

# UC Irvine

## UC Irvine Law Review

### Title

Disinformation Legislation and Freedom of Expression

### Permalink

<https://escholarship.org/uc/item/1x1532sm>

### Journal

UC Irvine Law Review , 10(2)

### ISSN

2327-4514

### Author

Nuñez, Fernando

### Publication Date

2020-03-01

# Disinformation Legislation and Freedom of Expression

Fernando Nuñez\*

*The rise of disinformation on social media has prompted governments around the world to enact legislation that may affect every person's right to freedom of opinion and expression. This Note seeks to explore the rise of disinformation and the legal framework that applies and to highlight some of the recent proposals to combat disinformation on social media.*

Introduction .....	784
I. The Problem of Disinformation on Social Media Platforms .....	785
II. Marketplace of Ideas as an Alternative to Legislation?.....	787
III. Legal Framework .....	789
A. United States .....	789
B. International.....	790
IV. Disinformation Legislation.....	791
A. Digital Media Literacy.....	792
B. Transparency Initiatives .....	792
1. Bots .....	793
2. Location .....	794
3. Advertisements .....	795
C. Platform Liability .....	796
Conclusion.....	797

---

\* J.D. University of California, Irvine School of Law. I would like to thank Professor David Kaye for his guidance both in writing this article and throughout my time at the International Justice Clinic. I would also like to thank the members of the *UC Irvine Law Review* for their assistance in preparing this article for publication.

## INTRODUCTION

The right to freedom of opinion and expression is a central pillar of democratic societies and a guarantor of free and fair electoral processes. Disinformation poses a significant threat to that right as it is increasingly used to influence elections and public perception around the world. An aggressive disinformation campaign can distort the “marketplace of ideas,” where truth is expected to triumph, by flooding the market with falsities. Although the problem of disinformation has been around for centuries, it has been compounded by the internet and popularity of social media platforms. The combination of social media and advancing technology has enabled the production of disinformation at a scale that threatens an individual’s right to receive and impart ideas.

In the United States, ninety-three percent of the adult population said they received at least some news online,<sup>1</sup> with over sixty percent of them getting their news from social media platforms.<sup>2</sup> These platforms have become a necessity around the world as well. It is estimated that eighty-six percent of Middle Eastern internet users rely upon social media platforms, compared to eighty-two percent in Latin America, seventy-six percent in Africa, sixty-six percent in Asia and the Pacific, and sixty-five percent in Europe.<sup>3</sup> This new trend of increasing reliance on social media for news is rising at the same time that trust in traditional media is declining.<sup>4</sup>

In recent years, legislation aimed at countering disinformation has surged around the world.<sup>5</sup> Governments are experimenting with different strategies, ranging from creating task forces to outright criminalizing the dissemination of false content.<sup>6</sup> While the issue of disinformation is deserving of the attention it is receiving, some governmental responses have posed significant problems of their own. Some have gone as far as to shut down the internet to prevent the spread of disinformation.<sup>7</sup>

---

1. *Digital News Fact Sheet*, PEW RES. CTR. (July 6, 2019), <http://www.journalism.org/fact-sheet/digital-news/> [http://www.journalism.org/fact-sheet/digital-news/].

2. Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RES. CTR. (May 26, 2016), <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/> [https://perma.cc/XPF5-Y2LA].

3. Jacob Poushter, *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies*, PEW RES. CTR. (Feb. 22, 2016), <http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/> [https://perma.cc/EDC8-J3HF].

4. Art Swift, *Americans’ Trust in Mass Media Sinks to New Low*, GALLUP (Sept. 14, 2016), <https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx> [https://perma.cc/H4HW-2BAK].

5. Daniel Funke & Daniela Flamini, *A Guide to Anti-Misinformation Actions Around the World*, POYNTER, <https://www.poynter.org/ifcn/anti-misinformation-actions/> [https://perma.cc/82VR-HXMD] (last updated Aug. 13, 2018).

6. *Id.*

7. Tinuola Dada & Peter Micek, *Launching STOP: The #KeepItOn Internet Shutdown Tracker*, ACCESS NOW (Sept. 7, 2017, 12:42 PM), <https://www.accessnow.org/keepiton-shutdown-tracker/> [https://perma.cc/JY68-S5J5].

In light of the rise in disinformation legislation, it is helpful to understand the freedom of expression framework that applies and examine the different approaches legislators can take while still conforming to their obligation to uphold the right to freedom of opinion and expression. This Note focuses on disinformation legislation as it relates to social media platforms and explores some of the potential options that would be appropriate to combat the spread of disinformation. Part I serves as context for disinformation, why it currently poses a problem, and how it will worsen with technological advances. Part II describes how the marketplace of ideas has failed to strike a balance with disinformation and, as a result, legislation may be necessary to restore the market's equilibrium. Part III briefly sets forth the relevant domestic and international legal frameworks from which any such disinformation legislation will need to conform. Part IV then analyzes some of the more popular legislation that has either been enacted by certain governments or proposed by various actors, including legislators, civil society, and international organizations.

Ultimately, legislation that is aimed at increasing people's digital media literacy and access to information on social media platforms is in line with freedom of expression principles and could reduce the impact of disinformation. Such an approach avoids the risk of governments or private actors wielding the power to determine what is truthful and places that power with the users. However, more aggressive legislation may be viable depending on how it is implemented. Legislation that seeks to hold social media platforms liable for failing to remove content can conform to freedom of expression principles if it satisfies due process requirements and does not overly burden the platforms or encourage them to be overinclusive with their removals.

#### I. THE PROBLEM OF DISINFORMATION ON SOCIAL MEDIA PLATFORMS

As it is used in this Note, disinformation is defined as the deliberate creation and sharing of information that one knows is false.<sup>8</sup> Disinformation is often confused with misinformation, which is the inadvertent sharing of false information,<sup>9</sup> and both are frequently lumped together under the all-encompassing term "fake news." However, it is important to recognize the distinction between the three terms to properly address the unique threats that each one presents. Disinformation is a more serious threat to freedom of expression because it is information that is deliberately created to mislead and influence the public, unlike

---

8. European Comm'n High Level Expert Grp. on Fake News & Online Disinformation, *A Multi-dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation*, EUR. COMMISSION (Mar. 12, 2018), available at <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation> [<https://perma.cc/9LMJ-FDYR>] [hereinafter *A Multi-dimensional Approach to Disinformation*].

9. CHERILYN IRETON & JULIE POSETTI, UNESCO, JOURNALISM, 'FAKE NEWS' & DISINFORMATION 7 (2018), [https://en.unesco.org/sites/default/files/journalism\\_fake\\_news\\_disinformation\\_print\\_friendly\\_0\\_0.pdf](https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0_0.pdf) [<https://perma.cc/YHS4-QMWY>].

misinformation, which may be shared under a genuine belief that its contents are truthful.<sup>10</sup> Finally, using the term “fake news” is not helpful because it has been routinely used to describe subjectively unfavorable content or inaccurate content that is the result of mistake.<sup>11</sup>

Disinformation has been an issue that societies have dealt with for centuries. It can be traced as far back as the invention of the printing press in 1439.<sup>12</sup> But the technology that is available today amplifies the effects of disinformation to levels previously unimaginable.<sup>13</sup> For example, modern technology allows for the creation of convincing fake audio and video content known as “Deepfake.”<sup>14</sup> This type of content is made possible by automated computer programs that allow for users to superimpose images and videos onto unrelated images or videos.<sup>15</sup> This used to be a time-intensive and skilled task, but artificial intelligence programs have made it easier and more efficient for people to create convincing fake videos of individuals doing or saying things they never did.<sup>16</sup> The spread of this technology means that the line between fact and fiction may be blurred with negative consequences. Either people will believe the videos, which, if released at strategic times, could sway public opinion in an election or incite protests, or people will stop believing in videos as objective depictions of reality, which will make it increasingly difficult to prove a fact.

Modern technology has not only provided new, convincing, false content, it has also facilitated its dissemination. Social media platforms have made sharing content faster than ever by the retweeting, sharing, or reposting mechanisms they have implemented. This may not be a problem on its own, but recent research suggests that not all content spreads at the same rate. Research from Massachusetts Institute of Technology (MIT) suggests that false content spreads up to six times faster than factual content on social media sites and false news stories are seventy percent more likely to be shared.<sup>17</sup> Besides the users’ eagerness to share novel content with their circles, a contributing factor to the quick spread of false news is

---

10. *Id.*

11. *Id.* at 14.

12. Jacob Soll, *The Long and Brutal History of Fake News*, POLITICO (Dec. 18, 2016), <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535> [<https://perma.cc/4UFT-A9AT>].

13. IRETON & POSETTI, *supra* note 9, at 15.

14. *Id.* at 111.

15. Donie O’Sullivan, *When Seeing Is No Longer Believing: Inside the Pentagon’s Race Against Deepfake Videos*, CNN, <https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/> [<https://perma.cc/U927-Z9FM>] (last visited Jan. 3, 2020).

16. *Id.*

17. Peter Dizikes, *Study: On Twitter, False News Travels Faster Than True Stories*, MIT NEWS (Mar. 8, 2018), <http://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308> [<https://perma.cc/C25X-D3GA>].

that algorithms used in some platforms tend to suggest similar content from similar sources.<sup>18</sup>

The internet has also incentivized disinformation by turning it into a lucrative business.<sup>19</sup> Disinformation is cheap to produce because its producers do not have to engage in time-consuming cross-referencing or fact-checking. Disinformation also has nominal overhead costs to disseminate because posting on social media platforms is generally free. With the costs of production and distribution almost insignificant, a website dedicated to spreading disinformation can use online advertising to make a profit without the costs associated with investigation and verification that burden professional news agencies.<sup>20</sup> In 2016, a small town in Macedonia was discovered to be the source of many of these websites that were designed to profit from the U.S. presidential election.<sup>21</sup> One of the website's operators claimed to make \$2500 a day from this enterprise.<sup>22</sup> That business is preparing to continue its operations during the 2020 U.S. presidential election.<sup>23</sup>

## II. MARKETPLACE OF IDEAS AS AN ALTERNATIVE TO LEGISLATION?

In the marketplace of ideas, falsehoods compete with truth until truth prevails.<sup>24</sup> However, the market will not work if it has been compromised. As with most markets, “if the competitive nature of the market is eliminated . . . or if the market’s rationality is corrupted through socialization or propaganda, then the marketplace can no longer be trusted to properly value a particular good or service.”<sup>25</sup> The recent proliferation of disinformation on the internet seems to call into question the theory’s presuppositions and renders it inadequate to remedy the problem.

The marketplace of ideas rests on three pillars.<sup>26</sup> First, that ideas clash and are in constant dialogue with each other.<sup>27</sup> This requires the ideas to be accessible. Second, that the participants in the market are capable and willing to distinguish

18. *Id.*; Juan Carlos Escudero de Jesús, *Fake News and the Systemic Lie in the Marketplace of Ideas: A Judicial Problem?*, 87 REVISTA JURIDICA U.P.R. 1394, 1407 (2018).

19. Clare Duffy, *Websites That Peddle Disinformation Make Millions of Dollars in Ads, New Study Finds*, CNN (Aug. 18, 2019, 3:57 PM), <https://www.cnn.com/2019/08/18/tech/advertising-disinformation-money-reliable-sources/index.html> [<https://perma.cc/R7KY-28FF>].

20. *See id.*

21. Isa Soares, *The Fake News Machine: Inside a Town Gearing Up for 2020*, CNN, <https://money.cnn.com/interactive/media/the-macedonia-story/> [<https://perma.cc/VW93-G6AF>] (last visited Jan. 3, 2020).

22. *Id.*

23. *Id.*

24. The U.S. Supreme Court first introduced the concept of the marketplace of ideas through Justice Holmes’ dissent in *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting). It was later adopted by a majority of the Court in 1974 and has been reiterated ever since. *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 339 (1974).

25. Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 16 (1984).

26. Escudero de Jesús, *supra* note 18, at 1396.

27. *Id.*

truth from falsehood.<sup>28</sup> This is connected to the first pillar in that recipients of the information must be able to compare conflicting ideas and have the tools to enable them to choose the truthful idea, but it also requires them to be rational when making a determination in what to believe. Third, that falsehoods are accidental and that their promoters are in pursuit of the truth.<sup>29</sup> Proponents of ideas are not supposed to knowingly spread false ideas and manipulate others into believing them. Recent scholarship suggests that disinformation has undermined each pillar of the marketplace of ideas theory.<sup>30</sup>

Regarding the first pillar, many users of social media do not benefit from the clash of ideas that is supposed to occur in a healthy market because they have been isolated from opposing views.<sup>31</sup> In the political context, due in large part to the functionality of social media platforms, supporters of one candidate or party are unlikely to “follow” an opposing candidate or party, which makes them less likely to encounter contradictory ideas and further entrenches them in their beliefs.<sup>32</sup> This echo chamber problem worsens when it is combined with content that is targeted at its viewer’s confirmation bias, which results in sharing content based solely on the headline without verifying the underlying facts.<sup>33</sup>

Disinformation undermines the marketplace’s second pillar because once disinformation is believed, it is difficult to undo. Once a false story is disseminated, it is exceedingly difficult to counter it with the truth for two reasons. First, not all of the people that saw and were persuaded by the false story will subsequently encounter the truthful story. The lack of exposure to countervailing information that is central to the first pillar is implicated here. Second, and perhaps most problematic, of those people that do later encounter the truthful story, many of them are reluctant to adjust their beliefs.<sup>34</sup> A recent study showed that the vast majority of news readers maintained their beliefs based on a false story even after it was proven to them that what they read was false.<sup>35</sup> The marketplace simply does not work when its participants are reluctant to accept truth that contradicts their previously held beliefs.

Finally, the third pillar of the marketplace of ideas is directly contradicted by the recent rise of the disinformation business.<sup>36</sup> Actors who engage in spreading disinformation are rarely interested in having their ideas compete with the truth.<sup>37</sup>

---

28. *Id.*

29. *Id.*

30. *Id.*

31. *Id.* at 1407.

32. *Id.*

33. Filippo Menczer, *Fake Online News Spreads Through Social Echo Chambers*, SCI. AM. (Nov. 28, 2016), <https://www.scientificamerican.com/article/fake-online-news-spreads-through-social-echo-chambers/> [<https://perma.cc/Z2V3-RSG3>].

34. Emily Thorson, *Belief Echoes: The Persistent Effects of Corrected Misinformation*, 33 POL. COMM. 460, 477 (2016).

35. *Id.*

36. Soares, *supra* note 21.

37. *Id.*

They specifically target certain demographics and may only do so on a temporary basis.<sup>38</sup> Competition on the merits of an idea is not what they are after. Rather, disinformation often buries the truth by flooding the market with falsehoods. Disinformation is cheap to produce and disseminate on social media, whereas the truth usually requires verification and investigation, which are expensive endeavors.

### III. LEGAL FRAMEWORK

The marketplace of ideas has been ineffective to deal with the spread of disinformation. Thus, to ensure that the truth is discoverable, intervention by the state may be necessary. However, because some of the largest social media platforms are U.S.-based companies with a large presence around the world, the legislation that regulates them should account for its legal implications both in the United States and abroad so as not to overly burden the platforms with inconsistent and sometimes contradicting policies.

#### *A. United States*

The main concern for regulating disinformation on social media platforms is the First Amendment to the U.S. Constitution, which states that “Congress shall make no law . . . abridging the freedom of speech, or of the press . . .”<sup>39</sup> Under First Amendment jurisprudence, much of which developed when public debate occurred via pamphlets and speeches in public forums, courts have generally not been receptive towards regulations of speech that are based on the speech’s content, especially when the speech is political.<sup>40</sup> Content-based restrictions on speech are analyzed under the highest level of scrutiny, called “strict scrutiny,” and will only be upheld if they (1) further a compelling governmental interest and (2) are narrowly tailored to achieve that interest.<sup>41</sup> Thus, any regulations that are based on whether the speaker’s message is true or false are content-based and subject to strict scrutiny analysis, which almost always results in striking down the regulation.

The U.S. Supreme Court has allowed certain restrictions on speech that it determined does not deserve full First Amendment protection.<sup>42</sup> Those categories of speech include fraud, obscenities, defamation, and incitement.<sup>43</sup> The Court has been hesitant to add other categories to the list.<sup>44</sup> Disinformation intended to influence elections may be enough of a compelling state interest for the first

---

38. *See id.*

39. U.S. CONST. amend. I.

40. *Buckley v. Valeo*, 424 U.S. 1, 14 (1976) (“The First Amendment affords the broadest protection to such political expression in order ‘to assure [the] unfettered interchange of ideas for the bringing about of political and social changes desired by the people.’”) (quoting *Roth v. United States*, 354 U.S. 476, 484 (1957)).

41. *Reed v. Gilbert*, 135 S. Ct. 2218, 2226 (2015).

42. *See United States v. Alvarez*, 567 U.S. 709, 717 (2012).

43. *Id.*

44. *See id.* at 722.



prong.<sup>45</sup> However, even if that were the case, the regulation would have to be narrowly tailored to achieve that goal which presents major hurdles because it must be the least restrictive means of achieving the purported goal, a difficult standard to meet.

Due to First Amendment concerns of legislating content on social media platforms, much of the content regulation is left to the discretion of the platforms themselves by way of their community standards.<sup>46</sup> The platforms are free to remove content as they please without much government interference. That is in large part due to section 230(c) of the Communications Decency Act of 1996 (CDA), which states that “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”<sup>47</sup> As a result, the CDA immunizes sites such as Facebook and YouTube from liability arising out of the content posted by their users.<sup>48</sup> This statute was created during the early days of the internet to prevent interactive computer service providers from being subjected to countless lawsuits based on the content their users posted.<sup>49</sup> The rationale for such a blanket exemption from liability is to protect freedom of expression, which would be chilled because the platforms would overregulate their content out of fear of being sued.

### B. International

Article 19 of the International Covenant on Civil and Political Rights (ICCPR) provides the right for freedom of expression and opinion by stating that “[e]veryone shall have the right to hold opinions without interference.”<sup>50</sup> Article 19(2) further states that the right to freedom of expression includes the “freedom to seek, receive and impart information and ideas of all kinds.”<sup>51</sup> The United Nations Human Rights Council and General Assembly have stated that the freedoms people enjoy offline also apply online.<sup>52</sup>

---

45. For an analysis of this issue, see Richard L. Hasen, *Cheap Speech and What It Has Done (to American Democracy)*, 16 *FIRST AMEND. L. REV.* 200, 216 (2017).

46. See *Community Standards*, FACEBOOK, [https://www.facebook.com/communitystandards/hate\\_speech](https://www.facebook.com/communitystandards/hate_speech) [<https://perma.cc/Q45S-6FBH>] (last visited Jan. 3, 2020); *Ad Policy: Hateful Content*, TWITTER, <https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/hate-content.html> [<https://perma.cc/QY2W-MWNY>] (last visited Jan. 3, 2020); *Hate Speech Policy*, YOUTUBE, <https://support.google.com/youtube/answer/2801939?hl=en> [<https://perma.cc/3BAJ-CHMG>] (last visited Jan. 3, 2020).

47. 47 U.S.C. § 230 (2012).

48. See *id.*

49. Ashley Smith-Roberts, *Facebook, Fake News, and the First Amendment*, 95 *DENV. L. REV. ONLINE* 118 (2018).

50. International Covenant on Civil and Political Rights, art. 19, *adopted* Dec. 16, 1966, S. Treaty Doc. No. 95-20, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976) [hereinafter ICCPR].

51. *Id.*

52. See G.A. Res. 68/167 (Dec. 18, 2013); Human Rights Council Res. 26/13, U.N. Doc. A/HRC/RES/26/13 (June 26, 2014); Human Rights Council Res. 32/13, U.N. Doc. A/HRC/RES/32/13 (July 18, 2016); see also Human Rights Council Res. 32/38, U.N. Doc. A/HRC/32/38 (Nov. 5, 2016).

Under the ICCPR, states may adopt restrictions on content, but only as allowed by Article 19(3).<sup>53</sup> Under Article 19(3), any restriction on the right to freedom of expression must be provided by law and necessary to protect the rights or reputations of others and “[f]or the protection of national security or of public order . . . or of public health or morals.”<sup>54</sup> Such restrictions, however, must conform to the strict tests of necessity and proportionality.<sup>55</sup> Proportionality must ensure that the restrictions target a specific objective and do not unduly intrude upon the rights of targeted persons.<sup>56</sup> Finally, the restrictions must be “the least intrusive instrument amongst those which might achieve” the desired result.<sup>57</sup>

In the *2017 Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda*, the United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, and other intergovernmental experts concluded that “[g]eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including ‘false news’ or ‘non-objective information’ are incompatible . . . and should be abolished.”<sup>58</sup> Laws and regulations that require, or pressure, private entities to censor or remove content based on these vague and ambiguous criteria are a threat to the exercise of free expression.<sup>59</sup> That is because private entities are ill-equipped to monitor and regulate such content, and the possibility of facing high fines or the loss of the ability to operate can force platforms to over-regulate and disproportionately censor a wide range of permissible content.

#### IV. DISINFORMATION LEGISLATION

Legislating social media platforms based on disinformation comes with dangerous risks. It gives governments that are hostile towards the media a new means of silencing unfavorable opinions. The recent popularity of disinformation legislation comes as Freedom House’s 2017 report on global press freedoms found that media freedom was at its lowest point in thirteen years and that there were “unprecedented threats to journalists and media outlets in major democracies and new moves by authoritarian states to control the media, including beyond their borders.”<sup>60</sup> Some governments have unfortunately capitalized on the rising

---

53. ICCPR, *supra* note 50, art. 19(3).

54. *Id.*

55. Human Rights Committee, Art. 19: Freedoms of Opinion and Expression, General Comment No. 34, ¶ 22, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011).

56. *Id.* at ¶¶ 34–35.

57. *Id.* at ¶ 34.

58. *Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda*, ORG. FOR SEC. & CO-OPERATION EUR. (Mar. 3, 2017), <https://www.osce.org/fom/302796> [<https://perma.cc/5MK7-F4PQ>] [hereinafter *Joint Declaration on Freedom of Expression*].

59. *Id.*

60. *Press Freedom’s Dark Horizon*, FREEDOM HOUSE, <https://freedomhouse.org/report/freedom-press/freedom-press-2017> [<https://perma.cc/A8LQ-BHSR>] (last visited Jan. 3, 2020).

concerns of disinformation and enacted or proposed legislation that is in violation of freedom of expression rights.<sup>61</sup> However, unchecked disinformation may also have serious negative effects on peoples' ability to enjoy their right to freedom of expression and can have destabilizing effects on democracies. Thus, it is helpful to determine what legislators can do to stop the spread of disinformation while not infringing on the human right of freedom of opinion and expression. In that light, the following legislative approaches have been proposed by legislators in the United States and abroad, civil society, and international organizations.

#### *A. Digital Media Literacy*

Perhaps the best way to stem the flow of disinformation and have the lowest risk of intruding on freedom of expression is by enacting legislation that is aimed at enhancing digital media literacy. Digital media literacy can be defined as “the capacity to exert critical thinking as to the productions, representations, languages (visuals, texts, sounds), audiences and communities characteristic of mainstream and social media.”<sup>62</sup> By enhancing their citizens' digital media literacy, states can not only reduce the effects of disinformation but they can also benefit from a more educated citizenry. An effective policy would be one that strives for a population of citizens that are both alert to the threat of disinformation and armed with the critical-thinking skills necessary to reduce its effects. A media-critical citizenry will have the added benefit of incentivizing media companies to produce higher quality content.

Media literacy initiatives should be holistic in their approach. They should incorporate a school curriculum with adequate teacher training. Curricula will inevitably vary, but focus should be placed on skills that allow people to distinguish truthful information from false information. The International Society for Technology and Education has set out model standards for teaching “[s]tudents [to] evaluate the accuracy, perspective, credibility and relevance of information, media, data or other resources.”<sup>63</sup> Furthermore, because digital media is constantly evolving, as evident by the advances of Deepfake technology, digital media literacy should be a lifelong process that accounts for the continuing education of adults. Finally, increasing digital media literacy should be a partnership between government, educational institutions, journalists, social media platforms, and civil society.

#### *B. Transparency Initiatives*

Legislation aimed at ensuring transparency could also be instrumental in combating disinformation, while upholding the right of freedom of expression.

---

61. *Id.*

62. *A Multi-dimensional Approach to Disinformation*, *supra* note 8, at 25 n.40.

63. *ISTE Standards for Students*, INT'L SOC'Y FOR TECH. IN EDUC., <https://www.iste.org/standards/for-students> [<https://perma.cc/E4MD-MC6L>] (last visited Jan. 3, 2020).

Transparency compliments media literacy because knowledge of how and why content is posted is necessary for any digital media literacy efforts to be effective. To critically assess the information they are encountering, users will need to know its source and the motivation behind its dissemination.

### 1. Bots

Legislation aimed at requiring platforms to clearly label bots has the potential of reducing the effectiveness of disinformation. Bots, in the social media context, can be defined as automated online accounts that interact with users or other automated accounts.<sup>64</sup> Bots have been used to “manufactur[e] consensus” by artificially increasing the visibility of content or users on social media platforms.<sup>65</sup> Such activity skews the popularity of ideas and gives them more credibility than they would otherwise get. A version of this phenomenon has been observed in the commercial context where armies of bots are often used to mislead consumers by participating in online reviews for products or businesses.<sup>66</sup>

Bots have been an issue for years, but their recent popularity in political disinformation campaigns has resulted in an increased interest from lawmakers. The 2016 U.S. presidential election has revealed the powerful influence that bots can have.<sup>67</sup> One bot on Twitter reportedly generated 1200 posts during the final presidential debate.<sup>68</sup> Such an outpour of content floods the conversation and makes it difficult for users to find quality content. This type of activity interferes with a person’s right to receive and impart information.

Bot presence is likely to increase as the skill needed to create them has been made accessible to anyone with a preliminary understanding of programming code.<sup>69</sup> But responses to the bot threat have to be carefully considered. Outright restrictions on bots would not only violate freedom of expression by restricting a person’s right to receive and impart information, but it would also prohibit “good” bots that can be used as fact checkers, to conduct research, or simply for entertainment. Proposals to ban “malicious” bots encounter the difficult issue of determining whether a bot is good or bad and determining who the appropriate entity to make that conclusion will be. Therefore, a more appropriate solution is to mark bots to allow the user to critically analyze the content they produce.

---

64. See Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 990 (2019).

65. Farhad Manjoo, *How Twitter Is Being Gamed to Feed Misinformation*, N.Y. TIMES (May 31, 2017), <https://www.nytimes.com/2017/05/31/technology/how-twitter-is-being-gamed-to-feed-misinformation.html> [<https://perma.cc/V4C3-MM5T>] (quoting Samuel Woolley).

66. Shareen Pathak, *Amazon Reviews Have a Bot Problem*, DIGIDAY (Sept. 18, 2017), <https://digiday.com/marketing/amazon-reviews-bot-problem/> [<https://perma.cc/PZF6-UYUN>].

67. Douglas Guilbeault & Samuel Woolley, *How Twitter Bots Are Shaping the Election*, ATLANTIC (Nov. 1, 2016), <https://www.theatlantic.com/technology/archive/2016/11/election-bots/506072/> [<https://perma.cc/3W99-CCJR>].

68. *Id.*

69. *Id.*

There are varying approaches to marking bots. One is for legislators to require the platforms to identify bots themselves and then give them some sort of distinguishing mark. This can be accomplished by requiring social media platforms to implement a program or algorithm to monitor bot-like behavior and then mark those accounts they catch. Another way to do so is to require the platforms to have their users register bot accounts or risk getting banned. This system would resemble Twitter's verified accounts.<sup>70</sup> Another, more aggressive, approach is to enact legislation that would require owners to have their bots disclose they are bots whenever they communicate with users. Some jurisdictions have already proposed or implemented versions of such legislation. For example, California recently passed a bill that subjects bot owners to liability.<sup>71</sup> The relevant portion of the bill reads:

It shall be unlawful for any person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. A person using a bot shall not be liable under this section if the person discloses that it is a bot.<sup>72</sup>

The law went into effect on July 1, 2019.<sup>73</sup> There may be some concern that this type of law is infringing on First Amendment rights by compelling speech.<sup>74</sup> However, it is worth noting that other laws require disclosure of certain content that is in the public interest, such as warning labels on bioengineered foods and labels for prescription drugs.<sup>75</sup>

## 2. Location

Many countries are concerned about disinformation emanating from foreign countries that aims to influence their citizens. Requiring platforms to determine or authenticate the geographical source of accounts or posts could potentially reduce the likelihood that a state's domestic users would be influenced by disinformation by foreign entities. Such a strategy would address the concerns of foreign interference in elections that have arisen in recent years.

Facebook is already experimenting with this idea after reports that Russian disinformation was spread during the 2016 U.S. presidential election by creating

---

70. *About Verified Accounts*, TWITTER, <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts> [<https://perma.cc/VL9T-PA2H>] (last visited Jan. 3, 2020).

71. CAL. BUS. & PROF. CODE § 17941 (West 2019).

72. *Id.*

73. *Id.*

74. For an analysis of the legality of legislation to compel bot identity disclosures, see Lamo & Calo, *supra* note 64.

75. 7 C.F.R. § 66.100 (2019) (disclosure requirement for bioengineered food); 21 C.F.R. § 290.5 (2003) (warning label requirement for prescription drugs).

political pages impersonating American citizens.<sup>76</sup> Facebook now requires pages of a “large US audience” to “secure their account with two-factor authentication and confirm their primary country location.”<sup>77</sup> Facebook also displays the primary country location of the people who manage those pages.<sup>78</sup> It plans to roll this system out to its other popular platform, Instagram, with a goal to “prevent organizations and individuals from creating accounts that mislead people about who they are or what they’re doing.”<sup>79</sup>

Legislation that aims to require social media platforms to adopt similar methods would complement any efforts of increasing digital media literacy, but it also raises some concerns of its efficacy and risks of violating privacy rights. First, such a policy would be difficult to implement because users could circumvent it via virtual private networks and claim to be from whichever country they wish. Second, any efforts to go beyond self-reporting and IP address locations would have to be delicate so as not to intrude on privacy rights or place journalists or other vulnerable groups in danger from revealing their location.

### 3. *Advertisements*

Legislation could also require social media platforms to disclose the source of online political advertisements. Investigations after the 2016 U.S. presidential election revealed that foreign governments invested large amounts of money in Facebook advertising that reached over ten million people.<sup>80</sup> In response, some platforms have already taken steps to make advertisements more transparent for their users.<sup>81</sup> Transparency in this arena is critical because of the possibility of targeted political disinformation campaigns that may surge in the days before an election. In those situations, by the time false content is countered with the truth, the election may be over and important decisions would have already been made based on disinformation.

Although the steps already taken by platforms are commendable, they are not uniform.<sup>82</sup> Some platforms provide more information than others and offer differing levels of accessibility to the relevant content.<sup>83</sup> Legislators may be

---

76. *New Authorization for Pages*, FACEBOOK (Aug. 10, 2018), <https://www.facebook.com/business/news/new-authorization-for-pages> [<https://perma.cc/DS96-ZT8Y>].

77. *Id.*

78. *Id.*

79. *Id.*

80. Mike Isaac & Scott Shane, *Facebook’s Russia-Linked Ads Came in Many Disguises*, N.Y. TIMES (Oct. 2, 2017), <https://www.nytimes.com/2017/10/02/technology/facebook-russia-ads.html> [<https://perma.cc/VND7-EGVM>].

81. See *Political Content*, TWITTER, <https://ads.twitter.com/transparency> [<https://perma.cc/U2WD-PB7H>] (last visited Jan. 3, 2020); Rob Leathern, *Shining a Light on Ads With Political Content*, FACEBOOK (May 24, 2018), <https://newsroom.fb.com/news/2018/05/ads-with-political-content/> [<https://perma.cc/2R9V-RLHW>].

82. Compare the policies for political ads by Facebook and Twitter, found at Leathern, *supra* note 81, and *Political Content*, *supra* note 81.

83. Leathern, *supra* note 81, and *Political Content*, *supra* note 81.

interested in enacting legislation that will require disclosure in an accessible manner. California has recently passed a law that does just that.<sup>84</sup> Set to take effect in 2020, the law requires platforms to include with each advertisement a disclosure of who paid for the advertisement and maintain records of advertisements disseminated on the platform.<sup>85</sup> Such advertising is already regulated on the radio, television, and satellite.<sup>86</sup>

On the other hand, legislation that aims to ban advertisement from foreign entities entirely may encounter First Amendment challenges in the United States and may violate international law's requirement that the law be necessary and proportional.<sup>87</sup> In the U.S. context, Professor Richard L. Hasen suggests that current case law would prevent outright bans on advertisements by foreign entities because current law only allows for the ban of foreign advertisement that supports a specific candidate, and not advertisements that are "issue advocacy."<sup>88</sup>

### C. Platform Liability

One of the most controversial legislative proposals is whether to make platforms liable for failing to remove disinformation. Germany has already enacted a law that strives to do just that.<sup>89</sup> The "Netzwerkdurchsetzungsgesetz law" imposes heavy fines (up to fifty million euros) on social media platforms with more than two million users for failing to remove "obviously illegal" posts.<sup>90</sup> This law has been widely criticized for its threats to freedom of expression.<sup>91</sup> The United Nations Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye, expressed concerns with the law for its lack of judicial oversight and breadth that could prompt social media platforms to remove lawful content out of fear of being fined.<sup>92</sup> In his Joint Declaration, he further stated:

Intermediaries should never be liable for any third party content relating to those services unless they specifically intervene in that content or refuse to obey an order adopted in accordance with due process guarantees by an

---

84. Assem. B. 2188, 2018 Leg., Reg. Sess. (Cal. 2018), [http://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill\\_id=201720180AB2188](http://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB2188) [<https://perma.cc/2JBX-52RG>].

85. *Id.*

86. See 11 C.F.R. § 110.11(b) (2014).

87. See Ariel L. Bendor & Tal Sela, *How Proportional Is Proportionality?*, 13 INT'L J. CONST. L. 530, 530 (2015) (defining the international law proportionality requirement).

88. Hasen, *supra* note 45, at 217; see 47 U.S.C. § 230 (2018).

89. *Germany Starts Enforcing Hate Speech Law*, B.B.C. (Jan. 1, 2018), <https://www.bbc.com/news/technology-42510868> [<https://perma.cc/9D6G-Z3AR>].

90. *Id.*

91. *Id.*

92. Letter from David Kaye, U.N. Special Rapporteur, to Angela Merkel, Chancellor of Germany (June 1, 2017), <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf> [<https://perma.cc/3TNX-7KR8>].

independent, impartial, authoritative oversight body (such as a court) to remove it and they have the technical capacity to do that.<sup>93</sup>

Notwithstanding the problematic legislation that has been enacted in Germany, it may be possible for legislation that holds social media platforms liable to conform with freedom of expression principles. For example, in the United States, the Digital Millennium Copyright Act (DMCA) has been an effective tool for victims of copyright infringement to get social media platforms to remove violative content.<sup>94</sup> A similar framework can be constructed to remove content that is demonstrably illegal—in line with common law torts. Furthermore, as algorithms and perceptual hashing technology advance, content would be easier to remove, which reduces the burden on the social media platforms. U.S. Senator Mark R. Warner has proposed requiring platforms to do so in the United States.<sup>95</sup> Senator Warner proposes to modify section 230 of the CDA to allow for holding social media platforms liable for failing to remove content after a victim has obtained a court order and noticed the platforms.<sup>96</sup> Depending on how such a policy is implemented, it could satisfy freedom of expression requirements of due process and not overburden platforms.

#### CONCLUSION

Disinformation presents a danger to democracies and undermines people's right to receive and impart information. At the same time, legislation that is not carefully crafted with freedom of expression rights in mind can also lead to concerning outcomes. Governments should refrain from enacting policies that place them or private actors in the position of determining what is true or false. That is partly due to the difficulty in determining what is objective truth.<sup>97</sup> More importantly, however, it prevents bad actors from abusing such a system by restricting content that they find unfavorable.

A more prudent approach for disinformation legislation on social media platforms is to enact policies that empower the users to make the truthful determination for themselves. This begins with an educated society that recognizes the risks of disinformation and is armed with the tools to spot it and come to their own conclusions. That is why digital media literacy should be at the forefront of disinformation legislation.

Literacy would not be effective if social media platforms hide the information necessary to determine the veracity of the content their users come across. That is

---

93. *Joint Declaration on Freedom of Expression*, *supra* note 58.

94. *See* 17 U.S.C. § 101-122. (2012).

95. SENATOR MARK R. WARNER, POTENTIAL POLICY PROPOSALS FOR REGULATION OF SOCIAL MEDIA AND TECHNOLOGY FIRMS (2018), <https://graphics.axios.com/pdf/PlatformPolicyPaper.pdf> [<https://perma.cc/2424-AGUA>].

96. *Id.*

97. Christopher T. Wonnell, *Truth and the Marketplace of Ideas*, 19 U.C. DAVIS L. REV. 669, 673 (1986).



why legislation that aims to make social media platforms more transparent is essential. Transparency regarding who and what users are interacting with is critical for the users to make an informed decision based on the motives of the content creator. Therefore, legislation that requires the disclosures for bots, advertisements, and the source's geographical location may reduce the effectiveness of disinformation campaigns.

A more aggressive and potentially problematic approach is to make social media platforms liable for failing to remove false content. Concerns with this approach surround the platforms' precautionary removal of otherwise legal content out of fear of being fined. It is also problematic in situations where the government or other actors are the decision makers in what gets to stay or get removed. Furthermore, approaches that do not incorporate due process mechanisms are not in line with freedom of expression principles. That is not to say that such an approach is not possible. As noted, in the United States, social media platforms have been required to remove certain content based on copyright claims. A similar framework can be implemented regarding false content as long as it conforms to freedom of expression principles. Such a policy would need due process through an impartial body, such as a court, and would need to avoid burdening social media platforms with removal. Suggested policies that would require a user to obtain a court order and notice the platform seem like plausible solutions to the issue when combined with technological advances in hashing technology that would make the removal of the content more efficient.