Platform Lie-ability

Social media platforms should be held liable and morally responsible for circulating

disinformation and harassment that leads to racism, slander, and offline violence. Section 230 of

the Communications Decency Act exempts platforms from legal responsibility for what is posted

to their sites, stating that "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"[1]. This section must be reformed, for social medias' freedom from liability allows for

the flow of unmoderated information in a world where billions of people are absorbing online

sources. 510,000 comments are posted on Facebook and 456,000 tweets are posted on Twitter

each minute[2]. Users exercise their freedom of speech at sometimes harmful lengths, and

platforms either ignore these posts or remove them too late due to neglect and fear of

overstepping. Section 230 allows platforms to "profit off of hosting harmful content without

having to bear the costs of cleaning it up"[3]. Social media, a potentially democratic tool, has

become a town square of hatred and lies, facilitating racism, religious slander, and harassment

that can incite violence if not controlled properly.

Social media fosters and amplifies misinformation in a landscape of blind trust in online

sources. The number of people getting their news from platforms like Twitter and Facebook

constantly increases, especially amongst younger generations, with 1 in 5 Americans saying that

they receive political news from social media instead of television, news websites, or print

media[4]. Citizens turn to these free platforms in an attempt to become informed, which becomes a

---

[1] https://www.eff.org/issues/cda230

[2] https://www.bernardmarr.com/default.asp?contentID=1438#:~:text=More%20than%20300%20million%20photos,posted%20and%20293%2C000%20statuses%20updated

[3] https://www.wired.com/story/section-230-reform-safe-tech-act/

[4]

problem when there is no rule governing platforms to take down information. Facebook, a site from which 36% of Americans get their news[5], refers to untrustworthy news sources 15% of the time[6]. Falsehoods on Twitter are 70% more likely to be retweeted than actual news[7]. Companies wait too long to remove harmful posts, allowing them to be spread virally before deletion. In 2020, rumors on Facebook spread false claims that left-wing activists started deadly wildfires in Oregon[8]. These unfounded posts stayed on the site for several days. Facebook only took them down after police and the FBI debunked the rumors. Disinformation and conspiracies aren't just available — they are promoted by social media's algorithms to gain more views. Facebook has started tweaking its services in attempt to keep users on their site, "amplifying the impact of fabricated and sensational stories"[9]. Former President Donald Trump thrived off the algorithm's amplification of incredulous claims, using Twitter as the "core platform for spreading disinformation about the election"[10]. Thirty-six of Trump's top 100 Twitter posts contained false election information[11]. Trump's third-most popular tweet, receiving 1.2 million favorites and 191 thousand retweets, stated he had "WON THIS ELECTION, BY A LOT," a lie that sparked the right-wing activists' storming of the United States Capital[12]. In 24 days post-election, Trump

https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/

[5] https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/

[6] https://www.forbes.com/sites/traversmark/2020/03/21/facebook-spreads-fake-news-faster-than-any-other-social-website-according-to-new-research/?sh=292acc6a6e1a

[7] https://www.nytimes.com/2018/10/29/technology/hate-on-social-media.html

[8] https://www.npr.org/2020/09/22/915555286/what-can-social-media-do-to-slowdown-the-spread-of-misinformation

[9] https://www.nytimes.com/2018/01/14/technology/facebook-news-feed-changes.html

[10] https://www.cnbc.com/2021/01/13/trump-tweets-legacy-of-lies-misinformation-distrust.html

[11] https://www.cnbc.com/2021/01/13/trump-tweets-legacy-of-lies-misinformation-distrust.html

[12] https://www.cnbc.com/2021/01/13/trump-tweets-legacy-of-lies-misinformation-distrust.html

racked up 200 Twitter flags for disinformation[13]. Trump's Twitter was only removed after the Capitol Riot on January 6th.

The misinformation that circulates on social media platforms has the potential to escalate into political and ethnic violence. Fake news is posted to rile emotions or create rumors which led some to take physical action. This was demonstrated through the riot at the Capitol following a stream of incendiary tweets from Trump to "fight like hell" for the "stolen" election[14], leading to the death of five people[15]. This happens internationally as well. Facebook played the key role inciting real-world violence in Myanmar after episodes of violence against Rohingya refugees. The United Nations launched an investigation and concluded that Facebook was a determining cause in the genocide. The site, which aided in "foment[ing] division and incit[ing] offline violence," went unpatrolled by Facebook's employees[16]. Facebook "missed a crescendo of posts and misinformation that helped to fuel modern ethnic cleansing in Myanmar" as it lacked content moderation[17]. Indonesia, Facebook's third largest market, is another country assaulted by unchecked misinformation[18]. In 2017, rumors circulated of gangs abducting children and selling their organs. This resulted in the lynching of outsiders in nine different villages across the country.

Unchecked rumors that are spread on platforms can also lead to violence against religious groups. In 2018, a rumor circulated on Facebook with assumptions that a man's death was tied to a Muslim plot to rid Sri Lanka of Buddhists[19]. These lies built up rage amongst the community

13 https://variety.com/2020/digital/news/twitter-trump-200-disputed-misleading-claims-election-1234841137/
14 https://apnews.com/article/fact-check-trump-us-capitol-remarks-221518bc174f9bc3dd6e108e653ed08d
15 https://www.nytimes.com/2021/01/11/us/who-died-in-capitol-building-attack.html
16 https://about.fb.com/news/2018/11/myanmar-hria/
17 https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html
18 https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html
19 https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html

and caused extremists to create Facebook Groups used to plan violent Muslim attacks as revenge. The posts were "'the embers beneath the ashes' of Sinhalese anger," leading to the burning of a man at the stake[20]. Facebook played the main role in every step from "rumor to killing" by repeatedly ignoring violent and hateful posts. Another false rumor that police had found 23,000 sterilization pills from a Muslim pharmacist in Ampara spiraled into the beating of an innocent restaurant owner, the destruction of his shop, and the burning of a local mosque[21]. An online syndicate, Saracen, was uncovered by Indonesian police in 2016; yet they continue to run unsupervised on Facebook[22]. They uploaded incendiary posts disparaging government officials, insulting religions and ethnicities, and spreading fake news. One post slandered a former governor of Jakarta with a doctored video of anti-Islam comments.

Unfiltered speech on social media causes harm by perpetuating communities of hate. Harassment spreads rapidly through social media. Pew Research found that 41% of Americans have experienced some form of online harassment[23]. "Communal hatreds" overrun Facebook walls without legal oversight by the government[24]. Some platforms have policies banning hate speech, yet content that falls under this category is often not removed promptly, if at all. Harassment on social media rarely gets checked unless attention is drawn. In 2018, there were 11,696 Instagram posts with the hashtag "#jewsdid911"[25]. This hashtag has been deleted, yet others still stand such as "#jewsdidtheholocaust" with 382 posts[26] and "#88", an abbreviation for a Nazi salute, with 1.3 million posts[27]. Actress Leslie Jones was the target of cruel hatred from

---

[20] https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html
[21] https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html
[22] https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html
[23] https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/
[24] https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html
[25] https://www.nytimes.com/2018/10/29/technology/hate-on-social-media.html
[26] https://www.instagram.com/explore/tags/jewsdidtheholocaust/
[27] https://www.instagram.com/explore/tags/88/

internet trolls after the release of the new *Ghostbusters* movie[28]. Users called her slurs, attacked her image, and compared her to an ape. Jack Dorsey, Twitter's CEO, reached out to Jones and suspended or banned multiple accounts. Jones said that "hate speech and freedom of speech are two different things," urging platforms to more promptly take down hate messages.

Online harassment can foreshadow escalations to real world violence. Social media gives extremists a platform to publicize their opinions, find others that share their views, and spew a constant stream of hateful propaganda[29]. White-supremacists have "used technology in a way that has been unbelievably effective at radicalizing people"[30]. Wade Michael Page is one man who used Gab, an online forum, to spread his hatred and find like-minded people, urging them to "stand and fight"[31]. Page raided a Sikh temple in a Milwaukee suburb and killed six people in 2012. Robert D. Bowers expressed his hatred of Jews on Gab and proceeded to kill eleven people at the Tree of Life Synagogue in Pittsburgh in 2018. The gaming site Discord was used by extremists to plan the Unite the Right rally against the removal of the Robert E. Lee statue in Charlottesville[32]. They discussed logistics of the protest, including what weapons they would bring and legality of ramming a car into a crowd, directly preceding the automobile-induced death of Heather Heyer[33]. Facebook promised to take down white supremacist content in March 2019, and in some cases they have done so[34]. Facebook and Twitter banned David Duke[35] and James Allsup[36], two alt-right men who had large presences on social media, yet other volatile

[28] https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html
[29] https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons
[30] https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/
[31] https://www.washingtonpost.com/world/national-security/excessive-drinking-cost-wade-michael-page-military-career-civilian-job/2012/08/07/274ccc7a-e095-11e1-a421-8bf0f0e5aa11_story.html?utm_term=.67317353a4f7&itid=lk_inline_manual_7
[32] https://publicintegrity.org/politics/social-media-where-voices-of-hate-find-a-place-to-preach/
[33] https://twitter.com/UR_Ninja/status/897979666610491394
[34] https://www.theguardian.com/technology/2019/mar/27/facebook-white-nationalism-hate-speech-ban
[35] https://www.cnn.com/2020/07/31/tech/david-duke-twitter-ban/index.html
[36] https://www.kxly.com/facebook-instagram-delete-controversial-figure-james-allsups-accounts/

users still remain. The Twitter user @SAgenocide is one racist account filled with South African

bashes and has aimed to "stop the brutal murders of whites"[37].

There exists United States code suggesting social media platforms should be held liable

for the posts of their users; but there is also a moral obligation that rests with the companies. The

United States law states that "providing 'communications equipment' to terrorist organizations is

materially aiding terrorist organizations... [social media is] consciously providing a

communication service for terrorists and giving them an audience"[38]. These platforms are aiding

and abetting to criminal activity when posts on their sites that incite violence are not removed.

Social media companies might also be held liable under Section 2101 in Title 18 of the United

States Code[39]. This title states that it is illegal to use "any facility of interstate or foreign

commerce...to incite a riot… to commit any act of violence in furtherance of a riot". This means

platforms should be held responsible for information posted that incites riots or violence. Social

media companies have a moral imperative to remove misinformation and harassment from their

platforms even when legal code is murky. Platforms have the ability to imperceptibly change our

behavior and opinions, crafting a world where "each person has their own reality with their own

facts"[40]. Facebook COO Sheryl Sandberg said that she believed Facebook has both a moral and

legal obligation to take down accounts that are incentivizing violence during a Capitol Hill

hearing in front of the Intelligence Committee in 2018[41]. Rafi Martina, when asked about

platforms' moral obligations, said that "Facebook is still an engine for incitement all over the

---

[37] https://twitter.com/SAgenocide

[38] http://www.fletcherforum.org/the-rostrum/2020/4/7/social-media-laws-amp-terrorism

[39] https://www.law.cornell.edu/uscode/text/18/2101

[40] https://www.netflix.com/watch/81254224?source=35

[41] https://www.intelligence.senate.gov/hearings/open-hearing-foreign-influence-operations%E2%80%99-use-social-media-platforms-company-witnesses#

world against vulnerable ethnic, religious, and racial minorities"[42]. A reform of Section 230 is important in protecting groups often marginalized by online sites and protecting individuals from harmful disinformation.

The growing attention towards the lack of platform liability has led to the development of potential amendments to Section 230. Senator Mark Warner proposed the SAFE TECH Act this February, which, if passed, would make platforms legally responsible for unremoved content exemplifying stalking, harassment, or intimidation[43]. This act will not cover the removal of disinformation from sites but is a start to placing responsibility on those who should be moderating content. It is up to platforms, whether a law is passed or not, to "exercise editorial discretion and do some curation to maintain civil discourse on the platform"[44]. Users themselves can aid in calling out platforms, as targeting misinformation is difficult to legally address. Martina says that consumers must put pressure on platforms and call for more protective policies in "ways that the government can't do". Users should also follow a variety of accounts to digest diverse information to ensure unbiased content[45]. The need for social media companies to handle misinformation and harassment can be a matter of life and death — Section 230 must be reformed to put more liability on platforms for the harmful lies and hate messages circulating on their sites.

---

[42] https://collab.its.virginia.edu/portal/site/8c6b3550-1548-4fa5-9f08-17064d2c4c8a/page/31c93c76-834b-4354-bb8b-36f16bf0e12b

[43] https://www.wired.com/story/section-230-reform-safe-tech-act/

[44] https://www.npr.org/2020/05/28/864410781/new-executive-order-to-expose-social-media-companies-to-more-liability-for-conte

[45] https://news.usc.edu/179176/how-to-help-stop-fake-news-misinformation-usc-experts/

Work Cited

Beckett, Lois. "Facebook to Ban White Nationalism and Separatism Content." *The Guardian*,

    Guardian News and Media, 27 Mar. 2019,

    www.theguardian.com/technology/2019/mar/27/facebook-white-nationalism-hate-

    speech-ban.

Bond, Shannon. "What Can Social Media Do To Slow Down The Spread Of Misinformation?"

    NPR Radio IQ, NPR, 22 Sept. 2020, www.npr.org/2020/09/22/915555286/what-can-

    social-media-do-to-slowdown-the-spread-of-misinformation.

Chang, Alisa. "New Executive Order To Expose Social Media Companies To More Liability For

    Content." *NPR Radio IQ*, NPR, 28 May 2020, www.npr.org/2020/05/28/864410781/new-

    executive-order-to-expose-social-media-companies-to-more-liability-for-conte.

Edelman, Gilad. "Finally, an Interesting Proposal for Section 230 Reform." *Wired*, Conde Nast,

    5 Feb. 2021, www.wired.com/story/section-230-reform-safe-tech-act/.

Effron, Oliver. "David Duke Has Been Banned from Twitter." *CNN*, Cable News Network, 31

    July 2020, www.cnn.com/2020/07/31/tech/david-duke-twitter-ban/index.html.

Frenkel, Sheera, et al. "In Some Countries, Facebook's Fiddling Has Magnified Fake News." *The*

    New York Times, The New York Times, 14 Jan. 2018,

    www.nytimes.com/2018/01/14/technology/facebook-news-feed-changes.html.

Frenkel, Sheera, et al. "On Instagram, 11,696 Examples of How Hate Thrives on Social Media."

    The New York Times, The New York Times, 29 Oct. 2018,

    www.nytimes.com/2018/10/29/technology/hate-on-social-media.html.

Gardner, Kianna. "Social Media: Where Voices of Hate Find a Place to Preach." *Center for*

    Public Integrity, 30 Aug. 2018, www.publicintegrity.org/politics/social-media-where-

    voices-of-hate-find-a-place-to-preach/.

Gersema, Emily. "How Americans Can Help Stop Fake News and Misinformation." *USC News*,

    University of Southern California, 1 Dec. 2020, www.news.usc.edu/179176/how-to-help-

    stop-fake-news-misinformation-usc-experts/.

Hatzipanagos, Rachel. "Perspective | How Online Hate Turns into Real-Life Violence." *The*

    Washington Post, WP Company, 30 Nov. 2018,

    www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-

    life-violence/.

Healy, Jack. "These Are the 5 People Who Died in the Capitol Riot." *The New York Times*, The

    New York Times, 11 Jan. 2021, www.nytimes.com/2021/01/11/us/who-died-in-capitol-

    building-attack.html.

"Hearings." *Hearings | Intelligence Committee*, U.S. Senate, 5 Sept. 2018,

    www.intelligence.senate.gov/hearings/open-hearing-foreign-influence-

    operations%E2%80%99-use-social-media-platforms-company-witnesses#.

"How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read."

    Bernard Marr, 123 Internet,

www.bernardmarr.com/default.asp?contentID=1438#:~:text=More%20than%20300%20
million%20photos,posted%20and%20293%2C000%20statuses%20updated.

"#Jewsdidtheholocaust." *Instagram*, www.instagram.com/explore/tags/jewsdidtheholocaust/.

Koczur, Shayna. "It's Time to Reevaluate Legal Standards for Social Media's Role in
Radicalization." The Fletcher Forum of World Affairs, Tufts University, 29 Apr. 2020,
www.fletcherforum.org/the-rostrum/2020/4/7/social-media-laws-amp-terrorism.

Laris, Michael, et al. "Excessive Drinking Cost Wade Michael Page Military Career, Civilian
Job." *The Washington Post*, WP Company, 7 Aug. 2012,
www.washingtonpost.com/world/national-security/excessive-drinking-cost-wade-
michael-page-military-career-civilian-job/2012/08/07/274ccc7a-e095-11e1-a421-
8bf0f0e5aa11_story.html?utm_term=.67317353a4f7&itid=lk_inline_manual_7.

Laub, Zachary. "Hate Speech on Social Media: Global Comparisons." *Council on Foreign
Relations*, Council on Foreign Relations, 7 June 2019, www.cfr.org/backgrounder/hate-
speech-social-media-global-comparisons.

Mackey, Aaron, and Corynne McSherry. "Section 230 of the Communications Decency Act."
Electronic Frontier Foundation, www.eff.org/issues/cda230.

Mitchell, Amy, et al. "Americans Who Mainly Get Their News on Social Media Are Less
Engaged, Less Knowledgeable." *Pew Research Center's Journalism Project,* Pew
Research Center, 27 Aug. 2020, www.journalism.org/2020/07/30/americans-who-mainly-
get-their-news-on-social-media-are-less-engaged-less-knowledgeable/.

Martina, Rafi. *The News Media Spring 21 Apr 7 Lecture*. *UVACollab*, University of Virginia, 7

    Apr. 2020, www.collab.its.virginia.edu/portal/site/8c6b3550-1548-4fa5-9f08-

    17064d2c4c8a/page/31c93c76-834b-4354-bb8b-36f16bf0e12b.

Orlowski, Jeff, director. *The Social Dilemma*. *Netflix*,

    www.netflix.com/watch/81254224?source=35.

Rattner, Nate. "Trump's Election Lies Were among His Most Popular Tweets." *CNBC*, CNBC,

    13 Jan. 2021, www.cnbc.com/2021/01/13/trump-tweets-legacy-of-lies-misinformation-

    distrust.html.

Robinson, Erin. "Facebook, Instagram Delete Controversial Figure James Allsup's Accounts."

    KXLY, Morgan Murphy Media, 21 Aug. 2019, www.kxly.com/facebook-instagram-

    delete-controversial-figure-james-allsups-accounts/.

@SAgenocide. "Stop white GEnocide" *Twitter*, Sept. 2010, www.twitter.com/SAgenocide/.

Shearer, Elisa, and Amy Mitchell. "News Use Across Social Media Platforms in 2020." *Pew*

    Research Center's Journalism Project, Pew Research Center, 25 Mar. 2021,

    www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/.

Spangler, Todd. "Twitter Has Flagged 200 of Trump's Posts as 'Disputed' or Misleading Since

    Election Day. Does It Make a Difference?" *Variety*, Variety Media, 27 Nov. 2020,

    variety.com/2020/digital/news/twitter-trump-200-disputed-misleading-claims-election-

    1234841137/.

Stevenson, Alexandra. "Facebook Admits It Was Used to Incite Violence in Myanmar." *The*

    New York Times, The New York Times, 6 Nov. 2018,

    www.nytimes.com/2018/11/06/technology/myanmar-facebook.html.

Taub, Amanda, and Max Fisher. "Where Countries Are Tinderboxes and Facebook Is a Match."

    The New York Times, The New York Times, 21 Apr. 2018,

    www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html.

Travers, Mark. "Facebook Spreads Fake News Faster Than Any Other Social Website,

    According To New Research." *Forbes*, Forbes Magazine, 21 Mar. 2020,

    www.forbes.com/sites/traversmark/2020/03/21/facebook-spreads-fake-news-faster-than-

    any-other-social-website-according-to-new-research/?sh=292acc6a6e1a.

United States, Congress, *Title 18*. Cornell Law School, 1968,

    www.law.cornell.edu/uscode/text/18/2101.

@UR_Ninja. "Attendees of #UniteTheRight in #Charlottesville discuss legality of plowing into

    protesters with a vehicle, less than one month before rally." *Twitter*, 16 Aug. 2017, 3:33

    p.m., www.twitter.com/UR_Ninja/status/897979666610491394.

Vogels, Emily A. "The State of Online Harassment." *Pew Research Center: Internet, Science &*

    Tech, Pew Research Center, 31 Mar. 2021,

    www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/.

Warofka, Alex. "An Independent Assessment of the Human Rights Impact of Facebook in

    Myanmar." *Facebook*, Facebook, 5 Nov. 2018,

    www.about.fb.com/news/2018/11/myanmar-hria/.

Woodward, Calvin. "AP FACT CHECK: Trump's Call to Action Distorted in Debate." *AP NEWS*, Associated Press, 13 Jan. 2021,

apnews.com/article/fact-check-trump-us-capitol-remarks-221518bc174f9bc3dd6e108e653ed08d.

"#88." *Instagram*, www.instagram.com/explore/tags/88/.