

**A MARKETPLACE (FAILURE) OF IDEAS: FOREIGN  
DISINFORMATION CAMPAIGNS IN AN AGE OF GENERATIVE A.I.  
AND A PROPORTIONALITY APPROACH TO FALSEHOODS**

*Michael K. Park\**

*America’s foreign adversaries continue to employ disinformation and deepfakes—aided by artificial intelligence (“A.I.”)—in furtherance of attempts to undermine U.S. democratic institutions, manipulate global political narratives, and reshape the U.S.-led international order. This Article identifies some of the threats that foreign-based disinformation campaigns present, including the growing concerns of A.I.-based disinformation designed to undermine U.S. democratic institutions. It examines how the current free speech framework, informed by the marketplace theory and a categorical approach to the First Amendment, is unable to adequately address modern disinformation tactics. In response to technological and political developments and the Supreme Court’s rigid categorical approach to speech, this Article makes the case for a reorientation of free speech analysis for foreign-based falsehoods and the adoption of a proportionality approach informed by the First Amendment’s democratic nature.*

**TABLE OF CONTENTS**

**I. INTRODUCTION .....46**  
**II. PREVAILING FOREIGN THREATS TO THE SPEECH  
MARKETPLACE: DISINFORMATION IN AN AGE OF GENERATIVE  
A.I.....48**  
**III. A MARKETPLACE (FAILURE) OF COMPETITION WITH  
FALSEHOOD .....55**  
**IV. A REORIENTATION OF SPEECH ANALYSIS:  
PROPORTIONALITY REVIEW AND DEMOCRATIC THEORY ..... 61**  
**V. CONCLUSION ..... 71**

---

\* Jerry & Karla Huse Professor, Assistant Professor at the University of Nebraska-Lincoln.

## I. INTRODUCTION

The current technological competition between the U.S. and China reflects the return of great power competition, where China has set forth an ambitious plan to lead the world in artificial intelligence (“A.I.”) by 2030.<sup>1</sup> Amidst this competition is an informational struggle over the shaping of global norms, governance, and public opinion. Chinese General Secretary Xi Jinping has been unequivocal with his stated global communication goals: “We will accelerate the development of China’s discourse and narrative systems, better tell China’s stories . . . and present a China that is credible, appealing, and respectable.”<sup>2</sup> But unlike Western democracies, authoritarian states like China and Russia exploit the relative freedoms of the West, including its commitment to the freedom of speech and press, to wage an asymmetrical informational battle. One powerful speech manipulation instrument in China’s arsenal is the use of economic incentives linked to access (or denial) of its large consumer market to induce foreign actors and industries to advance the Chinese Communist Party’s<sup>3</sup> political messaging. In addition, the Chinese government’s recent employment of A.I.-generated disinformation campaigns represents another powerful speech manipulation instrument and a shift in tactics in its informational struggle. Consequently, these information manipulation tactics are immensely challenging for the U.S. system of free speech to address.

For most of 2023, suspected foreign operatives affiliated with the Chinese government have used fake images generated by A.I. to impersonate American voters online in order to disseminate disinformation and spark debate on divisive issues ahead of the 2024

---

<sup>1</sup> Graham Webster et al., *China’s Plan to ‘Lead’ in AI: Purpose, Prospects and Problems*, NEW AM. FOUND. (Aug. 1, 2017), <https://www.newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-ai-purpose-prospects-and-problems/> [https://perma.cc/S4QW-6HWA].

<sup>2</sup> James T. Areddy, *New Ways to Tell China’s Story*, WALL ST. J. (Oct. 23, 2022, 6:53 AM), <https://www.wsj.com/livecoverage/china-xi-jinping-communist-party-congress/card/new-ways-to-tell-china-s-story-JXt9XFnnegpB7yzmhFNT> [https://perma.cc/LD42-FS4M (staff-uploaded, dark archive)].

<sup>3</sup> See Michael K. Park, *The Art of War: Global Speech Diktats and the Regulatory Challenges to Address Foreign Political Censorship*, 2 FLA. ENT. & SPORTS L. REV. 41, 48–52 (2022).

U.S. election.<sup>4</sup> The alleged influence network utilized Western social media platforms to upload the fake images for the purposes of “denigrating U.S. political figures and symbols,” which were then reposted or shared unwittingly by online users.<sup>5</sup> For example, after one of the deadliest wildfires wreaked havoc on the island of Maui in August 2023, false reports online asserted that the fires resulted from a “weather weapon” being tested by the U.S. government.<sup>6</sup> The images accompanying many of these posts were generated by A.I., and researchers from Microsoft and the RAND Corporation discovered that the posts originated from a Chinese influence operation; the use of A.I.-generated images by these influence networks makes them among the first to be used in a disinformation campaign.<sup>7</sup>

New generative-A.I. disinformation techniques will continue to proliferate, and American citizens will likely see a critical mass of disinformation ahead of the 2024 U.S. presidential election. Adopting such extraterritorial disinformation tactics is undoubtedly antithetical to the U.S.’s tradition of free speech and free inquiry. But the unfortunate reality is that the First Amendment, under the classic marketplace theory, offers little recourse to such instruments of falsehood, enabling the growth of disinformation, market-based speech control, and self-censorship. In other words, more truthful speech in the market is no longer the elixir for falsehoods.

This Article, however, will not address issues of market-based speech coercion. Instead, it will center on the threats that foreign-based disinformation campaigns present, focusing on China’s informational tactics—particularly its use of disinformation. This Article begins in Part II by identifying some common foreign state-based disinformation tactics, including the use of deepfakes<sup>8</sup> and the

---

<sup>4</sup> See Sean Lyngaas, *Suspected Chinese Operatives Using AI Generated Images to Spread Disinformation Among U.S. Voters, Microsoft Says*, CNN (Sept. 7, 2023, 11:54 AM), <https://www.cnn.com/2023/09/07/politics/chinese-operatives-ai-images-social-media/index.html> [<https://perma.cc/2BWJ-4CBA>].

<sup>5</sup> *Id.*

<sup>6</sup> David E. Sanger & Steven Lee Myers, *China Sows Disinformation About Hawaii Fires Using New Techniques*, N.Y. TIMES (Sept. 11, 2023), <https://www.nytimes.com/2023/09/11/us/politics/china-disinformation-ai.html> [<https://perma.cc/LXX4-85JB>] (staff-uploaded, dark archive).

<sup>7</sup> *Id.*

<sup>8</sup> A deepfake is “[a]ny of various media, *esp.* a video, that has been digitally manipulated to replace one person’s likeness convincingly with that of another, often

growing concerns with A.I.-based disinformation campaigns designed to sow civil discord and degrade U.S. democratic institutions. What follows next in Part III is an examination of how the current U.S. free speech framework, buttressed by the classic marketplace theory and a rigid categorical approach to the First Amendment, is unable to adequately address modern speech manipulation instruments that pose significant threats to public discourse and democratic institutions. In response to recent political and technological developments and to the Supreme Court's rigid categorical approach to speech cases, this Article makes the case for reorienting free speech analysis in Part IV. Further, it makes the case for adopting a proportionality balancing approach, informed by democratic theory, to prospective regulations of foreign state-based disinformation. Finally, a hypothetical regulation on deepfakes will be explored under proportionality review to demonstrate how such an approach can offer a more nuanced and comprehensive balancing approach that can provide legislative flexibility and constitutional means to address the harms and looming threats posed by deepfakes and other modern disinformation tools.

## II. PREVAILING FOREIGN THREATS TO THE SPEECH MARKETPLACE: DISINFORMATION IN AN AGE OF GENERATIVE A.I.

The term disinformation, adopted from the Russian word *dezinformatsia*,<sup>9</sup> refers to “all forms of false, inaccurate, or misleading information designed, presented, and promoted to intentionally cause public harm or for profit.”<sup>10</sup> Disinformation is distinguishable from misinformation: The former pertains to the deliberate dissemination of falsehoods, while the latter pertains to the inadvertent or

---

used maliciously to show someone doing something that he or she did not do.” *Deepfake*, OXFORD ENGLISH DICTIONARY (July 2023), <https://doi.org/10.1093/OED/7847968874> [<https://perma.cc/PA8C-EAXG>].

<sup>9</sup> RICHARD H. SHULTZ & ROY GODSON, *DEZINFORMATSIA: ACTIVE MEASURES IN SOVIET STRATEGY* 2, 36–37 (1984).

<sup>10</sup> EUR. COMM'N: DIRECTORATE-GENERAL FOR COMM'NS NETWORKS, CONTENT AND TECH., PUB. NO. KK-01-18-221-EN-C, *A MULTI-DIMENSIONAL APPROACH TO DISINFORMATION, REPORT OF THE EUROPEAN COMMISSION HIGH LEVEL GROUP ON FAKE NEWS AND ONLINE DISINFORMATION 3* (2018), [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=50271](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271) [<https://perma.cc/A5XB-X5X8>].

unintentional dissemination of inaccurate information.<sup>11</sup> With the political ascendance of Donald Trump and the victory of the Leave campaign referendum in the United Kingdom, 2016 has been marked by some scholars as the dawn of the post-truth era.<sup>12</sup> The characterization of the modern epoch as post-truth reflects the idea that disinformation is influential in shaping public opinions and affecting political engagement.<sup>13</sup>

Disinformation in the modern era poses a grave threat to democratic institutions because of how rapidly it spreads through online social networks.<sup>14</sup> An analysis of 126,000 false and true news stories on X (formerly Twitter) from its inception in 2006 to 2017 revealed that falsehoods spread “significantly farther, faster, deeper[,] and more broadly than the truth in all categories of information.”<sup>15</sup> The research showed that it took true stories *six* times as long as false stories to reach fifteen hundred people, and false *political* news was more viral and reached more people than any other category of false information.<sup>16</sup> While some online social networks have taken some steps to limit political content on their respective platforms,<sup>17</sup> social

---

<sup>11</sup> See Elenda Broda & Jesper Strömbäck, *Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review*, 48(2) ANNALS OF INTER. COMM. ASSOC. 139, 142 (2024), <https://www.tandfonline.com/doi/pdf/10.1080/23808985.2024.2323736> [<https://perma.cc/FZ6M-ZYCY> (staff-uploaded)].

<sup>12</sup> See Marta Pérez Escolar et al., *A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation*, 11(2) MEDIA & COMMUN 76, 77 (2023); see also LEE MCINTYRE, *POST-TRUTH* (2018); MATTHEW D'ANCONA, *POST-TRUTH: THE NEW WAR ON TRUTH AND HOW TO FIGHT BACK* (2017).

<sup>13</sup> See Pérez et al., *supra* note 12.

<sup>14</sup> Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCIENCE 1146, 1147 (2018).

<sup>15</sup> *Id.*

<sup>16</sup> *Id.* at 1148 (emphasis added).

<sup>17</sup> See META, *Our Approach to Political Content*, <https://transparency.fb.com/features/approach-to-political-content> [<https://perma.cc/XS8K-VCR3>] (last visited Feb. 17, 2024); see also TIKTOK, *TikTok's Stance on Political Ads*, <https://www.tiktok.com/creators/creator-portal/en-us/community-guidelines-and-safety/tiktoks-stance-on-political-ads/> [<https://perma.cc/DN29-BGZ8>] (last visited Feb. 17, 2024).

media remains as popular as ever: Facebook has three billion monthly active users, while WhatsApp and Instagram each have two billion.<sup>18</sup>

Prior to the COVID-19 pandemic, the Chinese Communist Party (“C.C.P.”) had primarily used foreign-facing information channels to broadcast a positive image of itself, the ruling party. This approach included measures such as paying for Chinese propaganda in Western publications; using content-sharing agreements with foreign media outlets, or directly purchasing such outlets;<sup>19</sup> and sponsoring online influencers<sup>20</sup> to promote pro-Beijing messages.<sup>21</sup> One of China’s most public-facing soft power tools has involved lending pandas to zoos around the world, including several in the U.S., as a form of “panda diplomacy.”<sup>22</sup> But as of this Article’s publication, most have been returned to China, and the end of 2024 nearly left the U.S. panda-free for the first time since rapprochement with China in 1972.<sup>23</sup> The timing is curious, and perhaps the pullout is in response to heightened tensions between the two countries, but the C.C.P. also appears to have shifted its propaganda and informational tactics following the

---

<sup>18</sup> See STATISTA, *Most Popular Social Networks Worldwide as of January 2024*, <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [<https://perma.cc/S2NJ-M6W6>] (last visited Feb. 17, 2024).

<sup>19</sup> Global Engagement Center Special Report, *How the People’s Republic of China Seeks to Reshape the Global Information Environment*, U.S. STATE DEP’T (Sept. 28, 2023), <https://www.state.gov/gec-special-report-how-the-peoples-republic-of-china-seeks-to-reshape-the-global-information-environment/> [<https://perma.cc/6NUA-EBSW>].

<sup>20</sup> Paul Mozur, et. al, *How Beijing Influences the Influencers*, N.Y. TIMES (Dec. 13, 2021), <https://www.nytimes.com/interactive/2021/12/13/technology/china-propaganda-youtube-influencers.html> [<https://perma.cc/HK4H-UFWR> (staff-uploaded, dark archive)].

<sup>21</sup> Louisa Lim & Julia Bergin, *Inside China’s Audacious Global Propaganda Campaign*, GUARDIAN (Dec. 7, 2018, 1:00 PM), <https://www.theguardian.com/news/2018/dec/07/china-plan-for-global-media-dominance-propaganda-xi-jinping> [<https://perma.cc/8KLJ-MYYD>].

<sup>22</sup> Aaron McNicholas, *China’s Adorable Diplomats*, WIRE CHINA (Dec. 31, 2023), <https://www.thewirechina.com/2023/12/31/chinas-adorable-diplomats-pandas-china/> [<https://perma.cc/PFR4-FWCW>].

<sup>23</sup> *Id.* But see Michael Crowley & Eric Lee, *Giant Pandas from China Return to National Zoo in Washington, D.C.*, N.Y. TIMES (Oct. 15, 2024), <https://www.nytimes.com/2024/10/15/us/politics/pandas-china-national-zoo.html> [<https://perma.cc/E3JK-5G3S> (staff-uploaded, dark archive)] (noting China sent two pandas to the Smithsonian’s National Zoo in an apparent “revival of . . . panda diplomacy”).

COVID-19 outbreak.<sup>24</sup> For instance, China-sponsored propaganda began pushing conspiracy theories linking the U.S. government with nefarious labs and weapons research. Since March 2020, Chinese government officials—employing online and traditional media channels—have promoted the theory that COVID-19 originated from a U.S. Army biological research facility in Maryland.<sup>25</sup> In late 2023, Meta (parent company of Facebook) discovered that groups linked to China’s law enforcement agencies conducted the largest ever known covert digital influence campaign focused on discrediting the West with disinformation across over fifty social media platforms.<sup>26</sup>

Thus, in addition to its attempts to soften its image abroad, the C.C.P. has employed disinformation tactics that appear to mirror Russian disinformation campaigns set on sowing civil discord and undermining U.S. politics.<sup>27</sup> Earlier disinformation campaigns, such as those targeting the 2016 U.S. presidential election, often involved human operators, or “trolls,” to author false stories.<sup>28</sup> For instance, employees of the Internet Research Agency—a Russian entity indicted for meddling in the 2016 election—worked twelve-hour shifts creating fake social media accounts and writing fake stories.<sup>29</sup> But with the

---

<sup>24</sup> See Global Engagement Center Special Report, *supra* note 19; see also Olivia Gazis, *Amid Covid-19 Outbreak, China Shifts to Use “Russian-Style” Disinformation Tactics*, CBS NEWS (Mar. 30, 2020, 3:19 PM), <https://www.cbsnews.com/news/coronavirus-covid-19-outbreak-china-russian-style-disinformation-tactics/> [https://perma.cc/L6LX-QSAX].

<sup>25</sup> Bret Schafer, *China Fires Back at Biden with Conspiracy Theories About Maryland Lab*, FOREIGN POL’Y (July 9, 2021, 8:36 PM), <https://foreignpolicy.com/2021/07/09/china-fires-back-at-biden-with-conspiracy-theories-about-maryland-lab/> [https://perma.cc/43TS-F3PS].

<sup>26</sup> See Mark Scott, *China Behind ‘Largest Ever’ Digital Influence Operation*, POLITICO (Aug. 29, 2023, 1:59 PM), <https://www.politico.eu/article/china-behind-largest-ever-digital-influence-operation-says-meta/> [https://perma.cc/2GFA-SYX5].

<sup>27</sup> See Sanger & Myers, *supra* note 6; see also Nomaan Merchant & Matthew Lee, *US Sees China Propaganda Efforts Becoming More Like Russia’s*, AP NEWS (Mar. 6, 2023, 12:07 AM), <https://apnews.com/article/china-russia-intelligence-foreign-influence-propaganda-0476f41aa932cd4850627a7b8984baa2> [https://perma.cc/459U-42TW (staff-uploaded)].

<sup>28</sup> See Neil MacFarquhar, *Inside the Russian Troll Factory: Zombies and a Breakneck Pace*, N.Y. TIMES (Feb. 18, 2018), <https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html> [https://perma.cc/MF8Y-P5VR (staff-uploaded, dark archive)].

<sup>29</sup> *Id.*

meteoric advances in generative A.I., 2016 seems like a bygone era. The unprecedented speed and scale that A.I. offers can be harnessed to flood online channels of communication with false content. Troll farms are no longer necessary; with the incorporation of large language models (“LLMs”), A.I. technologies today can easily generate text and images with stunning fidelity. An A.I.-generated fake image of an explosion near the Pentagon triggered a short-lived stock-market selloff in 2023.<sup>30</sup> This incident marked possibly the first time that an A.I.-generated image adversely affected values on the stock market and underscores fears of how the nefarious use of A.I. can be used to disrupt vital democratic institutions.<sup>31</sup>

In addition to China, hostile states, including Iran, North Korea, and Russia, are already leveraging this new technology to generate fake content on politically divisive topics in order to degrade Western democratic institutions.<sup>32</sup> New schemes of disinformation and more potent iterations of existing ones will undoubtedly be unleashed for the 2024 U.S. presidential election and beyond. For instance, A.I. has the potential to supercharge the practice of astroturfing—the obscuring of the origin of a message that would appear less credible if the true origin was revealed.<sup>33</sup> With the use of generative A.I., rogue

---

<sup>30</sup> Andrew Ross Sorkin et al., *An A.I.-Generated Spoof Rattles the Markets*, N.Y. TIMES (May 23, 2023), <https://www.nytimes.com/2023/05/23/business/ai-picture-stock-market.html> [https://perma.cc/N95J-ACYU (staff-uploaded, dark archive)]; see Davey Alba, *How Fake AI Photo of Pentagon Blast Went Viral and Briefly Spooked Stocks*, BLOOMBERG (May 23, 2023, 11:36 AM), <https://www.bloomberg.com/news/articles/2023-05-22/fake-ai-photo-of-pentagon-blast-goes-viral-trips-stocks-briefly?embedded-checkout=true> [https://perma.cc/G767-Z7ZG].

<sup>31</sup> See Alba, *supra* note 30.

<sup>32</sup> Clint Watts, *China, North Korea Pursue New Targets While Honing Cyber Capabilities*, MICROSOFT: MICROSOFT ON THE ISSUES (Sept. 7, 2023), <https://blogs.microsoft.com/on-the-issues/2023/09/07/digital-threats-cyberattacks-east-asia-china-north-korea/> [https://perma.cc/9SFS-BWTR]; Steven Lee Myers et al., *Iran Emerges as a Top Disinformation Threat in U.S. Presidential Race*, N.Y. TIMES (Sept. 4, 2024), <https://www.nytimes.com/2024/09/04/business/media/iran-disinformation-us-presidential-race.html> [https://perma.cc/H7B6-NQY5 (staff-uploaded, dark archive)]; Hailey Fuchs & Josh Gerstein, *Foreign Influence Efforts Are Circling the Presidential Election. Again.*, POLITICO (Sept. 7, 2024), <https://www.politico.com/news/2024/09/07/foreign-influence-2024-election-00177828> [https://perma.cc/XY32-3YDW (staff-uploaded)].

<sup>33</sup> DANIEL KLIMAN ET AL., DANGEROUS SYNERGIES: COUNTERING CHINESE AND RUSSIAN DIGITAL INFLUENCE OPERATIONS 11 (2020).



entities can now—by orders of magnitude—cheaply produce large numbers of fake accounts and “disguise an influence campaign by making it appear as though it were organically originating from local politicians, civil society organizations, or civilians, when the support is actually manufactured.”<sup>34</sup>

Moreover, it has become increasingly difficult to discern A.I.-generated content from reality. Generative models such as DALL-E and Stable Diffusion can produce synthetic imagery that is often visually indistinguishable from human-created content.<sup>35</sup> Some A.I. models now can produce realistic videos from simple text instructions.<sup>36</sup> Moreover, advancements in generative systems are outpacing the tools to verify the veracity of generative A.I. content.<sup>37</sup> Some of the latest research on tools used widely in academia to detect A.I.-generated text reveal that they are neither accurate nor reliable.<sup>38</sup> An analysis of fourteen of the most widely used tools, including Turnitin, PlagiarismCheck, OpenAI Text Classifier, and GPT Zero, found that all scored below an 80% rate of accuracy, resulting in false positives (diagnosing human-written text as A.I.-generated) and false negatives (diagnosing A.I.-generated texts as human-written); the authors of the study concluded that approximately one out of five A.I.-generated texts would be misattributed to humans.<sup>39</sup>

---

<sup>34</sup> *Id.*

<sup>35</sup> Xuandong Zhao et al., *Invisible Image Watermarks Are Provably Removable Using Generative AI 1* (June 2, 2023), <https://arxiv.org/pdf/2306.01953> [<https://perma.cc/JK7P-6PNH>] (digital preprint).

<sup>36</sup> *See Sora*, OPENAI, <https://openai.com/sora> [<https://perma.cc/VR3N-AMDS> (staff-uploaded)] (text-to-video AI model).

<sup>37</sup> *See* Lauren Leffer, *AI Audio Deepfakes Are Quickly Outpacing Detection*, SCIENT. AMER. (Jan. 26, 2024), <https://www.scientificamerican.com/article/ai-audio-deepfakes-are-quickly-outpacing-detection/> [<https://perma.cc/H6YH-7ZNQ>]. *See generally* <https://www.scientificamerican.com/article/ai-audio-deepfakes-are-quickly-outpacing-detection/> [<https://perma.cc/2AVS-EGAY>]; Hazem Ibrahim et al., *Perception, Performance and Detectability of Conversational Artificial Intelligence Across 32 University Courses*, 13 SCI. REP. 12187 (2023), <https://www.nature.com/articles/s41598-023-38964-3> [<https://perma.cc/K8EC-XU3E>] (detection of AI-generated content in university student submissions).

<sup>38</sup> *See* Debora Weber-Wulff et al., *Testing of Detection Tools for AI-Generated Text*, 19 INTER. J. FOR EDUC. INTEGRITY ART. 26, at 1 (2023), <https://doi.org/10.1007/s40979-023-00146-z> [<https://perma.cc/7B8A-RU7W>].

<sup>39</sup> *Id.* at 25.

Digital watermarking has emerged as one potential method of identifying A.I.-generated content. Digital watermarking involves the use of hidden information, a signal, embedded in digitally transmitted data that can be detected algorithmically to prove the authenticity of an image or text.<sup>40</sup> President Biden's recent Executive Order addressing standards for A.I. safety and security points to the practice of watermarking as a safeguard against A.I.-enabled fraud and deception.<sup>41</sup> While leading technology companies, including Alphabet, OpenAI, Meta, and Amazon, have recently pledged a commitment to develop watermarking mechanisms to prevent A.I. from being used for disinformation,<sup>42</sup> the latest research reveals that strong digital watermarks on images are difficult, if not impossible, to achieve.<sup>43</sup> The authors of one recent study note that invisible watermarks can be removed with little to no degradation in image quality.<sup>44</sup> This irony with A.I. models and detection tools underscores the challenges that lie ahead: Advances in generative models that compelled the need for digital watermarking have also made watermarking removal easier when such models are incorporated into an attack on them.<sup>45</sup>

A.I. researcher Mustafa Suleyman argues that the scale and speed of A.I. proliferation will follow the trajectories of the printing press and the Internet.<sup>46</sup> Furthermore, while Moore's Law holds that

---

<sup>40</sup> See Seyed Mehran Dibaji, *A Tutorial on Security and Privacy Challenges in CPS*, in SECURITY AND RESILIENCE OF CONTROL SYSTEMS: THEORY AND APPLICATIONS 137 (Hideaki Ishii & Quanyan Zhu eds., 2022); see also John Kirchenbauer et al., *A Watermark for Large Language Models 1* (Int'l Conf. on Mach. Learning, 2023) <https://arxiv.org/pdf/2301.10226> [<https://perma.cc/ZPD4-MXW2>].

<sup>41</sup> Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Oct. 30, 2023).

<sup>42</sup> Vittoria Elliott, *Big AI Won't Stop Election Deepfakes with Watermarks*, WIRED (July 27, 2023, 7:00 AM), <https://www.wired.com/story/ai-watermarking-misinformation/> [<https://perma.cc/7X5N-MWEZ>].

<sup>43</sup> Hanlin Zhang et al., *Watermarks in the Sand: Impossibility of Strong Watermarking for Generative Models 1* (Int'l Conf. on Mach. Learning, 2023), <https://arxiv.org/pdf/2301.10226> [<https://perma.cc/VDC2-NTJMJ>]; see also Xuandong Zhao et al., *Invisible Image Watermarks Are Provably Removable Using Generative AI*, (June 2, 2023) (digital preprint), <https://arxiv.org/pdf/2306.01953> [<https://perma.cc/JK7P-6PNH>].

<sup>44</sup> Zhao, *supra* note 35.

<sup>45</sup> *Id.*

<sup>46</sup> Mustafa Suleyman, *Containment for AI*, FOREIGN AFFAIRS, (Jan. 23, 2024), <https://www.foreignaffairs.com/world/containment-artificial-intelligence-mustafa-suleyman> [<https://perma.cc/M6FN-2S6N>].

computing power doubles every two years, “progress has been even faster in A.I., with trends of lower costs and improving capability ascending on a curve beyond anything seen with a technology before.”<sup>47</sup> A new generation of autonomous A.I. tools is on the horizon, including artificial general intelligence (“AGI”), which will have human-like cognitive abilities to self-teach and complete complex activities autonomously.<sup>48</sup> Yet, AGI will also usher in new forms of disinformation and cyberattacks—and, as the history of great power competition shows, an A.I. arms race will likely skew in favor of A.I. for offensive purposes over A.I.-defensive efforts.<sup>49</sup> Nvidia founder Jensen Huang described A.I. as a technology that society should not be “terrified” of; instead, he encourages the public to take advantage of it because society is at “the beginning of a new industrial revolution.”<sup>50</sup> Yet, rogue actors and governments will also take advantage of advances in A.I. for nefarious purposes, making disinformation attacks not only more difficult to recognize but also much easier to propagate.

### III. A MARKETPLACE (FAILURE) OF COMPETITION WITH FALSEHOOD

Much of contemporary First Amendment doctrine has developed from the classic theory of the marketplace of ideas, first referenced by Justice Oliver Wendell Holmes in the early 1900s.<sup>51</sup> According to this theory, “truth is discovered through its competition with falsehood for

---

<sup>47</sup> *Id.*

<sup>48</sup> *What Is Artificial General Intelligence (AGI)?* MCKINSEY & COMPANY (Mar. 21, 2024), <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-artificial-general-intelligence-agi#/> [<https://perma.cc/G2ZJ-RH68>].

<sup>49</sup> See Jeppe T. Jacobsen & Tobias Liebetrau, *Artificial Intelligence and Military Superiority*, in *ARTIFICIAL INTELLIGENCE AND INTERNATIONAL CONFLICT IN CYBERSPACE* 135 (Fabio Cristiano et al. eds., 1st ed. 2023).

<sup>50</sup> Mukesh Adhikary, *‘Beginning of a New Industrial Revolution’: Nvidia’s Jensen Huang Says There’s AI ‘Awakening’ in Every Country* (Feb. 13, 2024, 10:01 AM), *BUS. TODAY*, <https://www.businesstoday.in/technology/top-story/story/beginning-of-new-industrial-revolution-nvidias-jensen-huang-says-theres-an-ai-awakening-in-every-country-417256-2024-02-13> [<https://perma.cc/SP8H-YJBC>].

<sup>51</sup> See *Abrams v. United States*, 250 U.S. 616, 630 (1919); Robert Post, *Reconciling Theory and Doctrine in First Amendment Jurisprudence*, 88 *CALIF. L. REV.* 2353, 2355 (2000).

acceptance.”<sup>52</sup> As C. Edwin Baker noted, the U.S. Supreme Court has relied upon this theory on many occasions to determine which speech is protected and to provide “justification for their First Amendment ‘tests.’”<sup>53</sup> For instance, in *Hustler Magazine v. Falwell*,<sup>54</sup> the Court asserted that false statements of fact “are particularly valueless; they interfere with the truth-seeking function of the marketplace of ideas.”<sup>55</sup> While addressing First Amendment implications with advertising, the Supreme Court noted that advertising is not “stripped of all First Amendment protection” because “[t]he relationship of speech to the marketplace of products or of services does not make it valueless in the marketplace of ideas.”<sup>56</sup> In *United States v. Alvarez*,<sup>57</sup> the Supreme Court—again relying on the marketplace theory—left no doubt that false speech is subject to First Amendment scrutiny.<sup>58</sup>

The divided Court in *Alvarez* held that the prosecution of an individual under the Stolen Valor Act for falsely claiming he was awarded the Congressional Medal of Honor violated the First Amendment.<sup>59</sup> The Court pointed out that “some false statements are inevitable” in order to facilitate an open and robust discussion of views.<sup>60</sup> While the Court has asserted that false statements are not protected in a variety of contexts, such assertions were tied to a “legally cognizable harm associated with a false statement.”<sup>61</sup> In response to the government’s argument that some regulations on false speech have been found constitutionally permissible (e.g., false statements made to a government official, perjury, or impersonating a government official), the Court noted that such regulations are necessary to preserve the integrity of legal judgments and government processes,

---

<sup>52</sup> C. Edwin Baker, *Scope of the First Amendment Freedom of Speech*, 25 UCLA L. REV. 964, 967 (1978).

<sup>53</sup> *Id.* at 968.

<sup>54</sup> 485 U.S. 46 (1988).

<sup>55</sup> *Id.* at 52; *see also* *Time, Inc. v. Hill*, 385 U.S. 374, 406 (1967) (Harlan, J., concurring) (“‘The marketplace of ideas’ where it functions still remains the best testing ground for truth.”).

<sup>56</sup> *See* *Bigelow v. Virginia*, 421 U.S. 809, 826 (1975).

<sup>57</sup> 567 U.S. 709 (2012).

<sup>58</sup> *Id.* at 715.

<sup>59</sup> *Id.*

<sup>60</sup> *Id.* at 718.

<sup>61</sup> *Id.* at 718–19.

but censoring falsity alone would chill speech.<sup>62</sup> Ultimately, the Court held that the government failed to satisfy “exacting scrutiny”<sup>63</sup> and, relying on the market competition framework, declared:

The remedy for speech that is false is speech that is true. This is the ordinary course in a free society. The response to the unreasoned is the rational; to the uninformed, the enlightened; to the straightout lie, the simple truth. The theory of our Constitution is “that the best test of truth is the power of the thought to get itself accepted in the competition of the market.”<sup>64</sup>

Today, however, the assumptions undergirding the marketplace framework cannot be relied upon to address the scourge of modern A.I.-assisted disinformation campaigns that aim to degrade our public discourse. According to C. Edwin Baker, one of the presuppositions of the marketplace theory is that people’s “rational faculties must enable them to sort through the form and frequency of message presentation to evaluate the core notions.”<sup>65</sup> In other words, people must be able to correctly perceive the truth or reality. But advanced communication technologies, such as generative A.I., have created an environment that makes it much more difficult to do so,<sup>66</sup> especially when generative advancements are outpacing the detection tools that distinguish reality from the realistic.<sup>67</sup> Constitutional scholar Tim Wu aptly notes that today’s tools of speech-control, including content and distribution-driven campaigns, are incredibly challenging for the First Amendment to address, leaving the Amendment “increasingly irrelevant in its area of historic concern: the coercive control of political speech.”<sup>68</sup> Wu points out three assumptions that followed the early development of First Amendment jurisprudence: (1) an underlying premise of informational scarcity, (2) an assumption that listeners have adequate time and interest to be influenced by public

---

<sup>62</sup> *Id.* at 720–23.

<sup>63</sup> *Id.* at 726 (noting the lack of a causal link between the government’s stated interest in regulating lies about the Medal of Honor and the injury to be prevented: “The Government has not shown, and cannot show, why counterspeech would not suffice to achieve its interest. The facts of this case indicate that the dynamics of free speech, of counterspeech, of refutation, can overcome the lie.”).

<sup>64</sup> *Id.* at 727–28 (citation omitted) (quoting *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting)).

<sup>65</sup> See Baker, *supra* note 52, at 967.

<sup>66</sup> See *supra* notes 34–46 and accompanying discussion.

<sup>67</sup> See *supra* notes 36–40 and accompanying discussion.

<sup>68</sup> Tim Wu, *Is the First Amendment Obsolete?* 117 MICH. L. REV. 547, 549 (2018).

views, and (3) an assumption that the U.S. government is the main threat to the “marketplace of ideas.”<sup>69</sup> These assumptions now seem anachronistic in a digital landscape on the verge of an A.I. revolution. Today, foreign-based disinformation and speech-control tactics pose some of the greatest threats to our public discourse. Malign foreign actors can easily flood Western social media channels with A.I.-generated disinformation, exploit data voids with falsehoods, and even use economic measures to manipulate or censor political speech.<sup>70</sup>

The rationality assumption underpinning the classic marketplace theory also requires that a person’s social standing not impact the way they perceive or understand the world.<sup>71</sup> As Baker noted:

If perceptions are social creations and if people’s social experiences are radically different, then mere discussion would be inadequate for discovering what truth or which perspectives are correct or best; one could not hope that employing reason in discussion would provide an unbiased insight into reality.<sup>72</sup>

Yet, political bias affects how unsubstantiated claims are viewed. Research by Nicolas Berlinski et al. provides evidence that exposure to unsubstantiated claims of voter fraud degrades confidence in elections, especially along partisan lines.<sup>73</sup> Even more concerning, exposure to fact-checks that reveal their falsity does *not* significantly reduce the damage from these unfounded claims.<sup>74</sup> Research on foreign interference in U.S. elections also reveals that foreign intervention polarizes voters along partisan lines.<sup>75</sup> Political affiliation matters regarding how such events are perceived.<sup>76</sup> To demonstrate this relationship, a recent study by Michael Tomz and Jessica Weeks showed that Americans react more negatively toward foreign meddling when a foreign entity supports a political candidate from an opposing political party.<sup>77</sup> This study revealed a partisan double standard,

---

<sup>69</sup> *Id.* at 553–54.

<sup>70</sup> See Park, *supra* note 3; see also *supra* notes 24–43 and accompanying discussion.

<sup>71</sup> See Baker, *supra* note 52, at 967.

<sup>72</sup> *Id.*

<sup>73</sup> See Nicolas Berlinski et al., *The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections*, 10 J. OF EXP. POL. SCI. 34 (2023).

<sup>74</sup> *Id.*

<sup>75</sup> See Michael Tomz & Jessica L.P. Weeks, *Public Opinion and Foreign Electoral Intervention*, 114(3) AMER. POL. SCI. REV. 856, 863–64 (2020).

<sup>76</sup> *Id.*

<sup>77</sup> *Id.* at 863.

whereby Americans are more likely to distrust the results of an election or lose faith in democracy if a foreign entity endorses a candidate from the opposition party.<sup>78</sup> Another recent study showed that even exposure to news about foreign-based disinformation targeting U.S. politics can affect the perceived impact of disinformation's influence on others and indirectly influence confidence in electoral outcomes.<sup>79</sup>

Moreover, the flood of disinformation and misinformation surrounding COVID-19 and its lasting consequences on public health and public discourse offers another example of how the rationality assumption may no longer hold true. The scale of false information regarding the pandemic led to an “infodemic,” defined by the World Health Organization as “too much information or false and misleading information” that results in “confusion, risk-taking behaviors . . . and mistrust in health authorities.”<sup>80</sup>

Steven Wilson and Charles Wiysonge's research on global anti-vaccination efforts on social media showed that foreign online disinformation campaigns were highly associated with both a drop in mean vaccination coverage over time and negative social media content about vaccination.<sup>81</sup> Research from the Pew Center also showed that political partisanship affected views about the dangers of COVID-19, offering further evidence of how the rationality assumption has been undermined; more than 85% of Democrats (or those who lean Democrat) viewed the coronavirus outbreak as a major public health threat versus just 45% of Republican-leaning citizens.<sup>82</sup>

---

<sup>78</sup> *Id.* at 866.

<sup>79</sup> See Andrew R.N. Ross, Cristian Vaccari & Andrew Chadwick, *Russian Meddling in U.S. Elections: How News of Disinformation's Impact Can Affect Trust in Electoral Outcomes and Satisfaction with Democracy*, 25(6) MASS COMM. & SOC'Y 786, 802 (2022), <https://www.tandfonline.com/doi/full/10.1080/15205436.2022.2119871> [<https://perma.cc/25X9-LD7T> (staff-uploaded)].

<sup>80</sup> *Infodemic*, WORLD HEALTH ORG., [https://www.who.int/health-topics/infodemic#tab=tab\\_1](https://www.who.int/health-topics/infodemic#tab=tab_1) [<https://perma.cc/NE5H-9ABS>] (last visited Sept. 20, 2024).

<sup>81</sup> See Steven L. Wilson & Charles Wiysonge, *Social Media and Vaccine Hesitancy*, 5(10) BMJ GLOB. HEALTH 1, 5 (2020), <https://gh.bmj.com/content/bmjgh/5/10/e004206.full.pdf> [<https://perma.cc/47SF-6LHQ>].

<sup>82</sup> See Alec Tyson, *Republicans Remain Far Less Likely than Democrats to View Covid-19 as a Major Threat to Public Health*, PEW RSCH. CTR. (July 22, 2020), <https://www.pewresearch.org/short-reads/2020/07/22/republicans-remain-far-less-likely-than-democrats-to-view-covid-19-as-a-major-threat-to-public-health/> [<https://perma.cc/MJ4S-N5HL>].

Furthermore, research on false claims related to COVID-19 revealed that such claims spread by public figures (i.e., celebrities, high-ranking public officials) attracted a large and disproportionate level of social media engagement, while public figures spread only 20% of false claims related to COVID-19, such claims attracted a large majority of all social media engagement.<sup>83</sup>

The discovery of truth through its competition with falsehood is also undermined by conformity bias. Put simply, conformity bias “reflects the fact that completely separate from our rational judgments, [people] simply do not like to stick out from a pack.”<sup>84</sup> Solomon Asch’s conformity experiments demonstrated the powerful effect of group pressure on individual independence and submission.<sup>85</sup> Research also shows that repetition increases perceived truth.<sup>86</sup> One theory that explains this illusory-truth effect is that repeated statements are more easily processed, eliciting mental shortcuts or heuristics; in other words, people are more inclined to believe what is familiar to them.<sup>87</sup> According to Nobel laureate Daniel Kahneman, “[a] reliable way to make people believe in falsehoods is frequent repetition because familiarity is not easily distinguished from truth. Authoritarian institutions and marketers have always known this fact.”<sup>88</sup> Furthermore, much of the political discourse today has migrated to private social networking groups that offer end-to-end encryption (e.g., Signal, WhatsApp, and Telegram), making it not only difficult for law

---

<sup>83</sup> See J. Scott Brennen et al., *Types, Sources, and Claims of COVID-19 Misinformation*, REUTERS INST. FOR THE STUDY OF JOURNALISM (2020), <https://ora.ox.ac.uk/objects/uuid:178db677-fa8b-491d-beda-4bacdc9d7069/files/sk0698765w> [<https://perma.cc/3X4Y-4W9X>]; see also Michael A. Gisoni et al., *A Deadly Infodemic: Social Media and the Power of COVID-19 Misinformation*, 24(2) J. OF MED. INTERNET RSCH. (2022), <https://www.jmir.org/2022/2/e35552/> [<https://perma.cc/M4VB-F42C>].

<sup>84</sup> See CAILIN O’CONNOR & JAMES OWEN WEATHERALL, *THE MISINFORMATION AGE: HOW FALSE BELIEFS SPREAD* 84 (2020).

<sup>85</sup> See Solomon E. Asch, *Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority*, 70(9) PSYCH. MONOGRAPHS: GEN. & APPLIED 1 (1956).

<sup>86</sup> Adam L. Alter & Daniel M. Oppenheimer, *Predicting Short-Term Stock Fluctuations by Using Processing Fluency*, 103 PROC. OF THE NAT’L ACAD. OF SCIS. 9369 (2006); Amos Tversky and Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COG. PSYCH. 207 (1973).

<sup>87</sup> *Id.*

<sup>88</sup> DANIEL KAHNEMAN, *THINKING, FAST AND SLOW* 62 (2011).



enforcement to monitor the rhetoric of extremist groups,<sup>89</sup> but also resulting in less fact-checking, which undermines the test for truth in the competition of ideas. In sum, the longstanding implicit assumptions of the marketplace theory are no longer applicable in today's media environment to actualize the truth-seeking function of the theory. More true speech is no longer the optimal remedy for false speech.

#### IV. A REORIENTATION OF SPEECH ANALYSIS: PROPORTIONALITY REVIEW AND DEMOCRATIC THEORY

In a recent speech case involving a student's vulgar social media post, the Supreme Court again invoked the "marketplace of ideas" theory to explain why the First Amendment protected such speech: "Our representative democracy only works if we protect the 'marketplace of ideas.' This free exchange facilitates an informed public opinion . . . ."<sup>90</sup> But if the marketplace can no longer adequately facilitate the competition for truth and address disinformation generated by today's A.I.-assisted communication tools, perhaps it is time to revise the methodological approach to speech analysis.

What is urgently needed is an approach that can both adequately address countervailing interests and legal rationale while also providing greater consideration for the purposes of the First Amendment. However, the current Court—informed by the marketplace theory—has adopted an inflexible categorical approach to First Amendment cases and has rejected balancing speech values against countervailing interests.<sup>91</sup> In response to the government's proposal for the categorical balancing of the value of speech against its societal costs,

---

<sup>89</sup> See Jack Nicas et al., *Millions Flock to Telegram and Signal as Fears Grow over Big Tech*, N.Y. TIMES (Jan. 13, 2021), <https://www.nytimes.com/2021/01/13/technology/telegram-signal-apps-big-tech.html> [<https://perma.cc/5W2H-TE5T> (staff-uploaded, dark archive)]; see also Sophie Haigney, *How Group Chats Rule the World*, N.Y. TIMES MAGAZINE (Jan. 16, 2024), <https://www.nytimes.com/2024/01/16/magazine/group-chats.html> [<https://perma.cc/6FWN-C6A8> (staff-uploaded, dark archive)].

<sup>90</sup> *Mahanoy Area Sch. Dist. v. B.L.*, 594 U.S. 180, 190 (2021).

<sup>91</sup> For some recent examples, see *Barr v. Am. Ass'n of Pol. Consultants, Inc.*, 140 S. Ct. 2335, 2346 (2020) ("Content-based laws are subject to strict scrutiny."); *infra* notes 90–92 and accompanying discussion.

the Court has asserted “[a]s a free-floating test for First Amendment coverage, that sentence is startling and dangerous.”<sup>92</sup>

In *Reed v. Gilbert*,<sup>93</sup> the Court starkly rejected any consideration of a town’s objectives when it reviewed the town’s sign code that subjected ideological, political, and temporary directional signs to different restrictions.<sup>94</sup> Justice Thomas, writing for the majority, held that the sign code is content-based and “presumptively unconstitutional”; therefore, there is “no need to consider the government’s justifications or purposes for enacting the code to determine whether it [was] subject to strict scrutiny.”<sup>95</sup> Alexander Tsisis points out that the Court’s marketplace approach to free speech is linked to libertarianism, the philosophy that “[m]ore speech in the marketplace of ideas is considered invariably better . . . and government interference is met with suspicion as an affront to personal liberty.”<sup>96</sup> This has led to a formalistic approach that “relies on predetermined criteria rather than the totality of circumstances of specific litigants and conflicts.”<sup>97</sup>

Contemporary First Amendment speech analysis, informed by the marketplace of ideas theory and tied to a rigid formalistic approach, is ill-fitted to address the scourge of today’s A.I.-assisted disinformation schemes. When applying existing precedents, modifying them, or even “overturn[ing] them altogether,” Tsisis contends that the “application of an ancient constitution to issues about new information technologies and digital materials requires contextual flexibility.”<sup>98</sup> Instead of a judicial approach that boxes every speech issue into one of several categories, a legal methodology is needed that can incorporate context, counterarguments, the purposes of the law, and the Constitution’s provisions.

Former Supreme Court Justice Stephen Breyer has been an advocate of such a methodology, promoting an approach that examines whether the regulation imposes restrictions that are

---

<sup>92</sup> *United States v. Stevens*, 599 U.S. 460, 470 (2010).

<sup>93</sup> 576 U.S. 155 (2015).

<sup>94</sup> *Id.* at 164–65.

<sup>95</sup> *Id.*

<sup>96</sup> ALEXANDER TISIS, *FREE SPEECH IN THE BALANCE* 44 (2020).

<sup>97</sup> *Id.* at 43 (“Categoriality fails to adequately address counterarguments, often disregards conflicting legal reasons, and downplays facts in the record.”).

<sup>98</sup> *Id.* at 46.

“disproportionate” when balanced against speech-related benefits.<sup>99</sup> In *United States v. Alvarez*,<sup>100</sup> Justice Breyer reiterated his methodology for reviewing laws under the First Amendment, noting that the Court has taken into account

the seriousness of the speech-related harm the provision will likely cause, the nature and importance of the provision’s countervailing objectives, the extent to which the provision will tend to achieve those objectives, and whether there are other, less restrictive ways of doing so. Ultimately the Court has had to determine whether the statute works speech-related harm that is out of proportion to its justifications.<sup>101</sup>

Breyer’s proportionality analysis calls for an evaluation of all relevant public and private interests and the context of a challenge to a speech regulation.<sup>102</sup> Proportionality can be understood as a “legal principle, as a goal of government, and as a particular structured approach to judicial review.”<sup>103</sup> Vicki Jackson notes that proportionality embodies the idea that greater harms imposed by the government “should be justified by more weighty reasons and that more severe transgressions of the law be more harshly sanctioned than less severe ones.”<sup>104</sup>

Furthermore, proportionality review offers a method of analysis that permits legislatures to respond flexibly to the disinformation threats posed by A.I. and emerging communication technology. It avoids adopting broad, rigid constitutional rules that often prevent elected officials from creating much-needed regulation.<sup>105</sup> Proportionality review offers an analytical approach that can comprehensively analyze government methods aimed at today’s foreign-based disinformation schemes—partly because it looks to the

---

<sup>99</sup> See *Bartnicki v. Vopper*, 532 U.S. 514, 535–41 (2001) (Breyer, J., concurring); *U.S. v. Alvarez*, 567 U.S. 709, 730–39 (2012) (Breyer, J., concurring); see also STEPHEN BREYER, *ACTIVE LIBERTY: INTERPRETING OUR DEMOCRATIC CONSTITUTION* 49 (2005).

<sup>100</sup> 567 U.S. 709 (2012).

<sup>101</sup> *Id.* at 730 (2012) (Breyer, J., concurring) (“In doing so, it has examined speech-related harms, justifications, and potential alternatives.”).

<sup>102</sup> TESISIS, *supra* note 96, at 41.

<sup>103</sup> Vicki C. Jackson, *Constitutional Law in an Age of Proportionality*, 124 *YALE L.J.* 3094, 3098 (2015).

<sup>104</sup> *Id.*

<sup>105</sup> BREYER, *supra* note 99, at 42–43.

purposes of statutes and constitutional principles.<sup>106</sup> Breyer contends that constitutional interpretation and application should take greater account of the Constitution’s democratic nature—such as its call for participatory self-government and democratic deliberation—when courts interpret the Constitution.<sup>107</sup> The adoption of proportionality analysis should, therefore, be informed by the democratic theory of self-government. One of the foundational purposes of the First Amendment is to facilitate deliberation and participatory self-government.<sup>108</sup> This constitutional purpose is often associated with Alexander Meiklejohn and the democratic theory of the First Amendment.<sup>109</sup> This theory is understood “to protect the communicative processes necessary to disseminate the information and ideas required for citizens” to sustain a self-governing republic.<sup>110</sup> It developed from the idea that freedom of speech is necessary to produce educated and informed public opinion.<sup>111</sup>

Today’s communicative processes are under the threat of being hijacked by a plethora of A.I.-generated fakes of human voices, texts, images, and videos. As aforementioned, it will become increasingly difficult to distinguish content that is real from content that is generated by artificial intelligence. Thus, the U.S. has a vested interest in promoting a speech environment in which the citizenry receives the necessary (and *accurate*) information to fulfill its sovereign function.<sup>112</sup>

Yet government attempts to regulate foreign-based disinformation and deepfakes will likely be challenged as content-based speech

---

<sup>106</sup> STEPHEN BREYER, *READING THE CONSTITUTION: WHY I CHOSE PRAGMATISM, NOT TEXTUALISM* 10 (2024) (“Purpose-based approaches advance constitutional and congressional purposes, not the judge’s own agenda.”).

<sup>107</sup> See BREYER, *supra* note 99, at 21–34.

<sup>108</sup> See BREYER, *supra* note 99, at 46–47; see also CASS SUNSTEIN, *DEMOCRACY AND THE PROBLEM OF FREE SPEECH* 18–22 (1995).

<sup>109</sup> Alexander Meiklejohn, *The First Amendment is an Absolute*, SUP. CT. REV. 245, 263 (1961); see generally ALEXANDER MEIKLEJOHN, *FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT* (1948).

<sup>110</sup> See Post, *supra* note 51, at 2367.

<sup>111</sup> See ROBERT C. POST, *DEMOCRACY, EXPERTISE AND ACADEMIC FREEDOM* 35 (2012).

<sup>112</sup> See Mark Lloyd and Michael Park, *The Constitutional Case for Addressing Critical Information Needs*, in *THE COMMUNICATION CRISIS IN AMERICA AND HOW TO FIX IT* 235, 237–39 (Mark Lloyd & Lewis A. Friedland eds., 2016); BREYER, *supra* note 99, at 39–47.

restrictions. As a general matter, content-based restrictions are presumed invalid<sup>113</sup> and “[i]f a statute regulates speech based on its content, it must be narrowly-tailored to promote a compelling [g]overnment interest.”<sup>114</sup> As the Supreme Court in *Alvarez* noted, outside certain contexts, false statements receive First Amendment protection.<sup>115</sup> Thus, regulation of false media content, including fabricated news posts and foreign-based deepfakes, will likely be subject to strict categorical interpretation—and strict scrutiny—by the current Supreme Court.<sup>116</sup> The government would have to make a showing that the regulation satisfies a compelling government interest and that the means adopted are narrowly tailored to serve that interest.<sup>117</sup> But categorically boxing in legislative responses to foreign-based disinformation as “presumptively unconstitutional” leaves little to no room for legislative dynamism and responsiveness to address modern A.I.-generated threats to democratic institutions.

Instead of rubber-stamping content-limiting regulation as presumptively invalid, judicial review should consider the speaker’s interests subject to the law and countervailing government objectives. Take, for example, a hypothetical regulation that prohibits foreign state-based deepfake content, including fake human voices, images, and videos. Assume this regulation is further limited with a requirement that substantial harm be directly foreseeable or likely. The Supreme Court has asserted that “[a] law that is content-based on its face is subject to strict scrutiny regardless of the government’s benign-motive, content-neutral justification or lack of ‘animus toward the ideas contained’ in the regulated speech.”<sup>118</sup> Under the current categorical approach to speech analysis, such a regulation would be

---

<sup>113</sup> See *Ashcroft v. American Civil Liberties Union*, 542 U.S. 656, 660 (2004).

<sup>114</sup> *United States v. Playboy Ent. Grp., Inc.*, 529 U.S. 803, 813 (2000).

<sup>115</sup> *United States v. Alvarez*, 567 U.S. 709, 719 (2012) (“The Court has never endorsed the categorical rule the Government advances: that false statements receive no First Amendment protection”). The Court explained earlier that “false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversation, expression the First Amendment seeks to guarantee.” *Id.* at 718; see also *supra* notes 58–63 and accompanying discussion.

<sup>116</sup> *Barr v. American Ass’n of Pol. Consultants, Inc.*, 591 U.S. 610, 618 (2020) (“Content-based laws are subject to strict scrutiny.”).

<sup>117</sup> See *Ashcroft*, 542 U.S. at 660; *Playboy Ent. Grp. Inc.*, 529 U.S. at 813.

<sup>118</sup> *Reed v. Town of Gilbert*, 576 U.S. 155, 165 (2015) (quoting *Cincinnati v. Discovery Network, Inc.*, 507 U.S. 410, 429 (1993)).

deemed “content-based,” since it targets foreign states’ deepfakes and is presumptively unconstitutional under exacting judicial scrutiny. Thus, the current formalistic and rigid approach to judicial review hampers legislative attempts to respond to foreign malign influence operations where—due to A.I.—the barriers of entry have become increasingly small, accelerating propaganda and disinformation aimed at the public.<sup>119</sup>

However, under a proportionality balancing approach, all relevant public and private interests would be examined, including the context of a challenge to such a hypothetical regulation. On the one hand, such a regulation elicits some potential speech-related harms. One such harm is related to government overreach and politicized enforcement; another is tied to the right to receive foreign speech. Regarding potential bans on deepfakes, Bobby Chesney and Danielle Citron note that self-serving prosecutions and politically motivated enforcement are possible.<sup>120</sup> James Weinstein contends that society should be reluctant to entrust government officials with the power to determine “the truth or falsity of factual claims made in the often highly ideological context of public discourse, especially when the claims are factually complex or uncertain.”<sup>121</sup> The dislike of minority or unpopular viewpoints, coupled with the uncertainty behind a deepfake creator’s intent, may result in politicized enforcement.<sup>122</sup>

Furthermore, while foreign speakers *outside* the U.S. do not enjoy First Amendment protection, foreign speakers *inside* the country are afforded First Amendment rights.<sup>123</sup> Thus, foreign state principals who have lawfully entered and resided in the U.S. enjoy First Amendment rights to their false speech content.<sup>124</sup> Moreover, domestic audiences

---

<sup>119</sup> See Sean Lyngaas, *AI Will Allow More Foreign Influence Operations in 2024 Election, FBI Director Says*, CNN (Feb. 29, 2024), <https://www.cnn.com/2024/02/29/politics/ai-2024-election-fbi-director/index.html> [<https://perma.cc/2497-Y86A>].

<sup>120</sup> Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1789 (2019).

<sup>121</sup> James Weinstein, *Climate Change Disinformation, Citizen Competence, and the First Amendment*, 89 U. COLO. L. REV. 341, 351 (2018) (“There is even greater reason to distrust the ability of government officials to fairly and accurately determine the speaker’s state of mind in making the alleged false statement.”).

<sup>122</sup> See Chesney & Citron, *supra* note 120, at 1789.

<sup>123</sup> See *United States v. Verdugo-Urquidez*, 494 U.S. 259, 270–71 (1990); Joseph Thai, *The Right to Receive Foreign Speech*, 71 OKLA. L. REV. 269, 275 (2018).

<sup>124</sup> See Thai, *supra* note 123, at 275–76.

still have a right to receive foreign information and ideas, and according to Joseph Thai, this includes “the speech of hostile foreign governments advocating civil riot or violent upheaval.”<sup>125</sup> This speech-related ban can potentially limit the diversity of views in the marketplace, where, as the Roberts Court has asserted, “the remedy for speech that is false is speech that is true.”<sup>126</sup>

On the other side of this ledger are countervailing interests in deterring foreign state-based deepfakes designed to deceive and cause harm. As aforementioned, disinformation and influence campaigns by foreign adversaries have been linked to disruption and harm in areas such as public health<sup>127</sup> and the financial markets,<sup>128</sup> continuing to pursue malign aims of sowing discord and undermining the U.S. political system.<sup>129</sup> These disinformation campaigns by foreign actors pose great harm to the constitutional principle of democratic self-governance by degrading the speech ecosystem that facilitates representative government. It has been well-documented that Russian operatives engaged in “information warfare” leading up to the 2016 presidential election,<sup>130</sup> and that China-affiliated actors are currently harnessing A.I.-generated media as part of a broad disinformation campaign in furtherance of sowing civil discord, disparaging U.S. political officials, and influencing elections.<sup>131</sup> Chesney and Citron point out that the threats of synthetic media “have systemic dimensions,” including harms to democratic discourse, election

---

<sup>125</sup> *Id.* at 281; *see also* *Lamont v. Postmaster Gen. of the U.S.*, 381 U.S. 301, 306 (1965).

<sup>126</sup> *United States v. Alvarez*, 567 U.S. 709, 727 (2012).

<sup>127</sup> *See* INFODEMIC, *supra* note 80.

<sup>128</sup> *See* Sorkin et al., *supra* note 30; Alba, *supra* note 30.

<sup>129</sup> *See generally* Robert Mueller, *Report on the Investigation into Russian Interference in the 2016 Presidential Election*, U.S. DEPT. OF JUSTICE (2019), <https://www.justice.gov/archives/sco/file/1373816/dl> [<https://perma.cc/29JL-WZNA>].

<sup>130</sup> *Id.*

<sup>131</sup> *See* Watts, *supra* note 32; Clint Watts, *China Tests US Voter Fault Lines and Ramps AI Content to Boost its Geopolitical Interests*, MICROSOFT: MICROSOFT ON THE ISSUES (Apr. 4, 2024), <https://blogs.microsoft.com/on-the-issues/2024/04/04/china-ai-influence-elections-mtac-cybersecurity/> [<https://perma.cc/6L6K-7KMH>] (“There has been an increased use of Chinese AI-generated content in recent months, attempting to influence and sow division in the U.S. and elsewhere on a range of topics including: the train derailment in Kentucky in November 2023, the Maui wildfires in August 2023, the disposal of Japanese nuclear wastewater, drug use in the U.S. as well as immigration policies and racial tensions in the country.”).

manipulation, and weakened trust in democratic institutions such as journalism.<sup>132</sup> In *Alvarez*, the Court noted that constitutionally permissible regulations on false statements have either been linked to “legally cognizable harm” or the preservation of the integrity of legal judgments and government processes.<sup>133</sup> Similarly, the threats posed by A.I.-assisted synthetic media from foreign states pose a grave danger to the integrity of the U.S. political system and its democratic institutions. Although it is difficult to measure the direct impact of disinformation or deepfakes on human behavior, the threats posed by new A.I.-assisted tools of disinformation are further compounded by the fact that current detection tools do not provide the means to discern the real from the realistic.<sup>134</sup>

Moreover, foreign state-based disinformation and deepfakes intrinsically represent low-value speech in public discourse, especially when they are created with intent to cause harm. As Tsesis notes, there is Supreme Court precedent for the view that “the social interest in order and morality” outweigh the speech interests in utterances that carry very little social value.<sup>135</sup> Foreign state-based disinformation not only represents low-value speech; it arguably poses a national security threat as well. While foreign adversaries have previously used disinformation to sow civil discord and influence political operations, it is not difficult to imagine adversaries utilizing deepfake disinformation schemes involving the military. Such a dynamic is troubling because it can distort international relations and even elicit violent reactions from viewers.<sup>136</sup> Although regulations on disinformation from abroad invariably limit some forms of false content in the marketplace, the need to preserve the integrity of our democratic institutions from evolving threats of new media-manipulating tools is critical to self-governance. In weighing tradeoffs,

---

<sup>132</sup> See Chesney & Citron, *supra* note 120, at 1777–85.

<sup>133</sup> *United States v. Alvarez*, 567 U.S. 709, 719–21 (2012).

<sup>134</sup> See *supra* notes 37–45 and accompanying discussion.

<sup>135</sup> See TSESIS, *supra* note 96, at 16. While citing *Beauharnais v. Illinois*, 343 U.S. 250 (1942), and *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942), the author points out that the Supreme Court has previously engaged in the balancing of speech and countervailing interests and that threats to democratic order may outweigh speech interests. TSESIS, *supra* note 96, at 16. Furthermore, the author notes that free speech doctrines in obscenity, defamation and fraud reflect the idea that “well-defined social concerns” can outweigh “expressive liberty” interests. *Id.* at 15–18.

<sup>136</sup> See Chesney & Citron, *supra* note 120, at 1783.



the potential harm from such synthetic media designed to deceive is substantial, while the value of such false speech is relatively low and offers little contribution to core free speech values.

Although restrictions on a foreign speaker's speech may depend on whether they are physically in the U.S., the constitutional rights of aliens are generally not coextensive with those of citizens.<sup>137</sup> In the U.S. political context, the Supreme Court has articulated that the government may exclude foreign citizens from activities "intimately related to the process of democratic self-government."<sup>138</sup> In *Bluman v. Federal Election Commission*,<sup>139</sup> plaintiffs, foreign citizens lawfully residing in the U.S., challenged a federal ban restricting foreign citizens from making contributions to political candidates and political parties.<sup>140</sup> The D.C. District Court in *Bluman*—in an opinion written by then-circuit judge Brett Kavanaugh—held that, for First Amendment analysis, the U.S. has a compelling interest in "limiting the participation of foreign citizens in activities of American democratic self-government, and in thereby preventing foreign influence over the U.S. political process."<sup>141</sup> Granted, such precedent pertains to foreign citizen contributions and expenditures that enable a type of speech (although not pure speech), but it also stands for the premise that the government may make distinctions based on the foreign identity of the speaker—whether or not the speaker is abroad or residing in the U.S.

Further distinctions based on the foreign identity of the speaker have been made under First Amendment analysis for the purposes of

---

<sup>137</sup> See Thai, *supra* note 123, at 275–77; Toni M. Massaro, *Foreign Nationals, Electoral Spending, and the First Amendment*, 34 HARV. J. L. & PUB. POL'Y 663, 681 (2011); see also *Agency for Int'l Dev. v. All. for Open Soc'y Int'l*, 140 S. Ct. 2082, 2086 (2020) ("First, it is long settled as a matter of American constitutional law that foreign citizens outside U.S. territory do not possess rights under the U.S. Constitution."). *But see id.* at 2099 (Breyer, J., dissenting) ("The idea that foreign citizens abroad *never* have constitutional rights is not a 'bedrock' legal principle. At most, one might say that they are unlikely to enjoy very often extraterritorial protection under the Constitution.").

<sup>138</sup> *Bernal v. Fainter*, 467 U.S. 216, 220 (1984).

<sup>139</sup> 800 F. Supp. 2d 281 (2011), *aff'd*, 565 U.S. 1104 (2012).

<sup>140</sup> *Id.* at 285.

<sup>141</sup> *Id.* at 288. The court earlier asserted that "[i]t is fundamental to the definition of our national political community that foreign citizens do not have a constitutional right to participate in, and this may be excluded from, activities of democratic self-government." *Id.* at 288.

hindering foreign propaganda. Under the Foreign Agents Registration Act (“FARA”),<sup>142</sup> agents of foreign state principals in the U.S. must register as “agents” and comply with disclosure requirements—including the requirement that “any informational materials” transmitted on behalf of a foreign principal include a “copious statement” on such materials affirming the agent-principal relationship.”<sup>143</sup> In *Meese v. Keane*,<sup>144</sup> one of the few Supreme Court cases to analyze provisions of FARA, the Court addressed the issue of labeling requirements after a U.S. citizen sought to exhibit three Canadian films that the U.S. government labeled “political propaganda.”<sup>145</sup> The Court held that using the term “political propaganda” pursuant to FARA’s labeling requirements on speech materials did not violate the First Amendment.<sup>146</sup> The decisions in *Bluman* and *Meese* reiterate that the government may make distinctions based solely on the foreign identity of the speaker. In sum, the First Amendment has been interpreted to allow some regulations on foreign-linked speech in furtherance of democratic self-government and limiting foreign-based influence on public opinion.

On balance, a strong case can be made that the countervailing interests in preventing the injuries posed by foreign-based deepfakes—and the gravity of potential harms on public discourse and political processes—coupled with the difficulty of detecting synthetic media, outweigh the speech-related harm of limiting foreign-based deepfakes designed to cause private and public harm. However, proportionality analysis still requires examining whether a regulation advances the government’s interest and is the “least restrictive means” of doing so.<sup>147</sup> Potential First Amendment threats may be alleviated by finely tailoring a regulation on falsity to require the speaker’s knowledge of falsity or showing that the content caused, or is likely to cause, material harm.<sup>148</sup>

---

<sup>142</sup> 22 U.S.C. §§ 611–21.

<sup>143</sup> *Id.* §§ 612(a), 614(b).

<sup>144</sup> 481 U.S. 465 (1987).

<sup>145</sup> *See id.* at 467–68.

<sup>146</sup> *Id.*

<sup>147</sup> *See United States v. Alvarez*, 567 U.S. 709, 730 (2012) (Breyer, J., concurring).

<sup>148</sup> *See id.* at 737–38 (Breyer, J., concurring) (“Some lower courts have upheld the constitutionality of roughly comparable but narrowly-tailored statutes in political contexts); *United We Stand America, Inc. v. United We Stand, America N.Y., Inc.*,

If such tailoring is difficult to achieve, one alternative to a ban could involve the requirement that synthetic media be disclosed or labeled as a fake. However, if the law targets foreign state-based content, issues of accountability arise since such content is being created and distributed from abroad. The onus of labeling could be placed with social media platforms,<sup>149</sup> but the broad reading of Section 230 of the Communications Decency Act (“CDA”)<sup>150</sup> disincentivizes editorial activity or, as Danielle Citron notes, removes “incentives for better behavior by those in the best position to minimize harm.”<sup>151</sup> Nevertheless, regulations on foreign-based disinformation are favored because the speech’s low-value nature is outweighed by government objectives and the constitutional purposes of self-government.

The democratic theory of the First Amendment requires a judicial application of free speech principles that considers whether speech promotes public discourse pursuant to maintaining a democratic republic.<sup>152</sup> In sum, when interpreting legislative acts intended to deter foreign state-based speech that poses serious harm and disruptions to public discourse, democratic institutions, and political processes, a proportionality approach informed by democratic theory should give considerable weight to the constitutional purpose of self-government.

## V. CONCLUSION

Rival governments in China and Russia continue to employ deepfakes—often aided by A.I.—and other forms of disinformation in furtherance of a competition of ideas, seeking to undermine U.S. democratic institutions, manipulate global political narratives, and

---

128 F.3d 86, 93 (2d Cir. 1997) (upholding against a First Amendment challenge an application of the Lanham Act to a political organization); *Treasurer of the Comm. to Elect Gerald D. Lostracco v. Fox*, 389 N.W.2d 446, 449 (Mich. Ct. App. 1986) (upholding against a First Amendment challenge a statute prohibiting campaign material falsely claiming that one is an incumbent).

<sup>149</sup> See generally Abe Loven, Recent Development, *No Need to Reinvent the Wheel: Why the U.S. Should Implement Co-Regulatory Mechanisms to Regulate Deepfake Content on Internet Platforms*, 25 N.C. J.L. & TECH. 231 (2023).

<sup>150</sup> See 47 U.S.C. § 230.

<sup>151</sup> Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61, 118 (2009).

<sup>152</sup> See Post, *supra* note 51, at 2367. See generally BREYER, *supra* note 99 (contending that constitutional interpretation and application should take greater account of the Constitution’s democratic nature, which promotes participatory self-government and democratic deliberation); SUNSTEIN, *supra* note 108.

reshape the U.S.-led international order. Yet, the current free speech framework based on the marketplace theory cannot adequately address modern disinformation tactics and speech-control tools that are being deployed in this new period of great power competition. Truth no longer corrects for falsity in the marketplace of ideas. Moreover, the current categorical and formalistic approach to free speech cases rejects the balancing of speech values and government objectives, leaving little room for legislative dynamism to address foreign-state disinformation campaigns.

Instead of treating speech-related categories as “outcome determinative rules,” as Justice Breyer contends, such categories should be treated as “rules of thumb.”<sup>153</sup> Potential regulations that limit or deter foreign state-based falsity should be subject to proportionality review. Laws that regulate such (lower-value) speech from foreign state speakers “warrant[ ] neither near-automatic condemnation (as ‘strict scrutiny’ implies) nor near-automatic approval (as is implicit in ‘rational basis’ review).”<sup>154</sup> Furthermore, since the competition with falsehood under the marketplace theory can no longer be relied upon, a proportionality analysis should take greater account of the First Amendment’s democratic nature (i.e., its promotion of democratic deliberation) when interpreting speech-related regulations that address foreign-based false claims. If proportionality looks to the purposes of statutes and constitutional principles, then proportionality review should consider the state’s compelling interest in promoting a speech environment in which the citizenry receives necessary and accurate information it needs to fulfill its sovereign function.

---

<sup>153</sup> *Iancu v. Brunetti*, 588 U.S. 388, 402–03 (2019) (Breyer, J., concurring) (“After all, these rules are not absolute. The First Amendment is not the Tax Code. Indeed, even when we consider a regulation that is ostensibly ‘viewpoint discriminatory’ or that is subject to ‘strict scrutiny,’ we sometimes find the regulation to be constitutional after weighing the competing interests involved.”).

<sup>154</sup> *United States v. Alvarez*, 567 U.S. 709, 731 (2012) (Breyer, J., concurring).