politician's stances. The paper argues that more straightforward data such as a user's self-reported political stances are needed or a larger amount of information about a user's connectedness to a public page is necessary to draw conclusions. This argument is applied to not just politics but can also be applied to other topics that garner strong opinions. The paper considers a user as connected to a public page if the user is subscribed, has interacted with a post on the page, or has a friend that has shared or reposted content from the page. The study found a high

level of overlap among Russian users on Vkontakte between their self-reported political affiliations and the public pages they were connected to. This suggests a high level of echo chamber effect as the content users interact with mostly aligns with their self-reported political biases.

Richard Rogers states in [14] that deplatforming has gained particular interest recently as former President Trump became the highest-ranking user yet to be banned from certain social media websites following the events of January 6th, 2021. This paper's research attempts to determine if deplatforming users actually works in limiting the spread of their messages and broader societal impacts. It also questions if deplatforming truly purges extremist content from mainstream networks or simply drives extremist content and users to more niche-driven networks such as Gab or Parler. The study also found that after deplatforming, there has been a significant migration back to individual websites where the trend for years has seen content and interaction increasing centralize on a handful of social media platforms.

Matteo Cinelli et al. in [15] review echo chambers across social media platforms such as Gab, Facebook, Reddit, and Twitter. On Facebook and Twitter, the results demonstrate a higher level of homophilic clusters on Facebook and Twitter. It also finds high levels of news consumption segregation on Reddit and Facebook with the highest level of segregation on Facebook. This means that of all social networks studied, Facebook arguably has the highest levels of homophilic or like-minded user clusters coupled with news consumption being more specific to particular user groups. Group polarization theory states that as online groups continually reinforce opinions of the majority in a group, social media algorithms will continue to reinforce these beliefs and drive a group to more and more extreme positions.

Elizabeth Dubois and Grant Blank in [16] argue that research and warnings of social media echo chambers have been greatly overstated. One flaw of the many studies looking into echo chambers and polarizations is that they tend to focus on one platform at a time. Most likely due to limited information, it can be more difficult to conduct a study of social media users and the content they interact with across the multiple social media platforms they may use. This paper states that due to the many options users have in choosing social media, users may be exposed to differing opinions more frequently than realized thus making echo chambers not as great of a concern. By looking at a subset of social media users in the UK, this study reviews users who are both interested in politics and those who also consume news from multiple sources. The study states that as people use more than one social network and/or are exposed to media in many formats such as television, radio, and newspapers, that the effect of any one online social network is mitigated.

Lin Cai and Zihang Wang in [17] speak to algorithms as property for different organizations and the different protections that come with them. It discusses how big data and algorithms (or artificial intelligence) go hand in hand, describing big data as the foundation and algorithms or AI as the mechanism by which big data is useful. In the last few years, algorithms and AI have been created that can evolve on their own or adapt to new data patterns. These models are created to be adaptable and have many promising use cases. In the context of social media and human behavior, they can change along with changing human behavior. It makes the argument that copyright law only protects algorithmic code but not the actual idea or solution that it is solving for, allowing competitors and others to essentially legally copy the idea with different code. This ties into why we know so little about how algorithms work on social media sites. Since only the code is protected by law and not the idea, social media companies have little interest in publicly divulging the details of how the algorithm exactly works in fear a competitor could copy the idea with different code. They also argue on pages 60 and 61 that if the public understands how the algorithms work, users may change their behavior potentially at the detriment of the social media organization. Companies may be hesitant to reveal how content algorithms function since it may negatively impact their core business model in such a high-choice media environment. The paper notes that algorithmic bias is a real concern using the example of police stations using crime algorithms to increase police presence in certain areas which had the unintended effect of being racially and ethnically discriminatory.

Mark MacCarthy in [18] proposes a way for researchers and other qualified, independent individuals to have a way to review algorithms in the hope of providing an additional layer of transparency and oversight. Through this proposal the authors hope to instill more public trust in social media companies and other organizations utilizing often secretive algorithms. It also advocates for users to be able to make informed decisions through increased transparency and to let "consumer choice" put pressure on social media companies. One of the major issues with public distrust and extremist content spread is that so little is understood about how social media companies recommend and personalize this content. Giving consumers a larger voice and self-empowerment, it is argued, would decrease false information spread, echo chambers, and the impact of social media to bring harmful content online to the real world.

Kathleen M. Carley et al. in [19] speaks to an emerging area in cybersecurity which the authors refer to as "social cyber-security." At the crux of this emerging area is the ability for computers, the internet, and humans to coexist in a space where the goal is to create spaces where information is free from bias and attempts to sway public opinion through false information. Traditionally, cybersecurity has been

viewed through the lens of hacking, stealing information, spreading malware and other impactful scenarios. In social cybersecurity, the focus is on the manipulation of people through technology and the social, political, and cultural effects. It uses examples such as the Russian agenda in the 2016 election to spread false information in U.S. technology spaces such as social media to sow discord and further their cause. Social cybersecurity specifically focuses on policy which differs from traditional cybersecurity where technology is the primary focus. There is significant overlap, in some cases, between traditional cybersecurity and social cybersecurity but both fall under cybersecurity's larger umbrella. The authors also make the case that social cybersecurity is an area that merges many disciplines including but not limited to sociology, marketing, forensics, information warfare and social psychology. A social cybersecurity approach combines methodologies in wanting to study both how technology plays a role and also what its impact is in a social context. Also mentioned are some of the existing issues with researching within this space; namely access to data. On Twitter, for example, free data provided is sometimes difficult to obtain and is limited in nature creating an inherent bias from the start. Access to larger data sets is costly and provided through third parties. These sets also do not contain all meta data.

3. Review of Social Media Algorithms

While there are many social media organizations, in this paper, we will review the policies and algorithms of Facebook and Twitter. Facebook and Twitter algorithms can be generally bucketed into one of two categories: engagement drivers and content moderation. Both sites use a variety of tactics to increase engagement which in turn increases exposure to advertisements, both companies' primary revenue streams. At the crux of the issue is the secrecy around how exactly these algorithms are designed to function. We can take some educated guesses and speak to examples that have been publicized but the algorithms themselves are proprietary and the companies are generally unwilling to divulge trade secrets. There are a few things we do know, however, from Facebook's own Data Policy and their data export service which includes 48 different categories of data (located in Facebook Settings>Your Facebook information>Download your information). [20]

- User interactions such as likes, shares, retweets, and comments are constantly collected and stored as part of the company's profile of each user
- Accounts, pages, and groups that users have followed or are members of are all categorized and added to the user profile
- Demographic information freely given by users such as their name, age, location, sex, gender, marital status, political affiliation, race, ethnicity,

health status, and even family members is collected and attached to the user profile
- Engagement algorithms analyze all of this data that is attached to the user profile to further recommend content that the company believes may increase a user's engagement and time spent on the website
- Advertisement algorithms also utilize much of this data as part of an ad-targeting approach

In many cases, the results of this data collection, while uncomfortable for some users, is innocuous. Facebook and Twitter provide free services for users and, in turn, they realize revenue through advertisers paying for exposure to users. In some cases, advertisers and even Facebook and Twitter have crossed the line in terms of what the general public may consider to be an invasion of privacy or overreach. In late November 2021, Facebook announced it would no longer allow advertisers to target users based on race, ethnicity, health status, religion, sexual orientation, or political affiliation after significant pressure. [22]

Facebook and Twitter's other set of algorithms, while related to their content promotion algorithms, are algorithms created with the intention of moderating content. Facebook subcontracts its fact-checking to a third-party organization called the International Fact-Checker Network (IFCN) which is managed by the Poynter Institute.

4. Research Methodology

To conduct our research, we administered a Qualtrics survey to Facebook and Twitter users using a variety of methods including multiple choice responses and Likert scales. Participants were reached via email to Columbus State University TSYS School of Computer Science students as well as via Facebook and Twitter. Our population sample consisted of Facebook or Twitter users at least 18 years of age that had used Facebook or Twitter within the past year. Respondents were asked up to 18 questions and must have agreed to an informed consent notice before beginning. Data was collected anonymously and did not include IP addresses or other identifying information.

When determining our research methodology, we developed several questions that evolved into our research hypotheses. Our hypotheses are as follows:

- Hypothesis 1: Facebook and Twitter false information warning messages are effective in changing opinions.

- Hypothesis 2: If Facebook or Twitter revealed more information of how they use artificial intelligence to influence the time users spent on the websites and the content that was pushed to them, users would choose to spend less time on the websites.
- Hypothesis 3: Exposure to false information warning messages on Facebook and/or Twitter leads to users being more likely to have been vaccinated against COVID-19.
- Hypothesis 4: Facebook and Twitter users that reported not independently researching content marked as "misleading" or "false" were more likely to have not received a COVID-19 vaccination.

To answer the question of if social media disinformation campaigns have influenced the opinions, attitudes, and actions of Facebook and Twitter users, a sample research survey was conducted. This survey is quantitative in nature and set out to answer if Facebook and Twitter users' opinions regarding COVID-19, health safety protocols, and vaccination were influenced by content they saw on the platforms. Specifically, we wanted to know if Facebook and Twitter users were persuaded to be vaccinated against COVID-19 or not, if they were willing to follow health safety guidelines, and if Facebook or Twitter content influenced these real-world decisions.

4.1 Survey Analysis

In total, 100 participants began the survey but 3 reported not using Facebook or Twitter at all and were excluded from the results. 97 participants reported using Facebook or Twitter at least once in the past year and completed the survey in full. Of the 97 participants, 55 unique participants reported using both Facebook and Twitter and 42 reported using only one of the platforms. Prior to asking participants if they have seen examples of disinformation displayed on Facebook or Twitter or if they have seen "false information" warnings displayed, a series of demographic and baseline questions were asked. One of these asked if participants were vaccinated against COVID-19 either by personal choice or due to a job or other requirement. 83 of 97 participants noted that they were vaccinated with only 2 stating it was due to a job or other requirement and the others stating it was by personal choice. 14 stated they were not vaccinated by personal choice and no participants chose the option stating they had refused the COVID-19 vaccination due to a documented medical condition. These 14 participants were key to answering our question of if social media disinformation found on Facebook or Twitter had contributed to a participant's real-world decision to not receive a COVID-19 vaccination.

5. Results and Discussion

In this section, we will discuss the results obtained from our survey to provide insights on our research questions.

5.1 Hypothesis 1: Facebook and Twitter false information warning messages are effective in changing opinions.

In questions 13, 14, and 15, we asked respondents who have seen one of these messages if they had:
Q13) Read the Explanation
Q14) The message changed their opinion
Q15) Conducted independent research outside of Facebook or Twitter

For the purposes of these questions, we consider "Always" and "Frequently" to be stronger indications with "Sometimes" and "Never" being weaker indications. Most participants in Figure 1 indicated that they only "Sometimes" or "Never" read the explanation, that the explanation changed their opinions, or that they conducted independent research. For each of the three questions asked, "Sometimes" was the most selected answer possibly indicating a general apathy or lack of attention paid to the false information messages.

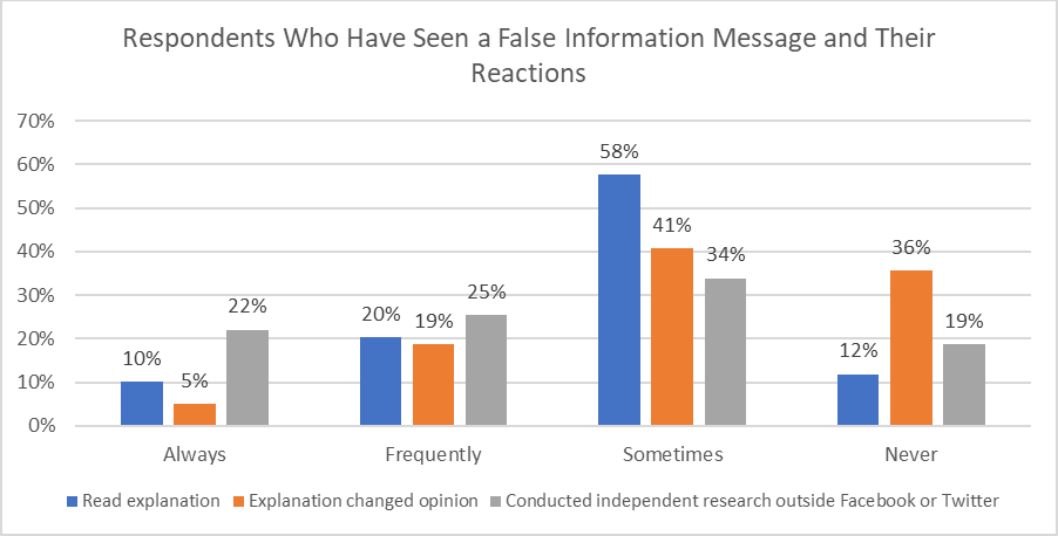

Figure 1: Actions taken by participants who have seen a false information message

5.2 Hypothesis 2: If Facebook or Twitter revealed more information of how they use artificial intelligence to influence the time users spent on the websites and the content that was pushed to them, users would choose to spend less time on the websites.

A slight majority of users said that their usage of Facebook and Twitter in Figures 2 and 3 would not change if the websites revealed more information of how they use algorithms using data points gathered on each user to manipulate the content users see. A large percentage of users also reported that they would use the websites less frequently with only 2 out of 97 respondents stating they would use the websites more frequently. Perhaps most interestingly, more frequent Twitter users across all usage distributions reported they would use the websites less frequently than the same distribution across Facebook users.

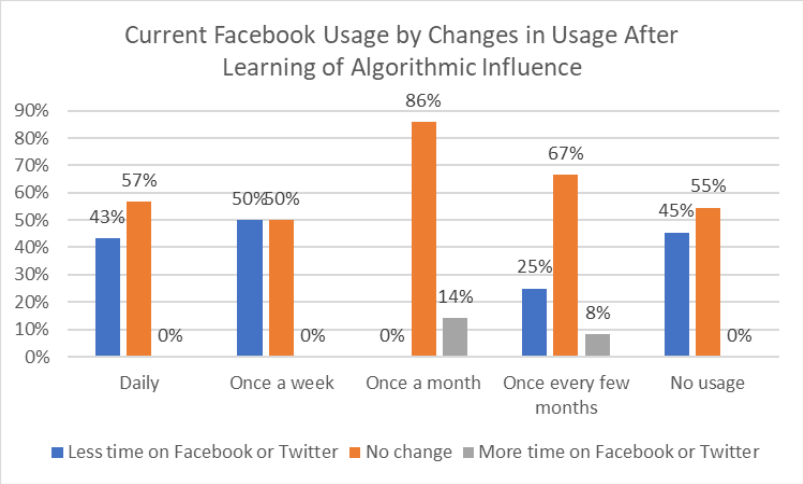

Figure 2: Changes in Facebook usage after revelation of algorithmic influence

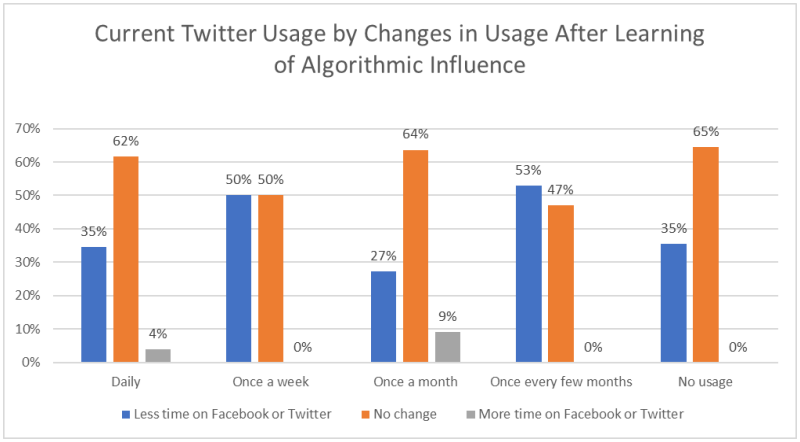

Figure 3: Changes in Twitter usage after revelation of algorithmic influence

5.3 Hypothesis 3: Exposure to false information warning messages on Facebook and/or Twitter leads to users being more likely to have been vaccinated against COVID-19.

Our results indicate nearly inverse results for respondents who have seen or not seen false information messages distributed by those who are vaccinated or not vaccinated against COVID-19 by personal choice as seen in Figure 4. 71% of those not vaccinated have not seen a false information message and 29% have. Conversely, 33% of those who are vaccinated have seen a false information message but 67% have. Only 2 respondents noted they were vaccinated against COVID-19 due to a job or other requirement and they were equally as likely to have seen a false information message. These results suggest that there may be a correlation between seeing false information messages and choosing to receive the COVID-19 vaccination.

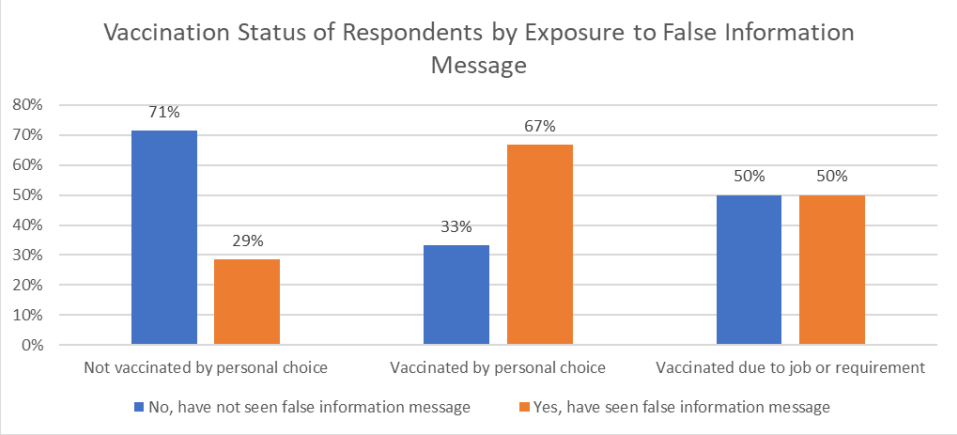

Figure 4: Vaccination status of respondents by exposure to false information

5.4 Hypothesis 4: Facebook and Twitter users that reported not independently researching content marked as "misleading" or "false" were more likely to have not received a COVID-19 vaccination

Only 4 respondents out of 97 were not vaccinated against COVID-19 by personal choice and have seen a false information warning message on Facebook or Twitter as seen in Figure 5. While a small number from our population, these 4 respondents noted that they only "Sometimes" or "Never" conducted independent research outside of Facebook or Twitter of any content marked with a false information warning message. Conversely, there was a much more equal distribution of frequency of independent research among those that were vaccinated against

COVID-19 by personal choice. As this is a small subset of users, a greater population sample may assist in answering this question with more confidence.

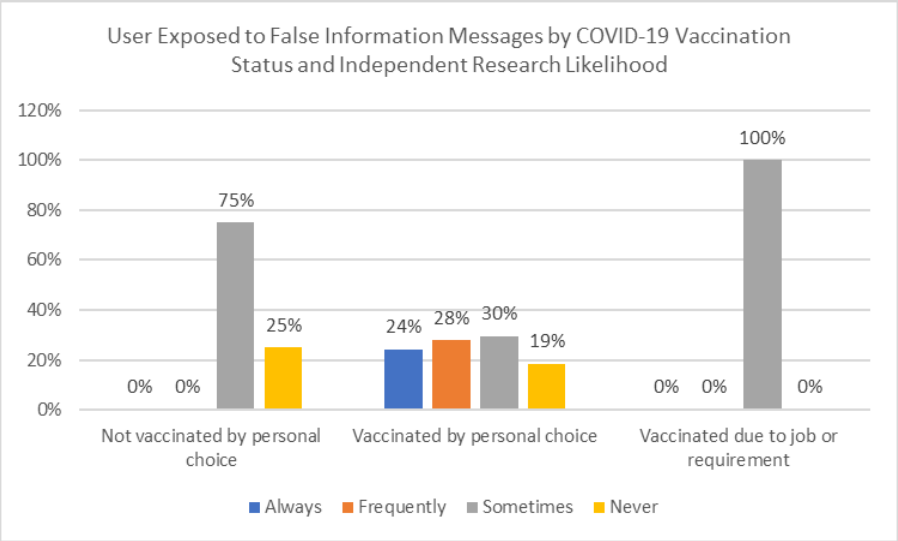

Figure 5: Covid19 vaccination status and exposure to false information

6. Conclusions

Our results suggest a correlation between COVID-19 vaccination status to having seen a false information warning message and responses to those messages. Respondents overwhelmingly selected "Sometimes" and "Never" as opposed to "Frequently" and "Always" meaning the majority of our respondents were indifferent at best to slightly dismissive at worst to Facebook and Twitter false information warning messages. Whether users who are vaccinated against COVID-19 spend more time on Facebook and Twitter and have a greater likelihood of seeing one of these messages or if another reason might explain how they are more likely to have seen one of these messages than those who reported not being vaccinated is unclear. What is clear from our results is that if Facebook or Twitter revealed details of how their algorithms are designed to influence the content displayed to users, most respondents indicated they would either use Facebook or Twitter less frequently or have no change. A very small minority indicated (2 out of 97 respondents) that learning more about these algorithms would cause them to spend more time on the websites. In addition to being intellectual property of Facebook and Twitter, fewer users spending less time on the websites could be another reason why they are unwilling to reveal additional information on their algorithms. Additionally, our results indicate a wider distribution of those who are

vaccinated and the frequency in which they independently research information marked with a false information warning message. While it appears from our results that those who are not vaccinated against COVID-19 do not research information independently at the same frequency as those that are vaccinated, our population was quite small so additional research with a larger sample may yield more confidence in results.

7. Acknowledgements

1. Peer edited by Dr. Radhouane Chouchane, associate professor of Computer Science at Morgan State University
2. Peer reviewed by Dr. Cameron Williams, assistant professor of Sociology at Columbus State University.
3. Beta-read by Jonathan DeYoung

References

[1] Peter M. Dahlgren. 2021. A critical review of filter bubbles and a comparison with selective exposure. Nordicom Review 42, 1 (2021), 15–33. DOI:http://dx.doi.org/10.2478/nor-2021-0002

[2] Jasper Schelling, Noortje van Eekelen, Ljsbrand van Veelen, Maarten van Hees, and Peter van der Putten. 2021. Bursting the Bubble. ACED (2021).

[3] M. Geetha Yadav, Rajasekhar Nennuri, N. Sairam, Y. Shiva Teja, Ganga Prasad 2021. Classifying Fake News Articles using Natural Language Processing and Supervised Learning Estimator. Annals of the Romanian Society for Cell Biology. 25, 6 (Jun. 2021), 6847–6856.

[4] David Lauer. 2021. Facebook's ethical failures are not accidental; they are part of the business model. AI and Ethics (2021). DOI:http://dx.doi.org/10.1007/s43681-021-00068-x

[5] Stephen Neely, Christina Eldredge, and Ron Sanders. 2021. Health Information Seeking Behaviors on Social Media During the COVID-19 Pandemic Among American Social Networking Site Users: Survey Study (Preprint). J Med Internet Res (2021). DOI:http://dx.doi.org/10.2196/preprints.29802

[6] Jianming Zhu, Peikun Ni, Guangmo Tong, Guoqing Wang, and Jun Huang. 2021. Influence Maximization Problem With Echo Chamber Effect in Social Network. IEEE Transactions on Computational Social Systems (2021), 1–9. DOI:http://dx.doi.org/10.1109/tcss.2021.3073064

[7] Fatimah Alzamzami and Abdulmotaleb El Saddik. 2021. Monitoring Cyber SentiHate Social Behavior During COVID-19 Pandemic in North America. IEEE Access 9 (2021), 91184–91208. DOI:http://dx.doi.org/10.1109/access.2021.3088410

[8] Felix Drinkall and Janet B. Pierrehumbert. 2021. Predicting COVID-19 cases using Reddit posts and other online resources. University of Oxford (2021).

[9] Christopher Whitfield, Yang Liu, and Mohd Anwar. 2021. Surveillance of COVID-19 Pandemic using Social Media: A Reddit Study in North Carolina. arXiv (June 2021)

[10] Joanne Chen Lyu, Eileen Le Han, and Garving K. Luli. 2021. COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. Journal of Medical Internet Research 23, 6 (2021). DOI:http://dx.doi.org/10.2196/24435

[11] Antonio F. Peralta, Matteo Neri, János Kertész, and Gerardo Iñiguez. 2021. The effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions. anXiv (May 2021).

[12] Matthew Andreotta et al. 2019. Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. SpringerLink (2019). DOI:http://dx.doi.org/10.31234/osf.io/bynz4

[13] Alexander Chkhartishvili and Ivan Kozitsin. 2018. Binary Separation Index for Echo Chamber Effect Measuring. 2018 Eleventh International Conference "Management of large-scale system development" (MLSD (2018). DOI:http://dx.doi.org/10.1109/mlsd.2018.8551823

[14] Richard Rogers. 2020. Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. European Journal of Communication 35, 3 (2020), 213–229. DOI:http://dx.doi.org/10.1177/0267323120922066

[15] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. Proceedings of the National Academy of Sciences 118, 9 (2021). DOI:http://dx.doi.org/10.1073/pnas.2023301118

[16] Elizabeth Dubois and Grant Blank. 2018. The echo chamber is overstated: the moderating effect of political interest and diverse media. Information, Communication & Society 21, 5 (2018), 729–745. DOI:http://dx.doi.org/10.1080/1369118x.2018.1428656

[17] Lin Cai and Zihang Wang. 2019. Coordination of Legal Protection of Algorithms and Intellectual Property System. Canadian Social Science 15, 6 (2019), 58–68. DOI:http://dx.doi.org/10.3968/11144

[18] Mark MacCarthy. 2020. Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry. SSRN Electronic Journal (2020). DOI:http://dx.doi.org/10.2139/ssrn.3615726

[19] Kathleen M. Carley, Guido Cervone, Nitin Agarwal, and Huan Liu. 2018. Social cyber-security. Social, Cultural, and Behavioral Modeling (2018), 389–394. DOI:http://dx.doi.org/10.1007/978-3-319-93372-6_42

[20] Data Policy. Facebook. (n.d.). Retrieved May 21, 2022, from https://m.facebook.com/privacy/explanation/

[21] Danielle Kehl, Priscilla Guo, and Samuel Kessler. 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing (2017). http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041

[22] Ortutay, B. (2021, November 9). Facebook parent meta to remove sensitive ad categories. AP NEWS. Retrieved from https://apnews.com/article/technology-business-misinformation-marriage-religion-aeb83a4553c8c9c6a5b26819fbdbb7e1

APPENDIX A: SURVEY QUESTIONS

Q1 What is your age?
- 18-25
- 26-35
- 36-50
- 51-65
- 66+

Q2 How often have you used Facebook in the past year on average?
- Not at all
- Once every few months
- Once a month
- Once a week
- Daily

Q3 How often have you used Twitter in the past year on average?
- Not at all
- Once every few months
- Once a month
- Once a week
- Daily

Q4 What is your gender?
- Man
- Woman
- Nonbinary
- Other

Q5 Do you identify as Hispanic or Latino?
- Yes
- No

Q6 How do you identify your race?
- American Indian, Alaska Native, or Indigenous
- Asian
- Black or African American
- Native Hawaiian or Pacific Islander
- White
- Other

Q7 What is your highest educational achievement?
- Some high school
- High school diploma/GED
- Some college
- Bachelor's degree
- Graduate degree
- Technical degree/certification

Q8 Which of the following COVID-19 guidelines do you regularly follow when in public? Check all that apply.
- Wear a mask
- Social distance
- Wash hands/use hand sanitizer
- Avoiding enclosed spaces with other people
- None

Q9 Have you received the COVID-19 vaccination either partially or fully?
- Yes, by personal choice
- Yes, due to job or other requirements

- No, by personal choice
- No, but due to a documented medical illness

Q10 If you did not receive the COVID-19 vaccination, did information found on Facebook or Twitter influence your decision?

- Not at all
- Somewhat
- Greatly
- Significantly

Q11 Have you seen a "fact checker" or "false information" message displayed on any COVID-19 related posts on Facebook or Twitter?

- Yes
- No

Q12 If you have seen a "fact checker" or "false information" message displayed on a Facebook post or Twitter tweet, how often did the following occur? An example of a Facebook message is displayed below.

	Never	Sometimes	Frequently	Always
Read the explanation				
Changed your opinion of the information presented				
Conducted independent research outside of Facebook or Twitter				

Q13 Which of the following kinds of posts or tweets do you recall seeing in the past year on Facebook or Twitter? Check all that apply.

- Health safety guidelines regarding the COVID-19 pandemic
- Safety of COVID-19 vaccines
- Negative effects of the COVID-19 vaccination on pregnancy or fertility
- Effectiveness of the COVID-19 vaccination
- COVID-19 vaccination causes human magnetism
- COVID-19 vaccination is a depopulation mechanism
- COVID-19 vaccination contains microchips or other technology
- None

Q14 How do you generally classify your political leaning?

- Democrat
- Independent
- Libertarian
- Republican
- Other

Q15 If Facebook or Twitter revealed details of how it uses artificial intelligence to influence your online behavior and the time you spend on the website, how would that impact the time you spend on the website?

- Less time on Facebook or Twitter
- No change
- More time on Facebook or Twitter

Q16 [Optional] (Optional) Please enter your email to be entered into a random drawing for a $20 Amazon gift card. Your email will only be used to contact you if you are chosen and will not be used to identify you.