

CONNECTICUT LAW REVIEW

VOLUME 55

JANUARY 2023

ONLINE EDITION

Note

Mitigating Misinformation on Social Media Platforms: Treating Section 230 of the Communications Decency Act as a Quid Pro Quo Benefit

MEGHAN E. McDERMOTT

The rise of misinformation on social media has prompted governments worldwide to enact legislation that may affect every person's right to freedom of opinion and expression. In the United States, combatting misinformation shares surprising bipartisan support in an ever-divided political landscape. While several proposals have emerged that would strip social media companies of the twenty-five-year-old law that shields them from lawsuits over content, it is unlikely that they would survive the seemingly insurmountable First Amendment scrutiny. Thus, an alternative to combatting misinformation is needed.

In an attempt to provide an alternative, this Note presents a model for mitigating misinformation. By dissecting the concept of misinformation, exploring the ways in which social media platforms can mitigate misinformation, and proposing the use of misinformation labels, this Note suggests using Section 230 of the Communications Decency Act as a quid pro quo benefit to incentivize social media platforms to mitigate misinformation.

NOTE CONTENTS

INTRODUCTION	3
I. MISINFORMATION & SOCIAL MEDIA PLATFORMS	4
A. DEFINING MISINFORMATION	5
B. THE RAGING RISE OF MISINFORMATION ON SOCIAL MEDIA PLATFORMS	5
II. LABELING MISINFORMATION	9
A. CURRENT MISINFORMATION LABELING EFFORTS BY SOCIAL MEDIA PLATFORMS	10
B. CATEGORIZING & LABELING TYPES OF MISINFORMATION WITH ALGORITHMS	12
III. SECTION 230 OF THE COMMUNICATIONS DECENCY ACT	15
A. A BRIEF OVERVIEW OF SECTION 230	16
B. FEDERAL PROPOSALS TO AMEND SECTION 230	18
IV. FCC REGULATION OF SOCIAL MEDIA MISINFORMATION	21
A. CREATING & ENFORCING MISINFORMATION LABELING STANDARDS THROUGH THE FCC	21
B. ENFORCING MISINFORMATION LABELING: SECTION 230 AS A QUID PRO QUO BENEFIT	23
C. THE CONSTITUTIONALITY OF ENFORCING MISINFORMATION LABELING	24
CONCLUSION	25



Mitigating Misinformation on Social Media Platforms: Treating Section 230 of the Communications Decency Act as a Quid Pro Quo Benefit

MEGHAN E. McDERMOTT *

INTRODUCTION

Social media has become an indispensable tool in today's society. Platforms such as Facebook, Twitter, and Instagram have allowed users to connect, share ideas, and receive news in remarkable ways. Still, the ease of communication these platforms provide has opened the door for the rampant abuse and manipulation of information. As a result, social media platforms have become breeding grounds for spreading misinformation.

The problem of misinformation was highlighted in 2016 when evidence showed that Russia was promoting falsehoods on social media platforms (like Facebook and Twitter) via internet bots and other means.¹ Since then, events such as the 2020 presidential election and the COVID-19 pandemic have emphasized this growing problem, ultimately forcing social media platforms to increase content moderation to ensure that the public is informed—and not misinformed—about these types of matters.²

While social media platforms engage in content moderation, the criteria for determining what constitutes harmful content is unclear to both users and employees tasked with removing it.³ The lack of required transparency has given social media platforms the flexibility to remove content as it suits them, usually in a way that maximizes their profits.⁴ In addition, the failing efforts to adequately mitigate misinformation have inspired little confidence in the ability of social media platforms to solve the problem independently. Thus, legislators, legal scholars, and the general public have requested a

* Author Bio

¹ STAFF OF S. COMM. ON INTEL., 116TH CONG., REP. ON RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE IN THE 2016 U.S. ELECTION 84–90 (Comm. Print 2020).

² Kari Paul, *Here Are All The Steps Social Media Made To Combat Misinformation. Will It Be Enough?*, THE GUARDIAN (Oct. 30, 2020, 1:00 AM), <https://www.theguardian.com/technology/2020/oct/29/here-are-all-the-steps-social-media-made-to-combat-misinformation-will-it-be-enough>.

³ See JASON A. GALLO & CLARE Y. CHO, CONG. RSCH. SERV., R46662, SOCIAL MEDIA: MISINFORMATION AND CONTENT MODERATION ISSUES FOR CONGRESS 1 (2021).

⁴ Sara Brown, *Social Media is Broken. A New Report Offers 25 Ways to Fix It*, MIT SLOAN SCH. MGMT. (June 30, 2021), <https://mitsloan.mit.edu/ideas-made-to-matter/social-media-broken-a-new-report-offers-25-ways-to-fix-it>.

more aggressive approach to content moderation.⁵

Although the misinformation surrounding the COVID-19 pandemic and the 2020 presidential election results have made clear the need for some level of content regulation for social platforms, finding a constitutional means of regulating misinformation on social media is no easy task. In 2021 and 2022, numerous legislative bills regarding social media regulation were introduced in Congress, most suggesting substantial amendments to Section 230 of the Communications Decency Act (“CDA”).⁶ Section 230 says interactive computer services—like social media platforms—are not responsible for the content (including misinformation) posted by their users.⁷ Amending Section 230 would likely reduce the spread of misinformation, but it would also have the practical effect of suppressing harmless, and even socially beneficial, discourse on social media.⁸ It is now a bipartisan policy goal to create a national regulatory approach to protect users, and the public, from abuse by the powerful social media companies that have become such a central part of the nation’s public life.⁹ The question left unanswered is: How can this be done?

In an attempt to answer this question, this Note presents a model for mitigating misinformation. Part I will define misinformation and address how the rise of social media has increased the presence of misinformation in today’s society. Part II will provide examples of how social media platforms currently label misinformation and how effective the labels are in mitigating misinformation. Part III will briefly overview Section 230 and discuss why the most recent legislative bills introduced to Congress seeking to amend Section 230 face constitutional challenges. Finally, Part IV will recommend regulating social media misinformation by leveraging Section 230 as a quid pro quo benefit to compel social media platforms to follow such guidelines.

I. MISINFORMATION & SOCIAL MEDIA PLATFORMS

Given the rise in misinformation on social media platforms over the years, it seems crucial for some type of outside regulation of such information. But, perhaps even more important, we need to identify how to use Section 230 to promote open communication channels while mitigating the spread of misinformation. To do so, Part I of this paper first defines

⁵ Matthew Feeney & Will Duffield, *A Year of Content Moderation and Section 230*, CATO INST. (Nov. 2, 2020, 3:14 PM), <https://www.cato.org/blog/year-content-moderation-section-230>.

⁶ See, e.g., Limiting Section 230 Immunity to Good Samaritans Act, H.R. 277, 117th Cong. (2021) (providing that Big Tech companies who want to receive Section 230 immunity must bind themselves contractually to a duty of good faith); Protect Speech Act, H.R. 3827, 117th Cong. (2021) (looking to narrow a platform’s ability to use Section 230 as a defense for content removal).

⁷ 47 U.S.C. § 230(c)(1).

⁸ Ryan Tracy, *Social Media’s Liability Shield Is Under Assault*, WALL ST. J. (Nov. 26, 2020, 10:00 AM), <https://www.wsj.com/articles/social-medias-liability-shield-is-under-assault-11606402800>.

⁹ GALLO & CHO, *supra* note 3.

misinformation and proceeds to discuss its popularity on social media platforms. The goal is to better understand a complex issue impacting all social media users and highlight misinformation's dangerous effects on society.

A. *Defining Misinformation*

Misinformation refers to false or out-of-context information presented as fact without an intent to deceive.¹⁰ As a result, people unknowingly spread misinformation, believing their sharing helps inform others.¹¹ For example, PizzaGate, a conspiracy theory about a pizzeria being involved in a child-trafficking ring run by Hilary Clinton, started rapidly spreading on multiple social media sites such as Facebook, Twitter, and Reddit.¹² Many users who shared, re-tweeted, or up-voted posts about the conspiracy did it because they thought it was true and wanted to inform friends and family about the scandal.¹³ However, promoting the PizzaGate conspiracy theory led a man to fire a rifle inside the alleged child-trafficking ring, which was actually just a pizza restaurant.¹⁴ The different types of misinformation found on social media, like the PizzaGate conspiracy theory, show how dangerous misinformation can genuinely be.

This paper will employ the term misinformation as an umbrella term to include all inaccurate or false information disseminated on social media platforms, including disinformation and misinformation. The various concepts covered in the umbrella term, such as disinformation, rumors, and fake news, all send some sort of inaccurate messaging that can cause a multitude of destructive effects through social media, especially when timely intervention is absent.

B. *The Raging Rise of Misinformation on Social Media Platforms*

Although the problem of misinformation dissemination has been around for centuries, the rise of social media platforms exacerbates the issue. The audience of social media users is massive. In 2020, seventy-one percent of U.S. adults got their news from at least one social media platform.¹⁵ Forty-eight percent of those adults say they get news from social media “often” or

¹⁰ CLAIRE WARDLE, FIRST DRAFT'S ESSENTIAL GUIDE TO: UNDERSTANDING INFORMATION DISORDER 8 (2019).

¹¹ *Id.*

¹² Eric Lipton, *Man Motivated by 'Pizzagate' Conspiracy Theory Arrested in Washington Gunfire*, N.Y. TIMES (Dec. 5, 2016), <https://www.nytimes.com/2016/12/05/us/pizzagate-comet-ping-pong-edgar-maddison-welch.html>; Amanda Robb, *Anatomy of a Fake News Scandal*, ROLLING STONE (Nov. 16, 2017, 3:07 PM), <https://www.rollingstone.com/feature/anatomy-of-a-fake-news-scandal-125877/>.

¹³ Robb, *supra* note 12.

¹⁴ Lipton, *supra* note 12.

¹⁵ MASON WALKER & KATERINA EVA MATSA, NEWS CONSUMPTION ACROSS SOCIAL MEDIA IN 2021, PEW RSCH. CTR. 3 (2021), <https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/>.

“sometimes.”¹⁶ And while two-thirds of those surveyed across all ages who are news consumers say they do not trust news on social media platforms,¹⁷ this lack of trust does not seem to affect their news consumption behavior. About fifty-three percent of U.S. adults get their news from social media platforms.¹⁸ Compared with some older news consumers, Gen Zs and millennials are more likely to identify social media platforms as one of their preferred methods to stay updated on news and current events.¹⁹ At the same time, they are less likely to express distrust in the news on social media platforms.²⁰ Yet, regardless of age, an analysis of over 4.5 million social media posts found that false news stories were seventy percent more likely to be shared than accurate news stories.²¹

The 2020 U.S. presidential election and the Coronavirus Disease 2019 (“COVID-19”) pandemic illustrate social media platforms’ contribution to the spread of misinformation. The 2020 presidential election was comprised of intimidating and misleading misinformation. As millions of Americans cast ballots in a chaotic and contentious presidential election, the volume of misinformation seen online was jaw-dropping.²² The Stopping Cyber Suppression program run by Common Cause—a grassroots group promoting democracy—identified close to 5,000 incidents of election misinformation between January and October of 2020.²³ Much of the election misinformation discovered by the Stopping Cyber Suppression program originated from a pro-Trump youth organization, Turning Point USA.²⁴ Turning Point USA employed “teenagers to disseminate misinformation, including attempts to discredit mail-in ballots” and

¹⁶ *Id.*

¹⁷ Brooke Auxier & Jana Arbanas, *Majority of News Consumers See “Fake News” as a Big Problem Today*, DELOITTE INSIGHTS (June 10, 2021), <https://www2.deloitte.com/us/en/insights/industry/technology/study-shows-news-consumers-consider-fake-news-a-big-problem.html>.

¹⁸ Elisa Shearer, *More Than Eight-in-Ten Americans Get News From Digital Devices*, PEW RSCH. CTR. (Jan. 12, 2021), <https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/>.

¹⁹ KEVIN WESTSCOTT ET AL., DIGITAL MEDIA TRENDS, 15TH EDITION, DELOITTE INSIGHTS 13 (2021), https://www2.deloitte.com/us/en/insights/industry/technology/digital-media-trends-consumption-habits-survey/summary-2021.html?icid=learn_more_content_click.

²⁰ Auxier & Arbanas, *supra* note 17.

²¹ Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCIENCE 1146, 1146 (2018).

²² Pam Fessler, *Robocalls, Rumors and Emails: Last-Minute Election Disinformation Floods Voters*, NAT’L PUB. RADIO (Oct. 24, 2020, 7:00 AM), <https://www.npr.org/2020/10/24/927300432/robocalls-rumors-and-emails-last-minute-election-disinformation-floods-voters>.

²³ *Id.*

²⁴ David Vance, *Turning Point USA Troll Farm Points to Need for Social Media Platforms to Strengthen Disinformation Protections*, COMMON CAUSE (Sept. 16, 2020, 1:12 PM), <https://www.commoncause.org/press-release/turning-point-usa-troll-farm-points-to-need-social-media-platforms-to-strengthen-disinformation/>.

downplay the health risks of COVID-19.²⁵ These attempts were all part of a “sprawling, yet secretive, campaign that experts say evades the guardrails put in place by social media companies to limit online disinformation of the sort used by Russia during the 2016 campaign.”²⁶

Such discoveries drove platforms such as Twitter, Facebook, and YouTube to take measures to counter misinformation. Some of these measures included adding labels to President Trump's misleading posts, notifying users that there was no immediate outcome to the presidential race, and changing algorithmic recommendations to limit users' ability to share falsehoods.²⁷ But since the 2020 election season, the misinformation tracking network for Common Cause has documented a dangerous trend; social media giants Facebook and Twitter appear to have dropped the ball on their commitments to police election misinformation.²⁸ Only seven months after President Biden was inaugurated, widespread propagation of false election claims continued to inundate social media platforms.²⁹ Yet, despite the ongoing spread of misinformation related to the 2020 election, social media platforms shied away from policing voting and election misinformation.³⁰ Facebook, in particular, banned political advertising after the 2020 presidential election, but has since lifted it.³¹

Social media platforms also failed to mitigate misinformation on the COVID-19 pandemic effectively. Similar to efforts taken during the 2020 presidential election, social media platforms implemented content moderation strategies, such as tagging or removing what they considered to be misinformation while promoting what they deemed reliable sources of information.³² For example, according to the VP of Global Affairs and Communications of Meta—formerly known as Facebook—Facebook and Instagram began removing COVID-19-related posts that made false claims

²⁵ *Id.*

²⁶ *Id.*

²⁷ Matt O'Brien & Mae Anderson, *Did Social Media Actually Counter Election Misinformation?*, ASSOCIATED PRESS NEWS (Nov. 4, 2020), <https://apnews.com/article/social-media-election-misinformation-632a5d93a6cc3ff37311a641d86bf5a1>.

²⁸ JESSE LITTLEWOOD & EMMA STEINER, TRENDING IN THE WRONG DIRECTION: SOCIAL MEDIA PLATFORMS' DECLINING ENFORCEMENT OF VOTING DISINFORMATION, COMMON CAUSE 1 (2021), https://www.commoncause.org/wp-content/uploads/2021/09/Disinfo_WhitePaper3.pdf. (“Although no one outside of the social media platforms themselves has access to enough data to say for certain how much effort platforms put into enforcement, there are many examples of social media posts generating high engagement on provably false claims similar to posts that were labeled or removed during the presidential election. What makes this even more problematic is that it is not just a failure of automated moderation: posts with false claims have been reported to the platforms as misinformation, but no action has been taken.”).

²⁹ *Id.* at 2.

³⁰ *Id.*

³¹ Samantha Lai, *Data Misuse and Disinformation: Technology and the 2022 Elections*, BROOKINGS (June 21, 2022), <https://www.brookings.edu/blog/techtank/2022/06/21/data-misuse-and-disinformation-technology-and-the-2022-elections/>.

³² GALLO & CHO, *supra* note 3, at 16–17.

about cures, treatments, the availability of essential services, or the location and severity of the outbreak in January 2020.³³ Efforts to reduce COVID-19 misinformation also included removing claims that physical distancing does not help prevent the spread of the coronavirus and banning ads that imply a product guarantees a cure or prevents people from contracting COVID-19.³⁴

However, claims that do not directly result in physical harm, such as conspiracy theories about the virus's origin, are not automatically removed from the platforms.³⁵ Instead, each claim undergoes a fact-checking process. Once a post is rated false by a fact-checker, the platforms reduce the post's distribution and show strong warning labels and notifications to people who still come across it, try to share it, or have already done so.³⁶ Nevertheless, an August 2020 report emphasized the shortcomings of this approach, revealing that out of 174 fact-checked COVID-19 health misinformation posts analyzed, only sixteen percent had warning labels.³⁷ And once the COVID-19 vaccine was released, misinformation surrounding the pandemic only increased.

As the U.S. Government began administering COVID-19 vaccines across the nation, false statements about the vaccine skyrocketed. Comments such as "[t]he Covid-19 vaccines will make you infertile," "Covid-19 vaccines don't work," and "[t]he government put a microchip in the Covid-19 vaccines"³⁸ began to pop up on social media. Despite all these statements being false, they have continued to spread like wildfire on social media platforms, causing confusion and leading people to decline COVID-19 vaccines; reject public health measures, such as masking and physical distancing; and use unproven treatments.³⁹ This failure is causing significant harm and undermining U.S. COVID-19 response and recovery efforts.⁴⁰

All the misinformation spread throughout 2020 and 2021 ultimately led to Congress holding numerous hearings to examine social media platforms' role in disseminating misinformation.⁴¹ Members of Congress introduced legislation to amend Section 230 of the CDA, which could affect the content moderation practices of interactive computer services, including social media platforms.⁴² Even the U.S. Department of Justice sent Congress a

³³ Nick Clegg, *Combating COVID-19 Misinformation Across Our Apps 5–6*, META (Mar. 25, 2020), <https://about.fb.com/news/2020/03/combating-covid-19-misinformation/>.

³⁴ *Id.*

³⁵ *Id.*

³⁶ *Id.*

³⁷ *Facebook's Algorithm: A Major Threat to Public Health*, AVAAZ (Aug. 19, 2020), https://avaazimages.avaaz.org/facebook_threat_health.pdf.

³⁸ Estelle Willie, *Misinformation Is the Biggest Threat to Ending This Pandemic*, ROCKEFELLER FOUND. (July 27, 2021), <https://www.rockefellerfoundation.org/blog/misinformation-is-the-biggest-threat-to-ending-this-pandemic/>.

³⁹ *Id.*

⁴⁰ *Id.*

⁴¹ GALLO & CHO, *supra* note 3, at 1.

⁴² *Id.* at 19.

review of Section 230 in 2020 identifying and outlining proposed reforms to the current legislation.⁴³

While some commentators identify potential benefits of amending Section 230, others have identified possible adverse consequences; thus, leaving Section 230 untouched to date.⁴⁴ Luckily, Section 230 may not have to be amended to address misinformation on social media platforms. Rather, Section 230 can be used as a quid pro quo benefit to compel social media platforms to mitigate the spread of misinformation by using misinformation labeling. For this reason, the following section will discuss current efforts made by social media platforms to reduce misinformation through labeling.

II. LABELING MISINFORMATION

Social media platforms maintain and enforce policies that ban users from posting certain content.⁴⁵ The platforms “may temporarily or permanently ban users that violate its policies, depending on the operator’s perspective on the severity of the users’ violation(s).”⁴⁶ But since there is no uniform standard for content moderation, misinformation moderation varies across social media sites.⁴⁷

Over the past five years, leading social media platforms, Twitter and Facebook, have changed their respective content moderation practices several times.⁴⁸ Notably, both platforms have altered their practices by taking a harder line with moderating misinformation.⁴⁹ However, research shows that labeling content works, and social media platforms have already started to label misinformation.⁵⁰ The attempts made by Twitter and

⁴³ See generally U.S. DEP’T JUST., SECTION 230 – NURTURING INNOVATION OR FOSTERING UNACCOUNTABILITY? (2020) (reviewing Section 230 of the CDA and providing recommendations for reform).

⁴⁴ Compare Lauren Rundall, Comment, *Don’t Break the Internet: § 230 and Its Role Within Today’s Modern Internet Era*, 5 BUS., ENTREPRENEURSHIP & TAX L. REV. 50, 52 (2021) (contending that the benefits of Section 230 outweigh the negative results its shield creates), with Natalie Annette Pagano, Comment, *The Indecency of the Communications Decency Act § 230: Unjust Immunity for Monstrous Social Media Platforms*, 39 PACE L. REV. 511, 537 (2018) (stating that Section 230 is insufficient and inadequate as it is currently drafted).

⁴⁵ Facebook Community Standards, META: TRANSPARENCY CTR., <https://transparency.fb.com/policies/community-standards/> (last visited Mar. 18, 2022); *The Twitter Rules*, TWITTER: HELP CTR., <https://help.twitter.com/en/rules-and-policies/twitter-rules> (last visited Mar. 18, 2022).

⁴⁶ GALLO & CHO, *supra* note 3, at 6.

⁴⁷ *Id.*

⁴⁸ See *Social Media’s Struggle with Self-Censorship*, THE ECONOMIST (Oct. 22, 2020), <https://www.economist.com/briefing/2020/10/22/social-medias-struggle-with-self-censorship> (discussing evolving methods of moderation).

⁴⁹ *Id.*

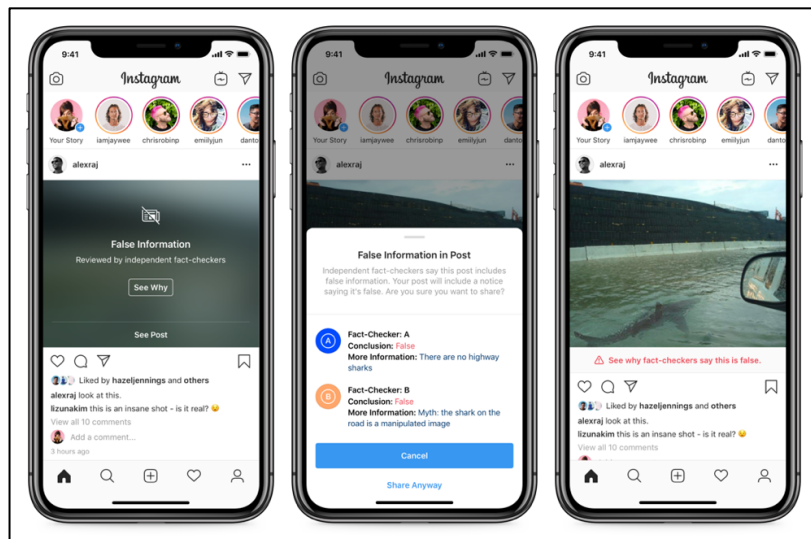
⁵⁰ See *Combating Misinformation on Instagram*, META: NEWSROOM (Dec. 16, 2019) [hereinafter *Combating Misinformation*], <https://about.fb.com/news/2019/12/combating-misinformation-on-instagram/> (discussing how Meta labels Instagram posts).

Facebook illustrate that moderating misinformation is possible—but labeling misinformation is how it will be done. Thus, Part II of this paper will highlight how social media platforms currently label misinformation.

A. Current Misinformation Labeling Efforts by Social Media Platforms

According to misinformation experts, the difficulty that average social media users have in identifying misinformation is one of the main reasons it spreads quickly.⁵¹ People who post misinformation are becoming so good at disguising misinformation that most people cannot tell the real news apart from the fake news. Thus, labels that identify the type of misinformation that a post spreads would be a substantial step in regulating misinformation on social media platforms.⁵²

This idea is not out of the ordinary. For example, in 2019, Facebook (now Meta) announced that it would begin to label misinformation on Instagram.⁵³ Although it would only label false information instead of all types of misinformation, the company provided its users with examples of how the social media platform identifies false information.⁵⁴



⁵¹ Peter Suci, *Spotting Misinformation On Social Media Is Increasingly Challenging*, FORBES (Aug. 2, 2021, 3:56 PM), <https://www.forbes.com/sites/petersuci/2021/08/02/spotting-misinformation-on-social-media-is-increasingly-challenging/?sh=8085d962771c>.

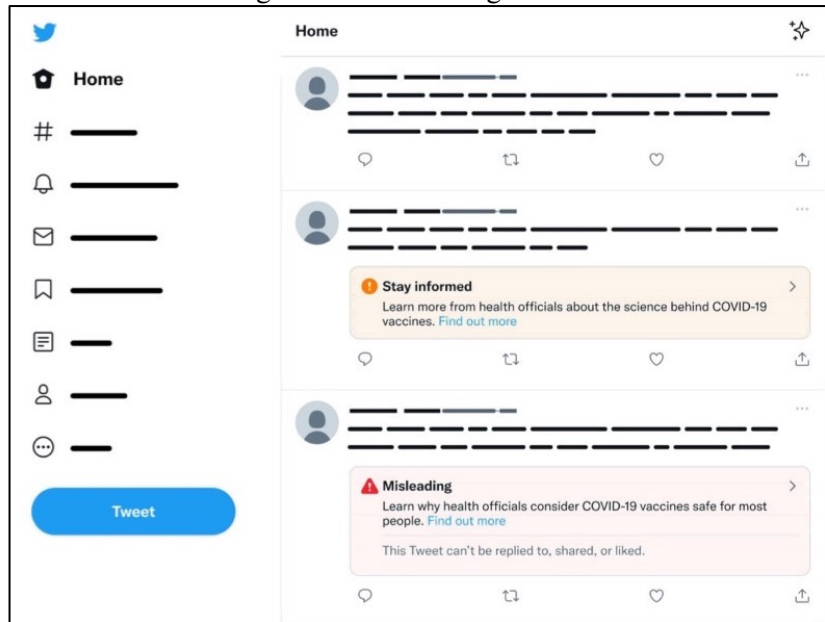
⁵² Emily Saltz et al., *It Matters How Platforms Label Manipulated Media. Here Are 12 Principles Designers Should Follow*, MEDIUM (June 9, 2020), <https://medium.com/swlh/it-matters-how-platforms-label-manipulated-media-here-are-12-principles-designers-should-follow-438b76546078>.

⁵³ *Combating Misinformation*, *supra* note 50.

⁵⁴ *Id.*

Figure 1: Instagram's Example of Labeling Manipulated Media⁵⁵

Following Facebook's lead, Twitter implemented a similar approach to identifying misinformation for users.⁵⁶ Thus far, Twitter's misinformation warning labels have shown "a seventeen percent increase in 'click-through-rate,' which means that more people clicked on the redesigned labels to read the information debunking false or misleading tweets."⁵⁷

Figure 2: Twitter's Example of Labeling Manipulated Media⁵⁸

Twitter's new misinformation warning labels are a helpful example of how social media platforms should warn users about misinformation. But Twitter did not just stop at misinformation labels. In January 2021, Twitter

⁵⁵ *Id.* This photo provided by Meta shows a screen with the label "false information" on Instagram. The labels obscure posts that have been debunked by Facebook's fact checkers and appear on posts in Stories and Instagram's main feed. Users will still be able to view the original post, but they will have to click "See Post" to get there.

⁵⁶ Yoel Roth & Nick Pickles, *Updating Our Approach to Misleading Information*, TWITTER: BLOG (May 11, 2020), https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.

⁵⁷ Barbara Ortutay, *Twitter Rolls Out Redesigned Misinformation Warning Labels*, ASSOCIATED PRESS (Nov. 16, 2021), <https://apnews.com/article/technology-business-media-social-media-misinformation-ac496a53fbc761146627fa534cb2f8d9>.

⁵⁸ Barbara Ortutay, *What's in a Tag? Twitter Revamps Misinformation Labels*, ASSOCIATED PRESS (July 1, 2021), <https://apnews.com/article/health-coronavirus-pandemic-election-2020-misinformation-technology-37cee761758f1072f91e1a362f769f5e> ("This photo provided by Twitter shows a screen that [displays] labels warning about misinformation."). The labels shown are provided as an example of how Twitter is redesigning its misinformation labels in an attempt to make them more useful and easier to notice, among other things.

introduced “Birdwatch,” a pilot program in the U.S. that uses “a new community-driven approach to help address misleading information on Twitter.”⁵⁹ “Birdwatch allows people to identify information in Tweets they believe is misleading and write notes that provide informative context.”⁶⁰ The program seeks to enlist Twitter’s users to flag and debunk misinformation on the social platform.⁶¹ Since Birdwatch’s launch, Twitter has “found that users are twenty to forty percent less likely to agree with a tweet’s claim after reading a Birdwatch note about it.”⁶² The social media platform “has also found that most of its user-submitted notes get solid marks for accuracy from professional fact-checkers.”⁶³

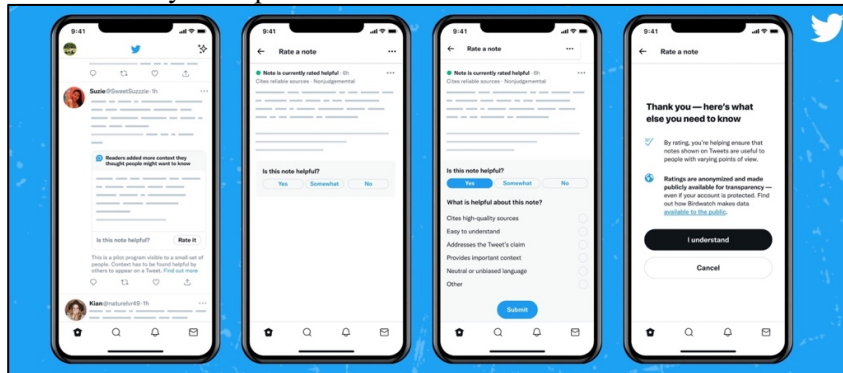


Figure 3: Birdwatch Misinformation Labeling on Twitter⁶⁴

While these attempts to minimize misinformation on social media platforms are steps in the right direction, social media platforms need to be incentivized to take the necessary steps to mitigate misinformation.

B. *Categorizing & Labeling Types of Misinformation with Algorithms*

Social media platforms rely on algorithms to organize content. These algorithms sort, index, curate, and prioritize user content, and suppress illegal and other content that social media platforms choose to moderate.⁶⁵ In addition, social media algorithms determine which content to deliver to

⁵⁹ Keith Coleman, *Introducing Birdwatch, a Community-Based Approach to Misinformation*, TWITTER: BLOG (Jan. 25, 2021), https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation.

⁶⁰ *Id.*

⁶¹ *Id.*

⁶² Will Oremus, *To Fight Misinformation, Twitter Expands Project to Let Users Fact-Check Each Other's Tweets*, WASH. POST (Mar. 3, 2022, 9:00 AM), <https://www.washingtonpost.com/technology/2022/03/03/twitter-birdwatch-fact-check-misinfo-test/>.

⁶³ *Id.*

⁶⁴ Keith Coleman, *Building a Better Birdwatch*, TWITTER: BLOG (Mar. 3, 2022), https://blog.twitter.com/en_us/topics/company/2022/building-a-better-birdwatch.

⁶⁵ Maria Alessandra Golino, *Algorithms in Social Media Platforms*, INST. FOR INTERNET & JUST. SOC'Y (Apr. 24, 2021), <https://www.internetjustsociety.org/algorithms-in-social-media-platforms>.

users based on their behavior.⁶⁶ Without these algorithms, social media platform regulation would be conducted by “sifting through all the content on an account-by-account basis.”⁶⁷ This type of regulation would be impractical, especially for platforms with hundreds or thousands of profiles, pages, and users.⁶⁸

Platforms like Facebook, Twitter, and YouTube “use algorithms to detect content suspected of violating their community standards.”⁶⁹ The algorithms “flag the questionable posts and refer them to human content moderators for evaluation.”⁷⁰ So, if social media platforms can design their algorithms to promote content that makes them the biggest profits, they can undoubtedly create better algorithms for labeling or flagging misinformation. For example, social media platforms can implement algorithms with “circuit breakers” that slow down the viral propagation of information on their network.⁷¹ Ultimately, a circuit breaker would be an imposition in the form of a hold; a hold would be placed on sharing information among network members once a certain number of shares had been reached, or if sharing was happening too fast.⁷² This method would not remove a story but would merely take the story out of the algorithms’ control and give that control to a human or, perhaps, a different set of algorithms with different objectives.⁷³

For example, and not as a recommendation, Facebook could implement an algorithm with circuit breakers for posts related to the COVID-19 pandemic. The algorithm would stop sharing the post to review its context or fact-check its information. Whether it be a human fact-checking it or another algorithm, the possibly misleading post would not spread for the time being. After the fact-checking, the post could go in two directions: (1) the information is accurate, and sharing can resume, or (2) the post contains misinformation and needs to be labeled accordingly.

In fact, Twitter has successfully implemented an algorithm to monitor misinformation on its platform. Currently, Twitter labels three types of misinformation: (1) “‘manipulated media,’ such as videos and audio that have been deceptively altered in ways that could cause real-world harm[.]” (2) “election and voting-related misinformation[.]” and (3) “false or

⁶⁶ *Id.*

⁶⁷ Brent Barnhart, *Everything You Need to Know About Social Media Algorithms*, SPROUT SOC. (Mar. 26, 2021), <https://sproutsocial.com/insights/social-media-algorithms/>.

⁶⁸ *Id.*

⁶⁹ Nina I. Brown, *Regulatory Goldilocks: Finding the Just and Right Fit for Content Moderation on Social Platforms*, 8 TEX. A&M L. REV. 451, 477 (2021).

⁷⁰ *Id.*


⁷¹ Shannon Bond, *Can Circuit Breakers Stop Viral Rumors on Facebook, Twitter?*, NAT’L PUB. RADIO (Sep. 22, 2020, 6:36 PM), <https://www.npr.org/2020/09/22/915676948/can-circuit-breakers-stop-viral-rumors-on-facebook-twitter>.

⁷² *Id.*

⁷³ *Id.* Ortutay, *Twitter Rolls Out Redesigned Misinformation Warning Labels*, *supra* note 57.

misleading tweets related to COVID-19.”⁷⁴ However, using an algorithm curated to identify and clearly label these three types of misinformation, Twitter has effectively increased the number of interactions users have with its misinformation warnings.⁷⁵

Twitter’s algorithm identifies tweets that fall under the types of misinformation listed above and then takes action based on three broad categories.⁷⁶ The first category, “Misleading [I]nformation,” includes “statements or assertions that have been confirmed to be false or misleading by subject-matter experts, such as public health authorities.”⁷⁷ The second category, “Disputed [C]laims,” includes “statements or assertions in which the accuracy, truthfulness, or credibility of the claim is contested or unknown.”⁷⁸ Lastly, the third category, “Unverified [C]laims,” includes “information (which could be true or false) that is unconfirmed at the time it is shared.”⁷⁹ Twitter’s algorithm uses the chart below to determine how a post should be labeled using these categories.



Misleading Information	Label	Removal
Disputed Claim	Label	Warning
Unverified Claim	No action	No action*
	Moderate	Severe
Propensity for Harm		

**We will continue to introduce new labels to provide context around different types of unverified claims and rumors as needed.*

Figure 4: How Twitter’s Manipulated Media Algorithm Labels Misinformation⁸⁰

⁷⁴ *Id.*

⁷⁵ *Id.*

⁷⁶ Roth & Pickles, *supra* note 56.

⁷⁷ *Id.*

⁷⁸ *Id.*

⁷⁹ *Id.*

⁸⁰ *Id.*

As the table shows, warnings may be applied to a Tweet depending on the propensity for harm and type of misinformation.⁸¹ And while Twitter only labels manipulated media misinformation, election and voting misinformation, and COVID-19 misinformation, the increased user interactions with the labels illustrate how effective algorithms can be and how impactful misinformation warning labels on social media platforms are. Thus, rather than amending Section 230 of the CDA, the government can compel social media platforms to label misinformation by leveraging Section 230 as a quid pro quo benefit.

III. SECTION 230 OF THE COMMUNICATIONS DECENCY ACT

Part I aimed to conceptualize misinformation and its current landscape in today's technology-filled society. It then outlined the complexity of misinformation and its influence on social media platforms. Specifically, this Note has examined misinformation's dangerous effects on society and the complex issue it creates for all social media users. Part II then highlighted an approach to reducing misinformation on social media platforms through labeling. While labeling does seem to be a promising method of mitigating misinformation, it seems crucial for some type of outside regulation to enforce the use of such labels. Perhaps, what may be even more important, is identifying how to use Section 230 to promote mitigating the spread of misinformation through labeling.

Section 230 of the Communications Decency Act spells out who is legally responsible for content on the internet. By providing a shield for social media platforms that host user-generated content, Section 230 has helped companies like Facebook and Twitter flourish.⁸² Regrettably, this liability protection also covers misinformation shared on these platforms by shielding social media platforms from liability for content created and shared by their users.⁸³ It permits websites, including social media platforms, to moderate user content without being liable for the content they host.⁸⁴ As social media platforms continue to be scrutinized by lawmakers over a number of issues,⁸⁵ one issue has been particularly popular—how to handle the spread of misinformation on social media platforms. To address the relationship between Section 230 and social media misinformation, this Part discusses current proposals to amend the consequential law. For that, a brief review of Section 230 is in order, so that we can better understand the impact of amending it.

⁸¹ *Id.*

⁸² Daisuke Wakabayashi, *Legal Shield for Social Media Is Targeted by Lawmakers*, N.Y. TIMES (Dec. 15, 2020), <https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html>.

⁸³ *Id.*

⁸⁴ *Id.*

⁸⁵ *Id.*

A. A Brief Overview of Section 230

Section 230 of the Communications Decency Act provides limited federal immunity to providers and users of interactive computer services.⁸⁶ Ordinarily, publishers are liable for what they publish. Yet, Section 230 precludes providers and users from being held liable for information provided by a third party.⁸⁷ By enacting Section 230, Congress intended to foster internet growth and development by minimizing governmental regulation and empowering internet companies to monitor objectionable content without the fear of incurring liability.⁸⁸ Courts have interpreted Section 230 in a way that prevents providers from being liable for third-party content.⁸⁹ For example, Section 230 has protected social media platforms from lawsuits that arise from their taking down of user-generated content.⁹⁰ Unfortunately, this application is exactly why misinformation distribution by social media platforms such as Facebook, Twitter, and YouTube occurs.

Section 230(c)(1) of the CDA states, “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”⁹¹ Section 230 sought to overturn *Stratton Oakmont v. Prodigy Services Co.*, where the court treated defamatory online comments as if they were in print, and held Prodigy liable because it moderated the online content.⁹² *Stratton Oakmont* raised concerns that platforms would be unsustainable if exposed to liability for the acts of any individual user and would be deterred from taking proactive action to filter for offensive content.⁹³ Hoping to incentivize voluntary content moderation, Congress designed Section 230 to encourage the free exchange of ideas and information over the internet while still fostering voluntary monitoring by internet service providers (“ISPs”).⁹⁴

Although the legislative history of Section 230 reflects an intent to overturn the result in *Stratton Oakmont* and incentivize content moderation, courts have applied Section 230(c)(1) broadly. The first federal court of appeals decision to examine the scope of Section 230(c)(1) was the Fourth

⁸⁶ 47 U.S.C. § 230(c)(1).

⁸⁷ *Id.*

⁸⁸ *See Id.* at § 230(b)(1–2) (outlining policy reasons for passing § 230).

⁸⁹ VALERIE C. BRANNON & ERIC N. HOLMES, CONG. RSCH. SERV., R46751, SECTION 230: AN OVERVIEW 8(2021).

⁹⁰ *Id.*

⁹¹ 47 U.S.C. § 230(c)(1).

⁹² *See generally* *Stratton Oakmont v. Prodigy Servs. Co.*, 1995 WL 323710 (N.Y. Sup. Ct., May 24, 1995) (holding that Prodigy, an online service which advertised itself as family-friendly and engaged in extensive filtering of inappropriate material, had taken on the role of a “publisher” and therefore was strictly liable for any defamatory user content, whether it knew about the content or not).

⁹³ *Id.*

⁹⁴ Ashley Johnson & Daniel Castro, *Overview of Section 230: What It Is, Why It Was Created, and What It Has Achieved*, INFO. TECH. & INNOVATION FOUND. 4–5 (2021) <https://itif.org/sites/default/files/2021-230-report-1.pdf>.

Circuit's 1997 decision in *Zeran v. America Online, Inc.*⁹⁵ In its holding, the Fourth Circuit claimed that in enacting Section 230, Congress was partly responding to concerns that online providers facing potential tort liability would just prohibit or remove user content instead of litigating its legality.⁹⁶ By shielding providers from that liability, Congress eliminated the incentive for providers to restrict user speech.⁹⁷

Since its publication, other courts have primarily adopted *Zeran*'s reasoning and broadly construed Section 230(c)(1).⁹⁸ For example, six years after *Zeran*, the Ninth Circuit in *Batzel v. Smith* concluded that the phrase "interactive computer services" in Section 230 was not limited to services providing access to the internet as in *Zeran* and earlier cases.⁹⁹ Instead, it also included "any information services or other systems" such as a listserv.¹⁰⁰ Subsequent cases also confirmed that users who were independent of an online service provider could invoke the protection of Section 230. For example, in 2008, the Fifth Circuit in *Doe v. MySpace* found that Section 230 immunity applied broadly to tort claims, not just those premised on defamation, as in the *Stratton Oakmont* decision.¹⁰¹ This broad view of Section 230 was followed a year later in *Barnes v. Yahoo!*.¹⁰² In *Barnes*, the Ninth Circuit clarified that the scope of Section 230 could

⁹⁵ *Zeran v. America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997) (where the Fourth Circuit Court of Appeals reasoned that Section 230 of the CDA provided broad immunity to Internet service providers (ISPs) from online libel suits). In its decision, the Fourth Circuit rejected the plaintiff's argument that the provision only worked to block publisher liability and chose not to impose liability on America Online ("AOL") for its failure to remove false information. *Id.*

⁹⁶ *Zeran*, 129 F.3d at 330–31; BRANNON & HOLMES, *supra* note 90, at 45–46.

⁹⁷ *Zeran*, 129 F.3d at 331.

⁹⁸ See generally Ben Ezra, Weinstein, & Co., Inc. v. Am. Online, Inc., 206 F.3d 980 (10th Cir. 2000) (holding that AOL was not an information content provider with respect to content on a website where AOL had collaborated with third parties to post stock market information that was sometimes inaccurate); Green v. Am. Online (AOL), 318 F.3d 465 (3d Cir. 2003) (holding that Section 230 immunized AOL from claims arising out of the transmission of information by unrelated third parties in an AOL chat room); Universal Commc'n Sys., Inc. v. Lycos, Inc., 478 F.3d 413 (1st Cir. 2007) (dismissing cyberstalking and security law claims because such claims, which sought to hold Lycos liable for its role in the publication of third party statements, were barred under Section 230); Johnson v. Arden, 614 F.3d 785 (8th Cir. 2010) (holding that the interactive computer service provider was entitled to Section 230(c) immunity where there was "no evidence that [provider] designed its website to be a portal for defamatory material or do anything to induce defamatory postings"); Klayman v. Zuckerberg, 753 F.3d 1354 (D.C. Cir. 2014) (holding that Facebook and Mark Zuckerberg are protected by § 230(c) of the CDA); Jones v. Dirty World Ent. Recordings LLC, 755 F.3d 398 (6th Cir. 2014) (holding defendant website and operator were not liable under the CDA because they did not materially contribute to the defamatory content); Ricci v. Teamsters Union Loc. 456, 781 F.3d 25 (2d Cir. 2015) (finding allegations that GoDaddy refused to remove defamatory content authored by a third party from its web servers did not withstand Section 230 immunity).

⁹⁹ *Batzel v. Smith*, 333 F.3d 1018, 1030 (9th Cir. 2003).

¹⁰⁰ *Id.*

¹⁰¹ See generally *Doe v. MySpace, Inc.* 528 F.3d 413 (5th Cir. 2008) (holding that Section 230 barred a negligence claim against MySpace for failing to implement safety measures that would prevent minors from lying about their age and creating "public" profiles that enabled them to communicate with sexual predators online).

¹⁰² *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1102 (9th Cir. 2009).

apply beyond causes of action sounding in tort to include any cause of action that “inherently requires the court to treat the defendant as the ‘publisher or speaker’ of content provided by another.”¹⁰³

While these decisions have established Section 230 as a broad shield for social media platforms and influenced the scope of available legal options in combating misinformation, critics argue that they have expanded Section 230’s protection too broadly by giving social media platforms too much freedom in their censoring decisions.¹⁰⁴ In fact, despite the Supreme Court’s recent decision to decline the review of Section 230’s scope, Justice Thomas agreed with these critics.¹⁰⁵ In a statement from Justice Thomas respecting the denial of certiorari, he posited that courts have interpreted the provision to confer far more immunity to online platforms than the law requires. He went on to suggest that the Supreme Court should reexamine the issue when a better case presents itself.¹⁰⁶ His statement indicates just how necessary a reevaluation of Section 230 is in today’s society.

B. *Federal Proposals to Amend Section 230*

At the time of Section 230’s enactment, less than eight percent of Americans had access to the internet, and those who did went online for an average of just thirty minutes a month.¹⁰⁷ Moreover, since social media platforms were practically non-existent at the time of its enactment, the spread of misinformation on these platforms was not an issue Section 230 sought to address.

By providing social media companies with immunity from liability for the content users post on the platform, Section 230 allows social media platforms to engage in good-faith moderation and set their own rules for

¹⁰³ *Id.*

¹⁰⁴ Matthew Ingram, *Section 230 Critics are Forgetting About the First Amendment*, COLUM. JOURNALISM REV. (July 29, 2021), https://www.cjr.org/the_media_today/Section-230-critics-are-forgetting-about-the-first-amendment.php.

¹⁰⁵ See *Malwarebytes, Inc. v. Enigma Software Grp. USA*, 141 S. Ct. 13 (2020) (“I agree with the Court’s decision not to take up this case. I write to explain why, in an appropriate case, we should consider whether the text of this increasingly important statute aligns with the current state of immunity enjoyed by Internet platforms.”).

¹⁰⁶ *Id.* at 12, 18 (“This petition asks us to interpret a provision commonly called §230, a federal law enacted in 1996 that gives Internet platforms immunity from some civil and criminal claims. . . . When Congress enacted the statute, most of today’s major Internet platforms did not exist. And in the 24 years since, we have never interpreted this provision. But many courts have construed the law broadly to confer sweeping immunity on some of the largest companies in the world. . . . Extending §230 immunity beyond the natural reading of the text can have serious consequences. Before giving companies immunity from civil claims . . . we should be certain that is what the law demands. Without the benefit of briefing on the merits, we need not decide today the correct interpretation of §230. But in an appropriate case, it behooves us to do so.”).

¹⁰⁷ Abbey Stemler, *What is Section 230 and What Lies Ahead for Social-Media Reform?*, SEATTLE TIMES (Aug. 3, 2021, 2:36 AM), <https://www.seattletimes.com/opinion/what-is-Section-230-and-what-lies-ahead-for-social-media-reform/>.

what content they will allow without fear of liability.¹⁰⁸ Without this immunity shield, social media platforms would face potential liability for anything users post on their websites. That liability alone could lead to these companies censoring free speech and heavily moderating content posted on the platform.¹⁰⁹

If Congress were to revoke Section 230, social media sites would be incentivized to delay and limit posts appearing on their website to complete a review of the material first, threatening how these platforms function.¹¹⁰ Thus, arguments that any alteration to the statute could cripple online discussion often hinder discussions of repealing or modifying Section 230.¹¹¹ Nevertheless, in light of today's technological advances, the legislative objectives that Section 230 once served present the most significant barrier to mitigating misinformation dissemination on social media platforms.

Legislators introduced several bills in the 116th and 117th Congress related to Section 230 and/or social media misinformation regulation.¹¹² Some proposals to amend Section 230 would have narrowed the scope of liability protection, such as only protecting the removal of specified categories of content.¹¹³ In contrast, other proposals would have allowed social media operators to be held liable for not removing objectionable content under certain conditions or in a timely fashion.¹¹⁴ Regardless, the recent proposals for amending, reconstructing, or simply eliminating Section 230 protections indicate that something needs to change.

Proposals to amend Section 230 undoubtedly involve risks to free speech. If social media platforms are liable for the content shared by their users, they will likely behave in risk-averse ways to remove content that creates or has the ability to create legal exposure.¹¹⁵ This could silence speakers and have a disproportionate impact on marginal voices. Following the misinformation mayhem surrounding COVID-19, Democratic Senator Amy Klobuchar proposed a bill to amend Section 230. Senator Klobuchar's

¹⁰⁸ Wakabayashi, *supra* note 82.

¹⁰⁹ Rundall, *supra* note 44, at 51.

¹¹⁰ *Id.* at 57.

¹¹¹ Wakabayashi, *supra* note 82.

¹¹² Rebecca Kern, *Push to Rein in Social Media Sweeps the States*, POLITICO (July 1, 2022, 4:30 AM), <https://www.politico.com/news/2022/07/01/social-media-sweeps-the-states-00043229>; see also GALLO & CHO, *supra* note 3, at 27–31 (summarizing the § 230 and misinformation legislation introduced in the 116th Congress).

¹¹³ See, e.g., Accountability for Online Firearms Marketplaces Act of 2021, S. 7752725775, 117th Cong. (2021) (seeking to remove Section 230 protections for online firearms marketplace); Preserving Political Speech Online Act, S. 2338, 117th Cong. (2021) (looking to limit acceptable reasons for “good faith” removal to content that is obscene, illegal, or excessively violent and introducing the idea of “bad faith” moderation, which it defines as blocking content on the grounds of race, religion, sex, national origin, or political affiliation or speech).

¹¹⁴ S. 2725; S. 2338.

¹¹⁵ GEORGE J. STIGLER CTR. FOR STUDY OF ECON. AND STATE, PROTECTING JOURNALISM IN THE AGE OF DIGITAL PLATFORMS 52 (2019).

Health Misinformation Act of 2021 seeks to carve out an exception for medical misinformation during a health crisis, such as the COVID-19 pandemic, making the platforms legally liable for distributing anything the government defines as untrue.¹¹⁶

A group of Republican members of Congress also proposed a suite of bills involving a host of carve-outs for Section 230.¹¹⁷ However, the carve-outs in these proposed bills aim to prevent platforms from removing certain content (primarily conservative speech) and forcing them to remove other content, such as cyber-bullying posts.¹¹⁸ For example, Republican Congressman Jim Jordan proposed the Protect Speech Act. The *Protect Speech Act* would amend Section 230 to “ensure that the immunity under such Section incentivizes online platforms to responsibly address illegal content while not immunizing the disparate treatment of ideological viewpoints and encouraging a vibrant, open, and competitive internet, and for other purposes.”¹¹⁹

Interestingly enough, almost a year after proposing the *Health Misinformation Act of 2021* to Congress, Senator Klobuchar introduced another misinformation bill; but this time, it specifically targeted social media platforms. In February 2022, Senator Klobuchar, alongside Republican Senator Cynthia Lummis, introduced a new bipartisan social media bill, entitled the Nudging Users to Drive Good Experiences on Social Media Act, or more commonly referred to as the Social Media NUDGE Act.¹²⁰ Similar to the proposal presented in this Note, the *Social Media NUDGE Act* attempts to craft and enforce content-neutral interventions that focus on reducing “social media addiction and the spread of harmful content.”¹²¹

¹¹⁶ Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021).

¹¹⁷ See, e.g., Limiting Section 230 Immunity to Good Samaritans Act, H.R. 277, 117th Cong. (2021) (providing that Big Tech companies who want to receive Section 230 immunity must bind themselves contractually to a duty of good faith); Disincentivizing Internet Service Censorship of Online Users and Restrictions on Speech and Expression Act, S. 2228, 117th Cong. (2021) (proposing to hold Big Tech responsible for complying with pre-existing obligations per Section 230 of the Communications Decency Act (CDA) of 1996 and clarify ambiguous terms that allow Big Tech to engage in censorship); Protect Speech Act, H.R. 3827, 117th Cong. (2021) (looking to narrow a platform’s ability to use Section 230 as a defense for content removal).

¹¹⁸ H.R. 3827.

¹¹⁹ *Id.*

¹²⁰ See Social Media NUDGE Act, S. 3608, 117th Cong. (2022) (proposing to require the Federal Trade Commission (FTC) to identify content-neutral platform interventions to reduce the harm of algorithmic amplification and social media addiction on covered platforms).

¹²¹ See generally Mark MacCarthy, Senator Klobuchar “Nudges” Social Media Companies to Improve Content Moderation, BROOKINGS (Feb. 23, 2022), <https://www.brookings.edu/blog/techtank/2022/02/23/senator-klobuchar-nudges-social-media-companies-to-improve-content-moderation/> (explaining that the NUDGE Act seeks to identify content-neutral intervention methods that larger social media companies could implement “to reduce the harms of algorithmic amplification and social media addiction” and to have the Federal Trade Commission determine which of the recommended social media interventions should be made mandatory).

On the surface, these bills seem to address worthwhile goals, given the prevalent spread of misinformation. However, some free-speech advocates argue that even well-intentioned laws like Klobuchar’s *Health Misinformation Act of 2021* and *Social Media NUDGE Act* could backfire and threaten our First Amendment right to free speech.¹²² For bills such as the *Health Misinformation Act of 2021* and the *Social Media NUDGE Act* to survive judicial scrutiny based on First Amendment principles, courts would need to expand First Amendment exceptions substantially.¹²³ An overly strict or severe amendment would impose too much liability to satisfy both interests. Legislators must balance the interests of keeping the internet a free, open channel of communication while keeping in mind the interests of objective and transparent misinformation regulation. As such, any proposal to mitigate misinformation by using Section 230 must take a First Amendment-friendly approach.

IV. FCC REGULATION OF SOCIAL MEDIA MISINFORMATION

The dangers of misinformation warrant some sort of regulatory model. Currently, social media platforms operate in the dark—they are accountable to no one. As discussed in Part II(B), social media platforms have failed to manage misinformation effectively thus far. Clearly, outside regulation is needed to effectively push social media platforms to address misinformation on their sites. The difficulty is in determining how misinformation regulation will be performed. Thus, this Part proposes using the FCC as an outside regulatory agency to promote misinformation mitigation on social media platforms. Specifically, it suggests granting the FCC the authority to enforce specific transparency standards on social media platforms or risk losing Section 230 immunity. Finally, this Part analyzes the constitutionality of the proposal in relation to First Amendment free speech protections.

A. *Creating & Enforcing Misinformation Labeling Standards Through the FCC*

Currently, social media platforms are not regulated by any federal agencies.¹²⁴ The rationale behind the lack of federal regulation relies heavily on Americans’ right to free speech as a fundamental governmental purpose since it promotes values central to the First Amendment. Therefore, assuring that any form of social media regulation does not infringe on First Amendment free speech rights is key to developing a fair, enforceable set of social media misinformation regulation standards. For this reason, this Note

¹²² Ingram, *supra* note 104.

¹²³ *Id.*

¹²⁴ VALERIE C. BRANNON, CONG. RSCH. SERV., LSB10309, REGULATING BIG TECH: LEGAL IMPLICATIONS 1 (2019). Although the FCC currently classifies broadband-internet access services as an “information service,” and subjects these service providers to a regulatory framework, it currently does not regulate social media platforms. *Id.*

proposes that Congress should introduce new legislation that grants the FCC the power to create and enforce misinformation labeling standards for social media platforms.

As an independent U.S. government agency overseen by Congress, the FCC is the primary authority for communications law, regulation, and technological innovation in the United States.¹²⁵ While the FCC's purview is expansive, it focuses a significant amount of its capabilities on "[r]evising media regulations so that new technologies flourish alongside diversity and localism[.]"¹²⁶ Since the Commission is responsible for "revising media regulations,"¹²⁷ the FCC seems to be the appropriate agency to enforce specific misinformation standards on social media platforms. Allowing the FCC to hold social media platforms accountable for mitigating misinformation would reduce concerns regarding infringing on free speech rights.¹²⁸ With this new oversight power, the FCC would write the rules and regulations regarding what is required in terms of content moderation.

Under this approach, the U.S. government would not be censoring free speech on social media platforms; instead, the government would grant the FCC the ability to enforce standards that promote the public interest. By employing misinformation labeling standards, social media platforms would not be taking down content; instead, they would provide users with more information about a post while respecting First Amendment free speech rights. Still, the FCC needs a set of guidelines to enforce accountability from social media platforms.

One approach may be allowing social media platforms to continue using the misinformation labels they have already integrated into their respective platforms. Another approach could entail the FCC creating its own labels that social media platforms must use in order to keep Section 230 protections. However, creating labels for all social media platforms to use would take much more time to implement. All platforms would need to integrate such labels into their already existing platforms. Not only would this take time, but it could also confuse users who have already begun interacting with pre-existing misinformation labels. Thus, allowing social media platforms to continue using the labels highlighted in Part II could be the most time-effective approach to reducing platform misinformation.

Either way, to ensure that the social media platforms are using misinformation labels effectively, a condition to keeping Section 230 protections could be reporting label click-through-rates or user interactions to the FCC throughout the year. If the data does not show a reduction in the amount of misinformation spread on the platform, maintaining Section 230

¹²⁵ *What We Do*, FED. COMM'NS COMM'N, <https://www.fcc.gov/about-fcc/what-we-do> (last visited Aug. 25, 2022).

¹²⁶ *Id.*

¹²⁷ *Id.*

¹²⁸ Nicholas Carr, *How to Fix Social Media*, THE NEW ATLANTIS, 2021, at 3, 16.

protections could then be at risk.

Although there are several approaches to compelling social media platforms to use misinformation labeling, this Note does not seek to provide an implementation-ready system to lessening misinformation. Instead, this Note seeks to suggest a First Amendment-friendly method of reducing misinformation by using content labels. Accordingly, the next section will discuss using Section 230 as a quid pro quo benefit to further oblige social media platforms to effectively label misinformation.

B. Enforcing Misinformation Labeling: Section 230 as a Quid Pro Quo Benefit

Social media platforms may not agree with the guidelines the FCC proposes. Disagreement over misinformation labeling guidelines may lead social media platforms to be non-compliant. Non-compliance would make the suggested method of mitigating misinformation ineffective. As a result, using Section 230 of the CDA as a quid pro quo benefit can provide the FCC with a way to compel social media platforms to adhere to misinformation labeling standards.

Section 230 immunity has protected social media platforms from significant legal liability and provided legal cover for the complicated decisions regarding content moderation.¹²⁹ Facebook and Twitter have recently cited it to defend themselves in court when users have sued after being barred from the platforms, making it clear that Section 230 is a valuable asset to most social media platforms.¹³⁰ Without Section 230, social media platforms would be subject to significant legal liability, and the ways in which the platforms function would change drastically. Therefore, this Note proposes allowing the FCC to set misinformation labeling standards that social media platforms must meet to reap the benefits Section 230 provides. And if social media platforms fail to meet the labeling standards, the FCC may revoke Section 230 immunity.

If the FCC were given the power to enforce misinformation labeling standards on social media platforms, the Commission would need to figure out how to measure social media platform compliance with the new standards. One recommendation for measuring misinformation is by tracking click-through rates of misinformation labels (like Twitter's current approach to measuring its own labeling effectiveness).¹³¹ However, this would require social media platforms to provide their private data to the FCC. Currently, social media platforms are not required to provide data to anyone. A critical starting point for effective regulation is data transparency:

¹²⁹ Wakabayashi, *supra* note 82.

¹³⁰ *Id.*

¹³¹ See Ortutay, *Twitter Rolls Out Redesigned Misinformation Warning Labels*, *supra* note 57 (explaining Twitter's use of "click-through rates" to measure the effectiveness of its misinformation labels).

full access to data for research not directed by social media platforms.¹³² Unlike Twitter, which has shared some information regarding the effectiveness of its misinformation labels, it is unclear how effective Facebook's misinformation labeling has been in mitigating misinformation on its platforms. To ensure all social media platforms effectively attempt to reduce misinformation, monitoring directed by a third-party outside of the influence of the platforms is essential.

By using the FCC as an outside regulatory agency that has the power to request data from social media platforms, the FCC could ensure that the misinformation mitigation methods that platforms use actually work to reduce misinformation. Platforms like Facebook have already suggested a reform of Section 230 that only provides immunity to social media platforms on the condition that they follow best practices.¹³³ If the FCC were to require platforms to meet specific misinformation labeling standards to keep Section 230 immunity, it is likely to compel the platforms to meet those standards.

C. *The Constitutionality of Enforcing Misinformation Labeling*

Although the recommendations above are worthwhile, they have no chance of being implemented unless they are constitutional. To grant the FCC the power to ensure social media platforms are effectively working towards mitigating misinformation, new legislation is necessary. But any legislative proposal that regulates online content, including content shared on social media platforms, may implicate the First Amendment to the extent that it burdens protected speech activity.¹³⁴

Both aspects of Section 230—granting providers immunity for hosting user content and for restricting content—were arguably intended to ensure that the government would not be responsible for striking a proper balance between these two goals.¹³⁵ Instead, private parties would decide whether content belonged online.¹³⁶ In this way, both aspects of Section 230 serve the First Amendment by shielding speech from government intervention. However, social media platforms are private parties and thus protected, but not bound, by First Amendment free speech rights. This simply means that

¹³² Shirin Ghaffary, *Facebook's Whistleblower Tells Congress How to Regulate Tech*, VOX (Oct. 5, 2021, 7:50 PM), <https://www.vox.com/recode/22711551/facebook-whistleblower-congress-hearing-regulation-mark-zuckerberg-frances-haugen-senator-blumenthal>. While data transparency is an issue that needs to be addressed in relation to social media platforms, it is beyond the scope of this paper.

¹³³ Elizabeth Culliford, *Analysis: If Facebook is the Problem, Is a Social Media Regulator the Fix?*, REUTERS (Oct. 6, 2021), <https://www.reuters.com/technology/if-facebook-is-problem-is-social-media-regulator-fix-2021-10-06/>.

¹³⁴ BRANNON & HOLMES, *supra* note **Error! Bookmark not defined.**, at 47.

¹³⁵ See 47 U.S.C. § 230(a)(4) (finding that the internet has “flourished, to the benefit of all Americans, with a minimum of government regulation”); *id.* § 230(b)(2) (stating that it is the policy of the United States “to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation”).

¹³⁶ *Id.*

social media platforms may restrict or regulate speech under their control.¹³⁷ Nevertheless, Section 230 is not constitutionally required.

Since Section 230 is not constitutionally required, revoking Section 230 protection can be done without violating the First Amendment.¹³⁸ Thus, an argument for restricting Section 230 immunity without raising any constitutional concerns can be made because Congress is not required to grant Section 230 immunity.¹³⁹ Accordingly, a successful misinformation regulation bill strategically uses the unique immunity shield Section 230 provides to the government's advantage. Introducing a bill that gives the FCC oversight of social media platforms does not impede users' free speech; instead, it provides the FCC with the power to oblige social media platforms to enforce effective misinformation labeling programs. By employing misinformation labeling standards, social media platforms would not be taking down content; instead, they would provide users with more information about the post while respecting First Amendment free speech rights.

Experts consider labeling misinformation a First Amendment-friendly response.¹⁴⁰ Instead of removing misinformed speech, labels counter the misinformation with more speech. More importantly, the labeling is done at a private company's initiative, not the government's direction.¹⁴¹ Moreover, today the research consensus among social scientists is that some fact-checking methods significantly reduce the prevalence of false beliefs.¹⁴² Thus, compelling social media platforms to adopt a First Amendment-friendly response to misinformation would balance the interests of regulating misinformation on social platforms and free speech rights.

CONCLUSION

Social media platforms are playing an ever-expanding role in shaping the contours of the information ecosystem today. These platforms have shouldered the burden of ensuring that the public is informed—and not misinformed—about matters affecting our democratic institutions in our elections and issues affecting our health and lives in the context of the pandemic. Yet, in the absence of federal or state regulation or guidance in the U.S., social media platforms have been left to address the misinformation issue on their own. And while platforms' intervention in the online marketplace of ideas is a start to mitigate misinformation, it is not without

¹³⁷ VALERIE C. BRANNON, CONG. RSCH. SERV., R45650, FREE SPEECH AND THE REGULATION OF SOCIAL MEDIA CONTENT (2019).

¹³⁸ *Gucci Am., Inc. v. Hall & Assocs.*, 135 F. Supp. 2d 409, 422 (S.D.N.Y. 2001).

¹³⁹ BRANNON & HOLMES, *supra* note 89, at 49.

¹⁴⁰ Emily Bazelon, *The First Amendment in the Age of Disinformation*, N.Y. TIMES MAG., (Oct. 13, 2020), <https://www.nytimes.com/2020/10/13/magazine/free-speech.html>.

¹⁴¹ *Id.*

¹⁴² *Id.*

its problems. The past few years clearly show us that social media platforms cannot be left to regulate misinformation. As discussed earlier, platforms like Facebook and Twitter need an incentive to maintain effective misinformation mitigation interventions. Without some sort of oversight, these platforms can get lazy in their efforts to reduce misinformation.

Relying on the government to regulate this space creates real threats of abuse and challenging First Amendment hurdles. Yet, at a time when public confidence in social media platforms is low and government interests in pursuing social media regulation are high, leaving control to social media platforms has provided an opportunity to expand the FCC by creating a division that can meaningfully regulate this space. But the success of an industry-wide self-governing model necessarily depends on incentives to participate, mechanisms for enforcement, and penalties for failure to comply. Clearly, social media platforms do not want to lose Section 230 protections. If they did in fact lose Section 230 protections, most platforms would be flooded with lawsuits. Thus, keeping Section 230 protections seems to be enough of an incentive to compel social media companies to take the necessary steps to mitigate misinformation.

Adopting a model similar to the suggested method proposed in this Note would promote order and efficiency within the industry and appeal to citizens and lawmakers who have been calling for change while avoiding hasty and problematic government regulation. Furthermore, such a solution avoids impeding First Amendment free speech rights and most of the challenges inherent in the decentralized and government-led frameworks. Therefore, granting the FCC the authority to regulate social media platforms and using Section 230 immunity as a quid pro quo benefit would likely result in disseminating accurate information and mitigating misinformation, ultimately benefitting both users and the social media platforms themselves.