

# BEYOND LIKES AND REPOSTS: SOCIAL MEDIA'S ROLE IN WARFARE

BRITNEY ARTEAGA \*

## I. INTRODUCTION

Social media has become a double-edged sword in our current digital era, bringing both serious advantages and disadvantages to the forefront of domestic and international discussions. Social media has become a weapon in political, cultural, and even military conflicts due to the spread of unregulated content, despite it enabling people to interact, share information, and voice their opinions. Social media has facilitated the spread of hate speech, propaganda, and false information, which in turn distorts realities and shapes narratives and opinions. Social media's effect on both a domestic and global scale have been catastrophic, ranging from promoting violence to further alienating societies.

This discussion will explore the complex role of social media in warfare, focusing on its utilization by various claimants, past trends, future trends and lastly a recommendation of how to fix the issue. Social media platforms have facilitated the spread of unprotected speech and disinformation campaigns, leading to the recruitment of supporters, incitement of violence, and promotion of terrorist propaganda. Additionally, the international protections for freedom of expression and the right to information create challenges in regulating harmful content, which raises concerns about how these rights are being exploited in the digital era.

## II. DELIMITATION OF THE PROBLEM

### A. THE RISE OF SOCIAL MEDIA

The rise of social media within society has become more prevalent in the past decade, pushing society into a digital era. While living in this new digital era, false information and unprotected speech have saturated our digital world.<sup>1</sup>

---

\* *Juris Doctor* Candidate, May 2026, St. Thomas University College of Law; B.S. in Political Science, Florida International University, 2023. This paper was written under the supervision of Professor Dr. Roza Pati.

<sup>1</sup> Abdul Rauf, *The Collapse of Certainty: Fake News and the Erosion of Truth*, HILAL ENG., Feb. 2025, at 55, 56 (explaining how misinformation and fake news has blurred the lines between truth and fiction to influence political opinions).

Social media dangers have been previously explored in regard to the psychological effects on individuals.<sup>2</sup> Social media has evolved from a casual and fun platform into a battleground for influence in political, military and cultural events.<sup>3</sup> In addition, the dangers of social media on a much larger scale, such as on a national and international level, have catastrophic effects. The dangers of social media on a much larger scale include: (1) swaying opinions, (2) shaping narratives, and (3) distorting reality.<sup>4</sup> Despite the dangers that arise from social media, users continue to be highly dependent on these social media platforms. Social media provides easy access to widespread information, whether true or false, and allows for weaponization of the content.<sup>5</sup>

Social media use is not limited to individuals, countries also utilize social media as tools and tactics for warfare.<sup>6</sup> Media as a warfare tactic can be seen from as early as the Cold War, when the Soviet Union would issue false information and control media messages with the goal of discrediting Western countries.<sup>7</sup> A more current utilization of social media in war can be seen in Russia. From 2016 to 2020, Russia utilized social media to create false narratives to weaken and create societal division within the United States.<sup>8</sup> Although Russia's weaponization of social media is not a new custom, their actions through social media have inspired other countries to utilize media for similar means.<sup>9</sup>

Social media platforms play a crucial role within warfare by allowing individuals to recruit supporters within the platform. In 2015, two individuals engaged in online messaging, through a social media platform, with an undercover

---

<sup>2</sup> Eugene Brusilovskiy et al., *Social Media Use, Community Participation and Psychological Well-Being Among Individuals with Serious Mental Illnesses*, in 65 *COMPUTS. IN HUM. BEHAV.* 232, 232–33 (Matthieu Guitton ed., 2016) (describing how severe mental illness may be due to an increased access to online interactions and online news consumption).

<sup>3</sup> Peter Singer & Emerson Brooking, *Why Social Media is the New Weapon in Modern Warfare*, KNOWLEDGE AT WHARTON (Jan. 17, 2019), <https://knowledge.wharton.upenn.edu/article/singer-weaponization-social-media/> [perma.cc/YLM8-7EVE] (stating social media is the “nervous system” of the modern world because aside from being part of our daily lives, social media also includes political campaigns to use in military operations and marketing wars).

<sup>4</sup> Rauf, *supra* note 1, at 56 (“Misinformation, disinformation, and fake news now permeate every corner of the digital landscape, influencing opinions, shaping narratives, and, at times, rewriting reality itself.”).

<sup>5</sup> Sarah Jacobs Gamberini, *Social Media Weaponization: The Biohazard of Russian Disinformation Campaigns*, 99 *JFQ*, 2020, at 4, 5 (explaining how countries have turned to media to weaponize information, including false information, during warfare to create societal division among other countries to invoke public crisis and trouble national security).

<sup>6</sup> *Id.* (discussing how Russia has weaponized information through media to “amplify societal divisions,” in the United States).

<sup>7</sup> *See id.* (explaining how the Soviet Union influenced other nations by using media to deliver false information, voice certain political views, and discredit Western influence).

<sup>8</sup> *Id.* (describing how Russia used media during the 2016 presidential election in the United States to spread false information with the goal of creating division among Americans).

<sup>9</sup> *Id.* (elaborating on how Russia plans to regain power following the fall of the Soviet Union by using media to manipulate narratives and exaggerate societal division by publishing false information).

FBI agent posing as an Islamic State of Iraq and the Levant (“ISIL”) member.<sup>10</sup> During this interaction on the social media platform, the two individuals expressed their desires to support and contribute to ISIL.<sup>11</sup> The FBI confronted the two individuals, and their families were shocked to find out their children supported ISIL, considering the minimal influence ISIL had in the region.<sup>12</sup> Despite limited physical influence in certain regions, media has become a primary element for terrorists to publish recruitment campaigns.<sup>13</sup>

In addition, social media platforms allow for a facilitated access to widespread information regarding terrorist propaganda.<sup>14</sup> Non-state actors, specifically terrorist groups, are able to operate seamlessly due to the anonymity the internet allows users to have.<sup>15</sup> In turn, social media platforms have exacerbated the impact of cyberattacks by facilitating the spread of fear, distorting the public’s opinion, and reaching mass audiences. Further, with no censorship or regulations on behalf of social media platforms, their inaction constitutes as compliance to the widespread access to false information and terrorist propaganda which is made available on their platforms.

Moreover, governments have also become dependent on social media use in war. The utilization of social media as a tool in warfare consists of creating political organizations and weaponizing content for propaganda.<sup>16</sup> Governments utilize social media by picking and choosing strategies to portray reality in a light most favorable to them.<sup>17</sup> Social media also acts as a search engine for governments by allowing them to learn about other countries political and social news. For example, some governments explore other governments

---

<sup>10</sup> Grace Callahan, *Attacking ISIL on Twitter: Addressing Ethical Responsibility in the Weaponization of Social Media*, 10 INTERSECT 1, 2 (2017) (describing how easily two students came across ISIL social media campaigns which were aimed at recruiting foreign supporters, despite the minimal influence in the region).

<sup>11</sup> *Id.* (explaining how two students simply logged into a social media application and were able to access terrorism support campaigns).

<sup>12</sup> *Id.*

<sup>13</sup> E.g., Rosanna E. Guadagno et al., *Social Influence in the Online Recruitment of Terrorists and Terrorist Sympathizers: Implications for Social Psychology Research*, 23 REVUE INTERNATIONALE DE PSYCHOLOGIE SOCIALE 25, 27 (2011) (explaining how Al- Qaeda has become a monumental force in the global war of terrorism due to their media use to recruit members and to spread political messages).

<sup>14</sup> Cf. Callahan, *supra* note 10, at 8–9 (elaborating how the use of Twitter (now known as “X”) exposes many individuals to propaganda and terrorist recruitment campaigns).

<sup>15</sup> Roza Pati, *Regulating Terrorist Activity in Cyberspace: Issues at Stake*, in RSCH. AGENDA FOR CYBERSECURITY L. & POL’Y 125–46 (Balleste et al. eds., 2025) (explaining how social media augments the distortion of reality when terrorist groups utilize the internet to hijack control systems).

<sup>16</sup> See KJ Boyte, *An Analysis of the Social-Media Technology, Tactics, and Narratives Used to Control Perception in the Propaganda War Over Ukraine*, 16 J. INFO. WARFARE 88, 88 (2017) (describing how social media has been utilized as a warfare tactic by influencing political views through the publication of false information).

<sup>17</sup> E.g., *id.* at 89 (illustrating how the Russian Federation, NATO, and the United States have used social media as a warfare tactic to control the perception of political events during the years of 2013–2015).

information on social media platforms to learn about their current strengths and weaknesses before, during, or after war times to exploit them.<sup>18</sup>

Non-state actors such as militias, business corporations, and individuals also utilize social media for hybrid warfare tactics. Critics have reported social media as the most advanced weapon in warfare among non-state actors.<sup>19</sup> Hybrid warfare has become more relevant now than ever in our current digital era. This strategic tactic incorporates military and non-military efforts, but may also solely rely on non-military methods.<sup>20</sup> Hybrid warfare may include cyberattacks and false information in social media campaign messages to destabilize opposing countries.<sup>21</sup> Islamic State of Iraq and Syria (“ISIS”), a militant group, utilizes hybrid warfare by spreading violent propaganda and videos on social media platforms and recruiting individuals that are influenced by their posts through communication on these platforms.<sup>22</sup>

## B. SOCIAL MEDIA FOR UNPROTECTED SPEECH

All individuals have the internationally protected right of freedom of expression and right to information, which includes the right to receive information through any form of media.<sup>23</sup> Because the right to information and freedom of expression are so broad, it leaves too much space for a misuse of this right within social media.<sup>24</sup> Social media has facilitated targeted speech against others by hosting unprotected speech.<sup>25</sup> Unprotected speech, such as hate speech, in social media includes “[g]eneralized racist, sexist, religious, anti-

---

<sup>18</sup> See Achmad Z. Huda et al., *Social Media as an Incubator of Youth Terrorism in Indonesia: Hybrid Threat and Warfare*, 11 J. KAJIAN INTERDISIPLINER ISLAM INDON. 21, 22 (2021) (describing how ISIS has used media to explore governmental divisions by spreading propaganda and targeting their influence); see also Mathew Lower & Thomas Hauschildt, *The Media as a tool of War: Propaganda in the Rwandan Genocide*, 2 HUM. SEC. CTR. 1, 2 (2014) (illustrating the several ways that propaganda in media sources such as newspapers and radios were targeting to incite violence during the Rwandan genocide).

<sup>19</sup> See Tilman Rodenhäuser & Samit D’Cunha, *Foghorns of War: IHL and Information Operations During Armed Conflict*, HUMANITARIAN L. & POL’Y (Oct. 12, 2023), <https://blogs.icrc.org/law-and-policy/2023/10/12/foghorns-of-war-ihl-and-information-operations-during-armed-conflict/> [perma.cc/4YH7-KPVB] (stating that information operations are used to influence adversaries during warfare by state and non-state actors).

<sup>20</sup> Huda et al., *supra* note 18 (explaining the warfare tactic of combining military and non-military strategies to explore other governmental information by spreading propaganda through media).

<sup>21</sup> *Id.* (describing how countries use hybrid warfare to weaken and destabilize other countries without solely relying on military tools).

<sup>22</sup> *Id.* (“[ISIS] has succeeded in exploiting so many of its enemy’s weaknesses that it has become widely known and continued to grow to this day.”).

<sup>23</sup> See also G.A. Res. 217 (XIX) A, Universal Declaration of Human Rights (Dec. 10, 1948) (commenting how individuals have the right to receive and impart information and ideas through any media regardless of frontiers).

<sup>24</sup> *Cf. id.* (granting access to information through any means regardless of frontiers, which inadvertently allows for false information to be accessed).

<sup>25</sup> See Joseph Walther, *Social Media and Online Hate*, in 45 CURRENT OP. IN PSYCH. (2022) (discussing how hate speech in social media is highly motivated by moral ground standing, political views and violence).

immigrant, gender, and sexual orientation-related insults, and verbal attacks based on ethnicity, political orientation, or other categorical characteristics.”<sup>26</sup>

Propaganda is also another form of unprotected speech that can be harvested in social media during warfare. In the early 1990s, propaganda in media sources played a crucial role in fostering ethnic and occupational hatred between the Hutus and the Tutsis.<sup>27</sup> The medias’ role in the Rwandan genocide amplified the existing distinctions, ethnic and occupational, between Hutus and Tutsis by broadcasting the locations of victims, urging the Hutus to take action against the Tutsis, and overall incitement of large-scale violence through false information.<sup>28</sup> The main takeaway from this devastating genocide is the monumental effect propaganda in the media had on provoking hatred and continuing warfare measures among both state and non-state actors.

### C. SOCIETAL BELIEF

The pressing issue of social media in war is now more disturbing than ever because of the societal belief that the right to information and freedom of expression will not be sabotaged by its use in warfare. This societal belief is rooted in many countries’ governing laws and in the general assembly of the United Nations.<sup>29</sup> The United States also recognized the right to information and freedom of expression as fundamental rights engrained in the First Amendment of the Constitution.<sup>30</sup> In the United States, regulation of such a fundamental right would be subject to the highest form of scrutiny.<sup>31</sup>

In addition, the United Nations has recognized the fact that freedom of expression and right to information are inalienable rights which lay the foundation for freedom, justice, and peace.<sup>32</sup> It is because this right is so rooted in a societal belief that it comes to a shock when used as a warfare tactic within social media. While these rights remain an internationally imperative concept, these rights

---

<sup>26</sup> *Id.*

<sup>27</sup> Lower & Hauschildt, *supra* note 18, at 3 (explaining how propaganda was a tool used during the Rwandan genocide to exaggerate the differences between Hutus and Tutsis with the intention of spreading fear).

<sup>28</sup> *Id.* (describing how newspapers and radios were inciting violence in the Rwandan genocide).

<sup>29</sup> See Universal Declaration of Human Rights, *supra* note 23 (stating individuals have the right to receive and impart information and ideas); see also U.S CONST. amend. I (“Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.”).

<sup>30</sup> U.S CONST. amend. I.

<sup>31</sup> See *Strict Scrutiny*, CORNELL L. SCH: LEGAL INFO. INST., [https://www.law.cornell.edu/wex/strict\\_scrutiny](https://www.law.cornell.edu/wex/strict_scrutiny) [perma.cc/UW4U-82AZ] (last updated Sep. 2024) (explaining how strict scrutiny must be applied when a fundamental right is at issue before the court).

<sup>32</sup> *E.g.*, *The Universal Right to Free Expression: An Interpretation of the Library Bill of Rights*, AM. LIBR. ASS’N, <https://www.ala.org/advocacy/intfreedom/librarybill/interpretations/universalright> [perma.cc/6LVY-ZM5Z] (last updated May 29, 2025) (discussing how freedom of expression is an essential human right that is affirmed in the Universal Declaration of Human Rights to promote worldwide peace).

have their limitations.<sup>33</sup> Based on the use of social media in warfare it is clear the problem lies with how broad the right to information and freedom of expression is, which leads governments and non-state actors to overstep on these rights.

### III. CONFLICTING CLAIMS

Having defined the impact social media has in warfare, it is critical to analyze the claims raised on a domestic and international level. Overall, it is believed that media escalates conflicts between countries whether directly or indirectly.<sup>34</sup> Media is often times used by countries to draw more emphasis than necessary on a certain political or social event, simply for the purpose of trying to gain more media exposure and supporters.<sup>35</sup> When viewed wholistically, social media in warfare impacts several claimants such as users in times of peace, users in times of war, legislatures, social media platforms, courts, and civil society. All claimants have important perspectives on this issue and each one plays a role on how social media is utilized in warfare.

#### A. USERS IN TIMES OF PEACE

Social media during times of peace has been linked to improved societal benefits.<sup>36</sup> Media has facilitated connection among individuals internationally, which has enabled individuals to adopt a feeling of belonging, support, and happiness.<sup>37</sup> Although peaceful measures within social media are often overshadowed by the radicalization and violence of campaigns filled with false

---

<sup>33</sup> *Freedom of Expression*, AMNESTY INT'L, <https://www.amnesty.org/en/what-we-do/freedom-of-expression> [perma.cc/U6QH-6LYX] (last visited Nov. 24, 2025) (“While international law protects free speech, there are instances where speech can legitimately [be] restricted under the same law – such as when it violates the rights of others, or, advocates hatred and incites discrimination or violence.”); see also Universal Declaration of Human Rights, *supra* note 23 (describing how individuals have a protected right to information and freedom of expression).

<sup>34</sup> Conflict Research Consortium Staff, *Summary of “The Role of Media in International Conflict,”* BEYOND INTRACTABILITY, <https://www.beyondintractability.org/artsum/young-role> [perma.cc/W2WT-EDWU] (last visited Nov. 24, 2025) (reporting how the media may sometimes escalate conflicts between countries instead of simply providing awareness due to the false information portrayed); cf. Roza Pati, *States Positive Obligation with Respect to human Trafficking: The European Court of Human Rights Breaks New Grounds in Rantsev v. Cyprus and Russia*, 29 BOS. U. INT'L L.J. 79, 141 (2011) (analyzing the gravity of human trafficking and its complexities for state regulation due to the limitations because one has a protected right to free movement; yet, offering a solution despite the harsh barriers need to overcome such limited right).

<sup>35</sup> Conflict Research Consortium Staff, *supra* note 34 (explaining how media platforms tend to focus on more dramatic and violent news, both in political and social events, because it attracts more attention to their site).

<sup>36</sup> Cf. Linda C. Ashar, *Social Media Impact: How Social Media Sites Affect Society*, AM. PUB. UNIV. (May 2, 2024), <https://www.apu.apus.edu/area-of-study/business-and-management/resources/how-social-media-sites-affect-society> [perma.cc/7X3J-8ST9] (describing a Harvard study conducted in 2020 which showed that routine use of social media positively affected society).

<sup>37</sup> *Id.* (elaborating on the “three health-related outcomes [of social media use during times of peace]: social well-being, positive mental health, and self-rated health”).

information, social media may be used in times of peace to counter the violence and false information.<sup>38</sup>

Researchers suggest that users in times of peace should utilize social media by demanding social media platforms to explain how their information, collected by media platforms, is being utilized and who has access to their information.<sup>39</sup> In times of peace, social media plays a crucial role for users by allowing: (1) easier communication among groups of individuals; (2) faster communication to bring attention to false information; (3) awareness about current conflicts; and (4) access to new data regarding populations and support groups.<sup>40</sup>

Social media bridges communication between individuals from all corners of the world, which in turn creates connections transcending borders.<sup>41</sup> Social media acts as a window into diverse cultures, traditions, and ethnicities, allowing individuals to explore and appreciate other communities across the globe.<sup>42</sup> This exposure harnesses a greater understanding, love, and curiosity which invokes an interest to learn about others around the world “without being restricted by the limitations of time and distance.”<sup>43</sup>

During the Rwandan genocide in 1994, media was utilized to reveal the location of victims, but after all the violence, media was utilized by peacemakers to influence the population in reconciliation.<sup>44</sup> The use of media post-violence in Rwanda illustrates two sides of the same coin, the violence and the peace-making ability media may have on individuals. After all, social media is a crucial tool used to motivate opinions, attitudes, and, most importantly,

---

<sup>38</sup> See Dr. Simone Bunse, *Social Media: A Tool for Peace or Conflict?*, STOCKHOLM INT’L PEACE RSCH. INST. (Aug. 20, 2021), <https://www.sipri.org/commentary/topical-background/2021/social-media-tool-peace-or-conflict> [perma.cc/49EP-PUKU] (explaining the different ways that social media may be utilized by users in times of peace).

<sup>39</sup> *Id.* (“[C]itizens should improve their resilience to disinformation, but also demand insight into the information collected about them by social media firms, how it is used and by whom.”).

<sup>40</sup> Cf. Thomas Zeitzoff, *How Social Media is Changing Conflict*, 61 J. CONFLICT RESOL. 1971, 1971–72 (2017) (discussing how social media may be utilized for conflict resolution in times of peace).

<sup>41</sup> Ashar, *supra* note 36 (“Social media sites provide opportunities for individuals to connect with friends, family, and like-minded individuals, even across geographical distances. These connections help maintain relationships and interests, fostering a sense of belonging, self-esteem, and support in users.”).

<sup>42</sup> Rebecca Swayer & Guo-Ming Chen, *The Impact of New Social Media on Intercultural Adaptation*, 21 U. RHODE ISLAND 151, 154 (2012) (explaining how social media has provided a manner for individuals to communicate without being restricted by time and distance).

<sup>43</sup> *Id.* (describing how communication without barriers provides for a sense of belonging through means of knowledge, behavior, and attitudes).

<sup>44</sup> Lower & Hauschildt, *supra* note 18, at 4 (noting that Rwandan radio broadcasts identified victims’ hiding places, thereby facilitating targeted killings during the genocide); see Vladimir Bratić, *Media Effects During Violent Conflict: Evaluating Media Contributions to Peace Building*, 5 CONFLICT & COMMUN ONLINE 1, 2 (2006) (describing the difference between media usage during the Rwandan genocide when used for violence versus when media was regained by peacemakers and then utilized for influencing harmony).

actions.<sup>45</sup> Specifically, in times of peace, media can play a powerful role in rebuilding and harvesting peaceful societies by individual usage.

#### B. USERS IN TIMES OF WAR

In our digital era, social media has unintentionally become a central platform for communication.<sup>46</sup> However, when free speech within media turns oppressive, it reveals a deeper issue, the absence of meaningful dialogue.<sup>47</sup> When communication loses its substance, it devolves into hollow words echoing within a chamber of like-minded voices.<sup>48</sup> The dangers posed by such an echo chamber is the amplification of existing fears and anxieties that resonates with a mass audience willing to listen, which further distorts the conversation. Social media has forged the path for persuasion and mass manipulation, which in turn grants access to vast audiences.<sup>49</sup> Both state and non-state actors can broadcast false information without regulation or thorough scrutiny.<sup>50</sup> Moreover, the inherent danger lies when the messages, specifically in times of war, are rooted in partial truths or, at times, outright falsehood.<sup>51</sup> These outright false narratives can shape perceptions of war that are far from reality which leads to the spread of misinformation.<sup>52</sup>

The war between Israel and Hamas has been widely covered on social media, pushing biased views on mass audiences regarding the affected communities. As these biased narratives, such as the one utilized in the conflict between Israel and Hamas, circulate the media, it creates polarized opinions that do not accurately reflect the complexities of current ongoing conflicts.<sup>53</sup> The false and

---

<sup>45</sup> Bratić, *supra* note 44, at 8 (concluding that social media influences communities' beliefs, opinions, and attitudes which eventually produces action).

<sup>46</sup> Cf. *The Evolution of Social Media: How Did it Begin, and Where Could it Go Next?*, MARYVILLE U. (May 28, 2020), <https://online.maryville.edu/blog/evolution-social-media/> [perma.cc/BAD3-ZZDU] (explaining that the main purpose of social media when it began was to connect individuals with friends and family when physically meeting was unlikely).

<sup>47</sup> Daniel Sanderson, *Voltaire on Freedom of Speech*, MEDIUM (Feb. 16, 2021), <https://planksip.medium.com/voltaire-on-freedom-of-speech-e6469d0a3b1e> [https://perma.cc/HLX5-VZXH] (analyzing French philosopher Voltaire's quote, "I disapprove of what you say, but will defend to the death your right to say it").

<sup>48</sup> See *id.* (interpreting Voltaire's famous quote as a warning that communication without substance fosters echo chambers).

<sup>49</sup> Cf. Stephen Wyatt, *The Weaponization of Information and Its Use Through Social Media*, CANADIAN FORCES COLL. 1, 2 (2019) (discussing how information on social media can reach millions of viewers instantly).

<sup>50</sup> *Id.* (stating that information published on social media receives little to no oversight and facilitates the spread of false information).

<sup>51</sup> *E.g., id.* at 4–5 (describing a 2013 cyberattack in which the Syrian Electronic Army hijacked the Associated Press's Twitter account to falsely report an explosion at the White House, causing a 150-point drop in the Dow Jones stock market that resulted in \$136 billion market value loss).

<sup>52</sup> Boyte, *supra* note 16, at 100 (explaining how social media has been used as a warfare tactic by distorting the truth by publishing false information).

<sup>53</sup> See Georgios Karagiorgos et al., *Gaza Through Whose Lens?*, CTR. FOR STRATEGIC & INT'L STUD. (Dec. 2023), <https://features.csis.org/gaza-through-whose-lens/index.html> [perma.cc/K3LJ-CMTW] (concluding that inaccurate portrayals of current events on social media has led to

biased information being spread on media by individuals “ha[s] had severe consequences for the Jewish-American, Palestinian-American, and Arab-American communities.”<sup>54</sup> False media coverage about the war has led to antisemitism, islamophobia, and hate crimes.<sup>55</sup>

In 2018, the United Nations reported that social media is a critical tool in war due to individuals exacerbating conflicts by posting hate speech.<sup>56</sup> Further, online violence plays an immense role in aggravating the ongoing genocide within Myanmar due to individuals posting radicalizing messages and repression.<sup>57</sup> “While the Myanmar military was committing crimes against humanity against the Rohingya, Meta was profiting from the echo chamber of hatred created by its hate-[spiraling] algorithms.”<sup>58</sup> This phenomenon is particularly evident in Myanmar, where social media has been significant in spreading radical messages, such as labeling Rohingyas as “invaders,” and echo chambers that have worsened the ongoing genocide.<sup>59</sup> Social media in war has served as a breeding ground for violence and repression.<sup>60</sup>

### C. LEGISLATURES

Social media, specifically social media platforms, have become an integral part of democracy.<sup>61</sup> It is because social media is crucial to the sense of democracy that legislatures have a crucial role in how it is utilized. Social media provides a platform, apart from mainstream media, that allows the spread of information from a different perspective. Moreover, legislatures have an interest in

---

polarization).

<sup>54</sup> *Id.* (stating that false and biased information on social media has produced concrete harms within such communities).

<sup>55</sup> *Id.* (linking misleading social media narratives about the war to specific real-world consequences).

<sup>56</sup> *E.g.*, Caroline Crystal, *Facebook, Telegram, and the Ongoing Struggle Against Online Hate Speech*, CARNEGIE ENDOWMENT FOR INT’L PEACE (Sep. 7, 2023), <https://carnegieendowment.org/research/2023/09/facebook-telegram-and-the-ongoing-struggle-against-online-hate-speech?lang=en> [perma.cc/GT3Y-37QE] (highlighting social media’s role in amplifying hate speech and deepening polarization during the 2017 Rohingya genocide in Myanmar’s Rakhine state).

<sup>57</sup> *Id.* (describing different types of online violence during the Myanmar conflict which show the lack of atrocity prevention within social media platforms).

<sup>58</sup> *Myanmar: Facebook’s Systems Promoted Violence Against Rohingya; Meta owes Reparations*, AMNESTY INT’L (Sep. 29, 2022), <https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebook-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/> [perma.cc/7MXF-VEGD].

<sup>59</sup> *Id.* (“In the months and years leading up to the atrocities, Facebook’s algorithms were intensifying a storm of hatred against the Rohingya which contributed to real-world violence[.]”).

<sup>60</sup> *Id.* (describing how social media has facilitated hate speech to be widespread among users by allowing echo chambers and radical messages by labeling ethnic groups as “invaders”).

<sup>61</sup> *E.g.*, Ricki-Lee Gerbrandt, *Threatening & Protecting Press Publishers and Journalism in the UK’s Regulation of Social Media Platforms*, J. OF MEDIA L. 68, 68–69 (2025) (explaining how journalists in media form an essential part of democracy during our digital era because they serve a key role in reporting information to a broader public for the purpose of informing and educating).

supervising and regulating social media usage because it can have a direct effect on their state governments.<sup>62</sup>

Legislatures are directly affected by unprotected speech on social media due to concerns over national security, the spread of misinformation, and the overall impact on public welfare.<sup>63</sup> With over 5 billion users globally, and an average daily usage of 2 hours, social media has become a dominant force in communication and information sharing.<sup>64</sup> This widespread usage highlights the crucial need for media regulation to prevent the dissemination of false information and protect the public. As social media continues to evolve, lawmakers face increasing pressure to balance free speech with the need to safeguard societal interests and maintain order.<sup>65</sup> Effective regulations can help mitigate the risks posed by unchecked content while ensuring that social media remains a space for constructive dialogue.<sup>66</sup>

#### D. SOCIAL MEDIA PLATFORMS

Social media platforms heavily monetize on the fact that their platforms are utilized as a main source of information.<sup>67</sup> Social media platforms essentially only cover one perspective when acting as a news outlet, which only increases the chances of creating echo chambers.<sup>68</sup> In fact, these platforms know this and

---

<sup>62</sup> E.g., Joshua A. Tucker & Jake Karr, *Why the Texas and Florida Social Media Cases are Important for Research Transparency*, CTR. FOR SOC. MEDIA AND POL. (Feb. 23, 2024), <https://csmapnyu.org/impact/policy/why-the-texas-and-florida-social-media-cases-are-important-for-research-transparency> [perma.cc/HM28-G3FF] (commenting on how the United States has national security concerns about the social media app TikTok, due to the immense impact other governments have had on Americans through the utilization of the app).

<sup>63</sup> E.g., *id.* (describing the fear of countries' legislatures regarding social media's impact on the younger generation's mental health); see also Brusilovskiy et al., *supra* note 2 (elaborating on the negative psychological ramifications social media platforms may have on individuals due to increased social media usage).

<sup>64</sup> Dave Chaffey, *Global Social Media Statistics Research Summary*, SMART INSIGHTS (Feb. 14, 2025), <https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/> [perma.cc/QVE3-FP6H] (stating how over half of the world now utilizes social media, making up over 60% of the world's population, with a daily usage of slightly over 2 hours).

<sup>65</sup> Tucker & Karr, *supra* note 62 (describing how state legislatures are unable to develop effective social media regulation without knowledge about the risks posed by social media platforms).

<sup>66</sup> Cf. *id.* (explaining how state legislatures are unsure of how to regulate these social media platforms due to social media policies implemented and the effects it would have culturally, socially, and politically); see also Bunse, *supra* note 38 (elaborating on how policymakers should create effective oversight and strict date management guidelines to help steer away from harmful social media usage).

<sup>67</sup> Cf. David Ardia et al., *Addressing the Decline of Local News, Rise of Platforms, and Spread of Mis- and Disinformation Online*, CTR. FOR INFO., TECH., & PUB. LIFE, <https://citap.unc.edu/news/local-news-platforms-mis-disinformation/> [https://perma.cc/E559-9YVJ] (last visited Nov. 24, 2025) (describing how traditional media sources, such as the Wall Street Journal, invest large sums of money into producing news reports, only for social media platforms to release nearly identical reports after traditional media has published their reports and endured the high costs).

<sup>68</sup> E.g., Karagiorgos et al., *supra* note 53 (explaining how social media platforms, instead of

instead of aiding the cause to eliminate echo chambers, social media platforms incentivize this behavior by creating specially tailored algorithms.<sup>69</sup> These algorithms create a “[t]unnel vision where many people, whatever their perspective, are getting fed the same [information] over and over again.”<sup>70</sup>

Social media companies make their profits by having users utilize their platforms for extended periods of time.<sup>71</sup> In doing so, social media platforms are able to sell more targeted advertising to companies wishing to issue inflammatory content containing false and hateful information.<sup>72</sup> This entire social media dynamic used to monetize content further generates more echo chambers.

The question arises, “Why is content not censored on social media platforms?” The question of platform censorship is a complex issue with multiple layers. The primary concern is that censorship could possibly infringe upon the internationally protected right to information and freedom of expression.<sup>73</sup> The concept of social media depends on the free exchange of ideas and opinions, despite a difference in ideas and opinions.<sup>74</sup> Additionally, censorship risks the suppression of legitimate perspectives based on political or corporate interests, which would inadvertently promote deeper echo chambers.<sup>75</sup>

#### E. COURTS

Courts have struggled with handling social media impact of warfare and genocides worldwide.<sup>76</sup> For instance, the International Criminal Court (“ICC”) has been struggling with how to hold Facebook accountable for their role in Myanmar.<sup>77</sup> Facebook’s facilitation of hate speech and false information

---

balancing both sides of the conflict within Gaza, chose to only showcase one perspective, which caused a false perception of reality).

<sup>69</sup> *Id.* (providing insight from Michelle Strucke, director of the Humanitarian Agenda and the Human Rights Initiative, on how social media algorithms foster “tunnel vision” by showing users the same perspectives repeatedly).

<sup>70</sup> *Id.*

<sup>71</sup> *Cf. Myanmar: Facebook’s Systems Promoted Violence Against Rohingya*, *supra* note 58 (explaining that Meta monetizes from prolonged user engagement on its platforms).

<sup>72</sup> *Id.* (describing how Meta profits from selling targeted advertising and that the amplification of inflammatory content, creating echo chambers reinforcing particular political viewpoints, is central to its business model).

<sup>73</sup> *See* Stephanie Walker, *What Are the Pros and Cons of Censorship on Social Media*, CHEKKEE (July 5, 2024), <https://chekkee.com/the-pros-and-cons-of-censorship> [<https://perma.cc/2UNP-JBMP>] (listing the potential harms of censorship on social media, including suppression of speech and potential abuse); *see also* Universal Declaration of Human Rights, *supra* note 23, at 19 (stating that individuals have the right to receive and impart information and ideas).

<sup>74</sup> *See* Walker, *supra* note 73 (“When certain topics are censored, it can limit the flow of open dialogue and impede the democratic exchange of dissenting ideas and opinions.”).

<sup>75</sup> *Id.* (explaining how the suppression of some perspectives may in fact create and deepen echo chambers because the platform would only tolerate certain constructive arguments).

<sup>76</sup> *See* Neema Hakim, *How Social Media Companies Could be Complicit in Incitement to Genocide*, 21 CHL J. INT’L L. 83, 87 (2020) (noting international courts’ confusion over how to address Facebook’s involvement in the Myanmar massacre).

<sup>77</sup> *Id.* (explaining how Facebook’s involvement has been “significant” but the United Nations did not identify the platform as criminally responsible).

significantly contributed to the ongoing persecution against the Rohingyas.<sup>78</sup> However, the ICC did not hold Facebook liable because of legal uncertainties of holding social media companies liable when their platforms are utilized to incite violence.<sup>79</sup>

Moreover, the ICC has adopted the Joint Criminal Enterprise (“JCE”) which is a method of criminal liability.<sup>80</sup> The JCE authorized courts to hold individuals accountable for acting as co-perpetrators.<sup>81</sup> The ICC allows for the prosecution of individuals that are indirectly perpetuating a crime.<sup>82</sup> ICC has established two other modes for imposing liability specifically for social media platform employees.<sup>83</sup> Firstly, aiding and abetting which refers to assisting in the crime of inciting genocide or harmful behavior.<sup>84</sup> Secondly, common purpose which applies when the group of individuals actually commit the crime.<sup>85</sup> Both forms of liability are seen as a form of being complicit to the crime. Here, the crime would be how social media platforms allow echo chambers and the spread of false information is utilized to target minority groups.<sup>86</sup> The Rome Statute outlines a set of guidelines specifically for the purpose of holding individuals criminally responsible for their incitement of crimes; thus, the Rome Statute controls ICC jurisprudence.<sup>87</sup>

---

<sup>78</sup> Cf. *Myanmar: Facebook’s Systems Promoted Violence Against Rohingya*, *supra* note 58 (stating that Meta’s algorithms intensified the ongoing violence and hatred against Rohingya rather than regulating harmful speech and false information); see also Hakim, *supra* note 76, at 87 (explaining that Facebook’s involvement was considered “significant” and useful to “spread hate.”).

<sup>79</sup> See Hakim, *supra* note 76, at 87 (explaining that confusion over Facebook’s liability raised broader questions about whether social media companies may be liable when their platforms are used to incite violence).

<sup>80</sup> See Kevin Aquilina & Klejda Mulaj, *The International Criminal Court and Responsibility for Mass Atrocities: Can JCE Enhance Capacity to Hold Masterminds Accountable?*, 27 CONTEMP. JUST. REV. 132, 134 (2024) (commenting on international courts adopting the Joint Criminal Enterprise methodology for imposition of liability).

<sup>81</sup> *Id.* (elaborating how the Joint Criminal Enterprise allows courts to impose liability for furthering crimes).

<sup>82</sup> Hakim, *supra* note 76, at 100 (“In addition to the principal perpetrator of a crime, ICL allows for the prosecution of other actors for the same crime through various modes of liability.”).

<sup>83</sup> *Id.* (describing there are two modes of liability that can be imposed upon for social media platform employees).

<sup>84</sup> *Id.* at 101 (explaining that aiding and abetting has dated back to 1945 when the liability was imposed for “‘accomplices participating in the formulation or execution of a common plan or conspiracy to commit’ a crime[.]”).

<sup>85</sup> *Id.* at 100 (explaining how imposing common purpose liability lowers the requirement for mens rea from purpose to knowledge); see also The Rome Statute of the International Criminal Court art. 25, July 17, 1998, 2187 U.N.T.S. 90 (establishing that a person may be held criminally responsible if they further the criminal activity or act with knowledge of the group’s intention to commit the crime).

<sup>86</sup> Cf. *Myanmar: Facebook’s Systems Promoted Violence Against Rohingya*, *supra* note 58 (stating how the social media platform, Meta, intensified the ongoing violence and hatred against Rohingya by shining more light on hateful speech and the false information).

<sup>87</sup> See The Rome Statute of the International Criminal Court, *supra* note 85 (explaining how the Rome Statute was adopted by the United Nations Diplomatic Conference of Plenipotentiaries for the establishment of an International Criminal Court on July 17, 1998).

## F. CIVIL SOCIETY

Social media provides a platform for civil societies and citizens to either promote peace or exacerbate conflicts. One positive outcome can be seen from a non-government organization, Caucasus Conflict Voices initiative, which aims to bridge cross border dialogue between Armenians and Azerbaijani citizens.<sup>88</sup> Armenians and Azerbaijani citizens have been in conflict over Nagorno-Karabakh territory since 1988.<sup>89</sup> However, the Caucasus Conflict Voices initiative has alleviated some tensions between the two groups.<sup>90</sup> The initiative fosters a dialogue between the two groups on the social media platform, Facebook, in order to showcase the peaceful coexistence between Armenians and Azerbaijani citizens in Georgia.<sup>91</sup> Although this initiative has not put an end to the ongoing conflict over Nagorno-Karabakh, it has highlighted the possibility of peace and added pressure on the states to push for negotiation.<sup>92</sup>

Humanitarian groups have also taken presence in social media platforms to help provide guidance, refuge, and protection. The United Nations High Commissioner for Refugees has utilized social media platforms such as Facebook, Twitter (now known as X), Instagram, and YouTube to raise awareness and promote their service to consumers and constituents, reaching a mass audience of 4,271,813 individuals.<sup>93</sup> These efforts made by humanitarian groups have aided in gathering donations and sharing success stories to inspire those who presently need it in times of conflict.<sup>94</sup> The efforts and success accomplished by these humanitarian groups illustrates how social media can be a critical tool for humanitarian aid.<sup>95</sup>

---

<sup>88</sup> Juraj Majcin, *Social Media Challenges to Peace-Making and What Can Be Done About Them*, 6 GRONINGEN J. OF INT'L L. 242, 248 (2018) (discussing the great example that Caucasus Conflict Voices Initiative has set forth by bringing together Armenians and Azerbaijani citizens, despite the growing conflict between the two groups over Nagorno-Karabakh).

<sup>89</sup> Center for Preventative Action, *Tensions Between Armenia and Azerbaijan*, COUNCIL ON FOREIGN REL. (Aug. 12, 2025), <https://www.cfr.org/global-conflict-tracker/conflict/nagorno-karabakh-conflict> [perma.cc/Y4LL-2WHE] (describing the conflict between Armenia and Azerbaijan and how the first conflict arose in 1988 when Nagorno-Karabakh declared its intention to join Armenia despite being located in Azerbaijan).

<sup>90</sup> Majcin, *supra* note 88, at 248 (explaining how the Caucasus Conflict Voices has positively affected the tensions between Armenian and Azerbaijani citizens).

<sup>91</sup> *Id.* at 249 (describing how the initiative works by creating dialogue on Facebook to showcase the peaceful coexistence between the two groups).

<sup>92</sup> *Id.* (“Such an alternative social narrative calling for peace instead of war can play an essential role in putting external pressure on negotiating parties to end the hostilities and come to a final agreement.”).

<sup>93</sup> Rubaiya Zannat, *Crisis Communication in the Ukraine-Russia Conflict: Use of Facebook by UN Humanitarian Organization* 19 (Dec. 2022) (M.A. thesis, University of Southern Mississippi) (describing the several social media platforms utilized by the United Nation High Commissioner for Refugees to aid in humanitarian relief).

<sup>94</sup> *See id.* at 20 (“Apart from promoting awareness and generating funds for a wide range of refugee groups around the globe, posts on this platform are also being used to spread information on how to get in touch with [United Nations High Commissioner for Refugees] services and other resources.”).

<sup>95</sup> *See id.* (explaining how these campaigns have been initiated in order to help find homes and shelters for displaced individuals during times of war).

#### IV. PAST TRENDS IN DECISIONS AND THEIR CONDITIONING FACTORS

Social media has been weaponized to facilitate the spread of hate speech and false information instead of being utilized for its intended purpose, worldwide communication.<sup>96</sup> The weaponization of social media has raised several concerns on an international, regional, and domestic level, prompting countries to attempt to regulate online speech. A regulation of online speech on social media platforms leads to less spread of false information, biased narratives, and propaganda overall. However, online speech regulation poses a significant challenge given that social media regulation requires constant monitoring.<sup>97</sup> The difficulties of online speech regulations are rooted in attempting to balance the scales of freedom of expression and right to information against curbing the dangers of online speech.<sup>98</sup> Making sure the balancing scales do not tip too much in one direction illustrates the complexities of social media regulation. Now we analyze what is being done at each level, internationally, regionally, and domestically, to combat the spread of hate speech and false information on social media platforms.

##### A. INTERNATIONAL

The Organization of American States (“OAS”) has established that every individual has the right to freedom of thought and expression, which encompasses the right to receive information without limitations.<sup>99</sup> The OAS is comprised of 34 independent states which have all ratified the American Convention of Human Rights.<sup>100</sup> In addition, the OAS has also stated that any promotion of hatred that incites violence would lead to a punishable offense by law.<sup>101</sup> The

---

<sup>96</sup> See Joseph Guay et al., *The Weaponization of Social Media: How Social Media Can Spark Violence and What Can Be Done About It*, MERCY CORPS 1, 7 (Nov. 2019), [https://www.mercy-corps.org/sites/default/files/2020-01/Weaponization\\_Social\\_Media\\_FINAL\\_Nov2019.pdf](https://www.mercy-corps.org/sites/default/files/2020-01/Weaponization_Social_Media_FINAL_Nov2019.pdf) [perma.cc/7495-CLV9] (describing how social media—with driving violence, persecution, and exploitation—possesses a unique ability to contribute to conflict and requires peacebuilders to adapt quickly to peacebuilding strategies).

<sup>97</sup> *But cf.* Janna Anderson & Lee Rainie, *The Future of Truth and Misinformation Online*, PEW RSCH. CTR. (Oct. 19, 2017), <https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/> [perma.cc/BFK9-5CAJ] (comparing the solution of online speech regulation on social media to the solution of high crimes by stating that both require an extensive amount of monitoring).

<sup>98</sup> See Anastasia Kozyreva et al., *Resolving Content Moderation Dilemmas Between Free Speech and Harmful Misinformation*, 120 PNAS 1, 2 (2023) (stating the complexities of social media regulation mainly concern balancing the fundamental right of freedom of expression and right to information with removing harmful content).

<sup>99</sup> Organization of American States, American Convention on Human Rights art. 13, Nov. 22, 1969, O.A.S.T.S. No. 36, 1144 U.N.T.S. 123 (“Everyone has the right to freedom of thought and expression.”).

<sup>100</sup> *Member States*, ORG. AM. STATES, <https://www.oas.org/ext/en/main/oas/member-states> [perma.cc/JTS2-CYYV] (last visited Nov. 24, 2025) (listing all 34 countries that are part of the Organization of American States).

<sup>101</sup> See American Convention on Human Rights art. 13, *supra* note 99 (stating the consequences of

OAS has made clear, through Article 13, their aim is to ensure that individuals retain their right to freedom of thought and expression.

Further, the Council of Europe has also spoken on regulating hate speech and false information in social media.<sup>102</sup> The Council of Europe has announced that member states may need to take certain actions to combat hate speech and false information.<sup>103</sup> These actions may include legislative reforms, educational initiatives, and possibly working with social media platforms to curb the spread of false information and hate speech. In 2013, the No Hate Speech Movement campaign was launched and led by the Council of Europe Youth Department in order to alleviate the ongoing issue of hate speech online and to promote human rights online.<sup>104</sup>

## B. REGIONAL

The global rise of online hate speech, false information, and harmful content within our digital era has prompted social media regulation internationally.<sup>105</sup> As a result, many countries have enacted laws to attempt to regulate this hateful speech and distorted truths.<sup>106</sup> However, the United States is not the only country that faces difficulties with speech regulation on social media platforms.<sup>107</sup> Although, many countries have attempted to mitigate these dangers,

---

using social media to incite violence); *see also, e.g.*, ORG. AM. STATES, MEDIA LITERACY AND DIGITAL SECURITY TWITTER BEST PRACTICES 9 (2019) (explaining how to change certain settings on social media platforms to protect your privacy interest and elaborates on digital literacy).

<sup>102</sup> *See generally Council of Europe on Hate Speech*, COUNCIL OF EUR., <https://www.coe.int/en/web/combating-hate-speech/council-of-europe-on-hate-speech> (last visited Nov. 24, 2025) (noting that the European Court of Human Rights may restrict hate speech and incitement to violence).

<sup>103</sup> *Id.* (explaining how the European council “may require the States to take certain positive measures.”).

<sup>104</sup> *See No Hate Speech Youth Campaign*, COUNCIL OF EUR., <https://www.coe.int/en/web/no-hate-campaign> (last visited Nov. 24, 2025) (elaborating on the No Hate Speech Youth Campaign, which is a national campaign that allows individuals to report hate speech, provides alternatives for taking action against hate speech, and supplies information on campaigns).

<sup>105</sup> *See, e.g.*, Zachary Laub, *Hate Speech on Social Media: Global Comparison*, COUNCIL ON FOREIGN REL. (June 7, 2019, at 3:51 EST), <https://www.cfr.org/background/hate-speech-social-media-global-comparisons> [perma.cc/Y8JN-VSYC] (explaining how the use of social media for false information has become international, affecting countries such as: United States, Latin America, Europe, Africa, Asia-Pacific, and the Middle East); *see also* DAVID KAYE, *SPEECH POLICE: THE GLOBAL STRUGGLE TO GOVERN THE INTERNET* (2019) (analyzing different international approaches to regulating online content).

<sup>106</sup> *See* Nicholas Barrett, *How Have Social Media Algorithms Changed the Way we Interact*, BBC (Oct. 12, 2024), <https://www.bbc.com/news/articles/cp8e4p4z97eo> [perma.cc/D7WX-ZL6Q] (explaining how social media platforms have been using algorithms for over 15 years, which illustrates how much social media platforms have slowly disrupted the free exchange of ideas between individuals); *see also* Evelyn Douek, *The Rise of Content Cartels*, KNIGHT FIRST AMEND. INST. (Feb. 11, 2020), <https://knightcolumbia.org/content/the-rise-of-content-cartels> [perma.cc/2YRC-FG63] (discussing how platform policies create a de facto system of private regulation across borders).

<sup>107</sup> *See, e.g.*, *Germany: Flawed Social Media Law*, HUM. RTS. WATCH (Feb. 14, 2018, at 12:01 EST), <https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law> [perma.cc/BS89-XN8A] (describing how Germany passed a law regulating online speech on social media platforms,

some laws have inadvertently caused more problems than they have solved, which further illustrates the complexity of speech regulation on social media platforms.<sup>108</sup>

The European Union has taken action against the spread of hate speech and false information, by implementing the Digital Services Act.<sup>109</sup> The Digital Services Act is a law aimed to curb the harmful effects of social media, specifically its role in distorting truths and spreading hate.<sup>110</sup> Further, the Digital Services Act also aims to guard against the growth of artificial intelligence within social media and the reduced content moderation of social media platforms.<sup>111</sup> The European Union has taken into account the regulation of not only social media platforms, but also search engines and internet service providers utilized by individuals in the European Union.<sup>112</sup>

The Digital Services Act affects platforms such as Facebook, Instagram, TikTok, Twitter (now known as X), and YouTube.<sup>113</sup> It requires these platforms

---

yet they faced difficulties because of the dangers of online speech regulation of social media platforms); see also Victor Claussen, *Fighting Hate Speech and Fake News. The Network Enforcement Act (NetzDG) in Germany in the Context of European Legislation*, 3 MEDIA L. 110, 111–12 (2018) (explaining how the regulation of freedom of expression in the context of the Network Enforcement Act is subject to the limitation imposed by the European Court of Human rights).

<sup>108</sup> See Claussen, *supra* note 107, at 115–16 (illustrating the dangers the new Network Enforcement Act posed by giving social media platforms too much discretion when it comes to picking and choosing what type of speech violates the Network Enforcement Act); see also *Germany: Flawed Social Media Law*, *supra* note 107 (arguing that the German law “can lead to unaccountable, over-broad censorship”).

<sup>109</sup> See Grant Baker, *The EU Digital Services Act: A Win for Transparency*, FREEDOM HOUSE (Apr. 4, 2024), <https://freedomhouse.org/article/eu-digital-services-act-win-transparency> [perma.cc/AFJ4-CCWK] (describing the European Union’s Digital Services Act as being implemented as of February 2024); see also LATHAM & WATKINS LLP, *THE DIGITAL SERVICES ACT: PRACTICAL IMPLICATIONS FOR ONLINE SERVICES AND PLATFORMS 1* (2023) (explaining the goal of the European Union, in implementing this law, is to create a “safer and more open digital market”).

<sup>110</sup> See Baker, *supra* note 109 (explaining how the Digital Services Act hopes to address the harms exacerbated by social media platforms which include “disinformation and harassment”); see also Aina Turillazzi et al., *The Digital Services Act: An Analysis of its Ethical, Legal, and Social Implications*, 15 L., INNOVATION & TECH. 84, 90–91 (2023) (expanding on the Digital Services Acts goal to protect against illegal content that is not subject to other speech regulations; therefore casting a wider net to protect the rights and freedoms of individuals on social media platforms).

<sup>111</sup> See Baker, *supra* note 109 (describing the increased utilization of artificial intelligence within social media platforms and how it further threatens the issue of information distortion); see also Samantha Bradshaw, *Influence Operations and Disinformation on Social Media*, CTR. FOR INT’L GOVERNANCE INFO. 41, 44–45 (2020) (explaining the rapid utilization of artificial intelligence and its contributions to the spread of false information).

<sup>112</sup> See Baker, *supra* note 109 (describing how the regulation will now aim to guard issues within “social media platforms, search engines, online marketplaces, and internet services providers”); see also LATHAM & WATKINS LLP, *supra* note 109, at 2 (outlining the scope of the DSA to include various intermediary services).

<sup>113</sup> See European Commission Press Release, *Digital Services Act: Commission Designates First Set of Very Large Online Platforms and Search Engines* (Apr. 25, 2023), [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_23\\_2413](https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413) [perma.cc/5EC4-VKWB] (listing the designated platforms including Facebook, Instagram, TikTok, and YouTube); see also Natasha Lomas, *Europe Names 19 Platforms That Must Report Algorithmic Risk Under DSA*, TECHCRUNCH (Apr. 25, 2023,

to restrict content that violates their own terms and conditions but goes a step further by also requiring these platforms to restrict content that violates the laws of the respective European Union member states.<sup>114</sup> These restrictions are subject to be appealed by users who have had their content removed or altered.<sup>115</sup> The Digital Services Act is monitored by “trusted flaggers” who explicitly point out content that is illegal.<sup>116</sup> A violation of the Digital Services Act results in a 6% fine of the social media platforms global revenue.<sup>117</sup>

### C. DOMESTIC

The United States has a deeply rooted tradition of protecting freedom of speech.<sup>118</sup> This fundamental right is protected under the First Amendment of the Constitution, which encompasses both: the right to information and the freedom of expression.<sup>119</sup> Given the foundational nature of these rights, the regulation of speech is an incredibly difficult and challenging task.<sup>120</sup> An attempt by the United States government to impose restrictions on the right to information or freedom of expression must satisfy the harsh standard of strict scrutiny.<sup>121</sup> Strict scrutiny is a legal standard that requires the government to prove a

---

at 6:38 PDT), <https://techcrunch.com/2023/04/25/europe-names-19-platforms-that-must-report-algorithmic-risks-under-dsa/> [perma.cc/9287-UH5H] (reporting on the platforms designated by the European Commission as being subject to the DSA).

<sup>114</sup> See LATHAM & WATKINS LLP, *supra* note 109, at 2 (describing the DSA’s rules for compliance); see also Turillazzi et al., *supra* note 110, at 90 (analyzing the DSA’s requirement for platforms to restrict illegal content).

<sup>115</sup> See LATHAM & WATKINS LLP, *supra* note 109, at 4 (explaining the mechanisms for user appeals and judicial overview under the DSA); see also *The Digital Services Act*, EUR. COMM’N, [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en) [perma.cc/8QXA-R3KN] (last visited Nov. 24, 2025) (noting that the act includes safeguards for users, including the ability to challenge content moderation decisions).

<sup>116</sup> See LATHAM & WATKINS LLP, *supra* note 109, at 5 (describing the enforcement role of “trusted flaggers”); see also *Trusted Flaggers Under the Digital Services Act (DSA)*, EUR. COMM’N, <https://digital-strategy.ec.europa.eu/en/policies/trusted-flaggers-under-dsa> [perma.cc/N66T-N8D4] (last visited Nov. 24, 2025) (explaining that trusted flaggers are expert entities whose notices of illegal content must be prioritized by online platforms).

<sup>117</sup> LATHAM & WATKINS LLP, *supra* note 109, at 4 (explaining the penalties for a violation of the Digital Services Act).

<sup>118</sup> See *Strict Scrutiny*, *supra* note 31 (stating that the First Amendment guarantees freedoms concerning expression and speech); see also Deborah Fisher, *Social Media and the First Amendment*, FREE SPEECH CTR. (Oct. 24, 2023), <https://firstamendment.mtsu.edu/article/social-media/> [perma.cc/K5QW-T2V7] (expanding on the limitations the Supreme Court has placed on speech regulation).

<sup>119</sup> See Fisher, *supra* note 118.

<sup>120</sup> See Victoria L. Killion, *Freedom of Speech: An Overview*, CONG. RSCH. SERV. (Sept. 13, 2024), <https://crsreports.congress.gov/product/pdf/R/R47986> [https://perma.cc/HTF6-TBZS] (explaining that when a regulation impacts the fundamental right of freedom of expression strict scrutiny is triggered).

<sup>121</sup> See *id.* (describing that when strict scrutiny is triggered the government has the burden of proving they have a compelling state interest, and the means employed are the least restrictive and narrowly tailored).

compelling interest for such a regulation, and that the action is narrowly tailored to achieve that interest in the least restrictive means available.<sup>122</sup>

However, despite the strong protections afforded by the First Amendment, the right to information and freedom of expression is not without limitations. The Supreme Court has recognized several exceptions where speech can be restricted without triggering strict scrutiny.<sup>123</sup> These exceptions are limited to five specific categories: (1) incitement; (2) fighting words that provoke immediate violence; (3) true threats of violence; (4) obscenity; and (5) defamation.<sup>124</sup> These narrow exceptions allow for the regulation of harmful or dangerous speech while maintaining the core principles of freedom of expression. However, these regulations are not sufficient to regulate all hate speech and distortions of truths online. Several states have attempted to further regulate online speech on social media platforms beyond the restrictions in the First Amendment.<sup>125</sup>

The United States has considered the dangers that social media poses, but more specifically, the dangers it poses on children. The United States passed a law aiming to protect children from utilizing social media without parental consent.<sup>126</sup> The Protecting Kids on Social Media Act was passed to require social media platforms to: (1) verify the age of the user; (2) not allow account creation without age verification; and (3) limit access to platforms for children under

---

<sup>122</sup> See *R.A.V. v. St. Paul*, 505 U.S. 377, 381 (1992) (explaining that an ordinance needs to be “narrowly tailored to serve a compelling governmental interest in protecting the community against bias-motivated threats”).

<sup>123</sup> See *Unprotected Speech Synopsis*, FIRE, <https://www.thefire.org/research-learn/unprotected-speech-synopsis> [perma.cc/9MYC-W3SU] (last visited Nov. 24, 2025) (describing the five exceptions to the fundamental right to information and freedom of expression); see also Caitlin Carlson, *Censoring Hate Speech in U.S. Social Media Content: Understanding the User’s Perspective*, 17 COMM’N L. REV. 24, 29 (2017) (elaborating on the regulations of online speech and emphasizing that if the speech does not fall within one of the exceptions, it will most likely be protected).

<sup>124</sup> Cf. *Unprotected Speech Synopsis*, *supra* note 123 (explaining why the five exceptions were established and are not afforded protections under the First Amendment); see also *Brandenburg v. Ohio*, 395 U.S. 444, 449 (1969) (establishing that incitement is not afforded protection under the First Amendment); see also *Chaplinsky v. New Hampshire*, 315 U.S. 568, 571–72 (1942) (holding that freedom of speech is not absolute and does not include “insulting or ‘fighting words’ -- those which by their very utterance inflict injury or tend to incite an immediate breach of the peace); see also *Virginia v. Black*, 538 U.S. 343, 360 (2003) (explaining why a prohibition on true threats “protects individuals from the fear of violence and the disruption that fear engenders, as well as from the possibility that the threatened violence will occur”); see also *Miller v. California*, 413 U.S. 15, 37–38 (1973) (explaining that obscenity does not have any First Amendment protection because it is “utterly without redeeming social importance” (citing *Chaplinsky*, 315 U.S. at 571–72)); see also *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 268 (1964) (explaining that defamatory language is not protected under the First Amendment if liable to cause “violence and disorder”).

<sup>125</sup> See, e.g., Joe Mullin, *2021 Was the Year Lawmakers Tried to Regulate Online Speech*, ELEC. FRONTIER FOUND. (Dec. 26, 2021), <https://www.eff.org/deeplinks/2021/12/2021-was-year-lawmakers-tried-regulate-online-speech> [perma.cc/7WF6-7GAM] (explaining that Florida passed a bill prohibiting the banning of presidential candidates from utilizing social media platforms and that Texas passed a bill prohibiting social media platforms from moderating content-based speech).

<sup>126</sup> See S. 1291, 118th Cong. (2023) (describing how the Protecting Kids on Social Media Act requires parental consent for children under the age of 13 to utilize social media).

thirteen years old.<sup>127</sup> This regulation will incentivize social media platforms to comply, because the penalty for non-compliance results in a civil penalty of \$10,000 multiplied by whichever is greater: (1) number of days the social media platform was in violation, or (2) the number of users harmed because of the violation.<sup>128</sup> In turn, the Protecting Kids on Social Media Act will curb the effect on children being subjected to hate speech and false information on social media platforms.

Moreover, hate speech on social media is more difficult to regulate in the United States because it concerns a fundamental right. In 1964, the Supreme Court of the United States held that the First Amendment extends protections to false statements so long as the statements have not been made with actual malice.<sup>129</sup> The court's reasoning for this extension was to promote open debate and further protect democratic society.<sup>130</sup> However, extending such protections creates dangers by casting too broad of a net. Policymakers now fear that this broad net over protected speech is facilitating hate speech, online threats, and the spread of dangerous information.<sup>131</sup>

In 1996, the Telecommunications Act shifted federal communication policies, it was one of the first acts to target speech.<sup>132</sup> The Telecommunications Act required v-chips in new televisions to allow parents to block certain channels from their minor children and “deregulated cable television service.”<sup>133</sup> The purpose of this act was to allow parents to control certain content that their children may be exposed to, specifically, content that included indecent

---

<sup>127</sup> See *id.* (stating the three requirements the Protecting Kids on Social Media Act imposes on social media platforms); see also *Social Media and Children 2024 Legislation*, NCSL (Nov. 22, 2024), <https://www.ncsl.org/technology-and-communication/social-media-and-children-2024-legislation> [<https://perma.cc/X5GG-BXT5>] (elaborating on the concerns that have come along with the rise of social media and how it may affect children's mental health).

<sup>128</sup> See Matthew Bergman, *What is the Protecting Kids on Social Media Act?*, SOC. MEDIA VICTIMS L. CTR. (Aug. 14, 2025), <https://socialmediavictims.org/congress/protecting-kids-on-social-media-act/> [<https://perma.cc/Q2A9-5WGV>] (expanding that violations may result in civil penalties of \$10,000 multiplied by either the number of days the social media company was in violation or by number of users harmed).

<sup>129</sup> See *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 262 (1964) (stating that a requirement for liability under libel includes proof of actual malice under Alabama law).

<sup>130</sup> See *id.* at 271–72 (reasoning that free debate must be protected if freedom of expression is to have “breathing space”).

<sup>131</sup> See MIKE GODWIN, *CYBER RIGHTS: DEFENDING FREE SPEECH IN THE DIGITAL AGE 17–21* (MIT Press ed., 2003) (stating that “fear breeds repression; and repression breeds hate”).

<sup>132</sup> See *Telecommunication Act of 1996*, FED. COMM'N COMM'N (June 20, 2013), <https://www.fcc.gov/general/telecommunications-act-1996> [[perma.cc/7M6G-4C2J](https://perma.cc/7M6G-4C2J)] (emphasizing how the Telecommunications Act was one of the first major acts to target telecommunication and its main goal was to foster competition within the industry).

<sup>133</sup> Ruth Ann Strickland, *Telecommunication Act of 1996*, FREE SPEECH CTR. (Jan. 1, 2009), <https://firstamendment.mtsu.edu/article/telecommunications-act-of-1996/> [[perma.cc/LE2P-GWE4](https://perma.cc/LE2P-GWE4)] (explaining how the provisions within the Telecommunications Act aimed to: deregulate cable television service, allow local telephone companies to provide service, and requires v-chips to facilitate parents' ability to regulate cable access to their children, and expand the telecommunication industry).

materials.<sup>134</sup> However, soon after the Telecommunications Act passed, it became one of the most controversial acts because free speech was directly implicated through the Communications Decency Act.<sup>135</sup> The Communications Decency Act was part of the Telecommunication Act of 1996 as an attempt to diminish access to indecent materials on the internet to minors.<sup>136</sup>

In fact, in the case of *Reno v. American Civil Liberties Union* in 1997, the Supreme Court established that the Internet as a medium was entitled to all the protections afforded by the First Amendment.<sup>137</sup> Because of *Reno*, the Telecommunications Act was deemed overbroad and encroached on protected speech.<sup>138</sup> The Telecommunications Act was soon invalidated because it was criminalizing speech that was afforded protection under the First Amendment of the United States Constitution.<sup>139</sup> This seminal case has defined the limits on regulating online speech even when the government has a compelling interest in protecting minors from obscenity. The invalidation of the Telecommunications Act on 1996 shines light on the ongoing struggle of balancing regulation of online speech and protecting fundamental rights.

The recently enacted Take It Down Act aims to combat against a new growing issue within our digital era, the publication of non-consensual intimate imagery (NCII), which has been facilitated by the developments in artificial intelligence and the misuse of social media platforms.<sup>140</sup> False information and hate speech have significantly contributed to the uproar in social media misconduct.<sup>141</sup> The Take It Down Act further illustrates the upward slope our generation is taking on utilizing social media for negative impacts as opposed to its intended affects. However, the Take It Down Act highlights a broader issue at

---

<sup>134</sup> See *id.* (describing how the Telecommunication Act of 1996 allowed parents to supervise the content their children watched).

<sup>135</sup> See *id.* (elaborating on how the Communication Decency Act, which was part of the Telecommunications Act, became an extremely controversial act because individuals expressed concerns about First Amendment violations).

<sup>136</sup> See Sara L. Zeigler, *Communications Decency Act and Section 230 (1996)*, FREE SPEECH CTR. (May 23, 2023), <https://firstamendment.mtsu.edu/article/communications-decency-act-and-section-230/> [perma.cc/BCN9-YUML] (explaining how the Communications Decency Act aimed to prohibit minors from accessing obscene and sexually explicit publications on the internet).

<sup>137</sup> See *Reno v. ACLU*, 521 U.S. 844, 851 (1997) (“Anyone with access to the internet may take advantage of a wide variety of communication and information retrieval methods.”).

<sup>138</sup> See *id.* at 846 (elaborating on how the Communications Decency Act suppressed a large amount of speech).

<sup>139</sup> See *id.* at 846–47 (explaining how the Communications Decency Act was not found to be narrowly tailored; therefore, the Communications Decency Act was deemed invalid because it did not pass strict scrutiny).

<sup>140</sup> See *Take It Down Act*, RAINN, <https://rainn.org/policy/take-it-down-act> (last visited Nov. 24, 2025) (describing how the Take It Down Act will compel platforms to remove NCII within 48 hours of a victim’s request and take all necessary steps to remove any duplicates of the imagery).

<sup>141</sup> See Caroline Brooks, *The Truth Behind Fake News and Politics on Social Media*, MICH. STATE UNIV. (June 2, 2020), <https://msutoday.msu.edu/news/2020/the-truth-behind-fake-news-and-politics-on-social-media> [perma.cc/VJ4Q-CS3W] (stating how fake news and hate speech is all over social media platforms which is extremely dangerous because society relies so heavily on social media now).

hand regarding social media platforms exploiting false information. Artificial intelligence has facilitated NCII, but it has also facilitated the generation of false information and the spread of it.<sup>142</sup> The danger arises when these artificial intelligence companies weaponize information to manipulate public perceptions and incite violence.<sup>143</sup>

The state of Ohio recently tried to regulate social media platforms when they passed the Parental Notification by Social Media Operators Act (“Act”).<sup>144</sup> The Act requires parental consent before allowing “any unemancipated minor under the age of sixteen, [to register] or [create] an account on their platform.”<sup>145</sup> However, the Act was found unconstitutional due to the First Amendment implications. The Act aimed to protect the utilization of social media platforms by minors and ensure that parents had notice of their child’s utilization of social media.<sup>146</sup> Although the state of Ohio had rational concerns regarding social media access to minors, the Act was found unconstitutional because it was not narrowly tailored.

Moreover, the United Kingdom has also acted on regulating online speech. The United Kingdom recently introduced the Online Safety Act aiming to protect minors and adults from unregulated speech on social media platforms.<sup>147</sup> The Online Safety Act holds social media platforms accountable for not removing illegal content when it appears.<sup>148</sup> For minors, social media platforms are required to prevent them from having access to inappropriate content.<sup>149</sup> As for

---

<sup>142</sup> See *Societal Risks and Well-Being*, NAT’L TELECOMM. & INFO. ADMIN., <https://www.ntia.gov/programs-and-initiatives/artificial-intelligence/open-model-weights-report/risks-benefits-of-dual-use-foundation-models-with-widely-available-model-weights/societal-risks-well-being> [perma.cc/RRP3-WB2M] (last visited Nov. 24, 2025) (explaining the dangers of artificial intelligence because it is able to create realistic imagery).

<sup>143</sup> See generally Mike Reilley, *Opinion: Navigating Accountability, the Roles and Risks of Using AI in Journalism*, RED LINE PROJECT (Nov. 9, 2024, at 10:41 PM), <https://redlineproject.news/2024/11/09/opinion-navigating-accountability-the-roles-and-risks-of-using-ai-in-journalism/> [perma.cc/RJY4-CBP7] (highlighting the issues with artificial intelligence which include, but are not limited to: data accuracy, false information, negative effects on organizations).

<sup>144</sup> See *NetChoice, LLC v. Yost*, 716 F. Supp. 3d 539, 546 (S.D. Ohio 2024) (discussing whether Ohio’s state law, Parental Notification by Social Media Operators Act, violates the First Amendment).

<sup>145</sup> See *id.*

<sup>146</sup> See *id.* (describing the State of Ohio’s interest behind the Parental Notification by Social Media Operation Act).

<sup>147</sup> See *Online Safety Act: Explainer*, DEP’T FOR SCI., INNOVATION & TECH. <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer> [perma.cc/TQU6-DQCQ] (last visited Nov. 24, 2025) (describing United Kingdom’s action against hate speech and false information); see also Peter Coe, *Tackling Online False Information in the United Kingdom: The Online Safety Act of 2023 and its Disconnection From Free Speech Law and Theory*, 15 J. OF MEDIA L. 213, 214 (2023) (explaining the United Kingdom’s recognition of the ongoing problem of dissemination of false information within social media platforms).

<sup>148</sup> See *Online Safety Act: Explainer*, *supra* note 147 (“[The Online Safety Act] puts a range of new duties on social media companies and search services, making them more responsible for their users’ safety on their platforms.”).

<sup>149</sup> See *id.* (“[Social media platforms] will be required to prevent children from accessing harmful and age-inappropriate content and provide parents and children with clear and accessible ways to

adults, social media platforms will need to be transparent as to whether content may be potentially harmful or not.<sup>150</sup> Social media platforms will be fined up to £18 million or 10% of their revenue if found in violation of the Online Safety Act.<sup>151</sup> However, the Online Safety Act only applies to social media platforms or search engines that receive significant users from the United Kingdom.<sup>152</sup>

In 2018, Germany's Network Enforcement Act went into effect, which aimed to regulate online social media speech.<sup>153</sup> The Network Enforcement Act compels social media platforms to remove harmful content, including hate speech and false information, from their platforms.<sup>154</sup> As a penalty for non-compliance the Network Enforcement Act imposes a 50 million euro fine on social media platforms.<sup>155</sup> The Network Enforcement Act has been a stepping stone toward the direction of limiting hate speech and false information on social media platforms. The Network Enforcement Act is also a step away from utilizing social media platforms as a tool for spreading propaganda, false information, and hate speech.

The dangers within the Network Enforcement Act include turning social media platforms into "overzealous" censors to avoid the imposed fines for violation of the Network Enforcement Act.<sup>156</sup> The Human Rights Watch critiqued the Network Enforcement Act because it places too high of a burden on social media platforms to sift through content and decide what speech violates the law.<sup>157</sup> This may include social media platforms making decisions motivated

---

report problems online when they do arise.").

<sup>150</sup> See *id.* (explaining how the Online Safety Act also protects adults because social media platforms will now be required to be more transparent about their content and give their users more control over what they want to see).

<sup>151</sup> See *id.* (describing how fines will be imposed against social media platforms who violate the Online Safety Act, including imposing criminal liability against senior managers of the social media platforms).

<sup>152</sup> See *id.* (explaining how the Online Safety Act is not limited to those social media platforms that are only based in the United Kingdom, but instead, social media platforms that have a "significant" number of users from the United Kingdom).

<sup>153</sup> See *Germany: Flawed Social Media Law*, *supra* note 107 (describing how Parliament approved the Network Enforcement Act on June 20, 2017, and it took effect on January 1, 2018); see also Thomas Kasakowskij et al., *Network Enforcement as Denunciation Endorsement? A Critical Study on Legal Enforcement in Social Media*, 46 *TELEMATICS & INFORMATICS* (2020) (explaining that Germany's intent behind the Network Enforcement Act was to combat defamation and incitement).

<sup>154</sup> See Kasakowskij et al., *supra* note 153 (explaining the type of social media content that is not allowed and would be a violation of the Network Enforcement Act).

<sup>155</sup> See *id.* (discussing the penalties social media platforms will face if found in violation of the Network Enforcement Act); see also Rachel Griffin, *New School Speech Regulation and Online Hate Speech: A Case Study of Germany's NetzDG*, *HAL OPEN SCI.* 1, 9–10 (2022) (explaining how the Network Enforcement Act threatens large platforms with heavy fines for "systematic failure to promptly remove illegal content").

<sup>156</sup> See Kasakowskij et al., *supra* note 153 (elaborating on the dangers of social media platforms being the main regulator when it comes to enforcement of the Network Enforcement Act); see also Griffin, *supra* note 155, at 15 (describing how the Network Enforcement Act restricts free speech and gives regulation power to private conglomerate corporations, resulting in too much power to these social media platforms).

<sup>157</sup> See Griffin, *supra* note 155, at 9 (explaining the dangers of social media platforms regulating

by political or biased views. Secondly, Human Rights Watch also criticizes the Network Enforcement Act because it lacks any judicial oversight; therefore, social media platforms will not be held accountable for wrongfully removing speech that is not hateful or distorted.<sup>158</sup> A lack of judicial oversight would leave users without a cause of action or opportunity to appeal. However, despite the many dangers posed by the German law, it has created widespread influence for other countries to implement laws to regulate social media.<sup>159</sup>

## V. FUTURE DECISIONS IN LIGHT OF CHANGED AND CHANGING CONDITIONS

Given the significant challenges in regulating online speech, legislators are shifting their focus to social media platform regulations to tackle this issue from a different angle. While hate speech and false information pose serious threats to society, directly regulating speech is a highly complex task. By regulating social media platforms and requiring transparency, users will be more aware of targeted hate speech and false information. Overall, social media platforms will become harder to saturate with false information and hate speech and, in turn, public perception will become harder to distort.

### A. FILTER BUBBLE TRANSPARENCY ACT

The Filter Bubble Transparency Act is a bill that has been introduced to Congress in order to combat echo chambers.<sup>160</sup> The bill requires large social media platforms to be more transparent on how they utilize artificial intelligence to curate algorithms for users.<sup>161</sup> The proposed bill requires social media platforms to allow users to choose between an algorithm generated by artificial intelligence or by the social media platform.<sup>162</sup> Additionally, the Filter Bubble

---

sensitive materials); *see also* Sandra Schmitz & Christian M. Berndt, *The German Act on Improving Law Enforcement on Social Networks (NetzDG): A Blunt Sword?* 13 (Working Paper No. 3306964/2018, 2018), <https://ssrn.com/abstract=3306964> [<http://dx.doi.org/10.2139/ssrn.3306964>] (explaining how one of the issues the Network Enforcement Act poses is granting an extensive interpretation power for content).

<sup>158</sup> *See Germany: Flawed Social Media Law*, *supra* note 107 (describing one of the critiques the Human Rights Watch has on the Network Enforcement Act); *see also* Griffin, *supra* note 155, at 15 (describing how the Network Enforcement Act restricts free speech and gives the regulation power to private conglomerate corporations).

<sup>159</sup> *Cf. Germany: Flawed Social Media Law*, *supra* note 107 (listing countries that have cited the Network Enforcement Act as support to help pass their own laws to help regulate speech on social media platforms including: Singapore, Philippines, Russia, Venezuela, and Kenya).

<sup>160</sup> *See* Filter Bubble Transparency Act, S. 2024, 117th Cong. (2021) (proposing user control over algorithmic curation of content).

<sup>161</sup> *See id.* (explaining what the Filter Bubble Transparency Act aims to do by giving users the option to utilize large social media platforms with knowledge on how their algorithms are curated).

<sup>162</sup> *See* Press Release, John Thune, U.S. Senator, Thune's Big Tech Algorithm Transparency Bill Unanimously Approved by Commerce Committee (July 27, 2023) (explaining how users will be given the choice to choose between the two types of algorithms if the bill passes).

Transparency Act essentially gives users the power to control how to encounter content on large social media platforms.

#### B. PLATFORM ACCOUNTABILITY AND CONSUMER TRANSPARENCY ACT

The Platform Accountability and Consumer Transparency Act (“PACT”) has also been introduced to Congress.<sup>163</sup> PACT requires social media platforms to provide access to “advertising libraries and disclos[e] . . . ‘highly disseminated’ or viral content.”<sup>164</sup> PACT also requires social media platforms to divulge how they create their recommendations and rankings on their platforms for users.<sup>165</sup> This newly introduced bill would allow for more transparency between social media platforms and their users and would allow for a decrease of social media platforms creating biased algorithms because they would need to disclose it to their users.

### VI. APPRAISAL, ALTERNATIVES, AND RECOMMENDATION OF SOLUTIONS

#### A. APPRAISAL

A recurring theme in addressing hate speech and false information is the undeniable reality that society is deeply immersed in a digital era, with this era only further expanding and becoming more intertwined with our day-to-day life. As social media continues to expand alongside society, it is critical to keep in mind that the utilization of these social media platforms should be for their intended purpose and not as a weaponization tool. Throughout our history, humans have demonstrated a habit to exploit new innovations in both the positive and the negative, which shines light on the importance of implementing safeguards to social media utilization.

It is evident that the dangers of social media utilization encompasses, but is not limited to: the spread of an immense amount of hate speech that leads to genocides; false information to distort reality; and even propaganda by large terrorist groups.<sup>166</sup> However, on the other side of that token, society has been able to communicate with individuals thousands of miles away; explore

---

<sup>163</sup> See Platform Accountability and Consumer Transparency (PACT) Act, S. 797, 117th Congress (2021) (proposing greater transparency in online content moderation).

<sup>164</sup> John Perrino, *Platform Accountability and Transparency Act Reintroduced in Senate*, STAN. UNIV. CYBER POL’Y CTR. (June 8, 2023), <https://www.techpolicy.press/platform-accountability-and-transparency-act-reintroduced-in-senate/> [<https://perma.cc/S9JP-QFQG>] (describing the bill and its new requirements for large social media platforms).

<sup>165</sup> *Id.*

<sup>166</sup> *Cf.* Lower, *supra* note 18, at 2 (describing how propaganda in media sources were utilized to incite violence which further escalated the Rwandan genocide).

different cultures around the world without leaving the comfort of our homes; and even keep in touch with our loved ones.<sup>167</sup>

Many countries have seen the dangers and beauties social media has to offer society, resulting in some countries taking a more proactive approach to combat the war with hate speech and false information; while some countries have focused more on censorship, as opposed to speech regulation, due to the complexities of balancing fundamental rights and unprotected speech.

### 1. Domestic Comparison

On a domestic level, the United States and Germany, both have different approaches as to how they choose to handle social media regulations. The United States has strong protections set in place to protect online speech regulation.<sup>168</sup> The United States has also given social media platforms immunity from being imposed liability when their platforms facilitate hate speech, false information, and propaganda.<sup>169</sup> However, the rise of social media within our society and its effect on younger generations have urged legislatures to curate acts to spread awareness for online safety. Yet, there are still no federal requirements for removal of hate speech or false information.

On the other hand, Germany has focused more on implementation and enforcement when regulating social media. Germany has set forth acts, such as the Network Enforcement Act, which requires platforms to take down content containing hate speech within 24 hours of receiving a complaint.<sup>170</sup> Unlike the United States, Germany enforces their acts by imposing high fines for non-compliance. Overall, German law allows for more regulation of online speech compared to the United States. While both countries have expressed their concerns regarding social media regulations, both take very different approaches.

### 2. Common Law Comparison

The United Kingdom and the United States both have common law legal systems, which calls for an interesting comparison between the two. As can be seen from their steps toward online speech regulation, both countries are primarily concerned with the online harm that directly affects minors. However, the United States is also concerned with not violating the fundamental rights of

---

<sup>167</sup> Cf. Sawyer, *supra* note 42, at 2 (elaborating on how social media has provided a manner for individuals to communicate and further connect with other individuals from the comfort of their own homes).

<sup>168</sup> See *Reno v. ACLU*, 521 U.S. 844, 868 (1997).

<sup>169</sup> See Rosie Moss, *The Future of Section 230 What Does it Mean For Consumers?*, NAT'L ASS'N OF ATT'Y GEN. (July 21, 2023), <https://www.naag.org/attorney-general-journal/the-future-of-section-230-what-does-it-mean-for-consumers/> [perma.cc/CXE7-42SA] (elaborating on the low impositions placed on social media platforms).

<sup>170</sup> See Kasakowskij, *supra* note 153 (explaining the requirements the Network Enforcement Act places on social media platforms).

freedom of speech, freedom of expression, and right to information as stated in the Constitution.<sup>171</sup>

Since the United States has more stringent procedures to regulate online speech, they focus on promoting parental involvement to intervene with the harms posed by social media.<sup>172</sup> Alternatively, the United Kingdom has taken a different approach by requiring platforms to mitigate the harms that users spread on their platforms. This approach allows the United Kingdom to mandate social media platforms to be safer for utilization by minors.

#### B. ALTERNATIVES AND RECOMMENDATIONS

The goal remains clear: stop the utilization of social media as a warfare tactic. To achieve this goal, countries should consider enacting new acts and regulations: to help increase transparency between social media platforms and users, to hold social media platforms accountable, and to establish an international cooperation plan. Transparency by social media platforms would allow users to take control of the content they wish to interact with and would aid users in avoiding false information. Moreover, holding social media platforms accountable would also incentivize platforms to take down hate speech, false information, and propaganda within reasonable time. Most importantly, an international cooperation plan is not only a recommendation to combat against social media utilization as a warfare tactic, but a necessity. Countries need to agree to work together to stop the spread of hate speech and false information. This international plan would include countries sharing technological advancements that have been created specifically for the purpose of combating against harm caused by social media platforms. Although regulating online speech is significantly challenging, it is critical to a broader objective: preventing the utilization of social media as a warfare tactic. Ultimately, social media truly extends beyond likes and reposts.

---

<sup>171</sup> See U.S. CONST. amend. I.

<sup>172</sup> See Ben Sperry, *A Coasean Analysis of Online Age-Verification and Parental Consent Regimes*, INT'L CTR. FOR L. & ECON. (Nov. 9, 2023), <https://laweconcenter.org/resources/a-coasean-analysis-of-online-age-verification-and-parental-consent-regimes/> [perma.cc/P2CB-ERGU] (describing how online safety may be reached if parents and children work together to avoid the dangers of social media).