FREE EXPRESSION RATIONALES AND THE PROBLEM OF DEEPFAKES WITHIN THE E.U. AND U.S. LEGAL SYSTEMS

Jared Schroeder†

Southern Methodist University

TABLE OF CONTENTS

ABSTRACT

Political actors are increasingly turning to deepfakes, realistic video clips that depict a person saying or doing something they never said or did, to mislead and manipulate public discourse. These video clips, which are becoming easier to create, are emerging as a global concern in online information environments that are already congested with false and misleading information. As nations seek to limit the influence of these fake videos, they are likely to encounter significant questions about freedom of expression. This article examines the legal rationales for free expression in the E.U. and U.S. legal systems, ultimately seeking to identify conceptual building blocks for how deepfakes that pose a threat to

† Jared Schroeder is an associate professor of journalism at Southern Methodist University.

democratic discourse can be limited without violating rights that are foundational in both regions.

INTRODUCTION

Deepfake enthusiasts have superimposed actor Nicholas Cage's likeness into countless movie scenes. His face replaces Amy Adams's in her role as Lois Lane in "Superman: Man of Steel" and it takes the place of Harrison Ford's character, Indiana Jones, in "Raiders of the Lost Ark."[1] Other deepfakes are more concerning. Buzzfeed published a fake video of President Barack Obama in 2018.[2] The clip was intended to alert people to misleading information by showing that such deceptive videos were possible.[3] The video, which was created using freely available software, featured believable footage of a former president calling Donald Trump a "dipshit" before later revealing actor and director Jordan Peele, like a puppeteer, voiced and created the president's actions in the clip.[4] In May 2019, someone manipulated video footage of U.S. Speaker of the House Nancy Pelosi to make it appear that her words were slurred and that she was not thinking clearly.[5] Trump shared the false video on Twitter and it received thousands of likes and retweets, despite widespread reports that it was a doctored clip.[6] While more doctored than deepfaked, the Pelosi clip added to growing concerns that the improving quality of deepfakes, along with the increasingly easy-to-use software, make these and other examples of intentionally misleading video depictions of public officials the heralds of what portends to be another disruptive development in a global, instantaneous information environment that is already

1. *See* Sam Haysom, *People are Using Face-Swapping Tech to Add Nicolas Cage to Random Movies and What is 2018*, MASHABLE (Jan. 31, 2018), https://mashable.com/2018/01/31/nicolas-cage-face-swapping-deepfakes/.

2. BuzzFeedVideo, *You Won't Believe What Obama Says in This Video!* 😉, YOUTUBE (Apr. 17, 2018) https://www.youtube.com/watch?v=cQ54GDm1eL0.

3. *See id.*; Craig Silverman, *How to Spot a Deepfake Like the Barak Obama-Jordan Peele Video*, BUZZFEED (Apr. 17, 2018), https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed.

4. *See id.*

5. *See* Drew Harwell, *Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media*, WASH. POST (May 24, 2019, 4:41 PM), https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/.

6. @realDonaldTrump, Twitter (May 23, 2019, 9:09 PM), https://twitter.com/realDonaldTrump/status/1131728912835383300. *See also* Makena Kelly, *Distorted Nancy Pelosi Videos Show Platforms Aren't Ready to Fight Dirty Campaign Tricks*, THE VERGE (May 24, 2019, 12:49 PM), https://www.theverge.com/2019/5/24/18637771/nancy-pelosi-congress-deepfake-video-facebook-twitter-youtube.

rife with intentionally false and misleading information.[7] In other words, deepfakes, video clips that depict people saying or doing something they never said or did, arrive as a fundamentally new challenge to truth and the flow of information in democratic society during a time when these foundational concerns are already being battered by a perfect storm of misinformation and disinformation.[8]

Of course, false information is nothing new. The environment in which information flows, however, as well as how individuals receive it, has shifted in fundamentally important ways in the networked era.[9] The channels of discourse have become more and more congested with intentionally false information.[10] So-called "fake news" reports have cascaded through virtual realms and across international borders to influence elections and contentious debates.[11] In particular, individuals and groups have worked to leverage the largely gatekeeperless, choice-rich, and global networked environment to create and share information that supports their beliefs and interests. Britain First, a far-right group that was banned from Facebook in April 2019, tweeted footage of Muslims "celebrating

7. *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1765–68 (2018) (explaining why deepfakes are "especially prone to going viral"); Paul Chadwick, *The Liar's Dividend, and Other Challenges of Deep-Fake News*, THE GUARDIAN (July 22, 2018, 2:00 PM), https://www.theguardian.com/commentisfree/2018/jul/22/deep-fake-news-donald-trump-vladimir-putin.

8. *See* John Villasenor, *Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth*, BROOKINGS (Feb. 14, 2019), https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/. *See also* Danielle Citron, *How Deepfakes Undermine Truth and Threaten Democracy*, TED (July 2019), https://www.ted.com/talks/danielle_citron_how_deepfakes_undermine_truth_and_threaten_democracy.

9. *See* SHERRY TURKLE, RECLAIMING CONVERSATION: THE POWER OF TALK IN A DIGITAL AGE 23 (2015); ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 18–19 (2000); CLAY SHIRKY, HERE COMES EVERYBODY: THE POWER OF ORGANIZING WITHOUT ORGANIZATIONS 22 (2008); MANUEL CASTELLS, NETWORKS OF OUTRAGE AND HOPE: SOCIAL MOVEMENTS IN THE INTERNET AGE 2–3 (2d ed. 2015).

10. *See* Erin Griffith, *Pro-Gun Russian Bots Flood Twitter After Parkland Shooting*, WIRED (Feb. 15, 2018, 2:00 PM), https://www.wired.com/story/pro-gun-russian-bots-flood-twitter-after-parkland-shooting/; Laura Silver, *Misinformation and Fears About its Impact are Pervasive in 11 Emerging Economies*, PEW RES. CTR. (May 13, 2019), https://www.pewresearch.org/fact-tank/2019/05/13/misinformation-and-fears-about-its-impact-are-pervasive-in-11-emerging-economies/; David M. J. Lazer, et al., *The Science of Fake News*, 359 SCI. 1094, 1095–96 (2018).

11. *See* Yuriy Gorodnichenko et al., *Social Media, Sentiment and Public Opinions: Evidence from #Brexit and #USElection* 28–29 (NBER Working Paper No. 24631, 2018), https://eml.berkeley.edu/~ygorodni/Brexit_Election.pdf.

the Paris terror attack in London" in 2017.[12] The only problem was the footage was of Pakistanis celebrating a World Cup win in cricket in 2009.[13] Similarly, American conspiracy theorist Alex Jones, whose content has been removed from Facebook, Twitter, and Spotify, has long spread rumors that the Sandy Hook school shooting in the U.S., which killed twenty children, was a hoax put on by those who seek stricter gun laws.[14] Despite the backlash these false messages often face, they spread six times faster than truthful information in online environments.[15] These falsities gain particular traction in heavily fragmented virtual communities, where individuals seldom encounter ideas that challenge their realities and where misleading information that supports dominant narratives is likely to be accepted and shared.[16] Legal rationales for free expression have struggled to adapt to these changes. As the information individuals use to construct their realities is increasingly digitalized, and as it shifts to virtual, rather than physical, forums, traditional legal assumptions about the flow of information have struggled.[17]

Furthermore, algorithms and bots are increasingly influencing the spread and prominence of true and false information across individuals' networks.[18] Massive corporations' proprietary algorithms are influencing

12. *See* Chris York, *Paul Golding, Britain First Leader, Posts Fake 'Paris Terror Attack Celebration Video'*, HUFFINGTON POST (April 21, 2017), https://www.huffingtonpost.co.uk/entry/paul-golding-britain-first-paris-attack_uk_58f9fbd2e4b00fa7de1363bc.

13. *See id.*

14. *See* Jack Nicas, *Alex Jones and Infowars Content is Removed From Apple, Facebook and YouTube*, N.Y. TIMES (Aug. 6, 2018), https://www.nytimes.com/2018/08/06/technology/infowars-alex-jones-apple-facebook-spotify.html.

15. *See* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCI. 1146, 1148 (2018).

16. *See* Itai Himelboim et al., *Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter*, 18 J. COMPUTER-MEDIATED COMM. 154, 156, 158, 171 (2013); 1 MANUEL CASTELLS, THE RISE OF THE NETWORK SOCIETY: THE INFORMATION AGE: ECONOMY, SOCIETY AND CULTURE 3–4 (2000); W. Lance Bennett & Shanto Iyengar, *A New Era of Minimal Effects? The Changing Foundations of Political Communication*, 58 J. COMM. 707, 720 (2008).

17. *See* CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 3–11 (2017); Jared Schroeder, *Marketplace Theory in the Age of AI Communicators*, 17 FIRST AMEND. L. REV. 22, 22–23 (2018).

18. *See, e.g.*, Khatya Chhor, *As French Media Went Dark, Bots and Far-Right Activists Drove #MacronLeaks*, FRANCE24 (Aug. 5, 2017), http://www.france24.com/en/20170508-french-media-blackout-bots-far-right-activists-wikileaks-pushed-macronleaks; Bence Kollanyi et al., *Bots and Automation Over Twitter During the U.S. Election* 1 (The Computational Propaganda Project Data Memo 2016.4, 2016), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2016/11/Data-Memo-US-Election.pdf; Jacob Shamsian, *There's a Bot on Tinder Trying To Influence Votes in the British Election*, BUS. INSIDER (June 8, 2017, 2:59 PM), http://www.businessinsider.com/united-kingdom-election-jeremy-corbyn-tinder-bot-labour-2017-6?r=UK&IR=T.

the information individuals encounter, often tailoring search outcomes and information suggestions to fit data about users' existing biases and the corporations' business interests.[19] To similar effect, bots are being used to push human communicators out of the information marketplace by producing massive amounts of information that can drown out human communicators and create a false sense that a truth has already coalesced around an idea when in reality it is just an army of bot accounts.[20] Deepfakes represent a particularly concerning addition to these emerging threats to democratic discourse because they incorporate traditionally credible audio and video into the types of false information that has already hindered the flow of information in the networked era.[21] They also strike at some of the greatest weaknesses within the dominant legal rationales for free expression in the E.U. and U.S. and raise questions about traditional arguments for safeguarding truth, falsity, and the flow of information.[22]

Absent legal precedents regarding deepfakes, within the E.U. or U.S. legal systems, this article seeks a potential pathway within the regions' free expression rationales for how deepfakes that damage democratic discourse can be limited. Jurists, in interpreting foundational freedoms in the E.U. and U.S., have rationalized safeguards for free expression in fundamentally different ways, thus laying substantially different groundworks for how to address the threats deepfakes pose to democratic discourse. At their cores, however, the systems share concerns about truth, falsity, and the flow of information. The U.S. system's rationales for safeguarding these matters has remained closely associated with libertarian interpretations of Enlightenment-founded ideas about the nature of truth and the rationality of the individual.[23] As Enlightenment thinker John Milton

19. *See generally* Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY (Tarleton Gillespie, Pablo J. Boczkowski, & Kristen A. Foot eds., 2014) (discussing algorithms and their impact on public discourse).

20. *See, e.g.* Molly K. McKew, *How Twitter Bots and Trump Fans Made #ReleaseTheMemo Go Viral*, POLITICO (Feb. 4, 2018), https://www.politico.com/magazine/story/2018/02/04/trump-twitter-russians-release-the-memo-216935; Chhor, *supra* note 18.

21. Robert Chesney & Danielle Keats Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882, 884 (2019).

22. *See* Chesney & Citron, *supra* note 21, at 887–88.

23. *See* Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 12–14 (1984); Fred S. Siebert, *The Libertarian Theory of the Press*, *in* FOUR THEORIES OF THE PRESS: THE AUTHORITARIAN, LIBERTARIAN, SOCIAL RESPONSIBILITY AND SOVIET COMMUNIST CONCEPTS OF WHAT THE PRESS SHOULD BE AND DO 50–57 (Fred S. Siebert, Theodore Peterson & Wilbur Schramm eds., 1956); Jared Schroeder, *Toward a Discursive*

explained, "[w]here there is much desire to learn, there of necessity will be much arguing, much writing, many opinions; for opinion in good men is but knowledge in the making."[24] Centuries later, in a landmark defamation case, U.S. Supreme Court Justice Louis Powell concluded, "there is no such thing as a false idea."[25] Thus, to avoid the persistent problem of government intervention in the discovery and acceptance of truth amidst a myriad of ideas, jurists have adopted the marketplace of ideas approach, which generally removes the government from limiting expression and has dominated U.S. jurisprudence for more than a century.[26]

Jurists in the E.U. have communicated similar concerns for the flow of information and the development of truth.[27] They have done so, however, by allowing the government to take more of a custodial role. In *Hertel v. Switzerland*, the European Court of Human Rights (ECtHR), explained, "[f]reedom of expression constitutes one of the essential foundations of a democratic society and one of the basic conditions for its progress and for each individual's self-fulfillment."[28] The Court continued, however, by recognizing there are limitations, "such are the demands of pluralism, tolerance and broadmindedness without which there is no 'democratic society.'"[29] Thus, ECtHR jurists weigh free expression rights with concerns for the public good and the quality of information available to citizens, often considering if the information the government seeks to restrict is "necessary in a democratic society."[30] Ultimately, the E.U. approach is markedly different, despite the common goals and

Marketplace of Ideas: Reimagining the Marketplace Metaphor in the Era of Social Media, Fake News, and Artificial Intelligence, 52 FIRST AMEND. STUD. 38, 39–40 (2018).

24. JOHN MILTON, AREOPAGITICA AND OF EDUCATION 45 (George Holland Sabine ed., 1951).

25. *See* Gertz v. Robert Welch, Inc., 418 U.S. 323, 339 (1974).

26. *See* W. Wat Hopkins, *The Supreme Court Defines the Marketplace of Ideas*, 73 JOURNALISM & MASS COMM. Q. 40, 40 (1996); Philip M. Napoli, *The Marketplace of Ideas Metaphor In Communications Regulation*, 49 J. OF COMM. 151, 151 (1999).

27. *See, e.g.*, Hertel v. Switzerland, no. 25181/94, § 46, ECHR 1998-VI, http://hudoc.echr.coe.int/eng?i=001-59366; Kobenter and Standard Verlags GmbH v. Austria, no. 60899/00, § 29, 2 Nov. 2006, http://hudoc.echr.coe.int/eng?i=001-77786; Instytut Ekonomichenykh Reform, TOV v. Ukraine, no. 61561/08, § 43, 2 June 2016, http://hudoc.echr.coe.int/eng?i=001-163354.

28. *Hertel*, no. 25181/94, § 46(i).

29. *Id.*

30. *Id.* § 31. The "necessary in a democratic society" criteria comes from Article 10 § 2. *See* COUNCIL OF EUR., EUROPEAN CONVENTION ON HUMAN RIGHTS 12 (1950), https://www.echr.coe.int/Documents/Convention_ENG.pdf [hereinafter ECHR]. Using the passage, the ECtHR often sets out to weigh expression's value to society against other concerns. *See, e.g.*, Fatih Tas v. Turkey (no. 2), no. 6813/09, § 13, 10 Oct. 2017, http://hudoc.echr.coe.int/eng?i=001-177868; Mouvement raëlien suisse v. Switzerland [GC], no. 16354/06, § 51–52, 13 July 2012, http://hudoc.echr.coe.int/eng?i=001-112165.

similar influences, than the U.S.'s, making the two ideal for examining the growing problem of deepfakes within the flow of information in democratic society.

Part I of this article defines deepfakes, examines the threats they pose to discourse in democratic society, and surveys existing laws and tech-firm policies regarding deepfakes. Part II analyzes the U.S. Supreme Court's use of the marketplace rationale for free expression, as well as criticisms of the theory and the impact of the Court's decision in *U.S. v. Alvarez* and other decisions that have dealt with falsity. Part III examines truth rationales in the E.U. legal system, focusing on the expectation that information be a public good, and the particular role the government plays as a custodian of discourse. The article concludes by drawing out conceptual building blocks regarding truth and falsity safeguards to propose an avenue through which regulators can limit deepfakes without undercutting safeguards for freedom of expression.

I. A DEEPER LOOK AT DEEPFAKES

Aside from Nicholas Cage video clips, early deepfake creators focused on placing celebrities' faces on porn stars' bodies to create realistic pornographic videos.[31] Such appropriations of individuals' identities in pornography led to the term "deepfakes," which benefit from increasingly nuanced neural networks that take images and clips of video and sound and replicate them, allowing creators to then superimpose the subject into an existing work.[32] This section defines deepfakes and deepfake technology. It also considers the potential threats deepfakes pose to democratic discourse and the measures that have been taken to curb their impact.

A. Fake Tools

A deepfake is a video clip that is "constructed to make a person appear to say or do something that they never said or did."[33] Such twenty-first-century, digital puppeteering portends to have a powerful impact on democratic discourse because, as one group of scholars concluded, despite skepticism about text and photography, "[w]e tend to place . . . trust in the voices we know and the videos we watch."[34] Deepfakes are emerging in discourse, and becoming more believable, as a result of

31. *See* Douglas Harris, *Deepfakes: False Pornography is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 101 (2019).

32. *See* Marc Jonathan Blitz*, Lies, Line Drawing, and (Deep) Fake News*, 71 OKLA. L. REV. 59, 61–62 (2018); Chesney & Citron, *supra* note 7, at 1758.

33. Villasenor, *supra* note 8.

34. *See* Jan Kietzmann et al., *Deepfakes: Trick or Treat?*, 63 BUS. HORIZONS 135, 136 (2020).

advancements in machine learning, particularly in deep neural networks, which can be trained by studying a large sample of images.[35] By studying many images of a celebrity, for example, the network can be trained, like a human artist who replicates the Mona Lisa, to identify characteristics and replicated them. While the Defense Advanced Research Projects Agency (DARPA), the research arm of the U.S. military, is spending millions to develop tools that detect deepfakes, their technologists admit that if a creator uses Generative Adversarial Networks (GANs), their work will in most cases be able to circumvent any watermarking or detection software.[36] By using GANs, a deepfake creator pairs two neural networks together.[37] One network is used to create the deepfake, while the other, the adversarial network, assesses the quality of the clip.[38] The two networks interact to refine the deepfake until it is becomes extremely realistic. Legal scholars Robert Chesney and Danielle Citron found, "[g]rowing sophistication of the GAN approach is sure to lead to the production of increasingly convincing deepfakes."[39]

Deepfake creation software is becoming easier to use for people who do not have advanced programming skills. FakeApp 2.0, one of the free programs, is available via a link in Reddit, which is where the celebrity pornography fakes appeared in late 2017.[40] Countless video tutorials on YouTube demonstrate how to use the program.[41] The software automates dropping faces from video clips and, once supplied with content, it can "train" itself in creating the deepfakes using the materials the user provides.[42] These capabilities raise important questions about the nature of deepfakes. They are not inherently bad for the flow of information and

35. *See* Chesney & Citron, *supra* note 7, at 1759.

36. *See* Will Knight, *The US Military is Funding an Effort to Catch Deepfakes and Other AI Trickery*, MIT TECH. REV. (May 23, 2018), https://www.technologyreview.com/s/611146/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/; William Turton & Matthew Justus, *"Deepfake" Videos Like that Gal Gadot Porn are Only Getting More Convincing—and More Dangerous*, VICE (Aug. 27, 2018, 2:00 PM), https://news.vice.com/en_us/article/qvm97q/deepfake-videos-like-that-gal-gadot-porn-are-only-getting-more-convincing-and-more-dangerous.

37. *See* Chesney & Citron, *supra* note 7, at 1760.

38. *See id.*

39. *Id.*

40. *See generally FakeApp Download Links and How-To Guide,* REDDIT, https://www.reddit.com/r/GifFakes/comments/7xv91x/fakeapp_download_links_and_howto_guide/ (last visited June 2, 2020). Reddit removed links to the deep fake pornography from its pages.

41. *See, e.g.*, Minipa, *FakeApp2.2.0 Tutorial – Installation and Usage (How to Put Nicholas Cage's Face on Various People)*, YOUTUBE (Mar. 7, 2018), https://www.youtube.com/watch?v=DVpWzVRjVRY.

42. *See* Chesney & Citron, *supra* note 7, at 1763.

the development of truth.[43] Deepfakes that jokingly place different actors into famous films are more like video memes than intentionally deceptive statements of truthful information. The humor is in the parody, which the creator *wants* the audience to notice. In this and other instances, deepfake technologies have educational, artistic, and social commentary potential.[44] As with most tools, the problems arise with how people use them. Deepfakes that include a world leader saying or doing something they never said or did for the purpose of helping or harming their credibility are, by their nature, different and more concerning than an accidentally communicated false fact or an unpopular or dangerous idea that emerges as part of a discussion. These types of deepfakes are instead a potentially more concerning extension of "fake news," which, by its nature includes the intent to mislead recipients.[45]

B. Deep Threats

Deepfakes raise a myriad of legal concerns. They can be used to harm or exploit individuals' reputations and profit from their likenesses without their permission, as has been the case when celebrities' images have been imported into pornographic videos.[46] They also raise significant copyright questions.[47] This article, however, focuses on societal, rather than personal or financial, harms as they can manifest in two ways: (1) The creation of widely believable false statements or facts that mislead and distort the search for consensus or shared truth in society and (2) the unique power of video clips to provide credibility to false messages and to enrage and enflame citizens into action.

The Brexit vote and U.S. presidential election in 2016, though not influenced by deepfakes, illustrated the growing power of intentionally false information, when distributed through people's virtual, generally ideologically homogenous, networks, to outstrip truthful information in spread and permanence. The Brexit campaign was characterized by multiple instances of intentionally false information. The Leave.EU party released a video that purported to show migrants attacking women in London.[48] The instances were staged, but still received hundreds of thousands

43. *See id.* at 1769 (discussing the educational value of deepfakes).

44. *See id.* at 1770–71.

45. *See* Edson C. Tandoc Jr. et al., *Defining "Fake News": A Typology of Scholarly Definitions*, 6 DIGITAL JOURNALISM 1, 2 (2017).

46. *See* Harris, *supra* note 31, at 101.

47. *See id.* at 107–11 (discussing copyright infringement and fair use of deepfakes).

48. *See* Benjamin Kentish, *Pro-Brexit Leave.EU Group Accused of Faking Videos and Forging Images of Migrants Committing Crimes*, INDEP. (Apr. 16, 2019),

of interactions on Facebook.[49] Rumors that the U.K. would save £350 million per week by leaving the E.U. and that Turkey would soon join the Union, which would lead to more immigration into the U.K., also flowed through online spaces, despite being false and widely questioned by mainstream media.[50] Similarly, during the U.S. presidential election, fake reports that Pope Francis had endorsed Trump flowed through Facebook, reaching nearly one million engagements before Election Day.[51] At the same time, a false report that President Obama had banned the "Pledge of Allegiance" in schools circulated on Facebook, where it received 2.1 million shares.[52] Soon a Facebook group named "For America" created a "save the pledge" campaign, which received thousands of shares and likes on Facebook as citizens took a patriotic stand against the president's attack on democracy—which never actually happened.[53]

While false information has always existed, the emergence of networked spaces has removed most of the factors that limited false information's ability to flow through society. Those who seek to influence discourse about major issues have found success online because they do not have to work through traditional media gatekeepers to spread their ideas.[54] Thus, misleading reports, even if widely debunked in mainstream media, can thrive if they fit within the existing narratives found in fragmented and often echo-chamber-like virtual communities.[55] In India, for

https://www.independent.co.uk/news/uk/politics/brexit-leave-eu-faking-forging-videos-images-illegal-migrants-violent-crime-aaron-banks-a8873461.html.

49. *See Revealed: How Leave.EU Faked Migrant Footage*, 4 NEWS (Apr. 16, 2019), https://www.channel4.com/news/revealed-how-leave-eu-faked-migrant-footage.

50. *See* Shehab Khan, *Final Say: The Misinformation that was Told About Brexit During and After the Referendum*, INDEP. (July 28, 2018), https://www.independent.co.uk/news/uk/politics/final-say-brexit-referendum-lies-boris-johnson-leave-campaign-remain-a8466751.html.

51. *See* Hannah Richie, *Read All About It: The Biggest Fake News Stories of 2016*, CNBC (Dec. 30, 2016, 2:04 AM), https://www.cnbc.com/2016/12/30/read-all-about-it-the-biggest-fake-news-stories-of-2016.html.

52. *See* Craig Silverman, *Here Are 50 of the Biggest Fake News Hits on Facebook From 2016*, BUZZFEED (Dec. 30, 2016, 9:12 AM), https://www.buzzfeednews.com/article/craigsilverman/top-fake-news-of-2016.

53. The false story originated on ABC.com.com, which is no longer active. *See* Silverman, *supra* note 51; *Obama Signs Executive Order Banning the Pledge of Allegiance in Schools Nationwide,* SNOPES (Aug. 16, 2016), https://www.snopes.com/fact-check/pledge-of-allegiance-ban/. While the "Save the Pledge" campaign on ForAmerica.com is also no longer available, see https://foramerica.org/ for examples of false and misleading information.

54. *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 15–16 (2018); CASTELLS, *supra* note 9, at 2–7.

example, a 2018 report on false information found it is most effective if it supports Hindu superiority, national progress, or the prowess and ability of Prime Minister Narendra Modi.[56] Truthfulness was not among the features of effective messages. This emerging reality means truth and falsity can vary from community to community, and their competing truths may never meet. In some online communities, it is a certainty that President Obama was born in Kenya and faked his birth certificate.[57] In many others, however, the birther movement has been debunked.[58] Both groups claim hold of absolute truth on the matter.[59] Though the narratives might change from nation to nation, partisans are establishing a playbook for effectively spreading false information via virtual communities. It is concerning that deepfakes may soon become a crucial part of such playbooks, where certain actors can use deepfakes to manipulate people's fears and reinforce their biases to achieve certain aims.

In particular, deepfakes carry unique, and particularly concerning, attributes into such a playbook. Text-based reports are second-hand accounts.[60] They are told through the eyes of the author.[61] Videos carry added believability because they "allow people to become firsthand witnesses of an event, sparing them the need to decide whether to trust someone else's account . . . ."[62] Thus, in a media environment that is fraught with intentionally fabricated and misleading information, deepfakes represent a more powerful tool for those who seek to weave disinformation into democratic discourse. A 2019 European Parliament report on disinformation and political propaganda concluded "'deep fakes' present an even more difficult problem than manipulated textual content, as they are

55. *See* SUNSTEIN, *supra* note 17, at 9–11 (discussing the consequences of such echochambers). For a broader discussion, see JARED SCHROEDER, THE PRESS CLAUSE AND DIGITAL TECHNOLOGY'S FOURTH WAVE: MEDIA LAW AND THE SYMBIOTIC WEB 62–65 (2018).

56. *See* SANTANU CHAKRABARTI ET AL., DUTY, IDENTITY, CREDIBILITY: FAKE NEWS AND THE ORDINARY CITIZEN IN INDIA 72 (2018), http://downloads.bbc.co.uk/mediacentre/duty-identity-credibility.pdf.

57. *See, e.g.*, Cleve R. Wootson Jr., *Joe Arpaio is Back—And Brought His Undying Obama Birther Theory With Him*, WASH. POST (Jan. 11, 2018, 11:36 AM), https://www.washingtonpost.com/news/powerpost/wp/2018/01/11/joe-arpaio-is-back-and-brought-his-undying-obama-birther-theory-with-him/.

58. *See, e.g.*, David Mikkelson, *Is Barack Obama's Birth Certificate Fake?*, SNOPES (Aug. 27, 2011), https://www.snopes.com/fact-check/birth-certificate/.

59. *See, e.g.*, Wootson, *supra* note 57 (discussing politician Joe Arpaio's statement that he has evidence former president Obama's birth certificate is fake in 2018).

60. *See* Robert Chesney & Danielle Citron, *Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics,* FOREIGN AFF., (Jan./Feb. 2019), https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war.

61. *See id*.

62. *See id*.

more likely to trigger strong emotions . . . ."[63] This more visceral connection to the information, combined with findings that misinformation that involves videos is more likely to be shared and remembered, raises grave concerns about the type of information citizens will have access to when seeking to make decisions and govern themselves.[64]

C. Laws and Corporate Policies

Lawmakers and social media corporations have taken notice of the threats deepfakes pose to democratic discourse. While future efforts to address the threats deepfakes pose will be shaped by the free expression rationales that are analyzed in this article, it is also important to consider the steps that have already been taken. No federal laws, in the United States or Europe, exist. Two deepfake-related laws have been proposed in the U.S. Congress, both of which have failed to gain traction.[65] The Deepfakes Accountability Act, which was filed in June 2019, was created to address "the spread of disinformation through restrictions on deep-fake video alteration technology."[66] The bill would criminalize communicating an intentionally false "personation record," unless the fake nature of the message is conspicuously disclosed.[67] The bill includes an exemption for video clips that a rational person would know is faked, which would seemingly safeguard Nicholas Cage parody and satire clips.[68] The bill also allows the government to create such clips for public safety and national security purposes.[69]

While this bill collects dust in committee, Texas and California have passed laws that address political deepfakes.[70] Texas's law, which went

63. JUDIT BAYER ET AL., DISINFORMATION AND PROPAGANDA—IMPACT ON THE FUNCTIONING OF THE RULE OF LAW IN THE EU AND ITS MEMBER STATES 118 (2019), http://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf.

64. *See* JOSHUA A. TUCKER ET AL., SOCIAL MEDIA, POLITICAL POLARIZATION, AND POLITICAL DISINFORMATION: A REVIEW OF THE SCIENTIFIC LITERATURE 48 (2018), https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf.

65. *See* Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. (2019); Malicious Deep Fake Prohibition Act of 2018, S. 3805, 115th Cong. (2018).

66. *See* H.R. 3230.

67. *See id*. § 1041(f).

68. *See id.* § 1041(j)(1)(E).

69. *See id.* § 1041(j)(1)(F).

70. *See* TEX. ELEC. CODE ANN. § 255.004(d)–(e) (West, Westlaw through 2019 Reg. Sess.); CAL. ELEC. CODE § 20010 (Deering, LEXIS through 2020 Reg. Sess.). *See also* CAL. CIV. CODE § 1708.86 (Deering, LEXIS through 2020 Reg. Sess.) (addressing revenge pornography, including provisions that would criminalize deepfake pornography). Similarly,

into effect in September 2019, amends the state's election code to criminalize deepfakes that are created "to injure a candidate or influence the result of an election."[71] Importantly, and likely problematically, the law implicates both the creator and the publisher of the video, likely creating liability for forums such as YouTube and Facebook.[72] Such broad language, despite the amendment being fewer than one hundred words, will likely raise legal concerns regarding free expression precedents within the U.S., as well as conflicts with section 230 of the Communications Decency Act (CDA), which generally immunizes online forums from liability for third-party content.[73] California's law is far more expansive. The law provides a vehicle for recourse for political candidates whose campaigns have been damaged by deepfakes. It defines a deepfake as an "image or audio or video recording [that] would falsely appear to a reasonable person to be authentic."[74] The California law explicitly exempts traditional news media, as well as satire and parody.[75] It also states the law does not negate CDA 230 immunities.[76] While these exemptions do more to clarify the law's scope, they also create significant grey areas, as those accused of creating and disseminating deepfakes can seek refuge in claiming they are a news outlet or their work is parodic.[77] Both states' deepfakes policies have yet to be tested in the courts. As discussed in Part III, however, the U.S.'s free expression rationales mean such attempts to limit deepfakes, particularly within the political realm, are unlikely to succeed.[78] The laws, as with any federal effort by the U.S. or E.U. member states, will likely struggle because they fail to account for the difficulty in identifying the creators of such clips and the jurisdictional limits that keep out-of-nation actors from being held liable.

Social media firms have the power to address both of these shortcomings by constructing forum-specific policies regarding deepfakes. Facebook and Twitter have sought user, as well as expert, input in crafting policies that address deepfakes.[79] In January 2020, Facebook unveiled

Virginia revised its state revenge pornography law to include deepfakes. *See* VA. CODE ANN. § 18.2-386.2(a) (LEXIS through 2019 Reg. Sess.).

71. *See* TEX. ELEC. CODE ANN. § 255.004(d).

72. *See id.* § 255.004(d)(1)–(2).

73. *See* 47 U.S.C. § 230(c)(1) (2020).

74. *See* CAL. ELEC. CODE § 20010(e)(1).

75. *See id.* § 20010(d)(2), (5).

76. *Id.* § 20010(d)(1).

77. *See id.* § 20010(d)(5) ("This section does not apply to materially deceptive audio or visual media that constitutes satire or parody.").

78. See *infra* Part III.

79. *See* Monika Bickert, *Enforcing Against Manipulated Media*, FACEBOOK (Jan. 6, 2020), https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/; Del Harvey, *Help*

its first policy regarding deepfakes, explaining it will block content that is manipulated in ways that would deceive and average person "into thinking that a subject of the video said words that they did not actually say" and the video is the product of artificial intelligence.[80] As with the state laws, Facebook's policy, as well as those that Twitter is constructing, have not been tested regarding their effectiveness. The corporations' policies, however, do not face the type of constitutional scrutiny that Texas's, California's, and any federal effort in the U.S. will encounter. Similarly, they circumvent E.U.-wide and member-state policies that might emerge in coming years.[81]

II. THE MARKETPLACE, FALSITY, AND THE U.S. SUPREME COURT

Concerns about truth and falsity, as well as how the Supreme Court has communicated how it understands freedom of expression, have historically been wrapped up in the marketplace of ideas metaphor.[82] The metaphor has been used in nearly every type of free expression case and by justices from a variety of legal philosophies.[83] Marketplace theory has become justices' dominant tool for communicating how they rationalize free expression, and therefore the primary concept in the U.S. legal regime that must be contended with when considering limitations on deepfakes that pose a threat to democratic discourse.[84] Despite such a prominent role in the U.S. free-expression system, justices have never explicitly defined what they mean when they use the metaphor.[85] Justice Oliver Wendell Holmes's initial use of the marketplace concept, in his influential dissent in the *Abrams v. U.S.* sedition case in 1919, did not include any citations or references regarding the foundational assumptions of

Us Shape Our Approach to Synthetic and Manipulated Media, TWITTER (Nov. 11, 2019), https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback.html. Bickert is Facebook's vice president for global policy management. Harvey is Twitter's vice president for trust and safety.

80. *See* Bickert, *supra* note 79.

81. *See* Chris Meserole & Alina Polyakova, *Order from Chaos: The West is Ill-Prepared for the Wave of "Deep Fakes" That Artificial Intelligence Could Unleash*, BROOKINGS (May 25, 2018), https://www.brookings.edu/blog/order-from-chaos/2018/05/25/the-west-is-ill-prepared-for-the-wave-of-deep-fakes-that-artificial-intelligence-could-unleash/.

82. *See* C. EDWIN BAKER, HUMAN LIBERTY AND FREEDOM OF SPEECH 3–5 (1989); MATTHEW D. BUNKER, CRITIQUING FREE SPEECH: FIRST AMENDMENT THEORY AND THE CHALLENGE OF INTERDISCIPLINARY 11–14 (2001). *See generally* Hopkins, *supra* note 26 (discussing the Supreme Court's use of the marketplace of ideas model in the resolution of free-expression cases).

83. *See* Hopkins, *supra* note 26, at 41–43.

84. *See id.* at 40.

85. *Id.*

what he termed "the theory of our Constitution."[86] In what was the first Supreme Court opinion to advocate for free expression, he explained, "the ultimate good desired is better reached by free trade in ideas—that the best test of truth is the power of the thought to get itself accepted in the competition of the market . . . ."[87]

Over time, primarily starting in the 1960s, justices fused Enlightenment-funded rationales to the metaphor, contending truth will generally succeed and falsity will fail when the government has little or no role in the exchange of ideas.[88] Such a conclusion finds support in the similarities between the ideas Justice Holmes communicated in *Abrams* and Enlightenment thinker John Milton's conclusion in *Areopagitica* that "Truth be in the field, we do injuriously, by licensing and prohibiting to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the wors [sic], in a free and open encounter."[89] It also finds support in the associations between those who drafted the nation's founding documents, including the First Amendment, and Europe's Enlightenment. Thomas Jefferson, John Adams, Benjamin Franklin, and James Madison, who drafted the First Amendment, were all influenced by Enlightenment thinkers, such as Milton, John Locke, Adam Smith, and David Hume.[90] Of course, Enlightenment thought carries certain assumptions about the nature of truth, the rationality of individuals, and the purpose of democratic society.[91] Enlightenment thinkers generally understood truth as being universal and the same for everyone, citizens as being primarily rational and capable of making sense of the information they encounter, and society as something that should be structured to benefit such rational individuals.[92] These assumptions bear importance to the U.S.'s ability to

86. *See* Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

87. *See id.*

88. *See* Jared Schroeder, *Shifting the Metaphor: Examining Discursive Influences of the Supreme Court's Use of the Marketplace Metaphor in Twenty-First-Century Free Expression Cases*, 21 COMM. L. & POL'Y 383, 403 (2016). *See* also Siebert, *supra* note 23, at 43; ROBERT A. FERGUSON, THE AMERICAN ENLIGHTENMENT 1750–1820 28 (1994).

89. *See* MILTON, *supra* note 24, at 50.

90. *See* Roy Branson, *James Madison and the Scottish Enlightenment*, 40 J. HIST. IDEAS 235 (1979) (paralleling the ideas in Madison's thinking to the writing of David Hume, Adam Smith, John Millar, and Adam Ferguson to suggest they were a significant influence on Madison's thought); DARREN STALOFF, HAMILTON, ADAMS, JEFFERSON: THE POLITICS OF ENLIGHTENMENT AND THE AMERICAN FOUNDING 3–4 (2005). Jefferson's "life, liberty, and the pursuit of happiness" bears significant overlap with John Locke's "life, liberty, and estate" from his Second Treatise of Government. *See* Jared Schroeder, *Shifting the Metaphor: Examining Discursive Influences of the Supreme Court's Use of the Marketplace Metaphor in Twenty-First-Century Free Expression Cases*, 21 Comm. L. & Pol'y 383, 391 (2016).

91. *See* Siebert, *supra* note 23, at 40–41.

92. *See id*.

respond to the threats deepfakes pose to democratic society. In critiquing the Enlightenment's truth, rationality, and society assumptions, historian David Hollinger explained that Enlightenment thought, "blinded us to the uncertainties of knowledge by promoting an ideal of absolute scientific certainty" and "led us to suppose that all people are pretty much alike, thus blinding us to diversity."[93] These assumptions have been wedded to the marketplace approach and have become the foundations for how free expression has been rationalized in the U.S. They are also the basis for substantial criticism of the metaphor and its use.

A. The Marketplace's Effect on U.S. Free Expression

Marketplace assumptions have had a profound, instrumental influence on the Court's freedom-of-expression rationales. Of particular concern regarding deepfakes, justices have generally contended that more information, regardless of its impact on democratic society, is better than limitations on the flow of information.[94] Such an assumption is in some ways contradictory to European approaches, which have eked out a place for the government to act as a custodian of the flow of information.[95] The U.S. Supreme Court has seldom explicitly associated the marketplace metaphor with the Enlightenment-related assumptions with which it has become wedded. Chief Justice Rehnquist, in his dissent in the *Central Hudson Gas & Electric Corp. v. Public Service Commission* commercial-speech case, connected the metaphor to Milton and Adam Smith.[96] He contended, "from the Court's frequent reference to the 'marketplace of ideas,' which was deemed analogous to the commercial market in which a laissez-faire policy would lead to optimum economic decisionmaking under the guidance of the 'invisible hand.'"[97] Chief Justices Rehnquist was skeptical that applying the metaphor to advertisements, especially in

93. David A. Hollinger, *The Enlightenment and the Genealogy of Cultural Conflict in the United States*, *in* WHAT'S LEFT OF ENLIGHTENMENT: A POSTMODERN QUESTION 8–9 (Keith Michael Baker & Peter Hanns Reill eds., 2001).

94. *See* Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring); Virginia v. Black, 538 U.S. 343, 358 (2003) (first citing Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting); and then citing Texas v. Johnson, 491 U.S. 397, 414 (1989)); N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964); *Johnson*, 491 U.S. at 419–20.

95. Article 10 § 2 of the European Convention on Human Rights provides substantial support for such an approach. *See* ECHR, *supra* note 30, at 12. *See also* Animal Defenders International v. United Kingdom [GC], no. 48876/08, §§ 116–17, ECHR 2013 (extracts), http://hudoc.echr.coe.int/fre?i=001-119244; discussion *infra* Subsection X.

96. *See* 447 U.S. 557, 592 (1980) (Rehnquist, J., dissenting).

97. *Id.* (citing ADAM SMITH, THE WEALTH OF NATIONS (1776)).

the name of the marketplace of ideas, was what the authors of the First Amendment intended.[98]

In many instances, however, the Enlightenment assumptions, like puppeteers, remain in the background but are pulling the strings regarding the Court's free-expression rationales. Perhaps most prominently, the Court concluded in *New York Times v. Sullivan* in 1964 that "debate on public issues should be uninhibited, robust, and wide-open, and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials."[99] While the decision does not explicitly use the word "marketplace," justices referred to the metaphor in the passage that preceded its "uninhibited, robust, and wide-open" conclusion.[100] The Court explained the First Amendment was "fashioned to assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people."[101] Such an emphasis on more information, rather than less, was most famously outlined in Justice Brandeis's concurring opinion in *Whitney v. California*.[102] Justice Holmes, whose marketplace dissent was joined by Justice Brandeis in *Abrams* in 1919, signed onto Justice Brandeis's reasoning in *Whitney* in 1927.[103] Justice Brandeis explained, "[i]f there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence."[104]

Such arguments for more information call on Enlightenment assumptions about truth and rationality that are at the heart of the marketplace. It can only be contended that more information, rather than less, is beneficial to democratic society if it is assumed citizens are capable of processing information and separating truth from falsity. These assumptions are found in many of the cases in which justices explicitly discuss the marketplace of ideas. In *Lamont v. Postmaster General*, decided a year after *Sullivan*, Justice Brennan communicated concern for the well-being of the marketplace if Postal Service guidelines limited the ideas citizens encountered.[105] He contended, "[t]he dissemination of ideas can

98. *See id.* at 592–93.

99. *Sullivan*, 376 U.S. at 270 (first citing Terminiello v. Chicago, 337 U.S. 1, 4 (1949); and then citing De Jonge v. Oregon, 299 U.S. 353, 365 (1937)).

100. *See id.*

101. *Id*. at 269 (citing Roth v. United States, 354 U.S. 476, 484 (1957)).

102. 274 U.S. 357, 375–78 (1927) (Brandeis, J., concurring).

103. *See* Abrams v. United States, 250 U.S. 616, 631 (1919) (Holmes, J., dissenting); *Whitney*, 274 U.S. at 380 (Brandeis, J., concurring).

104. *Whitney*, 274 U.S. at 377.

105. *See* 381 U.S. 301, 308 (1965).

accomplish nothing if otherwise willing addressees are not free to receive and consider them. It would be a barren marketplace of ideas that had only sellers and no buyers."[106] Justices conveyed similar concerns in *Simon & Schuster, Inc. v. Members of New York State Crime Victims Board* in 1991.[107] The Court struck down the "Son of Sam" laws, which limited criminals' abilities to profit from book deals that outlined their crimes.[108] The Court concluded the law "raises the specter that the government may effectively drive certain ideas or viewpoints from the marketplace. The First Amendment presumptively places this sort of discrimination beyond the power of the government."[109] In *United States v. Rumely*, a 1953 case that involved an author who refused to reveal to Congress who purchased his controversial political books, Justice Douglas communicated similar concerns.[110] He contended, "this publisher bids for the minds of men in the market place of ideas. The aim of the historic struggle for a free press was 'to establish and preserve the right of the English people to full information in respect of the doings or misdoings of their government.'"[111] In each of these cases, justices called upon the marketplace metaphor in rationalizing why the laws conflicted with the First Amendment.[112] The laws, justices concluded, limited the flow of information, thus limiting rational individuals' pursuits for truth within a relatively unhindered marketplace of ideas.[113]

Of course, justices have not always been of one mind regarding how to apply the marketplace metaphor. In the Court's deeply divided decision in *Citizens United v. FEC,* Justice Kennedy, in writing for the Court, contended a federal law that placed limits on electioneering communications would harm the marketplace by keeping some ideas from entering.[114] Justice Stevens, in his dissent, contended the law was intended to protect the marketplace.[115] He explained the law "reflects a concern to *facilitate* First Amendment values by preserving some breathing room around the electoral 'marketplace' of ideas, the marketplace in which the actual people of this Nation determine how they will govern

106. *Id.*

107. *See* 502 U.S. 105, 115 (1991).

108. *See id.* at 108, 123.

109. *Id.* at 116.

110. *See* 345 U.S. 41, 42, 56–57 (1953).

111. *Id.* at 56 (quoting Grosjean v. Am. Press Co., 297 U.S. 233, 247 (1936)).

112. *See id.*; *Simon & Schuster, Inc.*, 502 U.S. at 116; Lamont v. Postmaster Gen., 381 U.S. 301, 308 (1965).

113. *See id.*

114. *See* 558 U.S. 310, 335–36.

115. *See id.* at 473 (Stevens, J., dissenting).

themselves."[116] A similar dichotomy arose in the *Virginia State Board of Pharmacy v. Virginia Citizens Consumer Council* commercial speech case in 1976.[117] The Court, in striking down the commonwealth's ban on certain types of advertising, concluded the "relationship of speech to the marketplace of products or of services does not make it valueless in the marketplace of ideas."[118] Chief Justice Rehnquist found the Court's use of the marketplace metaphor misguided.[119] He explained, the differences "between a seller hawking his wares and a buyer seeking to strike a bargain to the same plane as has been previously reserved for the free marketplace of ideas, are far reaching indeed."[120] Similarly, in *First National Bank v. Bellotti* in 1978, the Court struck down a Massachusetts law that limited corporations' participation in political debates about issues that were not directly related to their interests.[121] The Court reasoned the law limited a source of information from contributing ideas that could be valuable to public discourse.[122] Justice Powell, in writing for the Court, explained the bank had a "role in affording the public access to discussion, debate, and the dissemination of information and ideas."[123] Justice White dissented, reasoning the corporations' participation in discourse could be "seriously threatening [to] the role of the First Amendment as a guarantor of a free marketplace of ideas."[124] In each instance, justices interpreted the marketplace metaphor differently. Though *protecting* the marketplace found itself in the dissent in each example, justices still constructed rationales for why, in some instances, the marketplace can be shielded from certain types of information.

B. Cracks in the Marketplace's Foundation

Many legal scholars have not shared the Supreme Court's generally steadfast trust in the marketplace approach. In particular, scholars have questioned the validity of the theory's Enlightenment-based assumptions regarding the nature of truth, the rationality of citizens, and the structure of society.[125] First Amendment scholar C. Edwin Baker built a long list

116. *Id.*
117. *See* 425 U.S. 748 (1976).
118. *See id.* at 749, 760, 773.
119. *See id.* at 781 (Rehnquist, J., dissenting).
120. *Id.*
121. *See* 435 U.S. 765, 767–68, 795 (1978).
122. *See id.* at 783.
123. *See id.*
124. *Id.* at 810 (White, J., dissenting).
125. *See generally* BAKER, *supra* note 82 (arguing that the marketplace of ideas theory is not persuasive).

of concerns about the theory.[126] At the top of his list, however, was the objective nature of the marketplace's truth assumptions. He asked, "why bet that truth will be the consistent or even the usual winner?"[127] He also concluded "truth is not objective."[128] Other scholars have identified similar concerns with the theory's truth assumptions. Legal scholar Stanley Ingber concluded the marketplace approach hinges on the nature of truth.[129] He explained, "[i]f truth is not ascertainable or cannot be substantiated, the victory of truth in the marketplace is but an unprovable axiom."[130] Legal scholar Frederick Schauer wove concerns about the nature of truth into a broader criticism regarding how information reaches and is understood by individuals.[131] He explained, "our increasing knowledge about the process of idea transmission, reception, and acceptance makes it more and more difficult to accept the notation that truth has some inherent power to prevail in the marketplace of ideas . . . ."[132]

These concerns about the Enlightenment-funded assumptions of truth, which have come to undergird the marketplace approach, find support in Justice Holmes's legal writings and correspondence.[133] While Justice Holmes introduced the marketplace concept into the Court's lexicon, he did not understand truth as being objective. Justice Holmes rejected absolute truth, calling it a "mirage."[134] He contended that people's realities are guided by experience.[135] In *Natural Law*, an article Justice Holmes wrote a year before he penned the *Abrams* dissent, he declared, "[c]ertitude is not the test of certainty. We have been cock-sure of many things that were not so."[136] Instead, Justice Holmes described himself as a "*bet*tabilitarian," explaining "I believe that we can *bet* on the behavior of the universe in its contact with us. We bet we can know what it will be."[137] His characterization of truth as being something that is liquid and

126. *See id.* at 12–17.

127. *Id.* at 6, 12.

128. *Id.* at 12.

129. *See* Ingber, *supra* note 23, at 15.

130. *Id*.

131. *See* Frederick Schauer, *The Role of the People in First Amendment Theory*, 74 CAL. L. REV. 761, 776–77 (1986).

132. *Id.* at 777.

133. *See* THE ESSENTIAL HOLMES: SELECTIONS FROM THE LETTERS, SPEECHES, JUDICIAL OPINIONS, AND OTHER WRITINGS OF OLIVER WENDELL HOLMES, JR. 107 (Richard A Posner ed., 1992) [hereinafter THE ESSENTIAL HOLMES].

134. *See id.*

135. *See id.*

136. Oliver Wendell Holmes, *Natural Law*, 32 HARV. L. REV. 40, 40 (1918).

137. THE ESSENTIAL HOLMES, *supra* note 133, at 108 (emphasis in original). *See also* David Luban, *Justice Holmes and the Metaphysics of Judicial Restraint*, 44 DUKE L.J. 449, 474 n.

evolving, rather than solid and unchanging, was part of an exchange in letters between British jurist Frederick Pollock and himself in 1929.[138] Similar ideas can also be found in his *Abrams* dissent from a decade before, where he explained, "[e]very year if not every day we have to wager our salvation upon some prophecy based upon imperfect knowledge."[139] Thus, scholars' concerns regarding the truth assumptions that have come to be foundational to marketplace thought find support in the jurist who introduced the approach into the Court's vocabulary.[140]

Finally, such foundational questions about the truth assumptions that have been wedded to the theory relate with concerns about the rationality expectations that are also central to the marketplace. Individuals encounter information differently, depending on a variety of factors, such as their personal experiences, socioeconomic position, race, religion, and other socializing influences. At the same time, messages reach people with different strength and frequency.[141] For these reasons, Baker reasoned the marketplace's Enlightenment assumptions fail at every turn.[142] He explained, "discussion is often insufficient by itself to determine the choice among different paradigms." [143] He concluded this is "precisely because the value-oriented criteria—interests, desires, or aesthetics—which guide the development of perceptions, appear ungrounded, incapable of objective demonstration."[144] Such concerns about the nature of truth and the ability of rational individuals to evaluate messages become more urgent in an era where falsity travels faster than truth across instantaneous, global virtual forums and emerging technologies allow believable deepfake videos to influence democratic discourse.

C. Alvarez *and the Court's Complicated Relationship with Falsity*

In striking down a portion of the Stolen Valor Act, which criminalized falsely claiming to have earned military honors, the Supreme Court in *United States v. Alvarez* in 2012 concluded that protecting intentional falsehoods was necessary to safeguard democratic discourse and the

78 (1994); *The Holmes-Cohen Correspondence*, 9 J. HIST. IDEAS 3, 11 (Felix S. Cohen ed., 1948).

138. THE ESSENTIAL HOLMES, *supra* note 133, at 108.

139. Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

140. *See* Robert Post, *Reconciling Theory and Doctrine in First Amendment Jurisprudence*, 88 CAL. L. REV. 2353, 2360 (2000).

141. *See* Ingber, *supra* note 23, at 15; BAKER, *supra* note 80, at 7; Jerome A. Barron, *Access to the Press—A New First Amendment Right*, 80 HARV. L. REV. 1641, 1642 (1967).

142. *See* BAKER, *supra* note 82, at 12.

143. *Id.* at 13.

144. *Id.*

marketplace of ideas.[145] In a telling conclusion to the Court's opinion, Justice Kennedy explained,

> The Nation well knows that one of the costs of the First Amendment is that it protects the speech we detest as well as the speech we embrace. Though few might find respondent's statements anything but contemptible, his right to make those statements is protected by the Constitution's guarantee of freedom of speech and expression.146

Such a conclusion moved the Court into new territory. It elevated—even equated—intentionally false statements to the same or similar levels of protection as those that have traditionally been reserved for the most protected forms of speech. Justice Kennedy, in writing for the Court, characterized the Stolen Valor Act as a content-based restriction.[147] He explained, "[t]he statute seeks to control and suppress all false statements on this one subject in almost limitless times and settings. And it does so entirely without regard to whether the lie was made for the purpose of material gain."[148] The Court's conclusions make it the most relevant precedent regarding falsity as it applies to deepfakes. By their nature, deepfakes are false. They represent someone saying or doing something they never said or did and generally seek to mislead the audience into believing what they are seeing and hearing actually occurred.[149] Almost any restriction on them would be based on their content. Similarly, Xavier Alvarez made intentional and knowingly false statements of fact during a public meeting, with the likely intent of misleading his audience into holding him into higher esteem.[150]

While *Alvarez* provides a relevant and recent precedent that largely supports protections for deepfakes, the decision stands alone. No other precedent is as protective of intentionally false expression and other decisions openly conflict with its conclusions.[151] Falsity-related precedents that preceded *Alvarez* paint a far more complex picture, something like a

145. *See* 567 U.S. 709, 727 (2012); 18 U.S.C. § 704 (2012).

146. *Alvarez*, 567 U.S. at 729–30.

147. *See id.* at 717.

148. *Id.* at 722–23 (citing S.F. Arts & Athletics, Inc. v. U.S. Olympic Comm., 483 U.S. 522, 539–40 (1987)).

149. *See* Villasenor, *supra* note 8.

150. *See Alvarez*, 567 U.S. at 713–14.

151. *See e.g.*, Vikram David Amar & Alan Brownstein, *The Voracious First Amendment*: *Alvarez and Knox in the Context of the 2012 and Beyond*, 46 LOY. L.A. L. REV. 491, 498–99 (2013) (explaining how *Alvarez* did not give other courts a bright-line approach to follow); Rodney A. Smolla, *Categories, Tiers of Review, and the Roiling Sea of Free Speech Doctrine and Principle*: *A Methodological Critique of United States v. Alvarez*, 76 ALB. L. REV. 499, 499 (2012) (noting how the Court failed to set forth a clear example for subsequent cases to follow).

Jackson Pollock work, of protections for false information. In *Sullivan*, the Court cautioned against limiting accidental falsities in speech.[152] The Court recognized "[t]hat erroneous statement is inevitable in free debate, and that it must be protected if the freedoms of expression are to have the 'breathing space' that they 'need . . . to survive.'"[153] Of course, in terms of defamation, the Court articulated the actual malice standard, which allows public officials to succeed in lawsuits when there is, "knowledge that it was false or with reckless disregard of whether it was false or not."[154] Thus, the Court drew a line between unintentional and intentional falsities, indicating it understood planned attempts to damage reputation with false information differently than simple errors that occur in communication.[155] Similarly, in *Hustler Magazine v. Falwell*, the Court reasoned, "even though falsehoods have little value in and of themselves, they are 'nevertheless inevitable in free debate.'"[156] Also, in *Philadelphia Newspapers v. Hepps*, justices reasoned "the Court has been willing to insulate even *demonstrably* false speech from liability" when the intent of the message is to inform, rather than mislead.[157] The Court explained such protections are particularly important regarding matters of public concern.[158] In each case, justices rationalized protecting *unintentional* falsehoods in order to safeguard democratic discourse.

The Court has made far more statements against protections for false information. The Court's first decision that directly addressed the First Amendment, *Schenck v. United States* in 1919, included Justice Holmes's conclusion, "[t]he most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic."[159] While it is questionable such a conclusion remains valid a century later, his initial suggestion that intentionally false and misleading information that is a danger to others should not be protected has persisted. In *Gertz v. Robert Welch, Inc.*, which provided the commonly referenced conclusion that "there is no such thing as a false idea," justices made a distinction between unpopular ideas and false facts.[160] In the same passage, the

152. *See* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 271–72 (1964).

153. *Id.* (quoting NAACP v. Button, 371 U.S. 415, 433 (1963)).

154. *Id.* at 280.

155. *See id.*

156. 485 U.S. 46, 52 (1988) (quoting Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974)).

157. *See* 475 U.S. 767, 778 (1986) (first citing Garrison v. Louisiana, 379 U.S. 64, 75 (1964); and then citing *Gertz*, 418 U.S. at 347).

158. *See id.*

159. *See* 249 U.S. 47, 52 (1919).

160. *See* 418 U.S. at 339.

Court found, "there is no constitutional value in false statements of fact. Neither the intentional lie nor the careless error materially advances society's interest in 'uninhibited, robust, and wide-open' debate on public issues."[161] The Court explained that false information falls to the level of unprotected speech, ultimately linking it to the fighting words doctrine that emerged in *Chaplinsky v. New Hampshire*.[162] Justices came to similar conclusions in *Hustler*, explaining, in the same passage it rationalized safeguarding unintentional errors in discourse, that "[f]alse statements of fact are particularly valueless; they interfere with the truth-seeking function of the marketplace of ideas."[163] In the *Virginia State Board* commercial speech case, justices explained "[u]ntruthful speech, commercial or otherwise, has never been protected for its own sake."[164]Justices echoed similar conclusions in several other First Amendment cases.[165] In each instance, justices reasoned intentionally false information is not a public good. Quite the opposite, they found it harms the marketplace.

Thus, in *Alvarez*, the Court provided expansive, easily applicable rationales for protecting intentionally false information that is spread via deepfakes. *Alvarez*, however, remains a precedential island. The mainland of the Court's discourse regarding falsity communicates strong protections for unintentional errors and unpopular ideas, but conveyed that intentionally misleading information should receive lesser or no protection. Both approaches connect with differing interpretations of marketplace theory. *Alvarez* supports the approach that people are rational and able to discern truth from falsity, while decisions that suggest intentionally false information should not be protected overlap with concerns that the marketplace must be protected from misleading, untrue information.

III. TRUTH RATIONALES IN THE E.U. SYSTEM

While U.S. jurists have constructed their conceptualizations of the nature of truth and the flow of information upon the marketplace approach—despite the concerns of legal scholars—the ECtHR has communicated different understandings when it comes to safeguarding free

161. *Id.* at 340 (quoting N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964)).

162. *See id.* (quoting 315 U.S. 568, 572 (1942)).

163. Hustler Mag. v. Falwell, 485 U.S. 46, 52 (1988) (citing *Gertz*, 418 U.S. at 340, 344 n.9).

164. Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, 425 U.S. 748, 771 (1976) (first citing *Gertz*, 418 U.S. at 340; and then citing Konigsberg v. State Bar, 366 U.S. 36, 49 n.10 (1961)).

165. *See, e.g.*, Herbert v. Lando, 441 U.S. 153, 171 (1979); Garrison v. Louisiana, 379 U.S. 64, 75 (1964); Brown v. Hartlage, 456 U.S. 45, 60–61 (1982); Madigan v. Telemarketing Assocs., 538 U.S. 600, 620 (2003).

expression. ECtHR judges conveyed they understand information as a public good that must be cared for by lawmakers and judges alike.[166] They also communicated that free expression must be balanced with other human rights, such as the right to privacy and citizens' reputations.[167] Of course, significant aspects of the divergence between the two systems come from the nature of the foundational documents the jurists must interpret. The First Amendment is written in absolute terms. The 45-word statement does not include any exemptions—explicit or implied—for government limitations on free expression.[168] The European Convention on Human Rights (ECHR), however, includes a two-part protection for freedom of expression.[169] Section 1 is written in terms that are similar to the First Amendment's wording.[170] Section 2, however, creates "duties and responsibilities" for citizens.[171] Section 2 allows for limitations that "are necessary in a democratic society," including national security, public safety, and the "protection of health or morals."[172] While the U.S. Supreme Court has identified a few categories of expression that are not protected, despite the First Amendment's wording, the absence of such an explicit list of limitations in the governing document helps explain the diverging rationales within the legal systems. Furthermore, the ECtHR has communicated three foundational concerns that generally diverge from U.S.-related free-expression interpretations. In free expression cases, ECtHR jurists communicated they expect protected information to be a public good, rather than a detriment; that they would balance free expression with other social concerns; and that they viewed the government as the custodian of public discourse.[173]

166. *See* Pauliukienė and Pauliukas v. Lithuania, no. 18310/06, § 48, 5 Nov. 2013.

167. *See id.* § 50.

168. *See* U.S. CONST. amend. I.

169. *See* ECHR, *supra* note 30, at 12.

170. *See id.*

171. *See id.*

172. *Id.*

173. *See generally e.g.*, Aquilina v. Malta, no. 28040/08, 14 June 2011, http://hudoc.echr.coe.int/fre?i=001-58239 (balancing the benefits and detriments of freedom of expression); Kobenter and Standard Verlags GmbH v. Austria, no. 60899/00, 2 Nov. 2006, http://hudoc.echr.coe.int/eng?i=001-77786 (discussing freedom of expression and the importance of the press); Steel and Morris v. United Kingdom, no. 68416/01, ECHR 2005-II, http://hudoc.echr.coe.int/eng?i=001-68224 (emphasizing the importance of the dissemination of information to the public).

A. Expectation of Public Good

Judges conveyed understandings that information is a public good, and therefore should be protected.[174] They did so, however, in two ways. In many cases, they communicated an *assumption* that information is a public good.[175] In other instances, however, they communicated an *expectation* that information be a public good.[176] This second approach left room for more limitations on information that jurists found did not contribute to public discourse and was therefore not a public good. In *Aquilina and Others v. Malta*, a defamation application that involved a journalist, the court highlighted the important role of the press and the right of individuals to receive information.[177] The court, however, qualified these statements by explaining "the press must not overstep certain bounds" and the press's "duty is nevertheless to impart—in a manner consistent with its obligations and responsibilities—information and ideas on all matters of public interest."[178] In *Kobenter and Standard Verlags GmbH v. Austria*, another defamation application, the court again lauded the importance of the press, but emphasized "it must not overstep certain bounds."[179] The court continued by outlining the "obligations and responsibilities" communicators have regarding "matters of public interest."[180] In siding with the publishers, jurists found "the applicants complied with their duties, responsibilities, and diligence as a public 'watchdog.'"[181] The court found the information *met the expectation* of being a public good. Such expectations find their foundations in Section 2 of Article 10, where the document introduces the "duties and responsibilities" criteria, as well as the question of whether information that faces government limitations is "necessary in a democratic society."[182] Jurists often conveyed that this question was their central concern, providing what is perhaps a central building block for legal questions regarding deepfakes.[183]

174. *See Aquilina*, no. 28040/08, § 44.

175. *See e.g.*, *Kobenter*, no. 60899/00, § 29 (discussing the utility of freedom of expression and the public's inherent right to receive information).

176. *See e.g.*, *Aquilina*, no. 28040/08, at § 19.

177. *See* no. 28040/08, § 43.

178. *Id.*

179. *See Kobenter*, no. 60899/00, § 29.

180. *Id.*

181. *See id.* §§ 31, 33.

182. *See* ECHR, *supra* note 30, at 12.

183. *See e.g.*, *Aquilina*, no. 28040/08, § 43; *Kobenter*. no. 60899/00, § 29; Steel and Morris v. United Kingdom, no. 68416/01, § 87, ECHR 2005-II, http://hudoc.echr.coe.int/eng?i=001-68224.

In many instances, however, the courts did not include such qualifying language. Instead, they explicitly communicated an assumption that information is a public good. Such an approach bears similarities with U.S. legal approaches. In *Steel and Morris v. United Kingdom*, a case in which McDonald's sued activists for claims they circulated about the company, the court highlighted, "[t]he issues raised in the leaflet were matters of public interest and it was essential in a democracy that such matters be freely and openly discussed."[184] Later in the application, the court explained, "there exists a strong public interest in enabling such groups and individuals . . . to contribute to the public debate by disseminating information and ideas on matters of general public interest."[185] Similarly, in *Hertel v. Switzerland*, in which a commercial appliance group sued to have a researcher enjoined from communicating his findings that microwaves were harmful to public health, the court explained, "[f]reedom of expression constitutes one of the essential foundations of a democratic society and one of the basic conditions for its progress."[186] The court continued, recognizing the free-expression-qualifying passages from section 2 of Article 10, that free expression includes unpopular information, as well as content that can "offend, shock or disturb."[187] The court communicated similar findings in *Instytut Ekonomichnykh Reform, TOV v. Ukraine*, which involved accusations that were published regarding Ukrainian political leaders during a Constitutional crisis.[188] In finding for the publisher, the court reasoned "[f]reedom of expression is applicable not only to 'information' or 'ideas' that are favorably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any section of the community."[189] In each of these applications, the court communicated an understanding that information is a public good. Together, the *assumption* of information as a public good and the more qualified *expectation* that content benefit society provide crucial building blocks for how the courts can approach deepfakes in future cases.

B. Balanced with Other Concerns

Jurists consistently communicated that free expression rights should be balanced with myriad other concerns. Such a theme in Article 10-

184. *See Steel and Morris*, no. 68416/01, § 79.
185. *Id.* § 89.
186. *See* no. 25181/94, § 46, ECHR 1998-VI, http://hudoc.echr.coe.int/eng?i=001-59366.
187. *See id.*
188. *See* no. 61561/08, §§ 7–8, 2 June 2016, http://hudoc.echr.coe.int/eng?i=001-163354.
189. *Id.* § 43.

related applications can provide further building blocks when it comes to deepfakes. The court, however, was not consistent in the concerns with which it balanced free expression. In *Steel and Morris*, jurists communicated they must balance "the need to protect the applicants' rights to freedom of expression and the need to protect McDonald's rights and reputation."[190] The ECtHR raised a similar concern in *Kobenter*, concluding it must balance "the applicants' interest in disseminating information and ideas on matters of public interest" and "the interest of the judge concerned in protecting his reputation and the standing of the judiciary."[191] Jurists conveyed similar concerns in *Delfi AS v. Estonia*, which dealt with a forum provider's liability for defamatory third-party messages.[192] The court explained Article 10 and Article 8 "deserve equal respect."[193] In each of these cases, the court balanced the "important benefits" of free expression on the Internet with reputation-related concerns.[194]

Reputation, however, was not the only matter jurists placed on the scales in balancing free expression rights. In *Mouvement Raëlien Suisse v. Switzerland*, which dealt with government officials' refusal to allow a group to display its poster, the ECtHR weighed the government's right to "prevent crime, to protect health or morals and to protect the rights of others" against free expression.[195] The court found the government's concerns regarding crime, health, morals, and the rights of others outweighed free expression, thus upholding the ban on the poster.[196] The court concluded the poster was not a form of political speech, which would receive more protection.[197] Citing Section 2 of Article 10, the ECtHR found the government can limit free expression in "matters liable to offend intimate personal convictions within the sphere of morals or, especially, religion."[198] The court considered similar matters in *Hertel*, where it listed "demands of pluralism, tolerance, and broadmindedness without which there is no 'democratic society.'"[199] In *Fatih Taş v. Turkey (No. 2)*, the second of five cases in which the same Turkish publisher was convicted of crimes for criticizing the government and discussing the PKK (the

190. *See* no. 68416/01, § 95, ECHR 2005-II, http://hudoc.echr.coe.int/eng?i=001-68224.
191. *See* no. 60899/00, § 27, 2 Nov. 2006, http://hudoc.echr.coe.int/eng?i=001-77786.
192. *See* no. 64569/09, § 15, ECHR 2015, http://hudoc.echr.coe.int/eng?i=001-155105.
193. *Id.* § 110.
194. *See id.*; *Kobenter*, no. 60899/00, § 29.
195. *See* no. 16354/06, § 54, 13 July 2012, http://hudoc.echr.coe.int/eng?i=001-112165.
196. *See id*. §§ 76–77.
197. *See id*. §§ 61–62.
198. *Id.* § 61.
199. *See* Hertel v. Switzerland, no. 25181/94, § 46, ECHR 1998-VI, http://hudoc.echr.coe.int/eng?i=001-59366.

Kurdistan Workers' Party), the ECtHR considered whether Turkey's government had unfairly limited the author's expression.[200] In doing so, the court weighed the nation's concern for preventing violence against free expression.[201] The ECtHR overturned the author's conviction, finding, "the interference with the applicant's right to freedom of expression was not justified by 'relevant and sufficient' reasons."[202]

Thus, the ECtHR has placed concerns for other rights promised within the ECHR, as well as less explicitly written concerns, such as tolerance, health, and safety, at the center of free expression decisions. While doing so has muddied the boundaries of free expression safeguards in the E.U., it leaves significant avenues open for the courts to limit deepfakes.

C. Custodians of Discourse

Finally, jurists consistently communicated they understood the government as the protector of democratic discourse. Within this area, the ECtHR conveyed that E.U. member states' governments, as well as the courts, must at times step in to block some ideas and expression to safeguard the flow of ideas. This role as caretaker of public discourse finds substantial support in Section 2 of Article 10, which allows for limitations as are "necessary in a democratic society."[203] Such an approach is substantially different than the marketplace of ideas that has dominated U.S. understandings. The marketplace includes an inherent distrust of government actors within the realm of ideas, instead placing faith in individual citizens to discern between truth and falsity.[204] In *Animal Defenders International v. The United Kingdom*, the ECtHR upheld the U.K.'s right to halt the broadcast of an animal rights group's advertisement.[205] The court reasoned "the general measure was necessary to prevent the distortion of crucial public interest debates and, thereby, the undermining of the democratic process."[206] In *Mouvement Raëlien Suisse*, the Swiss Federal Court concluded blocking a poster was "justified by sufficient public-

200. *See* no. 6813/09, § 18, 10 Oct. 2017, http://hudoc.echr.coe.int/eng?i=001-177868. *See also* Ronan Ó Fathaigh, *Prosecution of a Publisher for 'Denigration' of Turkey Violated Article 10*, STRASBOURG OBSERVERS (Oct. 29, 2018), https://strasbourgobservers.com/2018/10/29/prosecution-of-publisher-for-denigrating-turkey-violated-article-10/.

201. *See Fatih Tas*, no. 6813/09, § 18.

202. *Id.*

203. *See* ECHR, *supra* note 30, at 12.

204. *See supra* Part II.

205. *See* no. 48876/08, § 125, ECHR 2013 (extracts), http://hudoc.echr.coe.int/eng?i=001-119244.

206. *Id.* § 116.

interest grounds."[207] The ECtHR agreed the poster "sought mainly to draw the attention of the public to the ideas and activities of a group with a supposedly religious connotation that was conveying a message claimed to be transmitted by extraterrestrials, referring for this purpose to a website address."[208] The court reasoned individual states are best suited to protect the health, welfare, and safety of their citizens, particularly as they relate with the information they encounter.[209]

In upholding the fines in *Delfi*, the court reasoned online communication has the unique capability to damage reputation and invade individuals' privacy.[210] Thus, despite *Delfi's* decision to remove the offensive comments when it was alerted, as part of the notice-and-takedown system it created, the court concluded the company did not do enough to protect reputation and the quality of discourse within its forums.[211] Similarly, in *Hertel*, despite overturning the injunction as a violation of the author's rights, the court carefully considered the quality and substance of the information that was published about the dangers of microwaved food.[212] The court found, "[t]he Government submitted that the interference in the applicant's freedom of expression was aimed at guaranteeing fair and free competition in the interests of society as a whole. It had therefore met a pressing social need."[213] While the court did not find the nation's decision to be, when balanced with other concerns, reasonable enough to halt his expression, the decision further conveys the extent to which jurists understood the role of government as a custodian of democratic discourse.

CONCLUSION

Deepfakes pose a threat to the already tenuous nature of truth in online spaces. As intentionally false written reports lead to misleading information during elections and fuel conspiracy theories that are often seamlessly accepted within people's echo-chamber-like networked communities, deepfake videos add a more visceral and traditionally more credible form of false and misleading information. The emergence of deepfakes as a tool for state and non-state actors to manipulate discourse

207. *See* Mouvement raëlien suisse v. Switzerland [GC], no. 16354/06, § 21, 13 July 2012, http://hudoc.echr.coe.int/eng?i=001-112165

208. *Id.* § 62.

209. *See id.* § 72.

210. *See* Delfi AS v. Estonia, no. 64569/09, § 110, ECHR 2015, http://hudoc.echr.coe.int/eng?i=001-155105.

211. *See id.* §§ 140–43.

212. *See* Hertel v. Switzerland, no. 25181/94, § 13, ECHR 1998-VI, http://hudoc.echr.coe.int/eng?i=001-59366.

213. *Id.* § 44.

threatens citizens' ability to believe their own eyes and undermines the work of the press in providing credible information to the public, since even truthful reports can be discredited by claims that they are deepfakes.[214] This article examined the potential threats deepfakes pose, as well as free expression rationales, particularly as they relate with truth and falsity, within the U.S. and E.U. legal systems. In analyzing how jurists in both systems have communicated how they understand free expression and the role of truth and falsity within democratic discourse, a few important concepts emerged as tools for how political deepfakes can be limited without undermining free expression rationales within the regions.

Jurists in the U.S. and E.U. conceptualized information as a *public good*. While the Supreme Court communicated it understood information as almost inherently of value, because of the marketplace assumptions that are at the heart of the nation's free-expression rationales, E.U. jurists conveyed similar understanding in some cases, but in others communicated an expectation that information be beneficial to the public.[215] Similarly, jurists in the E.U. and U.S. communicated concern for the flow of truthful, factual information in democratic society. They again did so in different ways. The ECtHR positioned the government as a type of custodian of public discourse. Doing so included the interpretation of Article 10-related free expression safeguards as something that must be protected and maintained. U.S. jurists wrestled with falsity and threats to discourse in different ways, but some of their rationales corresponded with ECtHR approaches. Supreme Court opinions generally interpreted the marketplace as allowing almost unfettered discourse, a conclusion that aligns with the theory's Enlightenment-related truth and rationality assumptions. Justices, however, dissented in a handful of cases, contending the laws in question would have *protected* the marketplace and its discursive processes from harmful information and influences.[216] In his dissent in *Citizens United*, Justice Stevens understood the campaign finance law in question as reflecting "a concern to facilitate First Amendment values by preserving some breathing room around the electoral 'marketplace' of

214. *See* Chesney & Citron, *supra* note 7, at 1785–86.

215. *See, e.g.*, Steel and Morris v. United Kingdom, no. 68416/01, § 79, ECHR 2005-II, http://hudoc.echr.coe.int/eng?i=001-68224; *Hertel*, no. 25181/94, § 46; Instytut Ekonomichenykh Reform, TOV v. Ukraine, no. 61561/08, § 43,2 June 2016, http://hudoc.echr.coe.int/eng?i=001-163354.

216. *See, e.g.*, First Nat'l Bank v. Bellotti, 435 U.S. 765, 810 (1978) (White, J., dissenting); Va. State Bd. Pharmacy v. Va. Citizens Consumer Counsel, Inc., 425 U.S. 748, 781 (1976) (Rehnquist, J., dissenting); Citizens United v. FEC, 558 U.S. 310, 473 (2010) (Stevens, J., dissenting).

ideas, the marketplace in which the actual people of this Nation determine how they will govern themselves."[217] Similarly, the Court has wrestled with protections for intentionally false information. Justices granted broad protection to false information in *Alvarez* in 2012, but have otherwise provided a mish-mash of precedential conclusions regarding intentionally false information.[218] Ultimately, the decisions whether to protect most information in the name of the marketplace and its assumptions or protect the marketplace from information that might harm the flow of information, as well as disagreements about protections for intentionally false information provide a narrow path for limitations on political deepfakes.

Such a narrow path can be thought of as opening the door for a *safeguarding principle*, which can be constructed from precedential findings within both legal systems. Using the *safeguarding principle*, laws can limit political deepfakes when they meet all of three criteria. First, the law narrowly defines a deepfake as a video clip that falsely attributes speech or actions to a person who never said or did the speech or actions in order to harm them politically or to confuse or manipulate democratic discourse. Any such law cannot define a deepfake more broadly or in a way that limits unintentional errors or unpopular ideas. Second, the communication being limited must include false statements of fact in that they attribute ideas, statements, or actions to those who did not communicate or commit them. Third, information must be a threat to democratic discourse. While this is the most subjective criteria, it requires laws address only deepfakes that damage the flow of information, rather than those that parody, comment, or challenge ideas. Such a final criterion requires carefully constructed means and ends.

Of course, precedent-based arguments can be made against any such *safeguarding principle*—in the E.U. or U.S. The three criteria, however, are not unprecedented. If we accept that deepfakes present a fundamentally new, unique, and dangerous threat to democratic discourse, calling upon precedential support protecting discourse, as the ECtHR did in expecting information to be a public good and calling on the government to act as a custodian of information or as Supreme Court justices did in dissents that contended some laws help protect the marketplace, provides a way forward. Similarly, precedents that limited free expression for intentionally false information, though there are decisions that contend otherwise, can operate as powerful allies. Such an approach protects broad swaths of deepfakes, such images of Nicolas Cage being superimposed

217. *Citizens United*, 558 U.S. at 473 (Stevens, J., dissenting).

218. *See* United States v. Alvarez, 567 U.S. 709, 729 (2012).

into films or Jordan Peele controlling the likeness and sound of President Obama to warn people about false information. It would allow, however, limitations of deepfakes that intentionally mislead the public about matters of public concern.