

Spring 2022

The Private Regulators of the New Public Square

Jagdeep Sekhon

Follow this and additional works at: https://repository.uchastings.edu/hastings_science_technology_law_journal



Part of the [Science and Technology Law Commons](#)

Recommended Citation

Jagdeep Sekhon, *The Private Regulators of the New Public Square*, 13 HASTINGS SCI. & TECH. L.J. 167 (2022).

Available at: https://repository.uchastings.edu/hastings_science_technology_law_journal/vol13/iss2/5

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Science and Technology Law Journal by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact wangangela@uchastings.edu.

The Private Regulators of the New Public Square

JAGDEEP SEKHON*

TABLE OF CONTENTS

Introduction	167
RoadMap	170
Part I: Current Social Media Policies	171
A. Importance of This Issue	172
Part II: Section 230 Debates	173
Part III: Platform Policing and our First Amendment Right	177
PART IV: Shadow Banning and Negative Impacts	180
PART V: Solutions	183
A. Transparency	183
B. Internal Review Boards	185
C. Diversity in Tech Companies	186
D. Legislative Solution	187
Conclusion	190

Introduction

Social media platforms in the past were used to share personal travel pictures and updates in users' lives. Today, these platforms have become a means of communication, a way for us to remain connected with the rest of the world. Some users have created businesses through these social media

* Jagdeep Sekhon is a graduate of UC Hastings Law Class of 2022. She will be joining Perkins Coie's Emerging Companies and Venture Capital Practice group. She would like to thank Alice Armitage for her constant support throughout the writing process and her family for pushing her to be the best version of herself. This article is focused on calling to attention the major issues with the lack of transparency on social media platforms and the negative impacts that marginalized communities face on these platforms. The idea of this article was formed when the Sikh community faced heavy censorship on social media platforms during the 2020 Farmers' Protest in India.

platforms where they sell products and services. This paper will analyze how social media platforms moderate the content that is posted on their websites. The main platforms of interest are Instagram and Facebook, along with some examples from Twitter and TikTok. Content is posted differently on each of these platforms and the ways users engage with their content is very different. TikTok is mainly a video platform where people post content in the form of videos with trending sounds and dances. Instagram on the other hand is platform where videos and photos are posted along with an option to post stories that disappear within 24 hours. The increase in users and user engagement comes with the need to regulate the content to make sure that the platforms provide a safe space for everyone. However, content regulation comes with a dark side, ie. shadow banning—a technique to moderate content posted on the social media platforms.¹ It is a way to ban people without informing them. This means that a user can continue to post, but their posts will be hidden from the rest of the users.²

The number of social media users has increased from 2.86 billion in 2017 to 3.78 billion in 2021 and is expected to increase to 4.41 billion by 2025.³ The growth of social media users has led to an increase in sharing of news-related content and an increase in the spread of misinformation. To combat the spread of misinformation, hate speech and cyberbullying, social media platforms have started to moderate the content that is posted online. Currently social media platforms use artificial intelligence (AI) or fact checkers to regulate their content.⁴ AI is used to track what users are posting in order to combat hate speech and cyberbullying.⁵ The AI picks up on words, pictures, and to some degree, context.⁶ However, AI is not properly trained in the diverse cultural terminology or trained to understand certain cultural norms.⁷ AI not being trained in cultural norms is one reason that content that complies with community guidelines is getting flagged and users

1. Samantha Cole, *Where Did the Concept of 'Shadow Banning' Come From?*, VICE: MOTHERBOARD TECH (July 31, 2018), <https://www.vice.com/en/article/a3q744/where-did-shadow-banning-come-from-trump-republicans-shadowbanned>.

2. *Id.*

3. STATISTICA RSCH. DEP'T, *Number of social network users worldwide from 2017 to 2025 (in billions)*, STATISTICA (Jan. 28, 2022), <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>.

4. Tilly Kenon, *How are social media platforms using AI?*, AI MAGAZINE: AI STRATEGY, (June 26, 2021), <https://aimagazine.com/ai-strategy/how-are-social-media-platforms-using-ai>.

5. *Id.*

6. *Id.*

7. Shirin Ghaffary, *The algorithms that detect hate speech online are biased against black people*, VOX (Aug. 15, 2019), <https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter>.

are getting punished without proper cause. There can be serious freedom of speech ramifications letting online social media platforms remove information or shadow ban accounts without proper accountability and transparency. Transparency refers to making it clear what kind of content is permissible, but also giving users a detailed description of why their content is being taken down. Transparency means understanding why certain content that does not appear to break community guidelines is being taken down. Why is certain content flagged? Transparency for the purpose of this paper means giving users an understanding of how the content is being policed.

The lack of transparency is evident because many users on social media are left wondering why they see less views after they post certain news-related content that is not breaking community guidelines.⁸ The lack of transparency impacts users negatively, by reducing attendance at grassroots events and decreasing money contributed to social causes. Further, the impact for some users can mean a drop in income.⁹ Social media for many users is a way to make money either through selling products, endorsements, or services. Therefore, the lack of transparency with what content violates community guidelines and gets shadow banned is important. Shadow banning is not a small issue. The consequences of accounts getting shadow banned is becoming a serious problem. The problem is not that content is being regulated, however, that human moderators or AI do not understand when something is news-related or hate speech. Therefore, they regulate the speech without really understanding its true meaning. This kind of content regulation may influence important political outcomes.

Section 230 of the Communications Decency Act (CDA 230) was a step in the right direction for moderating content on the internet when it was created in 1996. Section 230 states that “[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.”¹⁰ Section 230 means that “online intermediaries that host or republish speech are protected against a range of laws that might otherwise be used to hold them legally responsible for what others say.”¹¹ CDA 230 creates protection for the online platforms that has allowed modernization of the internet and free speech to

8. Rayhan Uddin, *Afghanistan: Muslim Instagram users complain about censorship*, MIDDLE EAST EYE (Aug. 27, 2021), <https://www.middleeasteye.net/news/afghanistan-instagram-muslim-users-complain-censorship-posts>.

9. Jesselyn Cook, *Instagram's CEO Says Shadow Banning 'Is Not A Thing.' That's Not True*, HUFFPOST (Feb. 25, 2020), https://www.huffpost.com/entry/instagram-shadow-banning-is-real_n_5e555175c5b63b9c9ce434b0 (last updated Feb. 27, 2020).

10. ELEC. FRONTIER FOUND., *CDA 230 The Most Important Law Protecting Internet Speech*, <https://www.eff.org/issues/cda230> (last visited Jan. 1, 2021).

11. *Id.*

grow,¹² but the overly broad reading of Section 230 has created another space for illegal activity. For example, Backpage is a hub created online to advertise illegal sex.¹³ The second provision of Section 230, has given online social media platforms broad authority to take down content that they find objectionable, whether such material is constitutionally protected.¹⁴

The answer to the current issues that social media platforms have is a multi-prong solution, starting with the idea that online social media platforms¹⁵ need to be more transparent with their users on what kind of content breaks community guidelines. The companies need to better address the vague community guidelines that social media platforms have listed on their websites.¹⁶ Another solution is to implement internal review boards in tandem with more sophisticated AI that can lead to better culturally aware content policing. Moreover, there needs to be a legislative approach to fix the problems with the way online social media platforms are operating. Lastly, there needs to be more diversity in the tech industry. Diversity in the executive C-Suite can give clarity to tech leaders on the cultural context of content that is getting shadow banned. The issues with tech are too complex to be solved by one solution; It requires a multi-prong approach.

RoadMap

This paper focuses on the issues regarding platform policing with Instagram and Facebook, which can be applied across all online social media platforms. Part I of this paper explores the current social media policies and the flaws that exist with them. These policies are vague, and the companies do not follow through with the guidelines that they create. Part II will focus on Section 230 of the Communications Decency Act.¹⁷ This section has arguably led to the growth of the internet, but it has also led to the issues we see today. The fact that courts and the legislature have not been taking action on the social media companies is partly attributed to the lenient way Section 230 was created. Part III dives into why the First Amendment needs to be applied to social media companies and private actors. Part IV discusses the

12. *Id.*

13. Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage: Revising Section 230 Immunity*, 2 GEO. L. TECH. REV. 453 (2018).

14. *Id.*

15. Mainly focused on Instagram and Facebook.

16. INSTAGRAM, *Community Guidelines*, <https://help.instagram.com/477434105621119> (last visited Nov. 15, 2012) ; FACEBOOK, *Instagram Community Guidelines*, <https://www.facebook.com/business/learn/lessons/facebooks-community-standards> (last visited Dec. 2021); Cook, *supra* note 9.

17. See ELEC. FRONTIER FOUND., *supra* note 10.

issues with social media platform, including shadow banning and the negative impacts it has on users and social media community. Lastly, Part V discusses the various solutions available to tech companies that may help fix the internal issues online social media platforms are facing and the reasons of why we need a legislative solution. For example, diversifying the C-Suite, creating internal review boards and creating better trained AI. The multi-prong approach will also need a strong legislative approach. We need Congress to legislate the technology sector in order to see real changes with the way social media platforms behave.

Part I: Current Social Media Policies

All online social media platforms have community guidelines they follow when taking down content from their websites; however, these guidelines are vague and can have the opposite impact than intended.¹⁸ Instagram, Facebook, and Twitter all use community guidelines to police information that is being posted on their platforms. For example, Instagram says that their goal is:

We want Instagram to continue to be an authentic and safe place for inspiration and expression. Help us foster this community. Post only your own photos and videos and always follow the law. Respect everyone on Instagram, don't spam people or post nudity.¹⁹

To achieve the above goal, Instagram has a list of definitions and examples that will be taken down. Some of those definitions are easy to understand, but others that are not as clear. For example, Instagram lists concrete examples of what nudity is and subsequently states that photos with nudity will be removed.²⁰ Instagram has defined nudity as: "This includes photos, videos, and some digitally-created content that show sexual intercourse, genitals, and close-ups of fully-nude buttocks."²¹ This definition can help a user understand what kind of content with nudity will be taken down. This definition provides some clarity because it contains a list of examples that can help the user understand the parameters of what nudity means to Instagram.

However, issues arise when there are vague guidelines without proper examples. For instance, Instagram lists cases where it may allow content that goes against community guidelines.

18. Faiza Patel and Laura Hecht-Felella, *Facebook's Content Moderation Rules Are a Mess*, BRENNAN CTR. FOR JUST. (Feb. 22, 2021), <https://www.brennancenter.org/our-work/analysis-opinion/facebooks-content-moderation-rules-are-mess>.

19. Instagram, *supra* note 15.

20. *See Id.*

21. *Id.*

In some cases, we allow content for public awareness which would otherwise go against our Community Guidelines – if it is newsworthy and in the public interest. We do this only after weighing the public interest value against the risk of harm and we look to international human rights standards to make these judgments.²²

The issue here is that, unlike the example where “nudity” has a definition with examples, the term “public interest” does not. The term “public interest” does not have any examples or a more detailed definition. This can make it hard for users to understand what public interest means for Instagram. Public interest can be defined in many ways. Who in the company defines the term “public interest”? Are the moderators adequately trained in the hundreds of cultural backgrounds that the specific user who has an Instagram account come from? Do these moderators come from diverse backgrounds? This matters because the term “public interest” can be different for everyone. The meaning can be different for a Black, Indian, Punjabi, or Hispanic user. It is this kind of language that causes issues for users. The vague language in social media platform guidelines is not just a problem for Instagram, but also for Twitter and Facebook. All the social media platforms have extensive community guidelines, yet users are still confused as to why their content is being taken down or shadow banned.²³

A. Importance of This Issue

Social media has become a big part of our current society; therefore, issues relating to content regulation can have a huge negative impact on users. Social media is how some people get their news, mobilize for social causes, remain connected, and make money.²⁴ Instagram, Facebook, and Tik-Tok have become a part of our daily lives whether we realize it or not. My friends and I personally follow accounts of small grassroot movements that track global issues. One way we get our news is through these accounts that show firsthand of what is happening in humanitarian crisis around the world. Anywhere in the world there is a social movement happening and it is being broadcasted on one of the many social media platforms. For example, what is happening in Ukraine is being spread on Instagram to help spread

22. Instagram, *supra* note 15.

23. See Michelle Santiago Cortes, *What is Shadow Banning & Why are TickTockers Complaining About It*, REFINERY29 (July 17, 2020), <https://www.refinery29.com/en-us/2020/07/9901461/what-is-shadow-banning-tik-tok>.

24. See Uddin, *supra* note 8; see also Cook, *supra* note 9.

more awareness.²⁵ People follow along not by just resharing posts but by donating to social causes. So, imagine when these social media accounts decide to censor that content, or they shadow ban an account that spreads urgent news.²⁶ Middle East Eye reports this is happening on numerous occasions when discussing issues related to the Muslim community.²⁷ It happened during the Afghanistan crisis in 2021 and during Israel's increased aggression against Palestine in the summer of 2021.²⁸ Shadow banning has real life consequences. As a society, we rely on these social media accounts now whether we like it or not and it's important that we hold them accountable for their wrongs. Along with holding them accountable, we as a society must demand real change.

To understand how the modern version of the internet has come to be, we need to go back in history and look at Section 230 of the Communications Decency Act. Section 230 is credited in helping the internet expand to the form we see it in today and arguably helped social media platforms achieve a massive expansion.

Part II: Section 230 Debates

Section 230 can be seen as the protector of freedom of speech on the internet because it protects online social media platforms from liability when users post improper content on their platforms. Section 230 states, "No 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."²⁹ This means that social media online platforms that host or republish speech are protected against laws that might be used against them to hold them legally responsible for what others say and do on their websites.³⁰ There are exceptions for some criminal and intellectual property types of claims.³¹

Section 230 is born out of *Stratton Oakmont, Inc. v. Prodigy Services Company*.³² In *Prodigy* an online service provider lost its protection as a

25. Drew Harwell and Rachel Lerman, *How Ukrainians have used social media to humiliate the Russians and rally the world*, WASH. POST (Mar. 1, 2022), <https://www.washingtonpost.com/technology/2022/03/01/social-media-ukraine-russia/>.

26. See Uddin, *supra* note 8.

27. *Id.*

28. *Id.*

29. Electronic Frontier Foundation, *supra* note 10.

30. *Id.*

31. *Id.*

32. *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995 WL 323710 (N.Y. Supr. Ct. May 23, 1995).

distributor because it did not filter inappropriate content on their website.³³ Even though the online message board was unaware that this content was posted, it became a publisher and became liable for the content.³⁴ The reason this was a bigger issue was that the operator had guidelines and screening programs to control the content that was posted. Representatives Christopher Cox and Ron Wyden were then motivated to create what we know today to be Section 230.³⁵

Section 230 also gives online social media platforms a high degree of immunity for when they take down content considered to be violating community guidelines. The speech could be protected by the First Amendment, but that doesn't matter. The First Amendment has limits when applied to private actors, so social media platforms get away with taking that speech down.³⁶ Overtime Section 230 has been read so broadly that it has allowed certain hate speech, and revenge porn sites to exist on the internet.³⁷ It has also allowed platforms free reign to shadow ban, close accounts and remove posts that they deem to be breaking their community guidelines. On one hand, it has allowed platform creators to become dictators by giving them the power to take whatever they want down and keep up what they want.³⁸ However, 230 has also allowed users freedom to post whatever is on their mind whether that be hate speech, images of people without their permission or misinformation.³⁹ Internal policies at Facebook show that unlike the rest of the Facebook community there are users that can violate company standards without facing any questions.⁴⁰

There are advantages of Section 230. Many argue that it has led to what we know as the modern-day internet.⁴¹ By taking away liability from platforms it has allowed the

operators breathing room to let users post information without overworking companies. The social media companies do not have to worry if

33. *Id.*

34. *Id.*

35. Citron, *supra* note 13.

36. David L. Hudson, Jr., *In the age of social media, expand the each of the first amendment*, ABA, https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/in-the-age-of-social-media-first-amendment/.

37. Mary Anne Franks, *Section 230 and the Anti-Social Contract*, LAWFARE (Feb. 22, 2021), <https://www.lawfareblog.com/section-230-and-anti-social-contract>.

38. Jeff Horwitz, *Facebook says its rules apply to all, company documents reveal a secret elite that's exempt*, THE WALL ST. J. (Sep. 13, 2021), https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353?mod=article_inline.

39. *Id.*

40. *Id.*

41. Citron, *supra* note 13.

someone posts something problematic; they as a company will not be held liable. The overworking accompanies the subjectivity of the company's own bias, but also requires companies to have huge number of resources to then police every single post that is put on the websites. It also allows platform creators not worry about legal ramifications from user generated content. For example, if someone makes a post on Facebook inciting violence which is illegal to do, then Facebook will not be liable for that speech. Facebook is not publishing the information as a company; it is simply allowing users to publish what they want. The user may be held liable, but not Facebook. This way Facebooks escapes liability.

It makes sense why Section 230 did not create more restrictions early on because the Internet was not as sophisticated as it is today and needed less regulation to get this version of the cyber space. Forcing the platforms to police content early on may not have led to the massive growth of Instagram and Facebook. Today users have created ways to build brands and careers out of the platform.⁴² That may have not been possible if the laws were too strict early on because there would be too many obstacles. There would be numerous regulations that the platform creators and users would have to abide by. There would always be a fear for the social media companies that they would be liable for user generated speech if Section 230 was not enacted.

Furthermore, aggressive platform policing runs into possible legal issues. Those legal issues can range from banning harmless hashtags to micromanaging accounts and spying on its users.⁴³ For instance when Israel's aggression had increased in the summer of 2021 accounts that were reporting bombings or force removals of residents that lived in Sheikh Jarrah neighborhood were getting censored.⁴⁴ Accounts that have never broken community guidelines were getting flagged when spreading news related content about the Israel and Palestine issue.⁴⁵ Supporters like Lon Safko argue that online social media platforms have a "moral and legal responsibility to deliver information clearly, organically, and without political censorship."⁴⁶ Protectors of the First Amendment argue that censorship on the internet is a

42. FIDM, *3 Reasons to Pursue a Career in Social Media*, <https://fidm.edu/en/news/Reasons+to+Pursue+a+Career+in+Social+Media>.

43. See Uddin, *supra* note 8.

44. *Id.*

45. *Id.*

46. Peter Suci, *Do Social Media Companies Have The Right To Silence The Masses – And Is This Censoring The Government?*, FORBES (Jan. 11, 2021), <https://www.forbes.com/sites/peter-suci/2021/01/11/do-social-media-companies-have-the-right-to-silence-the-masses—and-is-this-censoring-the-government/?sh=73a7bff248e2>.

violation of their First Amendment speech.⁴⁷ The idea is that even if people are posting things that are controversial it is the users' right to say whatever they want. Speech should not be taken down because others disagree with it.⁴⁸ Courts also have ruled in the belief that the CDA 230 reflected "Congress' desire to promote unfettered speech on the Internet."⁴⁹ This has led to Section 230 being read broadly and allowing for pages like Backpage or Dirty.com to exist.⁵⁰

Platforms not policing content on the internet can lead to sexual violence against women and hate speech websites/comments.⁵¹ The courts' broad interpretation of Section 230 leaves vulnerable communities at risk for harm on the internet.⁵² For example, the harm can be to women that have their private images exposed or hate speech against minority communities being circulated. As mentioned previously this is what is happening right now internally at companies like Facebook.⁵³ Since there is no uniform regulation on what kind of content cannot be posted or the kind of content that is harmful, the social media platform companies are creating their own internal processes which are not successful.⁵⁴ The current policies are harming some users.

While Section 230 may have been well-intentioned, and even necessary to help the internet mature, its broad grant of immunity decades later is problematic. We now have social media platforms taking advantage of that broad discretion and silencing speech.⁵⁵ The social media platforms regulate content on their platforms as per their convenience.⁵⁶ Section 230 has become a shield that platforms use to stifle speech which is hurting grassroots movements.⁵⁷ Movements like the Farmer's Movement or the Afghanistan evacuation crisis that get negatively impacted because the more people share about these issues, the less viewership the users' accounts get.⁵⁸ It doesn't help that these private actors move through the world without any regulation and any repercussions. It is unfair that content creators pay the price for representing

47. *Id.*

48. *Id.*

49. ELEC. FRONTIER FOUND., *supra* note 10.

50. Citron, *supra* note 13.

51. Franks, *supra* note 37.

52. *Id.*

53. Horwitz, *supra* note 38.

54. *Id.*

55. Uddin, *supra* note 8.

56. Horwitz, *supra* note 38.

57. *Id.*

58. *Id.*

their communities. The less views and Instagram shadow banning accounts can lead to a decrease in income for users' that use the platform to make a living due to less engagement from users.⁵⁹ It is making it harder to spread news to users and content that many use to stay updated on the current world issues.⁶⁰

Academics like Mary Anne Franks expose Section 230 as a problematic contract between the government and tech elites as a way to keep the power in the hands of elite whitemen.⁶¹ According to Mary Anne Franks, Section 230 of the CDA lets online social media platform operators know that the government "will not interfere when they disregard, encourage, or profit from the exploitation and abuse of vulnerable communities." Congress and courts fail to understand the degree that the internet can negatively impact people's lives. For example, "Backpage is a classifieds hub that hosts '80 percent of the online advertising for illegal commercial sex in the United States.'"⁶² Additionally, there is Dirty.com which is a platform devoted to spreading gossip mainly about college students.⁶³ Instead of protecting the victims, courts shield these websites because of the idea that the internet needs to promote "unfettered speech on the Internet."⁶⁴ However, once posted on the internet, content exists there forever. Negative comments, inappropriate pictures, and other harmful content can impact a person's chances of getting a job, romantic partner and taint their reputation. The fact that courts have turned a blind eye to this simply to protect speech is irresponsible. Courts need to push platforms to properly police harmful content on their platforms.⁶⁵ It can be hard to applaud growth on the internet when people's lives are impacted negatively and there is further promotion of systematic oppression.⁶⁶

Part III: Platform Policing and our First Amendment Right.

Social media platforms are known to censor speech on their platforms. This is done through shadow banning, censoring people's accounts and various other methods.⁶⁷ Professor David L. Hudson, Jr. argues that the First

59. Cook, *supra* note 9.

60. Uddin, *supra* note 8.

61. Franks, *supra* note 37.

62. Citron, *supra* note 13.

63. *Id.*

64. *Id.* at 459.

65. Franks, *supra* note 37.

66. *Id.*

67. Tanya Parmanand, *Shadowbanning: the newest form of social media censorship*, THE BOAR (Nov. 29, 2019), <https://theboar.org/2019/11/shadowbanning/>.

Amendment needs to be extended to private actors.⁶⁸ Traditional legal doctrine provides that those private actors are not constrained by the Constitution generally.⁶⁹ This doctrine is the “state action” doctrine which creates a zone of privacy and protects us from excessive governmental interference. This doctrine is limited when applied to private actors.⁷⁰ However, the fast growth of social media and the way it controls our communication demands that the First Amendment be extended to those companies.

Professor Hudson argues that the First Amendment is a protection of our right to freedom of expression which is enhanced through the marketplace of ideas and individual self-fulfillment.⁷¹ Powerful private actors can infringe on free expression rights just as much as public actors. Therefore, there needs to be an extension of the First Amendment to these private actors. Some states are already expanding the first amendment so that it applies to private actors.⁷² For example, the New Jersey Supreme Court has ruled that individuals have the right to challenge restrictive provisions of private homeowner associations; The state high court wrote:

“In New Jersey, an individual’s affirmative right to speak freely is protected not only from abridgement by government, but also from unreasonably restrictive and oppressive conduct by private entities in certain situations.”⁷³

Furthermore, Benjamin F. Jackson points out that users of social network websites deserve First Amendment protection because the use of social media encompasses the three of the interests in the First Amendment: freedom of speech, freedom of the press, and freedom of association.⁷⁴ Therefore, there needs to be constitutional protections that users have when they are using these social media websites.

It is important that our First Amendment rights be protected on the internet because the internet facilitates communication for millions of internet users. For example, social media is used to raise funds or bring awareness for any and every crisis that happens around the world.⁷⁵ Recently, nonprofits

68. Hudson, Jr., *supra* note 36.

69. *Id.*

70. *Id.*

71. *Id.*

72. *Id.*

73. *Id.*

74. Benjamin F. Jackson, *Censorship and Freedom of Expression in the Age of Facebook*, 44 N.M. L. REV. 121, 134 (2014).

75. Jacqueline Tabas, *How Nonprofits Can Use Social Media To Increase Donations And Boost Visibility*, FORBES (Mar. 6, 2021) <https://www.forbes.com/sites/allbusiness/2021/03/06/how-nonprofits-can-use-social-media-to-increase-donations-and-boost-visibility/?sh=6bf1b752bb7e>.

and individuals were collecting money for the crisis happening in Afghanistan.⁷⁶ However, the news and crucial links of raising money were suppressed by platform moderators.⁷⁷ Many people saw their Instagram accounts getting less traffic.⁷⁸ Middle East Eye reported that accounts with upward of 500,000 followers were getting banned and posts taken down even if the content that was being shared was from BBC or Al Jazeera.⁷⁹ If the censorship came from the government, then it would be violating the First Amendment. It would then offer individuals a route to receive some kind of government relief. However, there is no such option on these platforms. The social media companies are the judge and jury.

Professor Erwin Cherminsky has said that freedom of speech both helps people make better decisions and that there is an intrinsic benefit from people being able to express their views.⁸⁰ Therefore, any violation to the freedom of speech is harmful to these rights whether that violation be done by private or public actors.⁸¹ It becomes important that we start to compel these companies to protect our rights on the internet. The new era of the civil rights movement will be the fight for our rights on the internet.

There needs to be some censorship on the internet just like we have when speech is used in person. For example, in America there are categories of unprotected speech, regulations of where we can say or do certain things.⁸² It may be frightening that Facebook's Mark Zuckerberg and Twitter's Jack Dorsey say that they will moderate content using a combination of artificial intelligence and humans.⁸³ However, in this process, the AI will filter the obvious cases, and humans will handle the rest.⁸⁴ I am not arguing that we should have no content regulation; however, the issue is that the moderating accounts, without understanding proper context, are creating issues for users. Speech is being regulated without understanding the real meaning.

76. Haleluya Hadero & Glenn Gamboa, *How Instagram star helped rescue dozens from Afghanistan*, AP NEWS (Aug. 28, 2021), <https://apnews.com/article/africa-business-afghanistan-dcf861df931aa3a5441609adaf639b57>.

77. Uddin, *supra* note 8.

78. *Id.*

79. *Id.*

80. Hudson, Jr., *supra* note 36.

81. *Id.*

82. Freedom Forum Institute, *Which Types of Speech Are Not Protected By The First Amendment*, <https://www.freedomforuminstitute.org/about/faq/which-types-of-speech-are-not-protected-by-the-first-amendment/>.

83. Vasatn Dhar, *3 ways to increase social media platforms' transparency*, THE HILL (Nov. 11, 2020), <https://thehill.com/opinion/technology/525991-3-ways-to-increase-social-media-platforms-transparency?rl=1>.

84. *Id.*

Content does need to be filtered as we have seen that many social media accounts have become a haven for hate against minorities and women. Not only do we need the CEOs and internal company review boards to regulate speech properly, but we need our government to step in. Companies need a roadmap and Congress can provide that roadmap. Congress should work in tandem with the companies.

PART IV: Shadow Banning and Negative Impacts

Online social media platform operators realized after the 2016 Presidential elections that they need more proactive policing of accounts to help stop the spread of misinformation.⁸⁵ Platforms have turned to shadow banning to reach that happy medium of content policing. Shadow banning is defined “[A]s a person is shadow banned, their posts on a platform are rendered essentially invisible to everyone but themselves. Their experience using a site may not change — they feel like they are still posting normally — but other people cannot see the material they produce,” according to Monica Stephens, a Professor at the State University of New York at Buffalo, who studies social media.⁸⁶ Refinery29 also explains that “users’ views, likes, and comments aren’t reflective of the size of their following. Or people will suddenly experience a steep drop in engagement that will persist until the video that triggered it is removed or edited.”⁸⁷ Numerous reports have been filed against TikTok on behalf of black content creators that say their videos are taken down or see a loss in viewership even though their videos do not break the community guidelines that are made public.⁸⁸ For example Dill (@dillyonce) is a 25-year-old TikToker based out of Boston.⁸⁹ She says she’s never sworn in her videos or been cited for any violations, yet her videos get taken down regularly.⁹⁰ “For example, I was in Bali on vacation at this beautiful villa and I was recording a TikTok pretending to be the first Black LGBT+ contestant for the *Bachelorette in Paradise*,” she told Refinery29 over email.⁹¹ “The video lacked profanity, vulgarity, and music plagiarism yet upon

85. Tim Hwang, *Dealing with Disinformation: Evaluating the Case for CDA 230 Amendment*, <https://pacscenter.stanford.edu/wp-content/uploads/2018/08/SSRN-id3089442.pdf>.

86. Liam Stack, *What is a ‘shadow ban,’ and is twitter doing it to republican accounts?*, THE N.Y. TIMES (July 26, 2018), <https://www.nytimes.com/2018/07/26/us/politics/twitter-shadowbanning.html>.

87. Michelle Santiago Cortés, *What Is Shadow Banning & Why Are TikTokers Complaining About It?*, REFINERY29 (July 17, 2020), <https://www.refinery29.com/en-us/2020/07/9901461/what-is-shadow-banning-tik-tok>.

88. *Id.*

89. *Id.*

90. *Id.*

91. *Id.*

multiple uploads they were all removed for reasons I still don't understand; it's been very frustrating to be the victim of shadow banning."⁹²

The guidelines are available to anyone whether they have an account or not. Platform operators like Instagram and Facebook argue that they do not shadow ban, but the opposite is seen across the various social media platforms.⁹³ Shadow banning is important when it comes to stopping the spread of misinformation. Shadow banning allows the user to still have an active account, but their content reaches less people. The message such users are trying to spread won't have that same impact. Instead of banning the account which may run into bigger issues, shadow banning an account can be easier. However, issues arise when important content from grassroots organizations and other news related content is being shadow banned. The Farmers' Movement in India is one example. In November 2020 Indian Farmers took to Delhi, the capital of India and demanded that three controversial farm laws be repealed.⁹⁴ These laws were going to take away a lot of protections that farmers have when selling their goods.⁹⁵ This protest became one of the biggest protests in history.⁹⁶ Social media became one way that people on the ground were sharing news and progress about the movement. Many high-profile accounts and other news related accounts have been shadow banned when they share information about the Farmers' Movement in India.⁹⁷

Shadow banning ranges from firsthand videos of what is happening on the ground, daily updates of the protest, and information about the farm laws that are being implemented. Dr. Prabhjot Singh has devoted his whole Instagram account to calling out the atrocities that impact the Sikh Community and highlighting the Farmers' Movement.⁹⁸ He notes that in the beginning he was receiving a lot of traffic, views and likes on his post, but slowly that began to decrease. As the Farmers' Protest was gaining momentum in India the hashtag #Sikh was blocked on Instagram.⁹⁹ This is an issue because crucial

92. *Id.*

93. Cook, *supra* note 9.

94. *Farm laws: India farmers end protest after government accepts demands*, BBC, <https://www.bbc.com/news/world-asia-india-59566157>.

95. *Id.*

96. Mandeep Rai Dhillon, *The world's largest protest you've probably never heard of*, CNN (Mar. 6, 2021), <https://www.cnn.com/2021/03/06/opinions/india-farmer-protests-mandeep-rai-dhillon/index.html>.

97. Tom Raftery, *Supply Chain Continuous Improvement – A Chat with Prabhjot Singh*, THE DIGITAL SUPPLY CHAIN PODCAST (Mar. 30, 2021), <https://www.digitalsupplychainpodcast.com/354320/10183558>.

98. *Id.*

99. Sakshi Venkatraman, *Facebook and Instagram unblock #Sikh after restricting it for months*, NBC NEWS (June 9, 2021), <https://www.nbcnews.com/news/asian-america/facebook-instagram-unblock-sikh-after-%20restricting-it-months-n1227656>.

information about the protest was being shared by using this hashtag. There was no explanation given of why this hashtag was blocked or even why accounts similar to Dr. Singh's are being shadow banned. This is leaving many Sikh users feeling like their voice is being censored on the social media platforms.

The worst part of shadow banning is that the system is built on inequality. There are some high-profile accounts that can break all the rules they want and never get punished.¹⁰⁰ The Wall Street Journal reports that in private Facebook has built a system that exempts high-profile users from some or all of its rules.¹⁰¹ This phenomenon is coined whitelisted- where VIP users are shielded from enforcement actions and are allowed to post rule violating material pending reviews from Facebook.¹⁰² Those reviews in reality never come (XCheck).¹⁰³ This means that there are users than violate community guidelines but not face any repercussions.¹⁰⁴ There are many examples where this has happened.¹⁰⁵

For instance, international soccer star Neymar posted nude photos of a woman who had accused him of rape and this was shared widely before the content was removed by Facebook.¹⁰⁶ Furthermore, whitelist accounts shared information that Facebook's own fact checkers flagged as false and inflammatory.¹⁰⁷ This included information that vaccines are deadly, that Hillary Clinton had covered up "pedophile rings."¹⁰⁸ It was found that these users faced no consequences.¹⁰⁹ In another instance, some political whitelist users were sharing articles from an alternative-medicine websites claiming that a doctor in Berkeley, California, claims that chemotherapy doesn't work 97% of the time.¹¹⁰ Fact-checking organizations debunked the claim and noted both that the Science is misrepresented and that the doctor cited in the article died in 1978.¹¹¹ On the other hand, there are accounts that share grass-root news and crucial information around mobilization that see decreased engagement on their accounts, are shadow banned and sometimes their

100. Horwitz, *supra* note 38.

101. *Id.*

102. *Id.*

103. *Id.*

104. *Id.*

105. *Id.*

106. Horwitz, *supra* note 38.

107. *Id.*

108. *Id.*

109. *Id.*

110. *Id.*

111. Horwitz, *supra* note 38.

accounts get deleted altogether.¹¹² These whitelist users enjoy privileges that many others are not privy too. The number of users that are exempt from the so-called important guidelines has grown to 5.8 million users.¹¹³ Facebook has lied to not only the users but also to its own internal Oversight Board.¹¹⁴

It is clear that there are so many issues with the way online social media platforms are censoring information. They are shadow banning content that is not always harmful.¹¹⁵ Yet they are letting over five million users blatantly violate the rules.¹¹⁶ There is no uniformity to the way the rules are being applied.¹¹⁷ Such instances strengthen the argument that there needs to more transparency, and not just from whistle blowers or from secret investigations. We need transparency from the companies themselves. The social media platforms need to be held accountable because it can be a violation of our First Amendment right to freedom of speech which needs to be extended to social media platforms.

PART V: Solutions

A. Transparency

As the internet grows and more people join social media platforms, policing of these platforms is critical. Online social media platform policing is necessary to create a safeplace for users. However, social media platforms need accountability for the policing they do. The accountability is followed with transparency in guidelines, an internal review board, diversity in the C-suites, legislative solutions and better trained Artificial Intelligence that can detect context. The issues that the internet has today can only be tackled by a comprehensive list of solutions.

Professor Vasant Dhar is a Professor at the Stern School of Business at New York University, and he argues that “platforms should moderate accounts, not posts.”¹¹⁸ Professor Dhar says that if an account is not violating the law or platform’s rules then they should not be censored.¹¹⁹ This requires a deeper understanding of the user and not just a superficial one that the AI provides. Professor Dhar’s theory is comprehensive because it provides the AI with more context when determining if an account needs to be censored.

112. Uddin, *supra* note 8.

113. Horwitz, *supra* note 38.

114. *Id.*

115. Uddin, *supra* note 8.

116. Horwitz, *supra* note 38.

117. *Id.*

118. Dhar, *supra* note 83.

119. *Id.*

If the account is well connected and or publishes content a lot, then Professor Dhar mentions these accounts could be in a higher category.¹²⁰ Platforms must vet who they are hosting and why.¹²¹ At the moment, the platforms censor people if they share one or two controversial ideas.¹²² Instead, they need to see if someone is constantly posting controversial content, but then also assess if the posting is controversial or simply just spreading news.

Currently, each online social media platform has their own set of guidelines that they use to determine what kind of content stays on the website and what kind of content must be removed. For example, as mentioned above, Instagram lists as one of its guidelines when sharing photos, users have the right to share the photo, not post nude photos, no spam posting, not to post anything that is against the law (drugs, terrorism organized crime or hate groups), post news content without intense, graphic violence.¹²³ So, theoretically, if one follows these rules, then their post should not be taken down. However, the questions users are left with is then why is their content taken down when they follow the guidelines. At some points it is not clear as to what guideline is not followed because users are not given enough information about how they broke the community guideline. Even though the guidelines have examples, fact checkers that survey the posts can have a difficult time in understanding the context in which they are made. Twitter says that they do an analysis of context, but the data on how many posts they look at for context and what they do internally when these accounts do violate community guidelines is not available.¹²⁴ Without the context this leaves people in the dark when their posts are taken down. This then creates issues regarding whether social media platform is legitimate and creates user black lash. In today's society where every company is worries about reputation and image, if platforms cannot gain user trust, then people will leave the platforms and move on to the next big thing.¹²⁵ Therefore, offering transparency is the best thing for platforms.

120. *Id.*

121. *Id.*

122. *Id.*

123. Facebook, *supra* note 16.

124. Twitter Help Center, *Our approach to policy development and enforcement philosophy*, TWITTER, <https://help.twitter.com/en/rules-and-policies/enforcement-philosophy>.

125. Seoyeon Hong and Sookwang Klive Oh, *Why People Don't Use Facebook Anymore? An Investigation Into the Relationship Between the Big Five Personality Traits and the Motivation to Leave Facebook* (July 10, 2020), <https://internal-journal.frontiersin.org/articles/10.3389/fpsyg.2020.01497/full#%3A~%3Atext%3DThe%20number>.

B. Internal Review Boards

The other major reform that needs to happen for these online platforms is the Internal Review Board (IRB) similar to the medical field.¹²⁶ This would be a board that is internal in the tech company and would be used every time the company is coming out with a new AI program.¹²⁷ The technology companies would set it up with company funds and the board members will be properly trained. In order to make sure the board is nonbiased; it would have one to two neutral members that are not a part of the company.¹²⁸ These individuals would not be impacted in any way if the company does or does not implement the program. All these members should get cultural sensitivity training which will help them in understanding how different cultures operate and understanding the tech from that lens. The other members would be lawyers, engineers, individuals experienced in ethics, AI experts etc.²¹ This board can even have a focus group that can be used to test out the AI prior to putting it out in the market. One issue with the tech industry is that technology is put out before its seen what kind of impact it will have on users. Once the internal IBR disapproves the idea then it can be appealed until the issues with the tech are fixed. The negative impacts of AI content tracking are not yet fully realized. However, we have seen some of the negative impacts of shadow banning and taking down of accounts can have huge impacts on social movements. The IBR can help mitigate those risks especially because the IBR is part of multi-prong approach. The solutions require a human to take a look at the issue from a nuanced and granular level while the AI approaches issues from a broader level.

AI can be a great resource to quickly sort through content and get rid of anything that breaks community guidelines, but AI needs to be trained in context better.¹²⁹ The major issue with AI is that it can flag things as hate speech or spread of misinformation because it does not understand context. Tech experts need to create AI that can pick up on context. A team of USC researchers created a hate speech classifier that is more context sensitive.¹³⁰ This means the AI is not just picking on words, but also looking at how the

126. Reid Blackman, *"If Your Company Uses AI, it needs an Institutional Review Board."*, HARV. BUS. REV. (Apr. 1, 2021), <https://hbr.org/2021/04/if-your-company-uses-ai-it-needs-an-institutional-review-board>.

127. *Id.*

128. *Id.*

129. Citron, *supra* note 13.

130. Caitlin Dawson, *Context Reduces Racial Bias in Hate Speech Detection Algorithms*, USC VITERBI, <https://viterbischool.usc.edu/news/2020/07/context-reduces-racial-bias-in-hate-speech-detection-algorithms/#:~:text=A%20team%20of%20USC%20researchers,never%20mind%20artificial%20intelligence%20systems>.

words play into the post. So, for example just because the “N” word is in a post does not automatically break the rules. It’s look at the entire post and seeing the context in which the N word is used in. This makes it less likely to mistake a post containing a group identifier as hate speech. Currently AI is oversensitive to group identifiers like “black,” “gay,” or “transgender,” which are only indicators of hate speech when used in some settings.²³ The notion is not to get rid of AI and have people manually review every post that is flagged, but to train AI much better in conjunction with some manual review.

C. Diversity in Tech Companies

One cause of ineffectiveness in the tech industry is the lack of diversity. Diversity in the C-suite for tech companies is needed because the voice of minorities and women needs to be represented more. Minorities and women are the ones that can get impacted the most if platforms over police or do not police at all.¹³¹ Professor Mary Anne explains that in the 1990s, there were websites filled with content of sexual violence against women, harmful content from Ku Klux Klan and the Aryan Liberty Net members.¹³² A student from the University of Michigan posted comments on a message board about raping, torturing, and killing a female from his school.¹³³ However, the creators of Section 230 were not worried about creating a safe space for women on the internet.¹³⁴ They were more worried about government intervention in the cyberspace.¹³⁵ This set precedent for what is allowed on websites today, and we still see such websites/ comments thrive on some of the social platforms. Therefore, making sure the C-suites are filled with people from diverse backgrounds will help in spotting issues better and leads to more creative solutions.

The recent diversity changes in the tech industry are a big win for the companies. Recently Twitter announced Parag Agrawal as CEO joining the ranks of other South Asians CEOs for major tech companies.¹³⁶ Diversity is not the only solution to the problems in tech and social media, but it is a start.

131. Michel Martin, *Internet Harassment Of Women: When Haters Do More than Just Hate*, NPR, <https://www.npr.org/2014/01/08/260757625/internet-harassment-of-women-when-haters-do-more-than-just-hate>; Office of the High Commissioner, *Report: Online hate increasing against minorities, says expert*, U.N. HUM. RTS. (Mar. 23, 2021), <https://www.ohchr.org/en/stories/2021/03/report-online-hate-increasing-against-minorities-says-expert>.

132. Franks, *supra* note 37.

133. *Id.*

134. *Id.*

135. *Id.*

136. Nikhil Inamdar and Aparna Alluri, *Why Indian-born CEOs dominate Silicon Valley*, BBC NEWS (Dec. 4, 2021), <https://www.bbc.com/news/world-asia-india-59457015>.

In a Forbes' article, data from a global survey that the more gender diversity there is in a company, the higher the performance and higher the financial results.¹³⁷ Diversity helps to develop a connection with customers, and it can create trusts.¹³⁸ If people on the team look like the customers, more people are attracted to the product. Adding diversity to the C-suites and overall company will help build trust with customer, increase products, and avoid unnecessary censorship of content.

D. Legislative Solution

In order to create real change, we need the government to step in. Internal changes will only go so far. We cannot solve the issues with censorship, freedom of speech without some government intervention. Congress has been holding hearing after hearing regarding issues with how social media companies are operating.¹³⁹ Since then, it has been years and Congress has taken no action to curb the harmful impact that social media is having on our First Amendment rights, the negative impacts that social media has on children and young adults. The Wall Street articles describe research that indicates that new changes to the News Feed algorithm had inadvertently rewarded "misinformation, toxicity, and violent content."¹⁴⁰ Facebook describes the feed as "the constantly updating list of stories in the middle of your home page. The Feed is the first page of the social media platforms that includes status updates, photos, videos, links, app activity and likes from people, Pages, and groups that you follow on Facebook."¹⁴¹ The feed works in similar ways on other social media platforms. Despite knowing about these effects, Mark Zuckerberg resisted fixing the problem.¹⁴² After these hearings and evidence of the negative impact of social media Congress has left the companies to regulate themselves because there have been no new bills being passed after these hearings.¹⁴³ Thinking that these companies

137. FORBES TECH. COUNCIL, *12 Ways Diversity Makes a Difference in Tech*, FORBES (July 12, 2018), <https://www.forbes.com/sites/forbestechcouncil/2018/07/12/12-ways-diversity-makes-a-difference-in-tech/?sh=37b958f62bc6>.

138. *Id.*

139. Miles Parks and Shannon Bond, *5 Takeaways From Big Tech's Misinformation Hearing*, NPR, <https://www.npr.org/2021/03/25/981203566/5-takeaways-from-big-techs-misinformation-hearing>.

140. Keach Haggey and Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead*, THE WALL ST. J. (Sept. 15, 2021), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215?mod=article_inline.

141. FACEBOOK HELP CTR., *How Feed Works*, FACEBOOK, <https://www.facebook.com/help/1155510281178725>.

142. Haggey et. al, *supra* note 140.

143. Cecilia Kang, *Congress, Far From 'a Series of Tubes,' Is Still Nowhere Near Reining In Tech*, N.Y. TIMES, <https://www.nytimes.com/2021/12/11/business/congress-tech-regulation.html>.

may be able to fix the problems themselves is optimistic. Congress needs to step up and regulate the way these companies are doing business.¹⁴⁴ This is a pressing issue that needs to be addressed.

Congress needs to take the lead, like some other countries have, to combat the negative impacts of technology.¹⁴⁵ China started by limiting kids' video game time.¹⁴⁶ The reasoning being that children and teenagers need a real break. I do not think Congress needs to take such drastic measures, but it must do something. Even after changes to the algorithms, researchers have noted in internal memos that "Misinformation, toxicity, and violent content are inordinately prevalent among reshares."¹⁴⁷ This inadvertently makes certain voices louder making them show up on users' feeds more.¹⁴⁸ Even though Facebook has said that it is "gradually expanding some tests to put less emphasis on signals such as how likely someone is to comment or share political content."¹⁴⁹ However, this is not going to change the problem. We do not need less news being shared on the internet. We need accurate news and grassroots organizers to not be penalized. Unfortunately, this is not what is happening on the social media platforms. This shift in Facebook to reduce political content being shared comes after it was criticized for playing a role in the January 6th riot in D.C.¹⁵⁰

Professor Dhar has also called on Congress to regulate social media platforms in hopes of creating more transparency. Professor Dhar equates the transparency in social media to the regulation of the U.S. financial system. He argues transparency is the reason that the financial system is the most trusted in the world.¹⁵¹ It's true because, as he mentions, without proper oversight, these financial institutions have manipulated markets, behaved in unethical behavior, destroyed audit trails, and participated in shady financial arrangements.¹⁵² This was seen in the 2008 recession.¹⁵³ Professor Dhar

144. Haggey et. al, *supra* note 140.

145. Sharon Pruitt-Young, *In China, Kids Are Limited To Playing Video Games For Only 3 Hours Per Week*, NPR, <https://www.npr.org/2021/08/30/1032489883/china-kids-video-games-limits>.

146. *Id.*

147. Haggey et. al, *supra* note 140.

148. *Id.*

149. *Id.*

150. Craig Silverman, Craig Timberg, and Jeremy B. Merrill, *Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan 6 Attack, Records*, PROPUBLICA (Jan. 4, 2022), <https://www.propublica.org/article/facebook-hosted-surge-of-misinformation-and-insurrection-threats-in-months-leading-up-to-jan-6-attack-records-show>.

151. Dhar, *supra* note 83.

152. *Id.*

153. *Id.*

argues that these platforms are new cowboys, but the risks are much higher if we don't reign them in.¹⁵⁴ He advocates that the platforms need to provide verifiable audits of the interventions they have taken.¹⁵⁵ The audits include targeting which is similar to what media companies are already required to do.¹⁵⁶ This way we can track what the algorithms are doing versus us being manipulated by the data these algorithms provide.¹⁵⁷ He argues this will help us understand what other kind of damage beside the negative effect of polarization and echo chambers are having on society.¹⁵⁸ These things are important to know because we need to understand as humans to what extent the choices we make are manipulated by social media platforms and how much psychological damage these platforms really do. We get to see further how problematic social media platforms are through documentaries like *The Social Dilemma*. In the documentary, former employees from the social media platforms explain how damaged things are in the way these companies run.¹⁵⁹

Currently, social media platforms have no real incentive to change their ways. The fines that these companies get or the cost that the companies have to bear flying their executives to D.C. can simply count as operating costs. It is the cost of doing business which can be recuperated through increasing prices for placing ads or various other mechanism. However, these audits will hold the CEOs more accountable. It will offer concrete proof of the impact the social media platforms have and make the information readily available to Congress so they can take proper action. This hopefully incentivizes them to share the information to Congress, but also to the users that are on the platforms.

Professor Dhar also proposes two other changes. One of which is that transparency of the platforms. As mentioned above this can done through audits. These audits would consist of reports of when the social media companies intervene when users violate community guidelines, what that intervention looks like, trails of important platform activity, why some users are targeted.¹⁶⁰ This will be helpful in creating laws to regulate them properly, but also so that their impact on society can be studied. To create effective laws, we need to know the impacts that the social media platforms have on society. These audits will help to see what kind of content is being prioritized

154. *Id.*

155. *Id.*

156. *Id.*

157. Dhar, *supra* note 83.

158. *Id.*

159. *THE SOCIAL DILEMMA* (Exposure Labs 2020).

160. Dhar, *supra* note 83.

and allow policy makers to reprimand the social media platforms if they go astray. Research now shows that densely connected networks can result in echo chambers that sow division and increase polarization.¹⁶¹ This polarization is what can cause problems especially when news is being shared online.

Congress taking the lead on regulating the tech industry brings its own issues. For instance, many policy makers do not even have an Instagram, Twitter or Facebook account. They don't understand the business model. However, the government cannot feign ignorance and decline to create proper laws. They can easily get up to date by hiring staffers to properly educate them on the topic. They can create committees that can dive into the issues in a more nuanced way. Simply not understanding technology is not an excuse as to why there is a lack of regulation.

Conclusion

An increase in social media users has led to many opportunities for news to spread quickly; however, it has also allowed the spread of misinformation. To combat the spread of misinformation, online social media platforms have started to police the content that is posted. Platform policing does serve a positive purpose. It prevents harmful content from reaching users. It prevents hate speech reaching vulnerable user and Section 230 shields platforms from liability when they take down this information, but also protected them if they are not able to. However, many times the information that is taken down can be crucial to grassroots movements. This is where there needs to be proper training of AI, diversification of the C-suites, legislative reform and implementation of the IRB. The multi-prong approach can create a safe space for minorities who are the victims of hate speech while also protecting women from sexual violence. The comprehensive changes can lead to more trust in users and legitimacy back to the platforms.

The current issues of shadow banning and the harmful effects that social media has on our freedom of speech is not something that can just be fixed internally. Social media companies not only need AI, but also more sophisticated employees that understand the various cultural backgrounds the users come from. This means that when content is flagged it's not just the AI looking at it, but employees that can decipher what is or is not harmful content.

Moreover, we need Congress to step in and regulate. We need Congress to create transparency laws which the tech companies can abide by. We need the government to create a framework that the social media companies can follow. Therefore, it is not just an inside job, but also something that our government needs to investigate. It is not enough to have fines since those

161. *Id.*

finer just become operating costs that social media companies are okay paying. We need to create regulations that can help users understand what the tech companies are doing and offer tech companies the rules under which they need to operate.
