

FILTERING FALSEHOODS: AN ANALYSIS OF THE
REGULATION OF FALSE INFORMATION ON SOCIAL MEDIA

by

ISAAC ORONSKY

A THESIS

Presented to the Department of Journalism and Communications
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Journalism

May 2025

An Abstract of the Thesis of

Isaac Oronsky for the degree of Bachelor of Arts
in the Department of Journalism and Communications to be taken June 2025

Filtering Falsehoods: An Analysis of the Regulation of False Information on Social Media

Approved: Professor *Thomas Bivins, Ph.D.*
Primary Thesis Advisor

The rapid spread of false information on social media presents significant danger for modern society. This thesis examines the evolving landscape of the regulatory efforts aimed at curbing the spread of misinformation and disinformation on these platforms. It explores the evolution of the news media landscape, highlighting the ongoing tensions between First Amendment protections and the dangers of misleading and/or false speech. This paper considers the existing methods of regulation by social media companies, the United States Government, and the U.S. Court systems, as well as other proposed methods by these three bodies. By evaluating these policies and case studies, this project presents some possible solutions to be used moving forward.

Acknowledgements

Thank you to Thomas Bivins, my Primary Thesis Advisor, for his guidance, humor, and flexibility. Thank you to Robert Mauro for his support. Thank you to my roommates for answering my many questions at all hours. Thank you to anyone that let me talk things through to them. And finally, thank you to my parents, Jodi and Carl Oronsky, for always telling me what I need to hear when I need to hear it, as well as my brother and grandparents for always pushing and supporting me. I couldn't have done this without you.

Table of Contents

| | |
|--|----|
| Introduction | 6 |
| January 6 th , 2021 | 6 |
| Misinformation and Disinformation | 8 |
| A History of News Media in the United States | 8 |
| Free Speech | 14 |
| Social Media and News Consumption | 16 |
| Methods | 19 |
| Methods of Regulation | 20 |
| Company Regulation | 20 |
| Government Regulation | 29 |
| Court Regulation | 35 |
| Conclusion | 38 |
| Bibliography | 42 |

List of Figures

| | |
|--|----|
| Figure 1.1: News consumption across platforms – percent of U.S. adults who ___ get news from... | 9 |
| Figure 1.2: News consumption across digital platforms – percent of U.S. adults who ___ get news from... | 10 |
| Figure 1.3: News consumption by social media site – percent of U.S. adults who regularly get news on each social media site. | 11 |
| Figure 2.1: An example of an alert that might be displayed on content containing false information published on Facebook as of March 2021. | 21 |
| Figure 2.2: Facebook content removal notices and explanations as of March 2021. | 21 |
| Figure 2.3: Context-providing measures on Facebook as of March 2021. | 22 |
| Figure 3: Community notes on X as of 2025. | 27 |

Introduction

January 6th, 2021

On January 6th, 2021, the U.S. Capitol building was breached in an attack for the first time in over 200 years (Elliott, 2021). The attack followed the gathering of tens of thousands of protesters on the day of the Congressional certification of the 2020 election results. By the end of the day, roughly 150 officers from surrounding departments and agencies were injured (Cameron, 2022). Seven people, including three Capitol police officers, died. Two other officers killed themselves in the days following.

Many of those present at the protest and subsequent attack were seen waving flags with President Trump's name on them, and/or wearing "Make America Great Again" regalia. Scholars agree that many of those at the Capitol that day believed that President Trump, not President Biden, had won the election (Pennycook and Rand 2021, Zadrozny, 2020). There were no credible indications of widespread election fraud then, and none have been uncovered since. The basis for this belief came from Trump himself. In the months leading up to the election, he repeatedly took to social media to spread false claims of fraud about the upcoming election process. This continued the morning of the 6th. At 8:17 a.m. EST, he tweeted: "States want to correct their votes, which they now know were based on irregularities and fraud, plus corrupt process never received legislative approval." By 10:30 a.m., the United States Capitol Police reported that between 25,000 and 35,000 people had gathered by the White House Ellipse in preparation for a "Save America" rally at which President Trump was set to speak, according to a 2021 Senate report. The speech, which began at 12:00 p.m., reiterated many of the same messages about election fraud that Trump had been spreading over social media over the previous months. President Trump finished speaking just after 1 p.m. By then, a crowd of

thousands had already amassed at the Capitol building, and some had made the first breach of the Capitol Police's perimeter around the building (U.S. Senate, 2021). Around that time, the building was put under lockdown, and the certification process had begun. Just after 1:30 p.m., Muriel Bowser, the Mayor of Washington D.C. called for National Guard support; and the protest was officially declared a riot at 1:49 p.m. EST. At 2:10 p.m., the rioters had breached the final barricade. Some smashed through Capitol windows a couple minutes later. Just before 2:15 p.m., the Senate (followed by the House) went into recess, and then-Vice President Mike Pence was evacuated. At 2:24 p.m., President Trump tweeted that Pence hadn't done what he should have to "give States a chance to certify a corrected set of facts, not the fraudulent or inaccurate ones which they were asked to previously certify." There was still no proof of election fraud. Over the next few hours, the Senate floor was breached, along with the Offices of then-Speaker Nancy Pelosi and other Representatives. As this was happening, President Trump continued to turn to social media to post unsupported claims of fraud. At 4:17 p.m. he posted a video in which he said "we had an election stolen from us. It was a landslide election and everyone knows it..." At 6:01 p.m., about 30 minutes after a perimeter was reestablished around the Capitol building and just after police confirmed that one of the rioters was dead, Trump tweeted: "These are the things and events that happen when a landslide victory is so unceremoniously and viciously stripped away..." He also told people to go home. The Capitol was declared secure at 8:00 pm., and Congress officially affirmed the results of the election at 3:47 a.m. on January 7th. Joe Biden was sworn in as the 46th President of the United States on January 20th.

An analysis of over 400 court filings for defendants arrested at the attack by Politifact found that hundreds of social media users, including President Trump and other high-profile Republicans, continually spread false information about voting machines, ballots, and election

results in the months leading up to and following the election. The analysis found that at least half of the defendants were influenced by this false information in their decision to attack the capitol.

Misinformation and Disinformation

The January 6th case study, though more public and disastrous than most, is not unique. Because of false information's clearly damaging effects, this project explores the approaches to its regulation on social media used by social media companies, the U.S. government, and the Courts. However, before understanding each method, it's worth defining "false information" a bit further. Scholars agree that false information is information that is either outright incorrect or so misleading that it could lead a lay person to believe or not believe something -- when the opposite is true. It can be broken down into two main categories -- misinformation and disinformation. Misinformation is false information that is created and/or spread unintentionally. On social media, this can manifest when someone reposts an article, photo, or video that contains any sort of false information, or by unknowingly creating and posting that information themselves. Disinformation, on the other hand, is false information that is spread deliberately. It's intended to mislead or misstate the facts of an issue, event, or story, and can be spread online in many of the same ways as misinformation. This project focuses on false information that is spread from people reporting on the news and/or newsworthy events.

A History of News Media in the United States

To understand how misinformation and disinformation affect modern news consumption on social media, one must first understand the basic history of news media in the United States. For over a century, news media was limited to articles found in newspapers. Eventually, radio and television also became major providers. Over the last two decades with the emergence of the

internet and social media, news consumption has largely gone virtual. According to data published in 2024 by Pew Research Center (Aubin et al. 2024), 86% of U.S. adults say they received their news either sometimes or often from a smartphone, computer, or tablet (Figure 1.1). Only 26% said that they got news from print, and 42% from radio. Since digital platforms are a product of technology advancements over the last three decades, these numbers represent a clear shift in news consumption habits from the inception of our country until now.

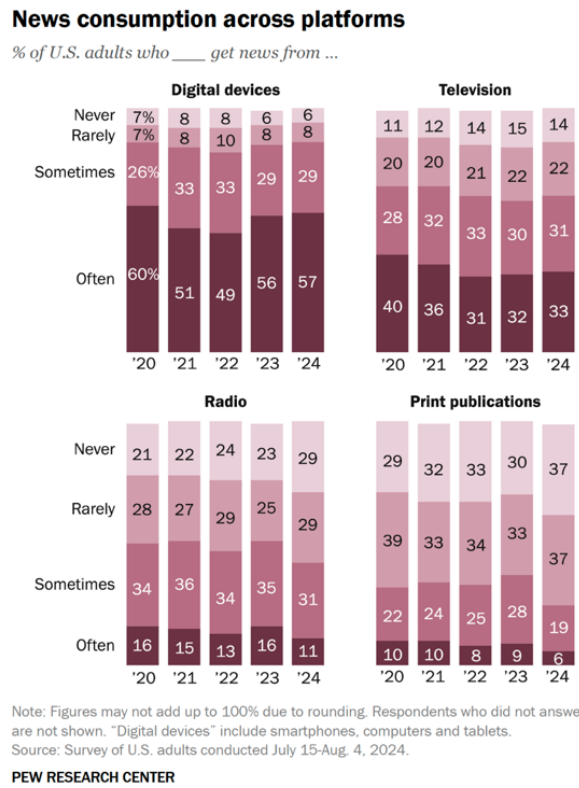


Figure 1.1: News consumption across platforms – percent of U.S. adults who ___ get news from...

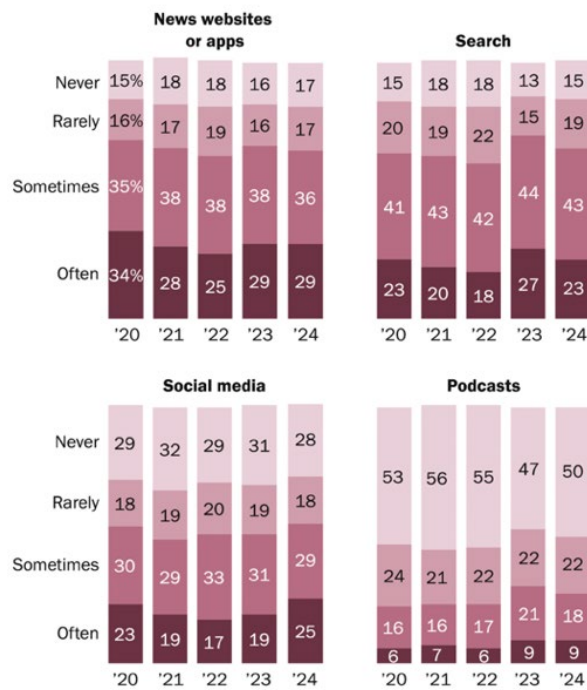
Courtesy of Pew Research Center (2024).

The study also looked at digital platforms specifically. When asked about their news consumption across digital platforms, 57% of respondents said that they at least sometimes got their news from social media (Figure 1.2). Further research investigated the percentage of U.S. adults that regularly get their news from various social media sites, and asked people which sites

they were using. Unsurprisingly, people were turning to Facebook the most, with 33% of people saying they got their news from that platform (Figure 1.3). YouTube and Instagram weren't far behind, with 32% and 20% of people regularly getting news from each platform, respectively (Aubin et al. 2024). TikTok and X (formerly Twitter) also represented some share of the use, at 17% and 12% each.

News consumption across digital platforms

% of U.S. adults who ___ get news from ...



Note: Figures may not add up to 100% due to rounding. Respondents who did not answer are not shown. Respondents who do not use the internet did not receive these questions; they are included with those who said "Never," along with those who do not say they get news from digital devices.
 Source: Survey of U.S. adults conducted July 15-Aug. 4, 2024.

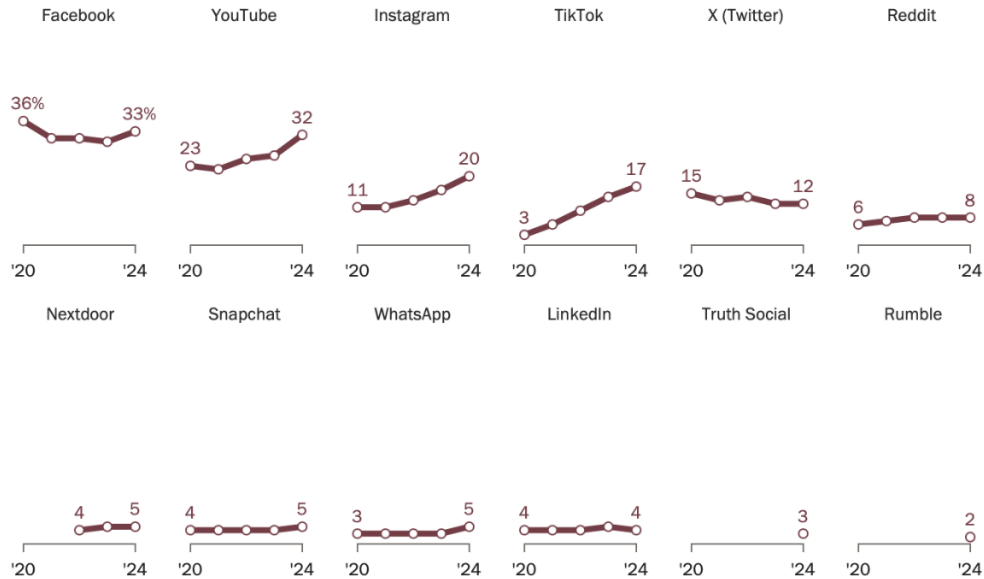
PEW RESEARCH CENTER

Figure 1.2: News consumption across digital platforms – percent of U.S. adults who ___ get news from...

Courtesy of Pew Research Center (2024).

News consumption by social media site

% of U.S. adults who **regularly** get news on each social media site



Source: Survey of U.S. adults conducted July 15-Aug. 4, 2024.

PEW RESEARCH CENTER

Figure 1.3: News consumption by social media site – percent of U.S. adults who regularly get news on each social media site.

Courtesy of Pew Research Center (2024).

Media Politics: A Citizen’s Guide, by Shanto Iyengar, explains how this change in news consumption has led to some dramatic shifts in how news is published (Iyengar, 2022). Iyengar, a political science professor at Stanford University, argues that traditional forms of media were either “point-to-point (between a single sender and recipient) or broadcast (between a single sender and multiple recipients).” For example, when news was published in newspapers or broadcast on television or the radio, a single news organization would communicate news to an audience. There was no real collaboration between the two parties – the audience was not afforded any opportunities to send any information back to the news organizations. Under this system, most people did not have the ability to broadcast any communication. The news organizations, through their journalist employees and editors, employed certain “gatekeeping” procedures, such as in-house fact checking, of their content (Iyengar 2022). Though every news

organization's individual fact-checking policy varied, people could assume that their news was being verified to a reasonable degree of certainty. If something published was later found to be wrong or misleading, people could reasonably assume that a public correction of some kind would be published by the news organization¹. But that changed with the introduction of the internet, Iyengar says.

“The development of the internet permitted simultaneous point-to-point and broadcast forms of communication for the first time and provided individual users with easy access to a huge audience. Every individual on the network of computers making up the World Wide Web is both a sender and a receiver. Any user of the internet can direct messages to individual recipients and at the same time harness the power of worldwide online networks to reach millions.”

The online networks Iyengar references have come in many forms, including social media platforms. As mentioned in the Pew Research study, Facebook has become one of the prominent sources of online news for many. The platform was launched in 2004 and has over 3 billion users worldwide (Dixon, 2025). X, launched two years later in 2006, has about 586 million users.

There are benefits to users of these sites being both “senders and receivers.” Since almost anyone can have influence, the spread of news is opened to a wider variety of perspectives (Iyengar, 2022). Older forms of news media were limited to just a few channels, and stories were told by just a few voices. This meant that news organizations, in addition to employing various fact-checking and gatekeeping procedures, were the ones influencing public conversation. Major news media got to have most of the say in determining what issues were spoken about on air (agenda setting), therefore determining what the country was thinking about (McCombs and Shaw, 1972). Now, anyone can partake in community-based journalism and be the one to break news if they have access to a cell phone and the internet. One of the first examples of the power

¹ This has, of course, also shifted in recent years. News organizations have aired false and/or misleading claims. Some publish corrections quietly, seemingly to hide them from their audience (Bump 2024).

of community social media journalism was the 2009 Hudson River plane crash landing in New York City. Heather Dueitt, a marketing consultant, wrote about her experience in a Forbes article titled *Sully: The Birth of Social Journalism and Ways to Approach New Social Platforms*. After hearing a splash from her office (as well as seeing a few Tweets mentioning a possible plane crash), she headed down from her office to the river. Though she recalls a crowd forming, she said there were no journalists when she first got to the scene – the river landing had just happened. She took out her phone and began taking photos and tweeting about what she saw. By the time she came back up to her office, her tweets had been screenshotted and published on CNN, the Wall Street Journal, ABC and NBC (Dueitt, 2017). The major news networks had to play catch-up by using social media content from Dueitt and others, because they couldn't break the news before it was published online. Community social media reporting has only grown since then. Firsthand experiences, photographs, and videos posted on social media have shaped movements. Many argue that if it were not for the ability for young people to organize on social media in the leadup and during the Arab Spring in 2011, the protests and uprisings (in Tunisia, Egypt, etc.) would not have been able to happen to the same effect (Hroub, 2015). There are also many that say that the 2020 Black Lives Matter protests and movement was bolstered by social media. The cell phone footage that captured the police murder of George Floyd was published on social media, and circulated rapidly (Burch and Eligon, 2020). Some of the audio from the video, such as Floyd's uttering of the words "I can't breathe," was used in chants, posters, and other forms of protest in-person and on social media. It was only after the video spread online that the major news media had access to the clip. Without social media, it's possible only a few would have seen the video, and the Black Lives Matter movement might not have been what it was.

Free Speech

In the United States, most speech is protected by the First Amendment, which reads:

“Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.”

However, the definition of “freedom of speech” hasn’t always been clear. Just a few years after the Bill of Rights was ratified and the First Amendment took effect, the United States Congress passed the Sedition Act of 1798. This law, passed when the U.S. was fearing war with France, included government-imposed wartime limits on speech. The law said that American citizens couldn’t “print, utter, or publish...any false, scandalous, and malicious writing” about the government. It expired three years later. Later, during WWI, congress passed the Espionage Act, which prohibited the interference with “the operation or success of the military or naval forces of the United States or to promote the success of its enemies and whoever when the United States is at war, shall willfully cause or attempt to cause insubordination, disloyalty, mutiny, refusal of duty, in the military or naval forces of the United States, or shall willfully obstruct the recruiting or enlistment service of the United States.” Both laws were used by the U.S. government to restrict speech during wartime.

Defining Speech Through the Courts

Since then, the Supreme Court has helped define what exactly “freedom of speech” entails through its rulings and published opinions. Early views of the court can be found in two cases: *Schenck v. United States* (1919), and *Abrams v. United States* (1919). In *Schenck*, we see a more restrictive view of the First Amendment. Towards the end of WWI, socialist Charles Schenck was arrested after handing out and mailing over 15,000 anti-war pamphlets. Schenck used the pamphlets to argue that the draft was in violation of the Thirteenth Amendment, which

prohibits involuntary servitude. He was charged with conspiracy to violate the Espionage Act, and the case eventually made its way to the Supreme Court. Schenck's lawyers argued that his arrest violated the First Amendment, and that Schenck's use of the pamphlets was protected speech. However, the arrest was ultimately upheld by a unanimous decision from the Court. Justice Oliver Wendell Holmes delivered the Court's opinion, which established the "clear and present danger" test (*Schenck v. United States*, 1919). This test allowed the U.S. government to restrict speech if it presented a clear and present danger. Holmes cited the fact that the country was in wartime when the speech was made, giving the U.S. government a bit more leeway to restrict Schenck's pamphlets. He compared the case to falsely yelling "fire" in a theater — speech that would also incite panic and/or danger. The clear and present danger test was a precedent for speech cases, and was referenced in another case the same year, *Abrams v. United States* (1919). In *Abrams*, two leaflets denouncing the US' sending of troops to Russia were dropped from a New York City window. According to the 7-2 decision by the court, the leaflets presented a "clear and present danger," and therefore were not protected by the First Amendment and *Abrams* lost the case. But this time, Justice Holmes disagreed with the outcome. In the dissenting opinion, he argued that the presence of only two leaflets didn't meet the standard outlined in *Schenck*. Because of that, *Abrams* shouldn't have been punished.

Though the Court in *Abrams* decided to again rule to censor speech, the dissenting opinion signaled later decisions that would give speech more protections. One of those Cases was *Brandenburg v. Ohio* (1969). In this case, Clarence Brandenburg, a leader of the Ku Klux Klan, invited a reporter to a rally in which members burned crosses and used speech that was derogatory towards minority groups, including Black people and Jews. Brandenburg was originally arrested and convicted under Ohio state law, which made it illegal to advocate for

"crime, sabotage, violence, or unlawful methods of terrorism as a means of accomplishing industrial or political reform" (*Brandenburg v. Ohio*, 1969). The case was appealed and eventually made its way to the Supreme Court. Contrary to the restrictions imposed in *Schenck* and *Abrams*, the Court sided with *Brandenburg*, deeming that the Ohio law and conviction had violated the Klan leader's right to free speech. The Court ruled that not only did speech have to contain a clear and present danger (as outlined in the *Schenck* case) to be restricted by the government, but also defined another factor – speech needed to incite “imminent lawless action.”

The Supreme Court continued to rule on First Amendment speech cases throughout the 20th century. *New York Times Co. v. Sullivan* (1964) established the concept of “actual malice” — the precedent that public figures needed to prove there was true malice in any defamatory statements made before being able to bring cases against those who made the statements. *New York Times Co. v. United States* (1971) limited “prior restraint” — U.S. government censorship of material before it was published in the name of national security. Both cases, in addition to others, provided decisions and standards that would further protect speech, especially as relating to news media.

Social Media and News Consumption

In the age of the internet, speech is everywhere. There are hundreds of different social media platforms, and people can post on them at any time. Many of the platforms are similar in that their goal is for people to connect through words, photos, or videos. However, they have a variety of different success and usage rates. Some come into popularity through trends but die out within a couple years such as Vine or Musical.ly, while others find themselves becoming a

mainstay of the public use. Each carves out some sort of niche in a community or group, and many provide a new feature for users.

Social media, by design, is largely reliant on advertising for revenue (Zenone et al., 2022). This means it is in platforms' best interest to keep users engaged. However, the strategies that social media companies use to keep users engaged can be problematic. Francis Haugen, a data scientist and former employee of Facebook, spoke on the issue in a 2021 interview with CBS News. She said that Facebook's algorithms work to keep users engaged by continuing to show them posts that have gotten high engagement levels from other users. Haugen cites false information as particularly engaging content, especially as it relates to violence and hate speech. This concept isn't new. It's been proven that people have a "negativity bias" – they are more likely to engage with news that is negative (Soroka et al., 2019). People have been shown to literally have a higher emotional response to negative news, keeping them more engaged with negative (and potentially false) content. According to Haugen, Facebook knows this and uses the phenomenon to increase their revenue stream: "if they change the algorithm to be safer, people will spend less time on the site, click on less ads, and they'll make less money" (Pelley, 2021).

There are other problems with the use of social media sites to publish news, as discussed previously. The internet and social media have contributed to the creation of a 24/7 news cycle (Iyengar, 2022). Due to the accessibility of the internet and social media, people no longer must wait for their morning newspaper or the 6:00 p.m. news to know the stories of the day – they can just open Facebook or X and scroll on their favorite news site's social accounts. And because anyone can report news, not just journalists, being the first to report the news can be nearly as important as reporting the news accurately. I argue that this is one of the main causes of the publishing of false information.

The increase in news content on social media has led to users having an increased choice in their news consumption. Users can selectively expose themselves to whatever news sites/sources they want through the following of certain accounts and organizations. People can also be subject to “confirmation bias” (Iyengar, 2022). This phenomenon occurs when people seek out information that aligns with their preexisting beliefs. I argue that the combination of selective exposure and confirmation bias likely amplifies the effects of misinformation and disinformation on social media. If false information more directly aligns with what someone already believes, they are more likely to consume that media instead of seeking out credible information and/or fact-checking their own beliefs. As a consequence, their beliefs become stronger – but not more accurate.

Methods

The bulk of this project focuses on understanding the various methods available to regulate speech on social media. While nearly all social media platforms can fall victim to their users spreading false information, this research largely focuses on what I'll call "the big two" – Meta (specifically Facebook) and X (formerly Twitter). These platforms are two of the most widely used, and they are the two that seem to make the most headlines for their problems with false information. I will divide my research into three parts: *Company Regulation*, *Government Regulation*, and *Court Regulation*. Each section will use past and current policies as well as example case studies to outline some of the ways that different entities have/are attempting to regulate content containing false information on social media. This project seeks to answer these three questions:

1. How has false speech on social media been regulated?
2. What do people think needs to be done (i.e., what are the current and past proposals)?
3. How can we improve the current systems?

I hope that by understanding the answers to these questions, I will be able to make suggestions as to some effective steps companies, the U.S. government, and the courts can use to combat false information online. The goal is not to come up with a perfect solution -- I don't believe there is one -- but rather to provide some suggestions for how to move forward in the battle to keep our communities safe from the spread of false information.

Methods of Regulation

Company Regulation

Because the misinformation and disinformation examined in this research has been published on social media platforms, it seems logical to think that it should be the responsibility of those social media companies to regulate the content. A July 2023 Pew Research Center study found that 65% of US adults believe that tech companies should be regulating false information on their platforms -- even if it means limiting the freedom of information (Aubin and Liedke, 2023). But what does that regulation look like?

Facebook

For Facebook, the answer is complicated. Since the 2016 Presidential election of Donald Trump, the company has changed its policies multiple times. During that year's election process, Facebook launched what was then a new fact-checking program. The program used third party organizations (such as The Associated Press) to review and/or take posts down that spread false information (Chow, 2025). According to March 2021 Congressional testimony by Facebook CEO Mark Zuckerberg, the platform worked with 80 independent third-party fact checkers to implement this system.

“If content is rated false by one of these third-party fact-checkers, we put a warning label on it. And based on one fact-check, we're able to kick off similarity detection methods that identify duplicates of debunked stories. When content is rated false, we significantly reduce its distribution; on average, this cuts future views by more than 80 percent.”

Some examples of the warning labels Zuckerberg references can be found in Figures 2.1, 2.2, and 2.3. Some of these labels must be clicked out of before viewing the post, while others simply provide a warning that certain posts might include false information alongside the content of the post. Some also provide links to information about why the post has been marked as false.

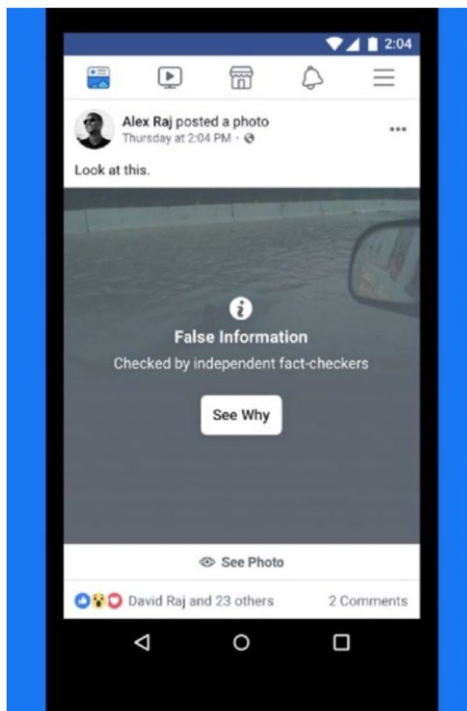


Figure 2.1: An example of an alert that might be displayed on content containing false information published on Facebook as of March 2021.

Courtesy of Facebook.

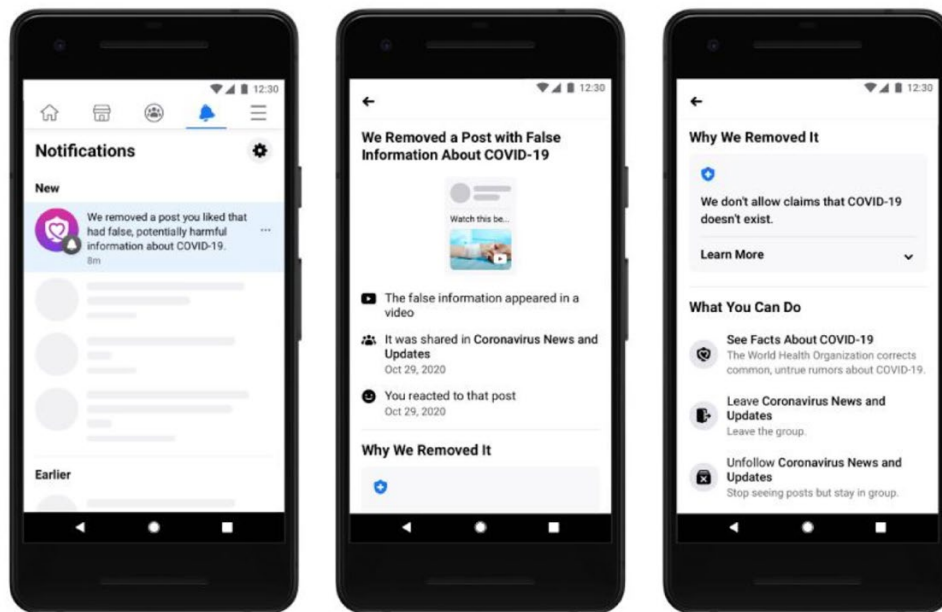


Figure 2.2: Facebook content removal notices and explanations as of March 2021.

Courtesy of Facebook.

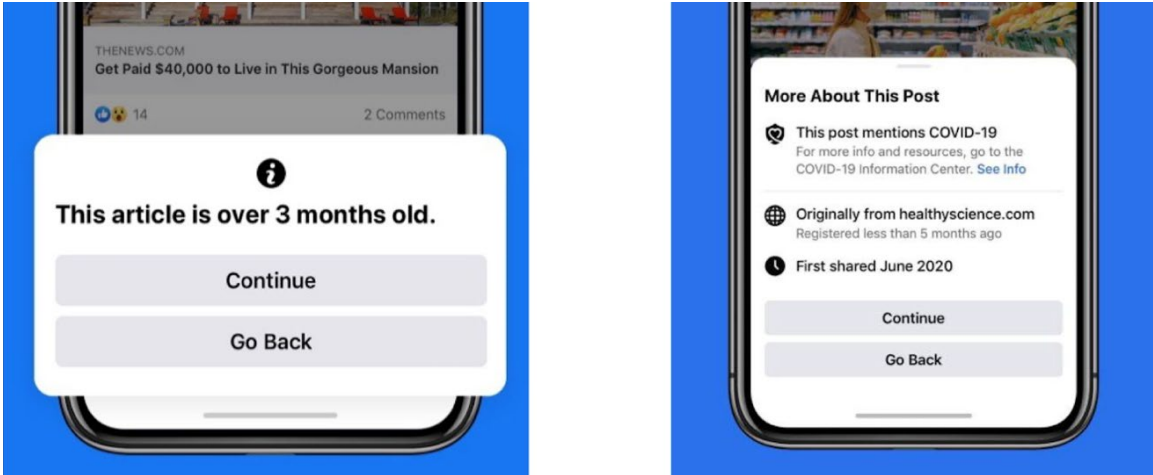


Figure 2.3: Context-providing measures on Facebook as of March 2021.

Courtesy of Facebook.

However, that system may not have worked as well as it was intended, according to a March 2021 study done by Avaaz. The study found that “the top 100 most popular false or misleading stories on Facebook, related to the 2020 elections, received an estimated 162 million views.” Of those 100 stories, 24% had no warning labels (Avaaz, 2021). While it is unknown how many views the posts would have had without the fact checking system, the results of this study seem to stand in contradiction to Zuckerberg’s claims in his testimony. Since that election, Facebook has also experimented with high profile account-banning. The most notable ban – that of President Donald Trump -- came after the January 6th, 2021 attack. The ban was conducted in response to statements posted on President Trump’s official accounts in the leadup, during, and following the attack². In a Facebook post published at 8:07 p.m. PST that day, Zuckerberg wrote:

“The shocking events of the last 24 hours clearly demonstrate that President Donald Trump intends to use his remaining time in office to undermine the peaceful and lawful transition of power to his elected successor, Joe Biden. His decision to use his platform to condone rather than condemn the actions of his supporters at the Capitol building has rightly disturbed people in the US and around the world. We removed these statements yesterday because we judged that

² Facebook initially banned President Trump’s accounts for 24 hours, before ultimately extending the block. (Oversight Board, 2024).

their effect -- and likely their intent -- would be to provoke further violence.... We believe the risks of allowing the President to continue to use our service during this period are simply too great. Therefore, we are extending the block we have placed on his Facebook and Instagram accounts indefinitely and for at least the next two weeks until the peaceful transition of power is complete.”

Facebook’s decision was sent to Meta’s Oversight Board on January 21st, 2021. The Board, built as an independent oversight body for the company, released a statement on May 5th of that year after conducting its review. The Board approved Meta’s 24-hour posting freeze but ruled that there needed to be a clearly defined timeframe on the “indefinite” suspension. Meta, Facebook’s parent company, decided on a two-year suspension, at which point they would reevaluate the situation. As part of its release, the board outlined the main factors that contributed to their decision. It cited two posts published by President Trump on the day of the incident. The first was from the video he posted at 4:21 p.m. EST, the full transcript of which read:

“I know your pain. I know you’re hurt. We had an election that was stolen from us. It was a landslide election, and everyone knows it, especially the other side, but you have to go home now. We have to have peace. We have to have law and order. We have to respect our great people in law and order. We don’t want anybody hurt. It’s a very tough period of time. There’s never been a time like this where such a thing happened, where they could take it away from all of us, from me, from you, from our country. This was a fraudulent election, but we can't play into the hands of these people. We have to have peace. So go home. We love you. You're very special. You've seen what happens. You see the way others are treated that are so bad and so evil. I know how you feel. But go home and go home in peace.”

Though the post also told his followers to leave the Capitol, President Trump continued to spread election fraud disinformation in this video. He cites “an election that was stolen from us,” calling it a “landslide,” which there was still no proof of. Due to the false information contained in this video (as well as incitement of violence), it was removed from Facebook at 5:41 p.m. EST for violating the company’s Community Standard on Dangerous Individuals and Organizations. The second post, published at 6:07 p.m., read:

“These are the things and events that happen when a sacred landslide election victory is so unceremoniously viciously stripped away from great patriots who have been badly unfairly treated for so long. Go home with love in peace. Remember this day forever!”

In this post, President Trump continued his disinformation campaign, even after seeing its effects. This was paramount in the decision to uphold the removal of his accounts. The Board also emphasized the earlier context of these two posts, which included a pattern of spreading false election information on Facebook in the months leading to the attack. This included an original post that suggested people should come to the rally in Washington D.C. on January 6th to “StopTheSteal” (Oversight Board 2021).

By the end of President Trump’s two-year suspension, the company determined that the “serious risk to public safety” that existed in January 2021 had receded, according to a statement (Clegg, 2023). At that point, President Trump’s accounts were reinstated. As of the publishing of this research, President Trump’s accounts remain up and running on Meta’s platforms.

In early 2025, Meta pivoted. Joel Kaplan, the company’s Chief Global Affairs Officer, released a statement paired with a video by Zuckerberg that outlined their new policies regarding fact checking. The release was divided into three parts. First, Kaplan announced that Meta would be moving away from the third-party fact checking system it had implemented in 2016. Instead, the company would be moving to a “community notes” system. According to the release, though the old fact-checking system began in good faith, it soon became obvious that “experts, like everyone else, have their own biases and perspectives.” Kaplan claimed that under the old system, too much content “that people would understand to be legitimate political speech and debate” was being censored. Instead, Meta would use X (which was already using a community notes system) as an example. For Facebook, the program would rely on the users of the platform to monitor content, by having them report potentially misleading content themselves. Under the

new system, posts that have been flagged are to be paired with a “less intrusive” label, as opposed to the full screen warnings that the old system employed (Kaplan, 2025). The second part of the release focused on what Meta called “allowing more speech.” Meta claimed that in its attempts to regulate content, it over-enforced rules. This led to unnecessary content censoring, which limited the amount of political debate possible on the platform. The release cited internal numbers from December 2024, saying that millions of pieces of content were removed daily (Kaplan, 2025). “While these actions account for less than 1% of content produced every day, we think one to two out of every 10 of these actions may have been mistakes (i.e., the content may not have actually violated our policies),” the release stated. This section also talked about how the company planned on limiting the use of automated rule-monitoring systems to scan only for “high severity” violations, such as terrorism and child exploitation, rather than all policy violations (Kaplan, 2025). This included reduced use to monitor false information. The final section said that Meta planned on phasing political/civic content back into feeds, after imposing limitations in 2021.

X (Twitter)

For X, the past decade has also led to some changes in false information regulation policies. X’s community notes program was implemented in 2022 (Kaplan, 2025). Originally used as a supplement to third-party fact-checking, X increased their use of community notes after the platform was bought by Elon Musk later that year (Wirtschafter and Majumder, 2023). Like the policies outlined by Meta, X’s community notes rely on contributions from users of the platform. Users are asked to write and rate notes on tweets to provide context, rather than depending solely on automated systems to do so. First, notes “start with the status of ‘needs more ratings,’” according to the company’s website. As seen in Figure 3.1, Users can then rate the

note, determining if it is “helpful” or not (X, 2025). What sets X apart from other community notes programs (such as Facebook, but also Reddit and Wikipedia) is in its tracking of note ratings from “different points of view” (Wirtschafter and Majumder, 2023). According to X’s website, the platform only considers a note “helpful” -- and therefore displays them on a post -- if enough people with “different perspectives” deem it so. But X’s determination of which users have different points of view is somewhat unclear. According to X’s webpage, users are determined to have different perspectives if they have rated posts differently in the past. The page claims that if people who have typically disagreed on their ratings of notes in the past agree that a given note is helpful, “it’s probably a good indicator the note is helpful to people from different points of view.” This means that when users originally write notes, their feedback isn’t public immediately. It is only once enough people that X determines have different perspectives agree that a note is “helpful.” that the note is displayed next to the post (as seen in Figure 3.2). If enough people determine the note to be “not helpful,” the note is not shown. The website says that the helpfulness status of each note is locked after two weeks. X releases public data on their community notes, which was used in a 2023 study of the program’s effectiveness (Wirtschafter and Majumder, 2023). The analysis relied on data from around 52,000 Community Notes, as well as 1.3 million ratings from over 27,000 contributors through February 2023.

“Based on this data, we find that most Community Notes contributors have never written a “helpful” note. Out of the nearly 7,000 unique contributors who have written notes, more than 5,500 (~80%) have never written a note locked as “helpful.” Of those who have written a note locked as “helpful,” 92% have done so four or fewer times. Only five users have written 50 or more “helpful” notes.”

Though the number has increased since Musk took over and X has pivoted to focus on the program, only 7% of all written notes have been determined “helpful.” There are also signs of a reduced partisan bias in “helpful” notes since the program expanded, which was originally a larger struggle for the company. Community notes do not force the deletion of any posts, and the

research showed that those who received Notes on their posts that were deemed “helpful” were no more likely to take down their posts than those who received Notes deemed “unhelpful” (Wirtschafter and Majumder, 2023). Based on these results from study, it seems as though the notes system does not have an impact on the subsequent behavior of users, and only a marginal impact on alerting users of posts containing false information.

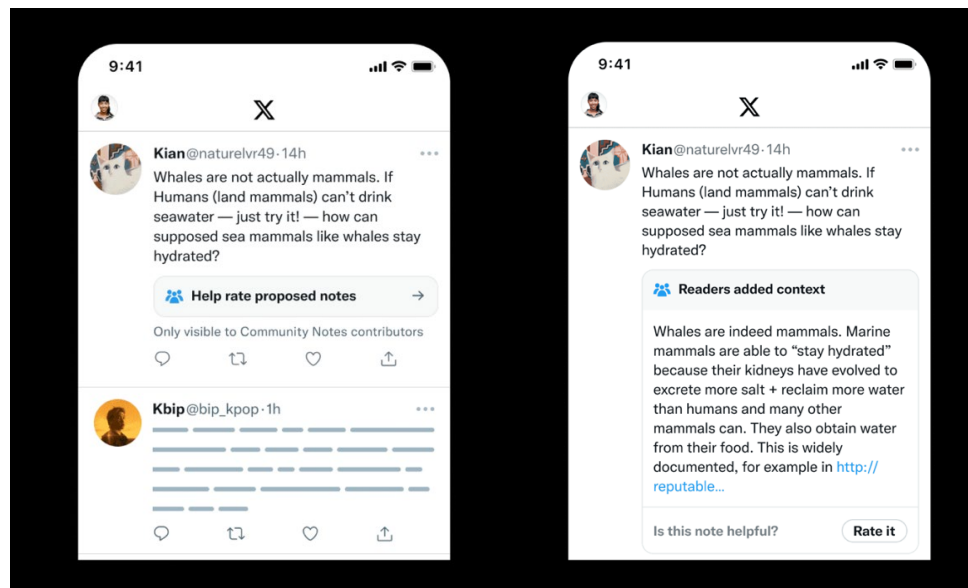


Figure 3: Community notes on X as of 2025.

The example on the left shows the screen that X users see before rating Community Notes. The example on the right shows a public Community Note that adds context to a tweet that contains possible false information. The bottom also prompts users to rate the helpfulness of the Note.

In line with Facebook/Meta, X also placed restrictions on President Trump’s accounts on its platform after the events on January 6th. That day, based on content posted containing similar false information to that published on Facebook, X locked President Trump out of his account for 12 hours (X, 2021). When he was let back in the next day, he continued tweeting falsities about the election and attack. On January 8th, X imposed a “permanent” ban on the @realDonaldTrump account in large part due to the President’s continued support of false

claims, according to a release posted on their website. The release cited two tweets from the 8th in their rationale behind the banning. The first tweet read:

“The 75,000,000 great American Patriots who voted for me, AMERICA FIRST, and MAKE AMERICA GREAT AGAIN, will have a GIANT VOICE long into the future. They will not be disrespected or treated unfairly in any way, shape or form!!!”

X determined that President Trump’s reference to his supporters having a “GIANT VOICE long into the future” and that them not being “disrespected or treated unfairly in any way, shape or form!!!” meant that he was likely to continue to support and spread false claims about him winning the election. The second post that the release cited read: “To all of those who have asked, I will not be going to the Inauguration on January 20th.” X determined that President Trump’s statement against going to the Inauguration again served as a confirmation to his supporters and followers that the election was not legitimate, a false claim. Soon after, X also permanently (at the time) suspended the @POTUS account, as well as the @TeamTrump campaign account³ (Denham, 2021). On January 13th, Jack Dorsey, the Chief Executive of X, tweeted about the ban. “I do not celebrate or feel pride in our having to ban @realDonaldTrump from Twitter, or how we got here.... I feel a ban is a failure of ours, ultimately, to promote healthy conversation. And a time for us to reflect on our operations and the environment around us,” Dorsey wrote in a thread on X (Paul, 2021).

In February 2023, a few months after Elon Musk bought the social media site, he reinstated President Trump’s X accounts (Jacobs, 2025). As of the publishing of this research, President Trump’s X accounts remain up and running.

³ In the days and months following the January 6th attack on the Capitol, various other social media platforms in addition to Facebook and X, such as Snapchat, YouTube, and Reddit, also banned President Trump’s account(s) on their platforms (Denham, 2021).

Government Regulation

As for governmental regulation of false information on social media, there are seemingly two schools of thought. First, that the U.S. government has a responsibility to protect its citizens from false (and potentially harmful) information through its own social media speech regulation. Second, that the U.S. Government has a responsibility to uphold free speech protections outlined in the First Amendment and should require social media companies allow users to publish information on their platforms freely. According to data from Pew Research Center, 55% of people believe the US government should be taking steps to regulate false information on social media (Aubin and Liedke 2023). The country is almost evenly split on the issue, which is likely why there hasn't been much Federal legislation passed. In fact, social media itself seems to be in a bit of a grey area for Congress — federal laws don't currently give an explicit definition to “online platform” (congress.gov, 2025). I argue that without this basic definition, members of Congress have been unable to agree on basic terms relating to the topic that any possible legislation would contain.

The Federal Communications Commission

When it comes to the regulation of broadcast speech, the U.S. Government's role is clear -- that responsibility lies with the Federal Communications Commission (FCC). The FCC is an independent government agency established in 1934. It is run by five commissioners, all appointed by the President, who serve for five-year terms and oversee several bureaus (FCC, 2025). The FCC makes three basic types of rules to regulate broadcast speech. Legislative rules create “legally binding rights and obligations for the agency and the public;” non-legislative rules interpret the meaning of other rules or statutes from the Commission or deliver policy statements (such as explaining specific fines); and Organizational and procedural rules simply

describe the agency's structure and outline how decisions are made (FCC, 2025). The commission makes rules authorized through Congressional statutes. These statutes can authorize rulemaking ranging from topics as simple as granting broadcasting licenses in the public interest, to content regulation. After a rule is submitted, it becomes available for public comment before ultimately being submitted to Congress. If lawmakers are unhappy with the rule, they can begin a process to overturn the rule. FCC rules may also be challenged in court, where they may be overturned or sent back to the Commission for further consideration.

One of the FCC's non-rulemaking bureaus deals with responding to complaints from the public. The claims often allege that "television and/or radio networks, stations, or their employees have broadcast extreme, incorrect, or somehow improper political, economic or social statements," according to the Commission's webpage.

"In some cases, the complaints allege that certain broadcast statements may endanger the United States or its people, or threaten our form of government, our economic system or established institutions like family or marriage. They say these statements are "un-American" and an abuse of freedom of speech. The FCC also receives complaints that some broadcast statements criticize, ridicule, "stereotype" or demean individuals or groups because of the religion, race, nationality, gender, gender identification, or sexual orientation, or other characteristics of the group or individual. Finally, many consumers complain that television or radio broadcasts are obscene, indecent, profane or otherwise offensive."

The FCC reviews these complains and works with other bureaus to conduct investigations. It will occasionally deliver fines to certain networks and/or broadcasts if it determines a rule was broken. But its regulation has a few parameters. The First Amendment bars the FCC from attempting to prevent the broadcasting of any point of view, and from outright censoring most broadcast material. The FCC is also held to the *Schenck* (1919) ruling and is therefore unable to make any regulations on statements that don't present a "clear and present danger of serious, substantive evil." However, the FCC does have some enforcement leeway in certain

circumstances. For example, though the First Amendment bars the Commission from outlawing the broadcast of indecent and profane material in its entirety, it is allowed to prohibit the material during certain hours of the day. Because there is a reasonable risk that children might be in the audience, the FCC prohibits indecent and profane material between 6 a.m. and 10 p.m. (FCC, 2025). The restrictions vary depending on if the broadcast is over-the-air or through cable/satellite.

None of these rules apply to social media. While people may be able to file FCC complaints about broadcast speech that may violate ethical standards regarding race, sexual orientation or gender identity, those same opportunities are not offered for speech found on social media. The FCC cannot restrict content posted on social media in any way.

Hearings and Information Gathering

In recent years, members have held hearings with leadership representatives from major tech companies to understand and work to combat misinformation and disinformation on their platforms. On March 25, 2021, three months after the attack on the Capitol, Members of the House Energy and Commerce Committee hosted Mark Zuckerberg (Meta), Jack Dorsey (X), and Sundar Pichai (Google) to discuss the spread of false information on their platforms, CNN's Brian Fung reported. Many members seemed upset at the tech CEOs, especially in the wake of not only the misinformation and disinformation campaigns relating to the attack on the Capitol but also surrounding COVID-19 public health information and vaccines. "You can take this content down. You can reduce the vision. You can fix this. But you choose not to," said Rep. Mike Doyle, chair of the House subcommittee on Communications and Technology at the time. "You have the means. But time after time you are picking engagement and profit over the health and safety of users," he said. Rep. Doyle's statement would suggest that he believes the

responsibility for the regulation of false information on social media belongs with the companies, but that isn't true for all lawmakers.

Common Carriers

For lawmakers who want existing company regulation to decrease in favor of a broader view of the First Amendment speech on social media, there is another solution. They believe that companies have biases and censor information unfairly, so they are in favor of social media platforms being designated as what are called “common carriers” (Huddleston, 2021). Common carriers are defined as entities that must provide their services without discrimination against people, groups, or content, because they provide an essential service (Villasenor 2022). The distinction draws from Section 230 of the Communications/Telecom Act of 1934 (updated in 1996) and would be another expansion on what the law covers to account for more modern technologies (Huddleston, 2021). Expansion of the law would allow for the government to limit the number of restrictions the owners of social media companies could impose on the users of their platforms. The key is considering the services that social media provides essential. Examples of common carriers are communications services like phone companies. Because they provide an essential service, they cannot refuse to allow people who talk about certain political or religious issues to use their services under Section 230. While the companies are permitted to take a stand for or against certain issues publicly, they cannot censor the speech of any users of their service that may disagree with those views. If social media sites were to be held as common carriers under Section 230, they likely wouldn't be able to regulate potentially hateful, racist, or ignorant speech from their users in almost any capacity. With a common carrier distinction, although the companies would continue to have First Amendment rights to free speech

themselves, any sort of regulation of content would infringe on the site's users' right to free speech.

Those in favor of a common carrier status argue that social media companies have marketed themselves as a method of communication, which is a common carrier service (Villasenor, 2022). They also argue that the platforms are not different from — but rather a modern manifestation of — other common carrier communications systems like landline and other cellular networks. Some also say that, because of the popularity and influence that social media sites have gained over the last decade, they hold immense political and economic power if they're able to restrict certain people and/or groups of people (Volokh, 2021). To ensure that one (or a select few of larger and/or more popular companies) don't have too much power, Section 230 should be expanded, and social media sites should be common carriers. President Trump (during his first term) hinted at applying this distinction by showing support for less regulation of content on social media. In a May 2020 Executive Order titled *Executive Order on Preventing Online Censorship*, he called out platforms like X for what he claimed was a political bias in their warning labels on certain content. The Executive Order suggested that the internet should be a place of “free and open debate,” and any sort of content regulation should be regarded as censorship of speech. It also directed the Attorney General to create a working group to address the issue, with the goal of eventually developing a proposal for Federal legislation that would achieve the views expressed in the Order. As of the publishing of this research, no working group was created to any large success, and there has been no major legislation passed that would officially designate social media as a common carrier under Section 230.

Those against common carrier distinctions point to the fact that social media is too far reaching to be subject to the standards required of a common carrier. Not only that, but those

standards find themselves outdated when considering the scope of modern technologies (Howell, 2024). Others say that for social media sites to be considered common carriers they would need to create a natural monopoly — which they don't (Huddleston, 2022). As this project has outlined, there are a multitude of social media platforms. Alternatives for popular platforms pop up all the time, and nearly anyone could create one of their own. There is also a belief among some that a common carrier distinction on social media would deter innovation in the field, because with an increase in government involvement there would likely be an increase in costs (Huddleston, 2022). These costs would both hinder innovation within the companies, and likely would also reduce the amount of competition the existing companies have. With less competition, there would likely be less incentive to improve their services or listen to users' complaints about the platforms.

Tech companies, including Facebook, have also given their thoughts on Section 230. In Zuckerberg's March 2021 testimony, he covered what he and Facebook believe the role of Section 230 should be in social media regulation. He said that he believes in a "thoughtful reform" of the law, one that requires social media companies to "demonstrate that they have systems in place for identifying unlawful content and removing it." However, he emphasized that he believes platforms should not be held liable for content that manages to avoid detection by those systems. This sentiment, as conveyed in 2021, stands in contrast with the company's recent actions. Community notes, as outlined by Facebook (using X as a model), seem to scale back the ability for fact-checking and content removal on the platform, as opposed to the platform's old system of using third-party fact checking systems.

Court Regulation

This project has already discussed some of the history of the U.S. Court systems and the First Amendment. When it comes to modern free speech considerations as it relates to false information on social media, there are a variety of methods currently available to respond to various forms of disinformation in the courts. One way that is achieved is through the bringing of defamation cases. Defamation, a statement that harms someone's reputation, is a complicated issue. It includes both libel (written or broadcast statements) and slander (spoken statements) (Cornell Law, 2025). Depending on who publishes the information, where the information is published, and whose reputation is harmed by the information, there are different standards for lawsuits (Bivins, 2023). Generally, a plaintiff in a defamation case must prove four things: 1) the statement is false, and was reporting to be fact; 2) the statement was published or communicated to a third person; 3) fault on the part of the defendant; and 4) some harm (damages) caused to the reputation of the person or entity who is the subject of the statement.

The goal of this section is not to conduct a full legal analysis of cases in which the plaintiffs sued for defamation as a method of regulation of false information on social media. There are too many cases, and too much nuance. Instead, I'll use this section to describe how people and/or businesses can use the courts as a method of regulation when these other methods – companies or the government – don't work. There are two notable cases from recent years -- *Sandy Hook Families v Alex Jones*, and *US Dominion, Inc. v. Fox News Network, LLC*. In both cases the plaintiffs were suing for defamation. The two cases provide examples in which the plaintiffs were successful in getting monetary compensation for false information that was spread about them or their loved ones on social media. They can also help provide more important

context to both sides of the issue — those spreading the information, and those affected by the false claims. For each case, I'll cover the answers to these questions:

1. What information was spread?
2. What argument did the defendant use?
3. What was the outcome?

Sandy Hook Families v Alex Jones (2022) involved Alex Jones and the families of the Sandy Hook Elementary School shooting victims. The case centered around Jones' publishing of conspiracy theories claiming that the 2012 shooting was a hoax, staged by the government as part of a plot to take away Americans' guns (The BBC, 2018). Jones, the founder of the website Infowars, repeatedly spread these false claims online and on social media, suggesting that the families of the 26 victims were either actors or complicit in the staged event. The families filed lawsuits against Jones for defamation, arguing that Jones' statements had caused significant damage. Many reported feelings of distress and experiencing harassment, and threats from his followers (Collins et al. 2024). Jones argued that the First Amendment protected his speech, even though what he was saying was false. Ultimately Jones was ordered to pay nearly \$1.5 billion in damages.

In 2021, another high-profile defamation case was filed: *US Dominion, Inc. v. Fox News Network, LLC* (2023). In this case, Dominion Voting Systems, a company that makes electronic voting machines, accused Fox News of knowingly broadcasting and publishing false claims, including accusations that Dominion's machines were used to manipulate the election results in favor of Joe Biden. The company sued for defamation, saying that the spreading of disinformation damaged its reputation and led to tangible harm, such as death threats against Dominion employees, loss of business, and significant damage to its brand. They presented

evidence of Fox executives and anchors expressing doubts about the fraud claims behind the scenes but continuing to give them airtime on the network and on social media platforms. Fox denied the claims, arguing that it was merely reporting on allegations made by public figures, which is protected by the First Amendment (Folkenflik, 2023). However, Dominion uncovered internal communications from Fox News employees that suggested some executives at Fox had doubts about the fraud claims, including host Tucker Carlson and anchor Sean Hannity. The case was settled in April of 2023 with a roughly \$787 million settlement — just under half of the \$1.6 billion Dominion had originally asked for.

Conclusion

Due to the increased accessibility and popularity of the internet over the last quarter-century, the spread of false information on social media has become widespread. False information, including misinformation and disinformation, is harmful, and there have been attempts to try to combat that harm through content regulation. However, there is also still a desire by many to be mindful of how people choose to do so. They want to make sure that regulation doesn't go too far, and First Amendment free speech protections remain intact.

Before delving into my conclusions, it's worth revisiting the July 2023 Pew Research Center data relating to social media regulation by tech companies. The researchers found that the 65% of U.S. adults who believe that tech companies (as well as the 55% of U.S. adults who believe that the U.S. Government) should be regulating false information on their platforms believe that the regulation should happen *even if it means limiting the freedom of information* (Aubin and Liedke 2023). To an extent, I agree with the sentiment. There have been too many examples of false information on social media having incredibly dangerous effects. January 6th was one of our country's darkest days, and I argue that it could not have happened to the extent that it did without the ability for President Trump and other high-profile people to spread false information over social media. The false claims from Alex Jones and Fox News on their social media accounts were also shown to have great harm, and those were just two of the examples brought to court. Social media companies should – at least in-part – have a responsibility to regulate false information on their platforms to protect their users and our communities.

But this project would be remiss if it did not address the free speech argument. As a general principal, social media sites should not be banning accounts without clear reason. The President Trump example was an extreme one, and in many cases there is no need to completely

ban someone's account. The idea that Section 230 of the Communications Decency Act should be extended to social media sites to preserve free speech has also emerged. I understand the need to protect free speech. Censorship has been used to quiet opposing voices and control people too many times, and I understand the hesitation to allow any entity, whether it be the social media companies, the U.S. government, or the Courts, have any sort of control over speech. But expanding Section 230 is not the answer. The ability for a single user to reach millions of others in seconds differentiates social media from other common carriers, such as phone lines. The consequences that unregulated speech (or that with little regulation) can have on a society are just too great. Continuing the path of deregulation through designating social media as a common carrier would only increase that danger.

But it's also clear that a singular regulation method doesn't have overwhelming support, and that a solution that uses just one of the three entities discussed in this project will not work. Social media evolves too quickly, and there is just too much content posted on too many platforms. To decrease harm and ensure that public safety is protected as much as free speech is, we should employ a system of checks and balances. In essence, all three regulation methods I've discussed must work together to regulate false information.

First, companies should continue to explore regulation policies that fit each of their individual platforms. Each platform works differently, and these companies are privately owned – therefore they should continue to be allowed to create their own policies. They must collect and respond to data that examines the success of their policies in regulating false information. I recognize there likely isn't one correct answer, but I suggest that platforms implement a system that combines third-party fact checking with community notes. Third parties will likely have more knowledge and resources than the average person to be able to fact-check correctly, so they

should continue to be included in companies' processes. But the community notes system, though far from perfect in its current form, also has merits. Social media companies should be taking advantage of their users when it comes to regulating content. These are the people actively using the platforms, and they may be able to help the third parties as they look to flag and fact-check potentially false posts.

However, Haugen, the Facebook whistleblower, helped explain why this cannot be the only method of regulation (Pelley, 2021). Social media companies will always have a financial incentive to boost content that might be incorrect, misleading, or based on incorrect facts. Speech containing false information often evokes a strong emotional response and will likely always get high engagement, increasing revenue for social media companies (Soroka et al., 2019). Because of this, the U.S. Government must help with the regulation of false information as well.

Lawmakers should use the FCC as an example and create an independent agency that can create rules for and oversee speech on social media. Like the FCC, the purpose of this new body should not be to censor speech and/or to infringe on First Amendment protections, but rather to protect against egregious misuses of free speech that could endanger others. To do this, the new FCC-like body would work similarly to the current FCC. It would be available to accept complaints from people about potentially dangerous false posts that violate rules outlined by the agency. After receiving a complaint, the agency will conduct investigations into the severity of the violation and distribute fines for the offending party.

Finally, the courts should continue to be available to combat false information that does fall through the cracks and ends up harming people. As much as I would like to believe that company regulation policies paired with this new FCC-like government agency will be enough to curb the effects of false information on social media, the sheer amount of speech that is

published on these sites makes it impossible. There will continue to be instances false information harming people or groups, and the courts must continue to be an option for those who have been harmed to respond and get content removed.

I don't believe any of these ideas are radical. The problem with the current methods of regulation is that it seems to be unclear as to what entities have what power over the situation. Though it's understandable that there has been a learning curve for new technologies, now is the time to figure out a solution. By establishing a governmental body like the FCC that has some authority over online speech, I believe the current methods of false information regulation by companies, the government, and the Courts will be strengthened as viable ways to combat this problem.

Bibliography

- "Abrams v. United States." *Oyez*, www.oyez.org/cases/1900-1940/250us616.
- Aubin, Christopher St., and Jacob Liedke. "In US, most favor restricting false information, violent content online." *Pew Research Center*, 20 July 2023, <https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online/>.
- Aubin, Christopher St., and Jacob Liedke. "News Platform Fact Sheet, 2024." *Pew Research Center*, 17 September 2024, <https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/>.
- Bidar, Musadiq. "Trump's Facebook and Instagram accounts active again." *CBS News*, 9 February 2023, <https://www.cbsnews.com/news/donald-trump-facebook-instagram-active-today-2023-02-09/>.
- Bump, Philip. "Fox News quietly reports on a fact sheet correcting Fox News misinformation." *Washington Post*, <https://www.washingtonpost.com/politics/2024/10/09/fox-news-quietly-reports-fact-sheet-correcting-fox-news-misinformation/>.
- Bivins, Tom. *Mixed Media: Moral Distinctions in Advertising, Public Relations, and Journalism*. Routledge, 2023.
- "Capitol riots timeline: What happened on 6 January 2021?" *The BBC*, 1 August 2023, <https://www.bbc.com/news/world-us-canada-56004916>.
- Clegg, Nick. "Ending Suspension of Trump's Accounts with New Guardrails to Deter Repeat Offenses." *Meta*, 25 January 2023, <https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/>.
- Collins, Dave, et al. "Alex Jones v Sandy Hook families: All you need to know." *AP News*, 14 June 2024, <https://apnews.com/article/alex-jones-infowars-bankruptcy-sandy-hook-0c3576e3c4bd853ac2cc5342118fca8c>.
- "Defining and Regulating Online Platforms | Congress.gov | Library of Congress." *Congress.gov*, 25 August 2023, <https://www.congress.gov/crs-product/R47662>.
- Denham, Hannah. "Trump and his allies are banned from these platforms." *The Washington Post*, 14 January 2021, <https://www.washingtonpost.com/technology/2021/01/11/trump-banned-social-media/>.
- Dueitt, Heather. "Sully: The Birth of Social Journalism and Ways to Approach New Social Platforms." *Forbes*, 26 January 2017, <https://www.forbes.com/sites/forbescommunicationscouncil/2017/01/26/sully-the-birth-of-social-journalism-and-ways-to-approach-new-social-platforms/>.

- Dwoskin, Elizabeth, and Craig Timberg. "Misinformation dropped dramatically the week after Twitter banned Trump and some allies." *Washington Post*, 16 January 2021, <https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/>.
- Eligon, John. "Bystander Videos of George Floyd and Others Are Policing the Police (Published 2020)." *The New York Times*, 24 November 2021, <https://www.nytimes.com/2020/05/26/us/george-floyd-minneapolis-police.html>.
- Elliott, Philip. "The Breach of the Capitol Spooked Us — As It Should Have." *TIME*, 7 January 2021, <https://time.com/5927664/capital-siege-trump-supporters/>.
- "Examining the U.S. Capitol Attack: A Review of the Security, Planning and Response Failures on January 6th." *Rules.senate.gov*, 1 June 2021, <https://www.rules.senate.gov/imo/media/doc/Jan%206%20HSGAC%20Rules%20Report.pdf>.
- "Executive Order on Preventing Online Censorship." Trump White House Archives, 28 May 2020, <https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-preventing-online-censorship/>.
- "Facebook: From Election to Insurrection." AVAAZ, 18 March 2021, https://secure.avaaz.org/campaign/en/facebook_election_insurrection/.
- Folkenflik, David, and Mary Yang. "Fox News settles blockbuster defamation lawsuit with Dominion Voting Systems." *NPR*, 18 April 2023, <https://www.npr.org/2023/04/18/1170339114/fox-news-settles-blockbuster-defamation-lawsuit-with-dominion-voting-systems>.
- "Former President Trump's Suspension." Oversight Board, 5 May 2021, <https://www.oversightboard.com/decision/fb-691qamhj/>.
- Fung, Brian. "Facebook, Twitter and Google CEOs grilled by Congress on misinformation." *CNN*, 25 March 2021, <https://www.cnn.com/2021/03/25/tech/tech-ceos-hearing>.
- Howell, Bronwyn. "Should Internet Platforms Be Classified as Common Carriers?" *American Enterprise Institute*, 22 January 2024, <https://www.aei.org/technology-and-innovation/should-internet-platforms-be-classified-as-common-carriers/>.
- Huddleston, Jennifer. "Consequences of Classifying Elements of the Internet as a Common Carrier." 23 February 2021, <https://www.americanactionforum.org/insight/consequences-of-classifying-elements-of-the-internet-as-a-common-carrier/>.
- Iyengar, Shanto. *Media Politics: A Citizen's Guide*. W.W. Norton and Company, 2022.
- Kaplan, Joel. "More Speech and Fewer Mistakes." *Meta*, 7 January 2025, <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.
- "New York Times Company v. Sullivan." *Oyez*, www.oyez.org/cases/1963/39.

“Notes shown on X.” X, <https://communitynotes.x.com/guide/en/contributing/notes-on-twitter>.

Paul, Kari. “Twitter chief says Trump ban was right decision but sets 'dangerous precedent.’” *The Guardian*, 13 January 2021, <https://www.theguardian.com/technology/2021/jan/13/trump-twitter-ban-jack-dorsey-chief-executive>.

Pelley, Scott. “Facebook whistleblower Frances Haugen details company's misleading efforts on 60 Minutes.” *CBS News*, 4 October 2021, <https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-misinformation-public-60-minutes-2021-10-03/>.

Rosen, Guy. “Our Response to the Violence in Washington.” *Meta*, 6 Jan 2021, <https://about.fb.com/news/2021/01/responding-to-the-violence-in-washington-dc/>.

“Rulemaking Process.” Federal Communications Commission, <https://www.fcc.gov/about-fcc/rulemaking-process>.

“Sandy Hook families v Alex Jones: Defamation case explained.” *BBC.com*, 2018, <https://www.bbc.com/news/world-us-canada-45358890>.

“Schenck v. United States.” *Oyez*, www.oyez.org/cases/1900-1940/249us47.

“Social Media and News Fact Sheet, 2024.” *Pew Research Center*, 17 September 2024, <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>.

“(Social) Media and Politics and the Arab Spring Moment.” *European Institute of the Mediterranean*, <https://www.iemed.org/publication/social-media-and-politics-and-the-arab-spring-moment/>.

“Social Media: Content Dissemination and Moderation Practices | Congress.gov | Library of Congress.” *Congress.gov*, 20 March 2025, <https://www.congress.gov/crs-product/R46662>.

“US Dominion, Inc. v. Fox News Network, LLC.” Casetext, 16 December 2021, <https://casetext.com/case/us-dominion-inc-v-fox-news-network-llc-1>
<https://www.pbs.org/newshour/politics/fox-dominion-voting-systems-reach-settlement-over-false-election-claims>.

U.S. Government. “Communications Act of 1934: as amended by Telecom Act of 1996.” *Federal Communications Commission*, 19 June 1934, <https://transition.fcc.gov/Reports/1934new.pdf>.

Villasenor, John. “Social media companies and common carrier status: a primer.” *Brookings*, 27 October 2022, <https://www.brookings.edu/articles/social-media-companies-and-common-carrier-status-a-primer/>.

- Volokh, Eugene. "Social Media Platforms as Common Carriers?" *Reason Magazine*, 5 July 2021, <https://reason.com/volokh/2021/07/05/social-media-platforms-as-common-carriers-2/>.
- Volokh, Eugene. "Treating Social Media Platforms Like Common Carriers?" *eScholarship*, 28 May 2021, <https://www.journaloffreespeechlaw.org/volokh.pdf>.
- Wirtschafter, Valerie, and Sharanya Majumder. "Future Challenges for Online, Crowdsourced Content Moderation: Evidence from Twitter's Community Notes." *Journal of Online Trust and Safety*, 2023, <https://www.tsjournal.org/index.php/jots/article/view/139/57>.
- X. "Permanent suspension of @realDonaldTrump." *X*, 8 January 2021, https://blog.x.com/en_us/topics/company/2020/suspension.
- Zuckerberg, Mark. "Excited to announce that we're launching Meta AI..." Facebook, 18 Apr. 2024, <https://www.facebook.com/zuck/posts/10112681480907401>.
- Zuckerberg, Mark. *Hearing Before the United States House of Representatives Committee on Energy and Commerce Subcommittees on Consumer Protection*, 25 March 2021, <https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-U1.pdf>.