Yale Journal of Law & Technology
Volume 23, Fall 2020

DEEPFAKES AND OTHER NON-TESTIMONIAL FALSEHOODS: WHEN IS BELIEF MANIPULATION (NOT) FIRST AMENDMENT SPEECH?

MARC JONATHAN BLITZ*

* Alan Joseph Bennett Professor of Law, Oklahoma City University. The author would like to thank the following individuals for suggestions on previous drafts or helpful discussions about the arguments in the article: Rebecca Aviel, Woodrow Barfield, Jeff Blitz, Alan Chen, Bobby Chesney, Danielle Citron, Rebecca Green, Eric Laity, Art LeFrancois, Jared Schroeder, Joseph Thai, Jeremy Telman, Alexander Tsesis, Yari Wildheart, Kiel Brennan-Marquez, Adam Kolber, and Helen Norton, and participants at a workshop at the 2019 Freedom Expression Scholars Conference at Yale Law School.

INTRODUCTION

"I tell you there's something phony going on. There's something phony . . . about this whole Medal of Honor business."
~ Captain Bennett Marco in The Manchurian Candidate (1962)

Individuals are presumptively protected by the First Amendment when they deceive other people by making false statements. The Supreme Court made this clear in the 2012 case of *United States v. Alvarez.* The criminal defendant at the center of the case, Xavier Alvarez, had falsely insisted that he had won a Congressional Medal of Honor for his bravery in battle. In reality, he had never served in the military.[1] He was prosecuted for his false claim under Congress's Stolen Valor Act, which made it a federal crime for a person to "falsely represen[t] himself or herself, verbally or in writing, to have been awarded any decoration or medal authorized by Congress for the Armed Forces of the United States."[2] But the Court found this law unconstitutional. Lies such as that of Xavier Alvarez, Justice Kennedy said in the plurality opinion, are shielded by the First Amendment's free speech protection unless the government can show that that they are not merely false, but also harmful in ways that have traditionally provided the basis for liability.[3]

But imagine that a person wishing to do what Xavier Alvarez did—give others the false impression he had won a Medal of Honor—did so not by using false statements, but rather by creating fake evidence. Imagine he didn't want others to have to take his *own* word that he was a Congressional Medal of Honor winner—that he instead wanted them to be able to reach this conclusion by observing the world around them, and using evidence

---

[1] United States v. Alvarez, 567 U.S. 709, 719 (2012) (plurality opinion).
[2] 18 U.S.C.A. § 704(b); *Alvarez*, 567 U.S. at 719.
[3] *Alvarez*, 567 U.S. at 719.

that they could *see and examine for themselves*. He might have done so, for example, by wearing a fake Medal of Honor around his neck, or perhaps a real medal that had made its way through an Internet auction or two into Alvarez's possession.

Or he might have created a web site presenting itself an authoritative "register of military award winners" and added his name to an otherwise accurate list of award winners. Perhaps this web site could have posed as the creation of the Department of Defense or another government agency, or as that of a private organization which has the purpose of celebrating military accomplishments.

Or imagine that he took this deception a step further. Imagine that he wanted his audience to see, with their own eyes, not only the Medal he had purportedly earned, but a video of the ceremony in which President Reagan presented it to him. In past years, generating such a visual illusion would be difficult: Alvarez never actually received a Medal of Honor from President Reagan, and it has generally been quite difficult to produce a video of an event that never occurred. A major movie studio, perhaps, could do so: The 1994 film, Forrest Gump, used special effects to show President Lyndon B. Johnson awarding a Congressional Medal of Honor to its fictional title character (played by Tom Hanks).[4]

But that kind of fabrication has required immense time, effort, artistry, and expense. Technological development, however, is making the generation of fake videos far simpler. With a kind of machine learning known as "deep learning," a computer program can quickly teach itself to recreate a person's image or voice, manipulate it—like a puppeteer controlling a puppet—and blend it

[4] FORREST GUMP (Paramount Pictures 1994).

seamlessly into an environment the person never inhabited.[5] This kind of fake video or audio, a "deepfake," can be very difficult to distinguish from genuine camera footage.[6] As a consequence, the seemingly real political speech we see by a U.S. President or other world leader might be one that never occurred. Deepfake creators, in fact, have generated speeches of this kind to demonstrate the power of this technology including a video showing President Obama warning—in a speech he never gave—about the dangers of deepfakes,[7] another showing President Nixon announcing the failure of the 1969 Apollo mission to the moon and the death of the astronauts on that mission[8] and a fake Christmas speech by Queen Elizabeth II to mark the end of 2020.[9] In this form, video and audio recordings or transmissions are no longer a window into remote events. They are instead a portal through which we see a hyper-realistic world that is fabricated and fictional.

With this technology in hand, Alvarez could have produced

---

[5] Oscar Schwartz, *You Thought Fake News Was Bad? Deep Fakes Are Where Truth Goes to Die*, GUARDIAN (Nov. 12, 2018), https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth.

[6] *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Cal. L. Rev. 1753, 1759 (2019) (describing how artificial intelligence is making deepfakes far more difficult to identify as fakes); Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned,'* WASH. POST (June 12, 2019), https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned..

[7] *See* James Vincent, *Watch Jordan Peele Use AI to Make Barack Obama Deliver a PSA about Fake News*, VERGE (Apr. 17, 2018), https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed.

[8] Jeffrey Delviscio, *A Nixon Deepfake, a 'Moon Disaster' Speech and an Information Ecosystem at Risk*, SCI. AM. (July 20, 2020), https://www.scientificamerican.com/article/a-nixon-deepfake-a-moon-disaster-speech-and-an-information-ecosystem-at-risk1.

[9] Zamira Rahim, *Deepfake' Queen Delivers Alternative Christmas Speech, In Warning About Misinformation*, CNN (Dec. 25, 2020), https://www.cnn.com/2020/12/25/uk/deepfake-queen-speech-christmas-intl-gbr/index.html.

video evidence for his tall tale. He could have created a video showing brief clips of President Ronald Reagan awarding various medals, including Medals of Honor—with one clip showing Reagan placing one such Medal of Honor around Alvarez's neck after vividly describing Alvarez's selflessness and bravery in battle. If he wanted to make this more convincing, he might insert this video clip into a series of other genuine videos showing Reagan presenting military and civilian awards. Someone who recognized any of the other award winners as individuals who *have* won those awards, might then mistakenly assume this means that all of the clips in the video are authentic (just as someone who saw Alvarez's name added to an otherwise accurate list of Medal of Honor Winners on a web-based database might assume that this means Alvarez's name belongs there too).

In all of the above cases, he would deepen his deception. He could now back up his lie by telling his audience, "if you don't believe my statement, here is additional evidence to support it that you can examine for yourself. You can see the medal I received, examine an authoritative web site supports my claim, and that there is a video showing me receiving the Medal of Honor."

When individuals deepen deception in this way, moving from fake words to the creation of fake evidence, does First Amendment protection move with them? Does the First Amendment protect them not only when they insert falsity into their own words, as the Supreme Court held in *Alvarez*, but also when they find ways to introduce it into fabricated evidence such as a deepfake video? Where someone not only tells a verbal lie, but also—or instead—falsifies the kind of *external* evidence others would use to check the veracity of that lie, such as a web site apparently created by an independent source or a videorecording, does the First Amendment also protect this additional deception?

There has been relatively little analysis of this question in First Amendment case law or scholarship. But scholars and commentators have recently begun to offer initial answers to it as they have struggled in the past three years with the threats raised by deepfakes. They often assume that if, as the Supreme Court has held, the First Amendment protects verbal lies it should also protect visual lies that one finds in deepfakes.[10] Particularly in this era, when people communicate on social media not just by posting comments, but also by sharing video clips, video is a form of expression. So if the holding of *Alvarez* is that a speaker presumptively has a right to insert falsehood into her own expression, then it follows she has a right to do so when she expresses herself with a vivid image sequence rather than with comments. To use language used by courts in First Amendment cases, sharing videos is—at least in the early twenty-first century—an "inherently expressive" social practice.[11]

Indeed, a video like the hypothetical one described above, depicting Alvarez receiving a Medal of Honor, is not only arguably the equivalent of Alvarez's protected false statement. It might also be an instance of another kind of unquestionably protected speech:

---

[10] *See, e.g.*, Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J.L. & TECH. 1, 34–35 (2020) (arguing that deepfakes and the creation of deepfakes are "a protected First Amendment activity"); Chesney & Citron, *supra* note 6, at 1790-1792; Russell Spivak, *"Deepfakes:" The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 358 (2019) (arguing that because of First Amendment "to pass constitutional muster, deepfake regulations must fall into one of these exceptional categories" of speech that don't receive First Amendment protection."); Chesney & Citron, *supra* note 6, at 1790-1792 (stating that "[d]eep fakes implicate freedom of expression, even though they involve intentionally false statements" and applying *Alvarez's* framework to understand what regulation the First Amendment would allow); Shannon Reid, *The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections*, 23 U. PA. J. CONST. L. 209, 216 (noting that "privacy tort protections" against deepfakes "are vulnerable to a First Amendment defense that deepfakes are protected speech" and proposing a solution to this problem).

[11] Rumsfeld v. Forum for Acad. & Institutional Rights, Inc., 547 U.S. 47, 49 (2006).

artistic expression. The makers of the movie, Forrest Gump, clearly had a First Amendment right to create a vividly realistic scene of their protagonist in a fictional Medal of Honor ceremony. The government could not constitutionally have ordered them to remove that scene from the film. Why then shouldn't modern-day video-makers be able to exercise the same creativity on YouTube or other social media sites? Why shouldn't they be able to give vivid visual form to their own autobiographical fictions, whether this involves weaving themselves into a Medal of Honor ceremony, making themselves a hero of a World Series game, or placing themselves on a concert stage to accompany Nat King Cole, the Beatles, or perhaps Johann Sebastian Bach, John Dowland, or Hildegard of Bingen?

To be sure, as numerous commentators have pointed out, some deepfakes might be far less whimsical and potentially quite dangerous. Bobby Chesney and Daniel Citron, for example, consider the ways deepfakes might be used to defame or defraud others—or to undermine national security by showing viewers missile attacks or riots that never happened—and consider how law, policy, and technology can (and cannot) help counter those threats.[12] Rebecca Green has warned that deepfake technology provides a powerful new disinformation tool for influencing elections: Campaigns (or those who sympathize with them) might create fake footage of their opponents.[13] Green has thus proposed a ban on using deepfakes for what she calls "counterfeit campaign speech."[14] Texas and California have now adopted bans of this kind.[15]

---

[12] Chesney & Citron, *supra* note 6, at 1773-85.
[13] *See* Rebecca Green, *Counterfeit Campaign Speech*, 70 HASTINGS L.J. 1445, 1451 (2019)
[14] Id.
[15] *See* CAL. ELEC. CODE § 20010 (West 2020) (barring anyone from "distribut[ing] distribute, with actual malice, materially deceptive audio or visual media" and defining such media to include "an image or an audio or video recording of a

Congress has considered a law like this on campaign-related deepfakes[16] as well as other deepfake restrictions, such as the Malicious Deep Fake Prohibition Act of 2018, proposed by Senator Ben Sasse in the Senate,[17] and the DEEPFAKES Accountability Act of 2019 proposed by Representative Yvette Clark in the House.[18] It has already enacted legislation that orders government to report on the threats posed by deepfakes, and incentivizes technologists to develop methods of detecting them.[19] Virginia and California have enacted laws dealing with pornography created with deepfake technology or other technology for creating altered video.[20] Other states are also considering deepfake restrictions.[21]

---

candidate's appearance, speech, or conduct that has been intentionally manipulated" to appear authentic and alter the audience's understanding of the candidate); TEX. ELEC. CODE. § 255.004 (West 2020) ("A person commits an offense if the person, with intent to injure a candidate or influence the result of an election: (1) creates a deep fake video; and (2) causes the deep fake video to be published or distributed within 30 days of an election.").

[16] *See* Deepfakes in Federal Elections Prohibition Act, H.R. 6088, 116th Cong. (2021).

[17] *See* Malicious Deep Fake Prohibition Act of 2018, S. 3805 115th Cong. (2018).

[18] *See* Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R.3230, 116th Cong. (2019).

[19] *See* National Defense Authorization Act for Fiscal Year 2020, S. 1790, 116th Cong. (2020) (requiring the Executive to report of foreign weaponization of deepfakes and use deepfakes for election interference, and establishing reward for research on deepfake detection methods).

[20] *See* CAL. CIV. CODE § 1708.86 (barring intentional creation or disclosure "an audiovisual work that shows the depicted individual performing in the nude or appearing to engage in, or being subjected to, sexual conduct," where it was or should have been clear to the creator (and was known to a discloser) that the depicted individuals didn't consent); VA. CODE ANN. § 18.2-386.2 (unlawful dissemination or sale of images of another).

[21] *See, e.g.*, An Act to Protect Against Deep Fakes Used to Facilitate Criminal or Tortious Conduct, H. 3366, 191st General Court (Mass. 2019), which would expand the state's definition of identity fraud to criminalize the creation or distribution of deepfakes intended for use in otherwise criminal or tortious conduct; An Act to Amend the Civil Rights Law, in Relation to the Right of Privacy and the Right of Publicity, A.B. A8155B, 2017-2018 Leg. Sess. (N.Y.); David Robb, *SAG-AFTRA Expects NY Gov. Andrew Cuomo To Sign Law Banning "Deepfake" Porn Face-Swapping*, DEADLINE (July 28, 2020), https://deadline.com/2020/07/deepfakes-sag-aftra-expects-andrew-cuomo-to-sign-law-banning-face-swapping-porn-1202997577. *See also* Stephanie Salmons, *Bills*

But *United States v. Alvarez* might—even while protecting lies—leave government with sufficient leeway to combat the graver falsehoods one finds in deepfakes. Justice Kennedy's plurality opinion concluded that false speech cannot be punished merely because it is false—but *may* be punished when that falsity is accompanied by serious harms of a kind that have traditionally been regarded as a kind of "legally-cognizable harm."[22] The intimidation and distress generated by fake video of a missile attack may count as presenting such a harm. Deepfake restrictions may also be permissible under *Alvarez* even if they do not fit into a familiar category of measures that address legal harms. They can do so when the government has no other, less speech-restrictive way to address a "compelling government interest."[23] Although this constitutional hurdle—which courts call "strict" or "exacting" judicial scrutiny—is almost impossible for government to overcome, government may be able to do so when the alternative is to leave people subject to the fear and manipulation threatened by deepfakes of missile attacks, natural disasters, or false election claims.[24]

In short, then, this account of deepfakes' status sketches a framework for answering the question raised earlier: Deepfake videos *are* First Amendment expression, like the genuine camera videos they emulate. As such, they are presumptively protected by the First Amendment shield for false claims one finds in *Alvarez*. When that expression becomes harmful enough, however, that presumption is overcome. A deepfake is then, like falsely shouting

---

Target The Use of 'Deep Fakes' in Hawaii, HAWAII TRIBUNE HERALD, Feb. 2, 2021, https://www.hawaiitribune-herald.com/2021/02/02/hawaii-news/bills-target-the-use-of-deep-fakes-in-hawaii/ (discussing proposed laws to introduced to "protect the privacy of a person's likeness by adopting laws that prohibit the unauthorized use of deep fake technology.").

[22] United States v. Alvarez, 567 U.S. 709, 719 (plurality opinion).

[23] *Id*. at 720, 724.

[24] *Id*. at 725-26.

fire in a crowded theater, within the government's power to stop.[25] As Chesney and Citron write, deepfakes allow such "false cries" to "go viral."[26]

This article, however, presents a different analysis of deepfakes' First Amendment status—and that of other fabricated evidence. Deepfakes will often deserve less First Amendment protection than a verbal lie. But this isn't because they are a more harmful form of expression. It is because there are some uses of deepfakes that are not generally expression of the kind the First Amendment protects. They are in some respects at least partly outside the First Amendment's "coverage."[27] And the reason they have to be is that they would otherwise extend the "authorship" that the First Amendment provides to speakers beyond the sphere that the Constitution sets aside for it (and can afford to set aside for it).[28]

More specifically, the First Amendment gives us a right to determine the statements we make, the stories we tell, and the other artwork we create—and it reserves for us (not the government) the decisions about what content to put there, allowing us to place even false content. But it does not give us the right to make ourselves an "author" of all sources of information our audience relies upon to understand the world. It doesn't mean we can shape, and possibly falsify, sources of information that appear (to our audience) to arise from sources outside of our control. That is a key reason, this article argues, that deceivers do not have a First Amendment right to

---

[25] *See* Schenck v. United States, 249 U.S. 47, 52 (1919).
[26] Chesney & Citron, *supra* note 6, at 1781.
[27] The distinction between First Amendment "coverage" and "protection" was developed by Fredrick Schauer and is more fully explained in his book, FREE SPEECH: A PHILOSOPHICAL INQUIRY 89 (1982).
[28] For an interesting exploration of how we might understand First Amendment speech protections as covering acts of "authorship" see Derek E. Bambauer, *Copyright = Speech*, 65 EMORY L.J. 199, 200-203 (2015) (exploring how copyright law might illuminate questions of First Amendment coverage).

deepen their deception to the extent described earlier—that they don't have a right to bolster the falsity in their own words by falsifying multiple aspects of the audience's environment. The other examples of deceptive evidence I considered provide some support for this point. The First Amendment does not appear to prevent the government from making it illegal to create a fake Web register of Medal of Honor winners. As I will explain more fully later, all of the Justices in *Alvarez* appear to agree on that point.[29] Nor does the First Amendment appear to prevent government from preventing the forgery of military medals: 18 U.S.C.A. § 704(a) makes it illegal to "knowingly . . . manufactur[e] . . . any decoration or medal authorized by Congress for the armed forces of the United States except when authorized under regulations made pursuant to law."[30] Our freedom of speech in other words does not include a freedom to disguise one's own work as an authoritative database of medal winners or as an official medal. We should pause then, before assuming it includes a freedom to disguise one's own fictional biography as an authoritative visual record of past events.

This does not mean that deepfakes are entirely without First Amendment protection. Rather, what makes deepfakes a challenge for First Amendment analysis is that they straddle the line between the realm that the First Amendment reserves for authorship and the informational realm *external* to speakers' words (which they do *not* have a First Amendment right to shape). On the one hand, videos, including deepfakes, may well be a part of that sphere of authorship. Video is, of course, in many circumstances, a medium of artistic expression, and deepfake technology can play a role in such artistic

---

[29] *See infra* text accompanying notes 204-209.

[30] There is additional discussion of this and its implications for cases asking if wearing unearned medal can be prohibited consistent with the First Amendment in the text accompanying *infra* notes 295-298 and note 299.

expression. It might provide professional and amateur moviemakers another kind of special effects technology.[31] Not only it is a tool for professional or amateur filmmakers to tell fictional stories. It is a means by which authors can visually illustrate or embellish their arguments or claims. A video posted on social media, for example, might try to highlight the dangers of global warming by using deepfake technology or other special effects to depict a future Miami, New York, Los Angeles experiencing massive flooding.

On the other hand, audiences don't look to videos solely to find others' artistic work or argument. They treat some videos as a reliable visual record of events. When viewing the raw footage captured in a security camera, for example, we generally don't treat it as a narrative someone wants to tell or an argument someone wishes to make. Even when we realize that a video or audiotape might have been edited to reflect a certain perspective, we have generally been able to assume that not all aspects of the video can be fabricated from scratch. Because video and audio recordings, prior to the age of deepfakes, have been difficult to counterfeit, they have often been viewed as more reliable evidence of events than verbal reports. Police body cameras, for example, have been demanded by rights organizations and adopted by cities and states so that citizens, courts, and police departments might review evidence of police encounters that would otherwise leave no record.[32] As Regina Rini writes, video or audio recordings have often been viewed as a way to resolve the uncertainties generated when witnesses have

---

[31] *See* Marc Jonathan Blitz, *Lies, Line Drawing, and (Deep) Fake News*, 71 OKLA. L. REV. 59, 114 (2018).

[32] *See* Clare Foran, Manu Raju, Lauren Fox and Ted Barrett, *Senate Democrats Block GOP Police Reform Bill, Throwing Overhaul Effort into Flux,* CNN (June 24, 2020), https://www.cnn.com/2020/06/24/politics/senate-police-reform-bill/index.html, ("[T]he Democratic plan has a focus on setting national standards, such as mandates for federal uniformed officers to wear body cameras and banning chokeholds.").

conflicting memories, or politicians contest claims about what they have done.[33] Even in an age when viral disinformation social media campaigns have weakened many Americans' confidence about what is true, many writers still point to evidence of audio- and video-recordings as harder to spin or deny than verbal reports and many other types of evidence. Commentators have assumed, for example, that the audio-recordings of President Trump's statements—with journalist, Bob Woodward, about the dangers of Covid-19, for example, or with Georgia Secretary of State Brad Raffensperger about recalculating votes in the 2020 election—provide powerful evidence of Trump's actual words.[34]

These video and audio records, in other words, have an informational rather than artistic function. Like their artistic counterparts, such informational video or audio recordings are protected by the First Amendment. Free speech law not only protects our right to create and post videos that are works of art, it also gives us a right to post unaltered video footage for its purely informational value. For example, if reporters or other individuals obtain video footage captured from a cell phone camera or home security camera showing police arresting a suspect, they do not need to add any creative alterations to this video in order for it to receive First Amendment protection. Simply posting the video on social media, or showing it to a friend or acquaintance, is an act of First

---

[33] Regina Rini, *Deepfakes and the Epistemic Backstop*, 20 PHILOSOPHERS' IMPRINT 1 (Aug. 2020).

[34] *See, e.g.*, Amy Gardner, *'I Just Want to Find 11,780 Votes': In Extraordinary Hour-Long call, Trump Pressures Georgia Secretary of State to Recalculate the Vote in his Favor*, WASH. POST (Jan. 3, 2021), https://www.washingtonpost.com/politics/trump-raffensperger-call-georgia-vote/2021/01/03/d45acb92-4dc4-11eb-bda4-615aaefd0555_story.html; German Lopez, *New Audio Proves It: Trump Deliberately Deceived America About the Coronavirus*, VOX (Sept. 9, 2020), https://www.vox.com/future-perfect/2020/9/9/21429166/trump-woodward-rage-coronavirus-covid-19-pandemic.

Amendment expression—and government could not punish or censor the distribution of such a video. This would arguably be true even if the video had little relationship to a matter of public concern: As the Second Circuit observed in *Universal City Studios v. Corley*, the First Amendment protects "[e]ven dry information, devoid of advocacy, political relevance, or artistic expression."[35]

But that the First Amendment protects our right to share accurate records does not necessarily mean it also protects our right to *falsify* the same records—and use them to deceive rather than inform. Consider again the possibility that those seeking reliable evidence of who has won military awards will look to an official government-created list on the web. Justices Kennedy and Breyer both suggest such a "database" or "register" as a way government can counteract impostors who pretend to have won a Medal of Honor.[36] We have a First Amendment right to share such information with others—for example, by sending a web link that directs them to register or database. But that doesn't mean we also have a First Amendment right to create a fake government website and use it to deceive audiences. That we have a First Amendment right to share a newspaper article from the Wall Street Journal does not mean we have a right to create a fake edition of the Wall Street Journal and try to pass it off as genuine.[37] We may have a free speech right to share a navigation chart, aeronautical chart, or other map to educate others about its content. But that does not mean we have a right to falsify the information in it before it is provided to an audience that predictably relies on its accuracy.[38] In this respect, the First Amendment protection that covers sharing of a data source

---

[35] Universal City Studios, Inc. v. Corley, 273 F.3d 429, 446 (2d Cir. 2001)
[36] *Alvarez*, 567 U.S. at 729; *Id*. at 738 (Breyer, J., concurring).
[37] *See, e.g.*, Blitz, *supra* note 31, at 64, 104.
[38] *See infra* text accompanying notes 213-217.

(where one doesn't have a broad right of falsification) may differ from First Amendment protection of the speech one authors or creates (where one does).

Courts should therefore consider whether, and when, the same asymmetry exists between genuine video or audio records and the deepfakes that emulate them. Even when we have a First Amendment right to share authentic footage from a security camera, a police body camera, or footage someone's cell phone has captured of a public event, that does not necessarily mean we have an equally strong First Amendment right to share a deepfake designed to give fictional events the appearance of such genuine footage. It is this latter *deceptive* use of deepfakes—not the artistic use of the same technology—that most deeply concerns many of those who are writing, or proposing laws on, deepfakes—and it is this deceptive use of deepfakes which this article argues is largely outside the scope of First Amendment protection.[39]

Part I will more fully explain the difference between artistic and different deceptive uses of deepfakes. More specifically, I will provide three analogies for deepfakes—one of which (to fictional stories and movies) places deepfakes squarely within the First Amendment's coverage and another (to counterfeit objects or environments) places it quite firmly outside of it. Between these two, in a sense, is the analogy I have already explored above between deepfakes and false statements of fact. Applying *Alvarez* to deepfakes described by such an analogy would place them

---

[39] *See infra* Part III. Jared Schroeder offers a somewhat similar argument, drawing on both US and European law, that the First Amendment should be understood to allow for the application of what he calls a "safeguarding principle," allowing regulation of deepfakes where necessary to protect "democratic discourse" from "deepfakes that damage the flow of information, rather than those that parody, comment, or challenge ideas." *Free Expression Rationales and the Problem of Deepfakes within the EU and US Legal Systems*, 70 SYRACUSE L. REV. 1171, 1202 (2020).

together with artistic deepfakes in the realm of core First Amendment speech.

But Part II will look more closely at this approach to deepfakes—one that treats them as analogous to verbal lies, and thus shielded by the First Amendment protection that *Alvarez* extends to verbal lies. It identifies some difficulties with applying *Alvarez* to deepfakes and looks briefly at how these are related to broader critiques that scholars have made of *Alvarez*'s analysis of the First Amendment status of lying.

Part III then more clearly explains why there are reasons to retain *Alvarez*'s protection for verbal lies—but those reasons do not apply to intentionally deceptive deepfakes. Rather, when deepfakes deceive, they typically function as "non-testimonial falsehoods" akin to deceptions carried out with counterfeit objects or environments. The First Amendment, I will argue, should not protect a right to intentionally create or promulgate non-testimonial falsehoods. In fact, Part III will argue, a closer look at *Alvarez* reveals another better analogy for deepfake videos: They are less like verbal lies than they are like a fake government website, something the First Amendment would not give Xavier Alvarez (or anyone else) a right to create and use as a tool of deception.

Parts IV and V will then address a significant complication for this account of deepfakes' First Amendment status: The line between the protected artistic and expressive use of deepfakes and the unprotected deceptive use of deepfakes is a difficult one to mark. Changing technologies blur this boundary line. The rise of deepfake technology itself is perhaps the clearest example. It extends an artist's control over the content of her film. Where she might once have been able to introduce authorship into the film only with artistic choices about how to capture and edit camera footage of external events, she can—with deepfakes—dispense with the

external events and instead create the footage from her imagination (with help from artificial intelligence). Deepfakes may therefore unsettle assumptions that viewers bring to videos: In a deepfake-filled future, audiences may bring the same skepticism to every video record they see as that which listeners are currently expected to bring to verbal statements.

But Part IV will argue that this development should not lead courts to simply erase or abandon the First Amendment boundaries described earlier. Listeners may have to bring skepticism to speakers' possibly dishonest statements. But First Amendment law should not condemn them to live with the same doubts about other traditionally more reliable sources of information—simply because technology has given a potentially manipulative speaker ways to exercise control over those as well. Rather, they should follow the example of Fourth Amendment law cases that have engaged in "equilibrium-adjustment" to prevent technological changes from unduly shrinking either the private space individuals need to find shelter from government surveillance—or the public space that leaves law enforcement with the room it needs to conduct investigations. In the First Amendment context, courts should assure that changing technologies do not give individuals unfettered authorship in areas where this is inconsistent with reliance interests.

Part V looks at a related challenge. Video recordings often blur together artistic and informational dimensions of this medium. Documentarians, for example, tell stories about factual events, using actual footage of those events. But they also make numerous artistic choices in filming and editing that footage—and sometimes weave actors' reenactments of events or scenes they have helped to shape (rather than simply captured on camera). Those who post a video on social media will often edit it before doing so. Moreover, it may be difficult to tell whether a video that deceives its audience was

intended to do so. Consider again a video depicting a fictional Medal of Honor ceremony. It may well be unclear to viewers of a video whether the person creating or sharing it intended it to be seen as a real event or as a vivid fantasy. In political discourse as well as artistic expression, audiences should perhaps be expected to bring the same skepticism to shared videos as they do to others' statements. But as Part V explains, courts have not interpreted the First Amendment to leave audiences this helpless. They have allowed government to protect the integrity of certain kinds of authoritative records.

As Part VI explains, this doesn't mean that others' presentation of a deepfake as real is entirely without First Amendment protection. In an age where video and audio can already be easily edited in certain ways, there is frequently uncertainty on the part of viewers and listeners about where a camera-generated record ends and the video-makers' or -editors' own creative contribution begins. Even security camera footage can be folded into artistic, political or other expression—and when it is, government would face some First Amendment limits on regulation of how it is altered. This does not mean, however, that such First Amendment limits should leave government powerless to counter deceptions caused by fabricated non-testimonial evidence. Rather, courts should view such uses of deepfakes or other non-testimonial evidence as being in a First Amendment middle ground, where government can regulate them subject to "intermediate scrutiny" and "viewpoint neutrality" requirements designed to let government further its significant interest in preventing viewers' deception—while preventing it from using this interest as an excuse to target the expressive components of fabricated or altered videos. Often, the measures most likely to survive such a judicial analysis will likely be rules that require those who create or share deepfakes to *disclose*

that they are deepfakes—or measures that safeguard the effectiveness of authentication technologies and practices of private actors.

I. THREE ANALOGIES: DEEPFAKES AS FABRICATED REALITIES, CREATIVE FICTIONS, AND FALSE TESTIMONY

What does the Constitution permit government to do about deepfakes? A common assumption has been whatever government does in this regard will be constrained by the First Amendment because, however dangerous deepfakes may be, they constitute expression. As I have written in prior scholarship on the First Amendment status of deepfakes, "[g]iving the government too much power to control how we use image-altering technology risks empowering it not only to prevent thorough deception, but also to restrict how we tell stories or otherwise express ourselves with [such] technology."[40] Video recordings, in other words, are familiar tools for story-telling and other expression.

However, on closer examination, videos—and the deepfakes that emulate them—are more complex and multifaceted. They aren't *only* vehicles for artistic expression. They are also means by which speakers convey factual information that a speaker wishes others to believe and understand: They constitute the speakers' "testimony" or perhaps, sharing of information that comes unaccompanied by any assertion. And sometimes, they are not even that. They are records that come to viewers not from a speaker who creates and shapes the video but directly from a camera that has captured and stored the light that has etched upon it a record of external events. In this part, I take a closer look at each of these three roles that videos can play: (1) video recordings as raw footage that a viewer obtains directly from a camera or other machine, (2)

[40] Blitz, *supra* note 31, at 114.

video recordings as fictional story-telling or other art, (3) video recordings as testimony or a speaker's sharing of factual information.

A. Deepfakes as Fabricated Realities

In 2002, a couple beginning a vacation at the Las Vegas Hard Rock hotel and casino were horrified, upon entering their room, to find a homicide victim lying in a pool of blood.[41] Before they could leave the room, they were confronted by hotel security guards asking about the murder. This nightmarish experience did not last long. The actor, Ashton Kutcher, soon appeared and revealed that the corpse was fake and the security guards questioning them were actors.[42] The unsettling illusion the couple encountered was part of a pilot episode of a hidden-camera TV show called "Harassment."[43] They then sued the show's network, MTV, for fraud and emotional distress.[44]

Deepfakes, as I noted above, are often viewed as visual lies. But the experience just described provides an alternative analogy. Rather than see a deepfake video as a high-tech equivalent of a false statement like that in *Alvarez*, we might see it as a digital equivalent of such a false object or environment. Imagine, for example, that the couple in the scenario above doesn't see a fake corpse directly in front of them. They instead receive a video call, and see, on the screen, a digitally-generated, but very real-looking corpse on their front yard. The call comes from someone who appears (and speaks)

---

[41] Gary Susman, *MTV Is Sued Over Corpse Prank*, ENTERTAINMENT WEEKLY (June 13, 2002), https://ew.com/article/2002/06/13/mtv-sued-over-corpse-prank; *Couple Sue Over TV Corpse "Prank,"* BBC NEWS (June 13, 2002), http://news.bbc.co.uk/2/hi/entertainment/2042466.stm.

[42] Susman, *supra* note 41.

[43] While the show was canceled after the lawsuit over this incident, it was revived as the show, *Punk'd*, which played pranks on celebrities. *See Punk'D*, INTERNET MOVIE DATABASE (IMDb), at https://www.imdb.com/title/tt0361227/.

[44] Susman, *supra* note 41.

hysterically in the guise of one of their neighbors

If this analogy has any validity, what does it tell us about deepfakes' First Amendment status? Deceiving someone with such a physical fake crime scene and a fake corpse seems unlikely to count as First Amendment expression. One could, perhaps, make a case that such trickery counts as First Amendment expression when it is an integral part of performance art, theater performance, haunted house, or interactive game (in an Escape Room, for example)—or perhaps even a reality television show.[45] But although a TV show portraying such deception would be speech, the underlying deception it portrays likely isn't. The couple encountering the fake corpse did not view it as part of any communicative act. Nor did they see it as sculpture or any other kind of artistic expression.[46] The success of the prank depended on the couple believing the corpse was actually a corpse, *not* the creation of an artist, author, or other speaker. This likely places it outside the scope of the First Amendment: From the audience's perspective, the crime scene is not a message for them[47] nor part of any kind of "inherently expressive" social practice.[48]

This kind of deception then is *not* a First Amendment equivalent of Xavier Alvarez's false autobiographical story—or any other kind of lie or tall tale. And it is helpful to begin to understand why. As noted in the introduction, First Amendment gives me a right to author books, plays, movies, Tweets, and a host of other

---

[45] *See infra* Part V-B.
[46] Susman, *supra* note 41.
[47] The fake crime scene thus will not count as First Amendment expression under the "Spence test" that courts sometimes use to determine whether certain conduct constitutes expression. *See infra* text accompanying notes 124, 130-134.
[48] The fake crime scene thus will not count as First Amendment expression on the ground that it is part of an "inherently expressive" social practice. *See infra* text accompanying notes 125-126, 128-129.

communications.[49] It does *not* give me a right to act as the "author" of another person's environment or her perceptions of that environment. I would almost certainly not be engaging in First Amendment speech, for example, if I somehow induced in my audience a visual hallucination of a dead body.[50] If I manipulate her perceptions so that she sees something that is not there, I am not presenting her with a claim that she or others would perceive as coming *from me*. I am rather exercising control over her own perceptions in such a way that my "authorship" is hidden. I am manipulating her thinking rather than appealing to it.[51] If it is not First Amendment speech when I cause this kind of hallucination from the "inside" of a person's mind, then it is not clear why it would be First Amendment speech when I generate the equivalent kind of illusion from the "outside" with a fake crime scene.[52]

Why then, one might ask, should the digitally-fabricated reality in such a video call—or a similar scene in deepfake security

---

[49] *See* Schacht v. United States, 398 U.S. 58, 63 (1970) (finding unconstitutional the conviction of actor for wearing military uniform in a play); Burstyn v. Wilson, 343 U.S. 495 (1952) (motion pictures are First Amendment speech); Geiger v. Dell Publ'g Co., 719 F.2d 515, 516 (1st Cir. 1983) (finding that the First Amendment limits the circumstances under which book content on matters of public figure can be liable for defamation); ; Tobinick v. Novella, 848 F.3d 935, 952 (11th Cir. 2017) (First Amendment protection for blog posts).

[50] As Thomas Scanlon has argued, the First Amendment's free speech protection almost certainly do *not* extend to the subliminal influence over another person's thinking that some have worried (albeit without factual support) might be accomplished with advertisements or rock songs: Subliminal messages, then, are not "speech" under the First Amendment. *See* T.M. Scanlon, Jr., *Freedom of Expression and Categories of Expression*, 40 U. PITT. L. REV. 519 (1979).

[51] David A. Strauss argues that speech that generates a certain reaction in a listener should presumptively protected by the First Amendment only when it appeals to a listener's faculties in a way "a rational person would value," not when it instead uses "autonomy-invading manipulation." *Persuasion, Autonomy, and Freedom of Expression*, 91 COLUM. L. REV. 335, 355, 366 (1991).

[52] Neil Levy has argued that "[u]nless we can identify ethically relevant differences between internal and external interventions and alterations [in the mind], we ought to treat them on a par." *See* NEIL LEVY, NEUROETHICS: CHALLENGES FOR THE 21ST CENTURY 61-62 (2007). Levy is here arguing for ethical not constitutional parity. But his argument has force in understanding the First Amendment implications as well.

camera footage—count as any more expressive than its physical equivalent? That we have a right to insert false content into our own words (or other expression), even our own factual assertions or "testimony," doesn't necessarily mean we have a right to make others *see* what is not there. The First Amendment protects us when we persuade, inform, enlighten, or entertain each other by making claims, offering arguments, or expressing feelings in ways that an audience is invited to assess and react to. Matters are different when we seek to shape others' beliefs not by inviting them to consider claims in a communication or work of art but rather by engaging in a God-like shaping of their surrounding environment or their means of perceiving it.

To be sure, there has not been much analysis of the First Amendment status of counterfeit realities—and that is perhaps because it is not a simple matter for individuals to create them. The above-described MTV prank required a tremendous amount of work (and resources) to exert control over only a small portion of a person's surroundings. The production staff for the TV show, Harassment, had to create a replica of a corpse and recruit and pay actors to portray security personnel. Even with all this work, the illusion was confined to a single room. Nor is it easy for individuals to take on a false physical appearance or speak in another's voice. While they might use a uniform or ID badge to give themselves a status they do not really have—by imitating a police officer, for example—they can't easily use a physical guise to take on the appearance of a person's friend or relative.

But just as advances in computer and Internet technology give government officials and companies new ways to monitor our private activity, so they give individuals new ways to control our perception of our (increasingly digital) environment. Our reading habits become easier for book publishers and other companies to

monitor when we switch from reading physical books, in the privacy of our own home, to eBooks or other documents on Internet-connected computers.[53] It is likewise easier for others to manipulate our perceptions when our attention is focused on digital images and sounds rather than an in-person encounters with flesh-and-blood people and physical objects.[54] Our experience of reality is increasingly virtual in this way. Millions have had little choice but to move much of their activities online to minimize risk during the coronavirus pandemic in 2020. As one New York Times essay noted, they had to "buil[d] virtual world[s] to replace a broken physical one."[55] But significant migration to virtual settings occurred long before then. We can't hear the voices, or see the faces, of far-away family members, friends, or business colleagues without the extended perception made possible by computer and communications technologies.[56] In the future, developments in virtual reality may be able to let us feel as though we are physically present in a remote location.[57]

The same technologies that extend our perception make it more vulnerable. Hackers might intercept and edit our sources of

---

[53] *See* Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981 (1996)

[54] *See, e.g.,* PHILIP P. PURPURA, SECURITY LOSS AND PREVENTION: AN INTRODUCTION 245 (2013) (describing use of video hacking and alteration to "plant evidence").

[55] Kevin Roose, *The Coronavirus Crisis Is Showing Us How to Live Online*, N.Y. TIMES (Mar. 17. 2020), (updated Apr. 2, 2020), https://www.nytimes.com/2020/03/17/technology/coronavirus-how-to-live-online.html.

[56] *See* LORENZO CANTONI & STEFANO TARDINI, INTERNET 54 (2006) (describing how telephones and televisions create a sense of presence at far-away locations).

[57] *See* Wijnand Ijsslestein and Giuseppe Riva, *Being There: The Experience of Presence in Mediated*
Environments, *in* WIJNAND IJSSLESTEIN & GIUSEPPE RIVA, BEING THERE: CONCEPTS, EFFECTS, AND MEASUREMENT OF USE PRESENCE IN SYNTHETIC ENVIRONMENTS (2003) ("With the advent and improvement of immersive displays, computing and network technologies, and interactive computer graphics, we can create more accurate reproductions and/or simulations of reality than were previously possible.").

perceptual knowledge before they reach us. They might alter a video- or audio-recording en route to us from a camera or microphone (or while it is in storage, before we access it).[58] Someone on the other end of a phone call or Internet video-chat might speak to us in someone else's digitally-recreated voice or appearance.[59]

This analogy I have just drawn, suggesting deepfakes are more like counterfeit objects or environments than verbal lies, might immediately elicit an objection. There is a stark and constitutionally-significant difference, one might argue, between a fabricated *physical* environment and a video that falsely *depicts* that environment with vivid realism. We generally have little choice but to assume that the physical environment around us is real. We cannot expect to stay sane (or be regarded as such) if we move through life with a pervasive skepticism—in the grip of which we doubt the reality of every person and thing around us.[60] By contrast, we can and do often doubt that the *depictions* of reality offered to us by others—whether they come to us in the form of possibly dishonest statements, or photographs or videos that can be edited or

---

[58] *See* Sonia Klug, *A.I. Is Changing How You See the World*, ONE ZERO (Sept. 24, 2019) ("Our impressions of the world are heavily influenced by the images we see online.")

[59] *See* Catherine Stupp, *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case*, WALL ST. J. (Aug. 30, 2019), https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

[60] *See* Ernest Sosa, *Knowledge: Instrumental and Testimonial*, *in* JENNIFER LACKEY & ERNEST SOSA, THE EPISTEMOLOGY OF TESTIMONY 122 (2006) (stating that "[e]pistemically justified trust in our senses is a gift of natural evolution . . . We accept their deliverances at face value as a default stance, and properly so."). For examples of paranoid science fiction stories that illustrate the disorientation that arises with radical doubt of one's reality, *see* Marc Jonathan Blitz, *Freedom of 3D Thought: The First Amendment in Virtual Reality*, 30 CARDOZO L. REV. 1141, 1158-59, 1229-1230 (2008); Blitz, *supra* note 31, at 59-61.

doctored to present a false or misleading picture.[61] We are still entitled to trust (in most cases) the reality of what we see in front of us with our own eyes, or hear directly with our own ears. What we can no longer do is trust blindly in the accuracy of a photo, or video- or audio-footage, created *by someone else*. The fabrication or doctoring of video and photos, in other words, is not a God-like reshaping of our environment, and our perceptions of it. It is rather a familiar exercise of human creativity.

In this respect, one might argue, it *is* much like the tall tale Xavier Alvarez generated with words. Whether we hear a false statement from him claiming he has won a Medal of Honor, see a fake photograph showing him receiving such a Medal from President Reagan, or watch a deepfake video portraying this scene, we are acting as audiences for some means by which people *represent* the world—with words, photographs, or video recordings, all of which they can conceivably alter or edit. We are not watching a set of events unfold directly in front of us. If the First Amendment protects falsity even when we add it to factual representations (as we do in false statements), then why not also the factual representation that occurs in a video recording or video stream?

There are, however, reasons to doubt that video-recordings and the deepfakes that emulate them should always be treated by courts as works of authorship—into which their creators have a First Amendment right to insert falsity. This characterization of deepfakes is, after all, not the one that dominates the current discourse about them by journalists, technologists, and lawmakers.

---

[61] *See* Hany Farid, *Digital Doctoring: Can We Trust Photographs?* (2007), https://farid.berkeley.edu/downloads/publications/deception09.pdf, (pointing out that "photography lost its innocence many years ago" and that the "long history of photographic trickery" dates back to the early days of photography in the mid-nineteenth century).

Far from treating deepfakes as an unremarkable new twist on old-fashioned lying, many lawmakers and journalists portray them as radically different—with potentially catastrophic impacts. Technologist Aviv Ovadya describes deepfake technology as entailing "the distortion of reality itself."[62] Other articles likewise describe deepfakes and similar technologies as fundamentally destabilizing our sense of what is real. Many of them insist that, with the rise of deepfakes, "seeing is no longer believing."[63] One writer, for example, worries that "the onslaught of deepfakes" will either dangerously deceive—or have the consequence, also harmful, that the new "norm on the Internet may be to distrust everything." [64] These characterizations of deepfakes portray them as falsifying our perceptions, not merely the accounts we hear from others.

These accounts may exaggerate the extent to which deepfakes will undermine our sense of reality. Even in a world where cannot trust video evidence, we will still be far from a

---

[62] *See* Aviv Ovadya, *What's Worse Than Fake News? The Distortion of Reality Itself*, WASH. POST (Feb. 22, 2018), https://www.washingtonpost.com/news/theworldpost/wp/2018/02/22/digital-reality.

[63] *See, e.g.,* Noemie Kempf, *Seeing Is Not Believing: How Deepfakes Are About to Transform Our Reality*, MEDIUM (Oct. 23, 2019). https://medium.com/la-nouvelle-frontiere/seeing-is-not-believing-how-deepfakes-are-about-to-transform-our-reality-619312658152; Carolyn Purnell, *Do We All Still Agree that "Seeing Is Believing"?,* PSYCHOLOGY TODAY (June 23, 2020), https://www.psychologytoday.com/us/blog/making-sense/202006/do-we-all-still-agree-seeing-is-believing, (arguing deepfakes undermine our collective notions of truth); Rob Toews, *Deepfakes Are Going to Wreak Havoc on Society. We Are Not Prepared*, FORBES (May 25, 2020), https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared, (noting that "[e]xperts predict that deepfakes will be indistinguishable from real images before long."); *Video Manipulation Technology Poses Growing Threat to National Security, Experts Say*, WTVMNEWS (Jan. 29, 2019), https://www.wtvm.com/2019/01/30/video-manipulation-technology-poses-growing-threat-national-security-experts-say, ("Seeing is no longer believing")*;* William A. Galston, *Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics*, BROOKINGS INSTITUTION, Jan. 8. 2020

[64] *The Deepfake Threat*, COMPUTER BUS. REV. (June 20, 2018), https://techmonitor.ai/techonology/data/deepfake-crisis

Matrix-like environment where anything and everything can be an illusion. We may be able to adjust to showing greater skepticism towards videos, just we have survived the rise of Photoshop and other programs for creating realistic pictures. Ultimately, we won't know how destabilizing deepfakes will be until we can answer empirical questions about viewers' psychological responses to deepfakes, both in the present, and in future eras where deepfakes are more familiar (and thus, perhaps, less likely to deceive). There is at least some reason to think people can see deepfakes without believing they are true. When watching a dramatic film that uses special effects to make dragons look real, animate a deceased actor for a new role, or manipulate historical footage, we are able to recognize what we see as the fictional creation of a filmmaker despite its extraordinary realism. This at least suggests that we will also be able to do so when such special effects tools expand beyond movies—and begin to infiltrate into video we find on social media or receive in text messages from friends.

However, it is unclear how well we will continue to be able to identify video as a fake without the markers that normally identify such movies or television shows as fiction—and that shows people we know as real (and perhaps know personally) speaking and acting as they do when we normally see or encounter them. Those who write about virtual reality have already noted that the "sense of presence" it gives us may sometimes make virtual experiences *feel* very real even when we *know* they are not.[65] To the extent our

[65] One of the most distinctive elements of a virtual experience is that it creates a sense of "presence" in—and not just a sense that one is engaged in view of—the world. *See* Marc Jonathan Blitz, *The First Amendment Video Games and Virtual Training, in* WOODROW BARFIELD & MARC J. BLITZ, RESEARCH HANDBOOK ON THE LAW OF VIRTUAL AND AUGMENTED REALITY (2018); Guiseppe Riva et al., *Neuroscience of Virtual Reality: From Virtual Exposure to Embodied Medicine*, 22 CYBERPSYCHOL. BEHAV. SOC. NETWORKING 82, (2019) (noting that "VR

cognitive processes have developed around a strong tendency to assume that what appears directly in front of us is generally real, we may not be able to easily adjust to a world where this often isn't the case. This may be true of deepfakes as well as virtual reality. Cass Sunstein writes that deepfakes have "a unique kind of authenticity; they are more credible than merely verbal representations. In a sense, they are self-authenticating. The human mind does not easily dismiss them, and if it does, there is some part of it that remains convinced."[66]

Our psychological reaction may also be a function of how frequently we encounter deepfakes: If they occur somewhat rarely, we may be less likely to expect or recognize them as fake. If they are pervasive, we might instead view every video with skepticism (even if it takes on an immersive virtual-reality form). But although such skepticism may shield us from deception, it could leave us with paralyzing uncertainty in its place. As Chesney and Citron put it, we might prevent the "truth decay" that arises from successful deception only by entering into the "trust decay" that arises with constant doubt of all video evidence.[67] It is also important, of course, to see whether the trust any individuals have traditionally placed in video might be shifted to some other technology or social

---

system, like the brain, maintains a model (simulation) of the body and the space around it"): Pietro Cipresso et al., *The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature*, FRONTIERS. PSYCHOL. (Nov. 6, 2018), https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02086/full, ("Presence is a complex psychological feeling of "being there" in VR that involves the sensation and perception of physical presence, as well as the possibility to interact and react as if the user was in the real world."); ; *see also Hearing Before the Subcomm. on Science, Technology and Space*, 105th Cong. 14 (1997) (statement of S. Kicha Ganapathy, Member, Technical Staff, Multimedia Communications Research Laboratory, Bell Laboratories).

[66] Cass R. Sunstein, FALSEHOODS AND FREE SPEECH IN AN AGE OF DECEPTION 119 (2021).

[67] Chesney & Citron, *supra* note 6, at 1785-86.

marker for distinguishing true and false records of the past.

The answers to these questions, when they come, may well play an important role in the First Amendment analysis of deepfakes, and how likely we are to perceive them as possible deceptions and are able to navigate around them when we do. But First Amendment analysis doesn't have to simply await more confident answers to all of these empirical questions. There is a normative component a well as an empirical component to such analysis. In the Fourth Amendment context, as explained below,[68] courts have distinguished between spaces, like the home or a private business, where we can reasonably expect to be shielded from police surveillance, and others where we cannot expect such privacy. It is conceivable such judgments can and should be informed by empirical studies of where people actually *do* expect privacy —and some scholars have provided such studies.[69] But courts (as explained below) have marked out and maintained the division between private and public spaces in Fourth Amendment law without relying on such studies. They have instead relied on their sense of what our social practices have treated as private—and of how to extend these practices to emerging technologies. For example, the Supreme Court has made it clear that the Fourth Amendment treats the home as a private space even when a resident doesn't view it as such.[70] Social conventions and legal traditions set

---

[68] *See infra* Part IV.B.

[69] *See* Christopher Slobogin & Joseph E. Schumaker, *Reasonable Expectations of Privacy and Autonomy in Fourth Amendment Cases: An Empirical Look at "Understandings Recognized and Permitted by Society,"* 42 DUKE L.J. 727, 730 (1993) (describing the results of "a survey of 217 individuals to ascertain their understanding of the interests implicated by various types of police investigative techniques" in order to provide an empirically-informed data for understanding what expectations of privacy individuals regard as "reasonable").

[70] *See* Smith v. Maryland, 442 U.S. 735, 741 n.5 (1979) (noting that "if the Government were suddenly to announce on nationwide television that all homes henceforth would be subject to warrantless entry," no one would have a subjective

it aside as a space where individuals can expect privacy.

In First Amendment law, this article suggests, courts similarly have to mark out what spaces are appropriate for individuals to exercise largely unfettered, First Amendment-shielded authorship, and what realms of activity are instead *insulated against* such authorship, so that government may safeguard the reliance we place in the information we find there.

Consider three examples of situations where deepfake technology can invade and distort processes where we might normally trust video or audio footage as conveying an accurate picture of the world, and where courts may understandably view others' manipulations as non-expressive. First, imagine a case where deepfake technology is used not to produce a video that purports to be a recording that documents a past event—but rather a Skype or other Internet chat in which I receive a video call from someone who looks and speaks like an old friend from college—but is actually someone else in the digital guise of that friend's appearance and voice. As Aviv Ovadya writes, this kind of deception is one of the risks raised by emerging methods of video- and audio-fabrication: it can potentially allow someone to "simulat[e] your spouse's voice on the phone asking for a bank password,"[71] or perhaps to simply pry into your private life, or unsettle you, by drawing you into a more intimate conversation that you would have in a cold call from a stranger.

Second, imagine that I wish to examine a live feed from an automated security camera mounted on a telephone pole in my

---

expectation of privacy in one's home going forward, but suggesting that courts should continue to treat homes as spaces set aside by social conventions for privacy—and consider to treat entry into the home as a Fourth Amendment search requiring a warrant).

[71] Ovadya, *supra* note 63. 8

neighborhood, or perhaps from a drone that hovers above the streets of that neighborhood. This might be a government-operated camera. Or it might be a camera operated by a local business or private organization for the benefit of security-minded citizens in the neighborhood. If, when I view this live feed I am (without realizing it) misdirected to a fake live feed, I will likely view it as a genuine live feed from an automated recording process—not as a video with a possibly dishonest and mischievous author. The same confusion may arise in a case where a person is searching not for a CCTV or drone camera's live feed, but for stored footage online.[72] In other words, the viewer will not be viewing what the security video displays as any speaker's claim or narrative. She will rather see it as a kind of perception-at-a-distance—or "telepresence"—mediated by a machine rather than by human judgments about how to portray a certain scene.

Third, consider a situation that is a bit more distant from our current everyday experience than are video chats or viewing of security camera footage: the use of video-based "lifelogging." Anita Allen describes "lifelogging" as a technology that can "record and store everyday conversations, actions, and experiences of their users, enabling future replay and aiding remembrance."[73] As Allen notes, the increasing power of computer devices might soon "enable people fully and continuously to document their entire lives."[74]

In each of these cases, the video feed or recording does *not* typically serve as a vessel for conveying the message of an author or

---

[72] *See* Ingra Kroener, CCTV: A TECHNOLOGY UNDER THE RADAR? 62 (2016).
[73] Anita L. Allen, *Dredging Up the Past: Lifelogging, Memory, and Surveillance*, 75 U. CHI. L. REV. 47 (2008).
[74] *Id.* Not all lifelogging takes the form of videorecording—lifeloggers use all kinds of different data to construct moment-by-moment archives of their existence, or "quantify" their histories including GPS records and health data—but videorecordings are the kind of lifelogging most vulnerable to being fabricated or distorted by the kind of deepfake technology that is the focus of this article.

editor. The video, of course, could conceivably be hacked or technologically manipulated. Or it might be a fake feed that, with the aid of deepfake technology or other editing tools, displays events that aren't happening or haven't happened. Perhaps the video watcher is aware of that possibility. But it seems implausible to expect her to treat the video with the same kind of skepticism that the First Amendment law assumes listeners will bring to verbal claims, or even with the kind of skepticism that people bring to photographs.

It is hard to see how video chats can serve their purpose, for example, if we constantly have to doubt whether the friends, acquaintances, or associates we were speaking with are really who they appear to be. Nor can we rely on security camera feeds to tell us if there is a dangerous situation, or an incident requiring investigation, in a world where such feeds can always be faked. Video lifelogging likewise doesn't serve its function if its log doesn't capture genuine footage of actual events to an ongoing, *unmanipulated* surveillance record that creates a trail of data that is "self-produced."[75] Deepfake alterations of such lifelog data would therefore distort archives that are supposed to be an unfiltered record of one's life. To the extent we use such lifelogs, as Allen suggests we might, as "a less fallible and selective adjunct to human memory,"[76] deepfake alterations of such archives would distort that memory.

Deepfakes then may not shatter our sense of reality—or sever the connection between seeing and believing in all contexts—to the extent that many current accounts warn they will. But they will at least likely raise significant problems for our social

---

[75] *Id.* at 54.
[76] *Id*. at 50.

practices—and uses of mediated perception—that can't be cured solely by increased skepticism. It may well be the case that the solutions to these concerns lie in technological answers, rather than government restriction. We may, for example, find new technologies for creating trustworthy records of past events—or secure visual channels for watching them as they occur. We could then respond to the uncertainties that deepfakes generate by shifting our trust to new markers of authenticity.[77] But as I will argue more fully below,[78] even if we may ultimately be *able* to adjust to the threats that arise when technology lets authors—or other would-be manipulators—expand their control over digital sources of knowledge or digital channels of perception, that doesn't mean that the First Amendment should force us to do so. Just because technology gives speakers a *capacity* to author parts of their audiences' environment—physical or digital—doesn't mean they should have a First Amendment *right* to do so.

B. Deepfakes as Creative Fictions

If deepfakes seem largely if not entirely outside the scope of First Amendment expression when they are the equivalent of deceptive counterfeit objects or environments, they are squarely within it when they play a different role. Consider again what deepfakes allow video creators to do. Harnessing the power of artificial intelligence, deepfake technology allows an author to use, as raw material for visual storytelling, the images and voices of real people. As noted earlier, AI researchers have famously demonstrated this technology by creating videos showing Barack Obama, Richard Nixon, and other leaders delivering speeches they

---

[77] *See infra* text accompanying notes 285-287.
[78] *See infra* text accompanying 288-299.

never gave.[79]

Deepfakes are hardly the first technology that allows an author to transform flesh-and-blood individuals into characters in a story. Roughly 2400 years ago, Plato did precisely this when he made his teacher, Socrates, the protagonist of almost all of his dialogues.[80] Although some of Plato's accounts were based on actual events (like Socrates' execution) and likely on arguments actually made by Socrates, others used Socrates as a vehicle for Plato's own ideas. As Peter Adamson writes, "[t]he Socrates of the early dialogues is surely based on the real Socrates; yet he is also a fictional construct, used to explore certain philosophical theories."[81] Numerous authors have similarly made fictional characters out of other historical figures such as Julius Caesar and Cleopatra,[82] Julian the Apostate,[83] Shoeless Joe Jackson and J.D. Salinger,[84] Lenny Bruce,[85] Lucille Ball,[86] and Hillary Clinton.[87] With the rise of graphic novels, dramatic movies, and animations, creative minds have obtained other means to portray people doing things they never did.

All of these kinds of storytelling are staunchly protected by the First Amendment. Government may not censor literature. And during the twentieth- and twenty-first centuries, the Supreme Court has extended this First Amendment protection to other, more technologically-advanced forms of story-telling. Movies have

---

[79] *See supra* notes 7-8.
[80] *See* Giovanni Reale, From the Origins to Socrates: A History of Ancient Philosophy 195 (John R. Catan ed., trans., 1987).
[81] Peter Adamson, *Who Speaks for Socrates*, 122 PHILOSOPHY NOW (2017).
[82] *See* WILLIAM SHAKESPEARE, JULIUS CAESAR; *Julius Caesar: History vs. Drama*, BILL/SHAKESPEARE PROJECT (Nov. 26, 2014) (noting "fudging of facts" in Shakespeare's play).
[83] *See* GORE VIDAL, JULIAN: A NOVEL (Vintage International 2003).
[84] *See* WILLIAM KINSELLA, SHOELESS JOE (First Mariner Books1999).
[85] *See The Marvelous Mrs. Maisel* (Prime Video 2017-2020).
[86] *See* DARIN STRAUSS, THE QUEEN OF TUESDAY (2020).
[87] *See* CURTIS SITTENFELD, RODHAM: A NOVEL (2020).

received First Amendment protection since 1952 when the Supreme Court decided *Joseph Burstyn v. Wilson*.[88] In that case, the Court struck down a New York statute that "permit[ted] the banning of motion picture films on the ground that they [were] 'sacrilegious.'"[89] "[M]otion pictures," said the Court, "are a significant medium for the communication of ideas" and "may affect public attitudes and behavior in a variety of ways, ranging from direct espousal of a political or social doctrine to the subtle shaping of thought which characterizes all artistic expression."[90] In *Brown v. Entertainment Merchants Association*, the Court likewise found that video games constituted First Amendment expression for largely the same reason. "Like the protected books, plays, and movies that preceded them," it said, "video games communicate ideas—and even social messages—through many familiar literary devices (such as characters, dialogue, plot, and music) and through features distinctive to the medium (such as the player's interaction with the virtual world)."[91]

What lessons do these examples hold for deepfakes' First Amendment status? Where deepfake technology is used by filmmakers or video game creators as a tool of computer animation, it is just as entitled to First Amendment protection as are other tools of filmmaking. In fact, other forms of computer animation have already given filmmakers the power to seamlessly weave fictional characters and fictional events into historical footage. As mentioned in the introduction, the 1994 movie, Forrest Gump, featured vivid realistic scenes of its title character receiving a Medal of Honor from Lyndon B. Johnson and also showed him chatting with John F.

---

[88] Burstyn v. Wilson, 343 U.S. 495 (1952).
[89] *Id.* at 497.
[90] *Id.* at 501.
[91] Brown v. Entm't Merchants Ass'n, 564 U.S. 786, 790 (2011).

Kennedy.[92] Over a decade earlier, the movie Zelig told the story of a chameleonic 1920s celebrity, Leonard Zelig, in the form of a documentary that included film clips showing the protagonist standing beside Babe Ruth on the 1929 Yankees, participating in Al Capone's criminal enterprise, and conversing with Pope Pius XI.[93] Such cinematic reworking of history receives just as much First Amendment protection as any other form of artistic expression. And amateur video creators posting a deepfake-generated fantasy on YouTube would receive no less First Amendment protection than their professional counterparts.[94]

This does not mean that every deepfake that is the product of a video creator's imagination will count as protected speech. Even artistic expression that is typically shielded by the First Amendment might not be when it generates certain types of harms, or otherwise falls within certain traditionally recognized exceptions to free speech protection.[95] Obscene movies lack First Amendment protection.[96] So might a film that is designed to defame someone portrayed or referenced in it.[97] Deepfake videos in these categories would likewise be subject to criminal or civil liability. Rights of publicity might also be relevant.[98] They generally protect

---

[92] FORREST GUMP (Paramount 1994).
[93] ZELIG (Orion 1983).
[94] *See* Chesney & Citron, *supra* note 6, at 1769-71 (describing artistic uses of deepfakes as a form of special effects in movies).
[95] *See* United States v. Stevens, 559 U.S. 460, 468 (2010) (noting the First Amendment has long excluded a few historic categories of speech such obscenity, defamation, fraud, incitement, and speech integral to criminal conduct).
[96] *See* Paris Adult Theatre I v. Slaton, 413 U.S. 49, 57 (1973).
[97] *See* Lovingood v. Discovery Commc'ns, Inc., 800 F. App'x 840, 848-49 (11th Cir. 2020) (assuming that plaintiff could have established defamation in dramatic film portraying his participation in a NASA panel if he could establish filmmakers acted with "actual malice," even though on facts of the case, plaintiff failed to show that).
[98] *See* Jesse Lempel, *Combatting Deep Fakes Through the Right of Publicity*, LAWFARE (Mar. 30, 2018), https://www.lawfareblog.com/combatting-deep-fakes-through-right-publicity.

individuals against use of their name or likeness for commercial gain.[99] As Chesney and Citron point out, the "commercial-gain element sharply limits the utility of this model: the harms associated with deep fakes do not typically generate direct financial gain for their creators."[100] Some have nonetheless wondered if such a right may be invoked by celebrities and others depicted in one of the earliest and most common uses of deepfakes—the use of such technology to makes them appear as "actors" in pornography videos they had nothing to do with. Such deepfake pornography is typically different from deceptive uses of deepfakes—because it often comes with a disclaimer specifically identifying it as fiction. As Chesney and Citron emphasize, such use of deepfakes may nonetheless cause extraordinary harm—because, even with a disclaimer, the person depicted may feel humiliation and fear, and the "psychological damage may be profound" when "[f]ake sex videos force individuals into virtual sex, reducing them to sex objects." [101] First Amendment law allows room for suits based on emotional distress under certain circumstances—so it is possible such psychological harms could provide a basis for proscription.[102]

But these are exceptions to the general rule that artistic deepfakes would be as staunchly protected by the First Amendment as any other artistic creation. Moreover, they seem unlikely to raise concerns about "reality distortion." The use of deepfakes or similar special effects technology in movies like Forrest Gump doesn't

---

[99] *See generally* JENNIFER E. ROTHMAN, THE RIGHT OF PUBLICITY: PRIVACY REIMAGINED FOR A PUBLIC WORLD (2018).

[100] Chesney & Citron, *supra* note 6, at 1794. As they note, there is also a "gendered dimension of deep-fake
exploitation. In all likelihood, the majority of victims of fake sex videos will be female." *Id.* at 1773; *see also* Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1874 (2019).

[101] Chesney & Citron, *supra* note 6, at 1773.

[102] *See* Samantha H. Scheller, *A Picture Is Worth A Thousand Words: The Legal Implications of Revenge Porn*, 93 N.C. L. REV. 551, 582 (2015).

distort reality because there is no reality to distort; audiences *know* that the film is a work of imagination and the realism of the film just makes this fantastical story more vivid. Seeing *isn't* believing for audiences of such films. Someone who sees the CGI animation of tall blue aliens in the movie, Avatar, for example, knows that, despite the realism of these characters, they are the product of special effects—not actual footage captured from an alien world.[103] Even when a film is meant to recast a historical event (like the American revolution) as a dramatic film, its audience knows that what they see is a work of human authorship, not footage of the actual event.[104] Most audiences will likely know, watching a deepfake James Dean perform in the 2019 movie, Finding Jack, that the real James Dean has been dead for sixty-five years.[105]

This is true even where films reshape historical footage. Viewers of Forrest Gump and Zelig know that the scenes blending the lead characters into archival footage are fictions—not only because the protagonist in each of these movies is fictional, but because the actors who play those protagonists (Tom Hanks and Woody Allen) could not have participated as adults in events that took place during their childhood or before they were born.[106] Moreover, when viewers watch these movies, whether in a movie

---

[103] AVATAR (20th Century Fox 2009); FORREST GUMP (Paramount 1994).

[104] *See, e.g.*, THE PATRIOT (Columbia Pictures 2000); REVOLUTION (Warner Bros. 1985).

[105] *See* Dani Di Placido, *James Dean and the Rise of 'Deep Fake' Hollywood*, FORBES (Nov. 8, 2019, 3:22 PM), https://www.forbes.com/sites/danidiplacido/2019/11/08/james-dean-and-the-rise-of-deep-fake-hollywood/.

[106] *See* Lovingood v. Discovery Commc'ns, Inc., 800 F. App'x 840, 847 (11th Cir. 2020) ("A reasonable viewer would understand within the first two minutes that he is not watching a documentary film that consists mainly of historical footage and interviews with the historical figures. He would recognize the parts of the film that do use historical footage and understand that they are meant to depict literal history, and he would understand that most of the film uses actors to portray historical events with some amount of artistic license.").

theater or by streaming it on a computer, they see individuals credited with writing the story and script, with directing and editing the film, and with adding special effects. The creators of the film haven't made any effort to hide their authorship of the film. On the contrary, they claim credit for it.

That the First Amendment unquestionably shields artistic deepfakes, however, raises an initial difficulty for a simple rule that excludes deceptive deepfakes like those described above that present the cyberspace equivalent of counterfeit environments or objects. Outside of a movie theater or museum, it may be difficult to distinguish the two—and in the modern age, art isn't always confined to theaters and museums. Deepfakes—although they are essentially fictional animations—don't *look* like animations of the past. They look and sound exactly like footage captured by a camera and a microphone. And so, without context that provides other markings of their fictionality (such as credits identifying a scriptwriter and director), an artistic deepfake may well appear to be genuine camera footage.

Consider some examples. I said earlier that it would be hard to see how video chats could serve their purpose if those using them were constantly in doubt about who they were talking to. But far from treating live video chat interaction, or video messages, as entirely off-limits to artistic manipulation many users of services such as Snapchat, Facetime, Skype, and Zoom use "filters" to alter their appearance,[107] "virtual backgrounds" to make them appear in a different environment than the one they are in,[108] or "face

---

[107] *See* Matthew Cage, *How to Put Effect on Instagram Video Call*, SOMAG NEWS (Apr. 13, 2020), https://www.somagnews.com/put-effect-instagram-video-call/.

[108] *See* Teena Maddox, *Tips on Choosing a Realistic Zoom Virtual Background for Your Business Meetings*, TECHREPUBLIC (June 9, 2020),

swapping" software to exchange their face with that of another person (or that of a statute or painting, for that matter).[109] "Voice changer" apps can similarly modify the voice one speaks with on a telephone call or audio recording.[110] The existing versions of these filters are typically identified quite easily by listeners or viewers (Zoom users who see a cartoon kitten talking to them will know their interlocutor is using a filter).[111] But technological advances may make these filters harder to identify as such. New filters being offered for Zoom calls use deepfake technology to let you appear in the guise of a celebrity. As one article notes, while the filters are currently identifiable as fake "given a little more time and paired with a more convincing vocal" this might "be the beginning of a sinister new world of deepfake Zoombombing."[112] On the other hand, such deepfake use of Zoom could also be used for expressive purposes rather than for cybercrime or harassment.

Moreover, not all artistic films are fictional. Documentaries provide a cinematic examination of factual subject matter. But documentarians exercise an active and creative role in shaping the story that their film tells. And they sometimes shape the content of their films not only through their choices about how to shoot or edit

---

https://www.techrepublic.com/article/tips-on-choosing-a-realistic-zoom-virtual-background-for-your-business-meetings/.

[109] *See* Jason Hellerman*, How Will Disney's Face Swapping Change Hollywood?,* NO FILM SCHOOL (July 1, 2020), https://nofilmschool.com/faceswap-tech-disney; Jacek Naruniec et al., *High-Resolution Neural Face Swapping for Visual Effects*, 39 COMPUTER GRAPHICS F. (2020).

[110] Simon Hill & Paula Beaton, *The Best Voice- Changing Apps for Android and iOS*, DIGITAL TRENDS (Feb. 1, 2021), https://www.digitaltrends.com/mobile/best-voice-changer-apps/.

[111] In February 2021, for example, a lawyer mistakenly appeared at a Zoom court hearing with a filter that made him appear as a cat, but it was apparent to the judge that the lawyer was not really a cat. *See* Daniel Victor, *'I'm Not a Cat,' Says Lawyer Having Zoom Difficulties*, NY TIMES, Feb. 9, 2021, https://www.nytimes.com/2021/02/09/style/cat-lawyer-zoom.html.

[112] Stephanie Palmer Debrien, *Zoombombed by Elon? New Program Brings Deepfakes to Video Conferencing*, SMART COMPANY (Apr. 21, 2020), https://www.smartcompany.com.au/coronavirus/elon-musk-deepfake-zoom/.

it, but also in their interactions with the people or environment that is their subject. The documentary, Winged Migration, for example, captured stunning footage not only of the migratory journey of wild birds, but also the aerial gymnastics of birds the film-makers had trained, through "imprinting," to fly alongside the camera operators It also recreated a scene the film's director had seen of a bird damaged by an oil spill and did so with milk and vegetable dye rather than oil so as not to cause any harm to the bird.[113] The documentary, The Thin Blue Line, had the subjects of the documentary reenact some scenes that were never previously captured on video.[114] Deepfake technology can make it far easier to create vividly realistic artificial footage to fill gaps in what the film-maker has actually captured on camera, but it might, in the process, create confusion about what is real footage and what is a deepfake creation. Still, it would intuitively run counter to First Amendment doctrine to let government interfere with, and constrain, the artistic choices of documentary filmmakers (except where their documentary, or some aspect of it, is defamatory or falls into another category traditionally unprotected by the First Amendment).

C. Deepfakes as False Testimony

This Part has so far considered two analogies for deepfakes. First, they are, in some cases, like digital versions of fake objects or counterfeit environments. They emulate videos that are proxies for our perception—but then feed our perceptions false information

---

[113] Richard von Busack, *The Secret Life of Birds is Revealed in Jacques Perrin's Winged Migration*, METROACTIVE (May 15, 2003), http://www.metroactive.com/papers/metro/05.15.03/winged-0320.html.

[114] *See* Charles Musser, *The Thin Blue Line: A Radical Classic*, CRITERION (Mar. 25, 2015), https://www.criterion.com/current/posts/3500-the-thin-blue-line-a-radical-classic, (noting that "the Academy of Motion Picture Arts and Sciences refused to consider it for an Oscar due to its use of 'reenactments' and other heresies. Traditionalists at the Academy felt it should be evaluated as a fiction film because of its 'scripted content,' a phrase that doubtless also referred to [Director Erroll] Morris's stylized use of lighting, music, costuming, and camera work.").

about the world by making us see what is not there. Second, deepfakes can do, with greater realism, what literature and movies have done in the past: Give people a way to enter vivid fictional worlds, which sometimes imagine real historical figures taking fictional action.

There is also a third analogy that is in some sense a hybrid of these two—and it is the one I first considered in the introduction: A deepfake might be described as the visual equivalent of a verbal lie. It can be another, high-tech way to do what Xavier Alvarez did when he lied that he had won a Medal of Honor. This is a hybrid of the previous two analogies because it shares a feature with each. Like the artistic deepfake-creator, the lying deepfake creator offers a fiction that (unlike the deepfake that is a proxy for perception) clearly *has* an author of sorts. If we see a deepfake video of a Medal of Honor ceremony on YouTube, we will assume that *somebody* posted it, and perhaps edited it. But like the creator of a deepfake-as-counterfeit-reality, the person who creates or shares the video is not simply admitting that her deepfake is fiction. She is presenting it as an accurate record captured by camera. Where then do we place this kind of deepfake in our First Amendment analysis? Is it First Amendment expression, like the deepfake that is openly a fiction? Or is it likely outside of the First Amendment's coverage, like falsified security camera footage?

For those who would apply the *Alvarez* framework to it, is a kind of expression: Like the verbal lie that Xavier Alvarez told, the visual lie in a deepfake is First Amendment "speech." And there are a number of other court precedents that support that conclusion. Together, they might seem to strongly suggest that even when deepfake creators present their computer-generated fiction not as artistic fantasy but rather as fact, they simply move from one staunchly protected kind of First Amendment speech (that of

fictional storytelling) to another (that of visual communication of information and fact-based storytelling).

First, it is not only dramatic films, but also factual videos that courts often treat as receiving strong First Amendment protection.[115] Factual videos do things that unquestionably count as "speech" under the First Amendment when done with words. Journalists can convey information about a battle in Afghanistan, for example, not only by writing an article about it, but by filming it and then including the footage in a documentary, video diary or television account.[116] Thus, the Supreme Court didn't doubt, in *United States v. Stevens*, that when Congress restricted "visual [and] auditory depiction[s],'" of animal cruelty, "such as photographs, videos, or sound recordings," it was restricting First Amendment speech.[117] In *Bartnicki v. Vopper*, the Court likewise found that it was unconstitutional for government to impose civil liability on reporters and other individuals who had played on the radio an audio recording of a union leader talking about "go[ing] to [the] homes" and "blow[ing] off [the] front porches" of antagonists in a labor dispute.[118] Federal law made it illegal to disseminate any recordings of illegally-intercepted conversations—and subjected such dissemination to civil liability. But the Court found that the publication of the audiotape was protected "speech about a matter of public concern."[119] There is no reason that publishing a videorecording on a matter of public concern would be any less expressive.

---

[115] *See, e.g.*, THE HORNET'S NEST (High Road Entertainment 2014).
[116] *See* Taylor Lorenz*, People Can't Stop Watching Videos of Police and Protesters. That's the Idea*, N.Y. TIMES (June 2, 2020) (noting that "countless videos" of police at protest after George Floyd's killings "have been shared on social media").
[117] United States v. Stevens, 559 U.S. 460, 460, 468 (2010).
[118] Bartnicki v. Vopper, 532 U.S. 514, 518-519 (2001).
[119] *Id.* at 535.

Moreover, if camera footage or an audio recording relating facts to its viewers receives First Amendment protection, then so too do the techniques that creators of such footage use to generate it. This was what the Seventh Circuit concluded in *American Civil Liberties Union v. Alvarez*, when it held that citizens had a First Amendment right to record police officers' activities in public. "The act of making an audio or audiovisual recording," it said, "is necessarily included within the First Amendment's guarantee of speech and press rights as a corollary of the right to disseminate the resulting recording."[120] The First Amendment would not effectively protect our right to express ourselves through painting if it allowed government to regulate our use of paintbrushes and paint, nor could it protect our right to express ourselves in music if it permitted government to ban our ability to learn and play instruments.[121] Similarly, videos posted on social media sites are expressive, so the First Amendment's shield against government censorship of such recording should cover not only the sharing of such videos, but also the tools necessary to create them. Cameras and microphones are the most familiar examples of such a tool. But deepfake technology is another.

To the above arguments for treating all deepfakes we receive from others as First Amendment expression, one can add arguments that draw not only on precedent and intuition, but also on the more formal tests courts have used to determine if non-verbal activity falls within the "coverage" of the First Amendment's free speech clause—namely, asking (1) whether it is an activity that is

---

[120] Am. Civil Liberties Union of Illinois v. Alvarez, 679 F.3d 583, 595 (7th Cir. 2012).
[121] *Id*. at 596 (citing Anderson v. Hermosa Beach, 621 F.3d 1051, 1061–62 (9th Cir. 2010)).

"inherently expressive,"[122] (2) whether it satisfies the "Spence test."[123]

The idea behind the first of these two inquiries is that certain kinds of social practices are *always* expressive—or at least have an important expressive dimension—even if they are wordless or lack any kind of message. Abstract art and instrumental music, for example, count as First Amendment expression. The First Amendment, the Supreme Court has stressed, "unquestionably shield[s]" the "painting of Jackson Pollock" and the "music of Arnold Schoenberg."[124] It also protects the nonsensical verse of Lewis Carroll in *Jabberwocky*, because poetry is First Amendment expression.[125] Which non-verbal practices count as inherently expressive is a matter of social convention and shared understanding.[126] It seems clear from the way people post videos on social media that, in the early twenty-first century—sharing of videos is a recognized medium for expression. As Seth Kreimer writes, "[i]n the last two generations, emerging technology and social practice have made captured images part of our cultural and political discourse," and it is now clear that "[i]n the current state of the law and culture of discourse, captured images—like words inscribed on parchment--fall within the protection of 'freedom of

---

[122] *See* Rumsfeld v. Forum for Acad. & Institutional Rights, Inc., 547 U.S. 47, 66 (2006) (noting that First Amendment protection for non-verbal conduct covers conduct that is "inherently expressive"); Hurley v. Irish-Am. Gay, Lesbian, Bisexual Group of Boston, 515 U.S. 557, 568 (1995).
[123] *See* Spence v. Washington, 418 U.S. 405, 410-11 (1974).
[124] *Hurley*, 515 U.S. 557, 569 (1995).
[125] *Id.*
[126] *See* MARK V. TUSHNET, ALAN K. CHEN, & JOSEPH BLOCHER, FREE SPEECH BEYOND WORDS: THE SURPRISING REACH OF THE FIRST AMENDMENT 10 (2017) (stating that "perhaps the best [method of identifying what counts as First Amendment speech] is to identify the social practices and conventions that constitute human expression and communication.")

speech.'"[127] This is as true of the moving images posted on social media as it is of still photography.[128]

Even if there remains any doubt about whether sharing of a video or audio clip is inherently expressive, it might still count as First Amendment expression under what courts call the "*Spence* test," derived from the Supreme Court's analysis in *Spence v. Washington*.[129] In short, the *Spence* test allows even conduct that is arguably non-expressive to temporarily take on the status of expression in particular contexts where it becomes "sufficiently imbued with elements of communication."[130] Burning a pile of documents might not be expressive. But such an act of burning paper may be *transformed* into an act of communication when it is clearly meant as a protest. To determine if an act is "sufficiently imbued with elements of communication," courts ask whether those performing the act (1) inten[d] to express a "particularized message" in (2) "surrounding circumstances" in which their audience is likely to understand that message. In *Spence* itself, for example, the Court found that a college student had engaged in First Amendment speech when he displayed an American flag upside down from his apartment window (with a peace symbol affixed to the flag).[131] The student had hoped that this act would communicate a message: As he explained in his trial, he wanted to express his belief that the "killing" occurring in the Vietnam War was wrong and that "America stood for peace."[132] This hope was insufficient by itself to

---

[127] Seth F. Kreimer, *Pervasive Image Capture and the First Amendment: Memory, Discourse, and the Right to Record*, 159 U. PA. L. REV. 335, 373-74 (2011).
[128] *See* Animal Legal Def. Fund v. Wasden, 878 F.3d 1184, 1203 (9th Cir. 2018) (finding that a video and the act of creating it "is . . . an inherently expressive activity.")
[129] Spence v. Washington, 418 U.S. 405, 410-11(1974).
[130] *Id.* at 409.
[131] *Id.* at 406-408
[132] *Id.* at 408.

make the act expressive. But the student's intent did make his act expressive *when combined with* the context—namely a time in American life when anti-war protests were common. "It would have been difficult," said the Court, "for the great majority of citizens to miss the drift of appellant's point at the time that he made it."[133] Many videos posted on social media would likewise meet the *Spence* test. A video displaying individuals receiving a Congressional Medal of Honor, for example, might convey the message that Medal of Honor recipients deserve the respect of all Americans—and this will be even clearer if the person who shares the video adds a title or other text emphasizing this message.

And there is still another argument treating shared deepfakes as First Amendment speech: At least some false testimony furthers some of the same First Amendment interests as imaginative fiction. Xavier Alvarez's autobiographical lie, for instance, was a fiction he presented about himself. As David Han writes, individuals *all* necessarily engage in a process that is sometimes akin to fiction-writing as they construct the self-image that they present to others. "[A] fundamental component of being an autonomous individual is exercising control over who you are--and who you are is, to a significant extent, a function of who you define yourself to be to others."[134] Our "self-definition interest, by its very nature," he argues "assumes some element of deception" because "we constantly craft different personas to present to different audiences."[135] That kind of "craft[ing] of personas" is not unlike story-telling. Xavier Alvarez, on this view, was not simply a liar when he made his false statement about receiving a Medal of Honor.

---

[133] *Id.*
[134] David S. Han, *Autobiographical Lies and the First Amendment's Protection of Self-Defining Speech*, 87 N.Y.U. L. REV. 70, 99 (2012).
[135] *Id.*

He was an author of sorts, adding fictional elements to the narrative he was providing to others about himself and his history.[136]

All of these First Amendment arguments for protecting false testimony can be extended to deepfakes. Deepfake technology provides a powerful means of visual storytelling—and this is true whether the story is fictional or factual. An individual can use this tool to create a video that depicts the fictional adventures of fictional characters, the real actions of historical figures (for example, in a deepfake of the Lincoln-Douglas debates), or perhaps the fictional actions of a real person. Regardless of which of these types of stories the deepfake creator brings to life on a video, one might argue, and regardless of precisely how they combine fact and fiction, the deepfake-creator is engaged in a familiar form of First Amendment expression.

With the benefit of these three analogies for deepfakes, we might tell a preliminary story about their First Amendment status. First, when we receive video footage (or audio footage, for that matter) directly from a camera or computer, the video isn't First Amendment speech simply because it doesn't come from any speaker. The digital footage we extract or receive directly from a camera or computer serve is not coming to us from a speaker. It is not a vivid distillation of someone else's story or claim. It is simply the record that light etched on a camera's film or digital storage. When we view this record, it seems like a proxy for perception in part because we are likely to assume that it remains unaltered by another person unless we alter it ourselves. Where a manipulator does surreptitiously access or intercept it and alter it, that is less

---

[136] *Id.*

communication than sabotage.[137]

Matters are different when videos come to us not from machines but rather from human speakers. When they do, the video is the expression of that speaker—who has a First Amendment right to author it even in ways that their audience (or the state) may disapprove of, and to use deepfake technology in doing so. This is true when the speaker uses a video to tell a tale that is openly a fiction. But it is also true, under *Alvarez*, when the speaker uses a video to present a falsehood as a fact, something she might do with deceptive editing or, perhaps, with a deepfake.[138]

This is not, however, the account I will defend in the remainder of the article. The endpoints of the spectrum of deepfakes I've discussed *are* largely accurate. Deepfakes are outside the First Amendment's coverage when they are used to manipulate perceptions by altering stored camera footage. They are strongly protected by the First Amendment when they are works of art. But

---

[137] This is not to say that, in interactions with machines, that output of a machine will *never* count as First Amendment expression. Scholars have already explored how the product of artificial intelligence (AI) could count as First Amendment speech in some situations. *See, e.g.*, Toni M. Massaro and Helen Norton, *Siri-ously? Free Speech Rights and Artificial Intelligence* 110 NW. U. LAW REV. 1169, 1174 (2016); Toni M. Massaro, Helen Norton, and Margot E. Kaminski, *Siri-ously 2.0: What Artificial Intelligence Reveals about the First Amendment* 101 MINN. LAW REV. 2481, 2482-2483 (2017), 2482; Margot E. Kaminski, Authorship, *Disrupted: AI Authors in Copyright and First Amendment Law*, 51 UC DAVIS LAW REV. 589, 610 (2017); Ronald K. L. Collins and David M. Skover (eds.), ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE (2018). That AI can create the kinds of outputs that human speakers create—such as essays, paintings, or musical compositions—does not mean that machines such as thermometers and cameras are similarly creating speech when they passively capture information for human audiences to access. Even if we have a First Amendment right to receive or share the information a camera captures, that does not mean that either a camera (or someone who hacks into and manipulates it) is a speaker for First Amendment purposes.

[138] *See Guidelines for Ethical Video and Audio Editing*, RADIO TELEVISION DIGITAL NEWS FOUND. (RTDNA), https://www.rtdna.org/content/guidelines_for_ethical_video_and_audio_editing (setting our journalists' obligations to avoid editing that creates misleading footage).

we should hesitate to treat them as equally protected when they are used by a speaker to perpetuate a factual falsehood.

Before setting forth this article's alternative to the *Alvarez* framework, however, I will revisit this framework—and explain a little more clearly, some of the problems that arise in applying it to deepfakes (and some of the problems scholars have raised about it more generally).

II. *UNITED STATES V. ALVAREZ*—AND DEEPFAKES AS VISUAL LIES

A. Reconciling Deepfake Dangers and Benefits

Deepfakes, as I noted earlier, can be dangerous. Fake videos of a fictional terrorist attack or missile attack might start a war or generate riots.[139] Fake videos of embarrassing individual behavior can provide new, more damaging forms of defamation or tools for blackmail.[140] They can also, I have noted above, hijack our perceptions and make us see what is not there. They can lead us to doubt what we see—perhaps not everywhere, but in numerous settings where we have relied on video to connect us to others, watch remote events unfold, or find solid evidence of past events.

How do we leave room for government to combat such dangers while leaving individuals with the freedom to use deepfakes for expressive or other artistic purposes? As I have written in earlier scholarship on deepfakes, the "video-altering technology that allows individuals to undermine each other's grasp of what is real" can "provide moviemakers with yet another tool to create the special effects that can make narrative films feel real to an audience. Thus, the same technology that might lose First Amendment protection when it fabricates news might merit robust First Amendment protection when it is, like other tools of modern filmmaking, a

---

[139] Chesney & Citron, *supra* note 6, at 1176.
[140] *Id*. at 1791-94.

means of telling a story."[141] What First Amendment framework can protect the First Amendment value of artistic and other expressive deepfakes while leaving government with leeway to counter the problems they raise, and protect the integrity of the video evidence they threaten?

I will ultimately argue that courts, in addressing this challenge, should look to a variant of the same framework that they have used in other situations where expressive conduct the First Amendment protects becomes intertwined with non-expressive impacts on our social life that government has to regulate.[142] Courts addressing such issues have used various approaches for dealing with what one might call First Amendment "middle grounds:" conduct that straddles the boundary line that separates protected First amendment expression from unprotected non-speech conduct. The most familiar of these frameworks is the "intermediate scrutiny" that the court applies when expression and conduct become intertwined in "expressive conduct," such as burning a draft card.[143] Courts apply that and other tests to let government target the physical or other non-speech harms it is responsible for protecting against (like the destruction that can arise from burning a record) while preventing it from using that government responsibility as an excuse to target the political or other messages intertwined with such harms. Under *R.A.V. v. St. Paul*, it applies the same kind of framework to government regulation of incitement, true threats, commercial speech, and other categories of unprotected

---

[141] Blitz, *supra* note 31, at 113-14.

[142] *See* Marc Jonathan Blitz, *Free Speech, Occupational Speech, and Psychotherapy*, 44 HOFSTRA L. REV. 681, 694 (2016) ("free speech law often has to deal with realms of human action where government's presence is necessary to assure individuals' health and safety but is simultaneously dangerous to their intellectual liberty and autonomy.").

[143] *See* United States v. O'Brien, 391 U.S. 367, 379–80 (1968); *see also infra* Part VI.C.2.

or less protected speech: Government may target the intimidation and potential for violence generated by a threat but may not punish *only* those threats that carry disfavored ideological or political content.[144] Deepfakes, I will suggest, merit the same First Amendment response: The First Amendment should permit government to counter the deception they can cause without letting it target the artistic or other expressive uses in it, and certainly without targeting deepfakes on the basis of the views they further rather than the deception they cause.[145] In many cases, this will entail letting government impose a disclosure requirement: It can let a deepfake creator create and disseminate an artistic or other video, but blunt the deceptive risk by requiring the deepfake-maker to disclose that it is a deepfake.[146]

Before describing this First Amendment framework for deepfakes more fully, this article will more carefully explain why and when deepfakes have a non-expressive dimension—and why there is a need for a framework proposed here. Part I, after all, only identified a fairly narrow category of non-expressive deepfakes—namely, videos that come to us directly from machines rather than from human speakers.[147] Most of the deepfakes that writers worry will usher in a new, more unsettling age of "fake news" are posted or disseminated by others, so one may wonder how these deepfakes are in any sense non-expressive. Moreover, if deceptive deepfakes are analogous to verbal lies, then under *Alvarez*, deepfakes are protected even when they are deceptive. But this Part and the next argue that there are problems with this equation between verbal lies and deepfakes.

---

[144] *See* R.A.V. v. St. Paul, 505 U.S. 377, 381 (1992); *see also infra* Part VI.C.1.
[145] *See infra* Part VI.
[146] See *infra* Part VI.B.
[147] See *supra* Part I.A.

B. Alvarez*, Lies, and Deepfakes*

First Amendment analyses of deepfakes often begin by analogizing them to lies—and then considering how *United States v. Alvarez* might apply to this kind of lie.[148] So it is useful to review the reasoning of the Justices who made up the majority in *United States v. Alvarez*, and the two key opinions that comprised it. In *Alvarez*, the four Justices joining Justice Kennedy's plurality opinion and the two joining Justice Breyer's opinion all took the position that false statements had value and could only be punished not merely because they are false, but only when they are harmful in some other way. In Kennedy's view, lying could only count as unprotected speech if it was accompanied by some "legally cognizable harm" of a kind that had traditionally placed speech outside of the First Amendment scope.[149] For example, lying is unprotected if it creates the kind of harm to reputation that would make it defamation, or the kind of harm to financial or other material interests that would make it fraud. Or if it harms basic operations of government, for example, by giving a person the false impression that a citizen was a law enforcement officer. The harms accompanying such speech make it something other than pure discourse: They make speech into a kind of activity which, like much non-speech conduct, can cause financial devastation or (in the case of a fake police officer) subject someone to unjustified coercion. In the absence of such harm, however, false factual claims remain as staunchly protected—in Kennedy's view—as objectionable doctrines or ideologies (which government generally cannot censor). Such false claims, said Kennedy, should typically be met with criticism by fellow citizens rather than suppression by

---

[148] *See supra* note 10.
[149] *See* United States v. Alvarez, 567 U.S. 709, 719 (2012) (plurality opinions).

government.[150] It is only in the very rare case that government can satisfy "exacting scrutiny" under the First Amendment (which other cases refer to as "strict scrutiny") that such a speech restriction is permissible. To overcome this extremely high hurdle, government has to show that restriction of certain speech is the only way to achieve a "compelling" interest—that is, one of the most extraordinary weight—and that there is no way it can achieve this interest while restricting less speech.[151]

Justice Breyer was willing to give government more leeway to restrict factually false claims. Rather than assume such claims receive near-absolute protection except when they fall into a few long-recognized categories of "legally cognizable harm," Breyer reasoned that false claims almost always have the potential to generate some kind of harm—and that these harms may well be unfamiliar harms that don't easily fit into historically-recognized categories. The harms that the Stolen Valor Act sought to address, for example, were a bit different from those that had long provided government with justification for regulating false speech: Allowing lies like those of Alvarez to convince unwitting listeners would "dilute[e] the value of [military] awards" and make it impossible for "the Nation [to] fully honor those who have sacrificed so much for their country's honor."[152] Still, Justice Breyer said, government should be allowed to address such harms if it does so in a way that takes account of the First Amendment interests threatened by government speech restriction—and does not do disproportionate damage to such interests.[153] In other words, whereas Justice Kennedy had said government can only restrict false speech if it

---

[150] *Id*. at 727 ("The remedy for speech that is false is speech that is true.").
[151] *Id*. at 724-726.
[152] *Id*. at 737 (Breyer, J., concurring).
[153] *Id*. at 730-731, 739.

meets strict scrutiny, Justice Breyer demanded only that it meet intermediate scrutiny. But Breyer's opinion still placed some significant demands on government restriction of lying: He noted that when statutes prohibiting or penalizing a lie have been viewed as constitutional, they have tended to have certain features that "limit the scope of their application, sometimes by requiring proof of specific harm to identifiable victims; sometimes by specifying that the lies be made in contexts in which a tangible harm to others is especially likely to occur; and sometimes by limiting the prohibited lies to those that are particularly likely to produce harm."[154]

For Justice Alito and the two other dissenters in *Alvarez* (Scalia and Thomas), the First Amendment offers no protection for most verifiably false factual statements. "Time and again," said the dissent, "this Court has recognized that as a general matter false factual statements possess no intrinsic First Amendment value."[155] Moreover, Alito stressed, the Stolen Valor Act only imposed a penalty on lying about a "narrow category of false representations about objective facts that can almost always be proved or disproved with near certainty" and "facts that are squarely within the speaker's personal knowledge."[156] Moreover, Xavier Alvarez's lie was not a claim "about philosophy, religion, history, the social sciences, the arts, and other matters of public concern."[157] It was a false claim about a purely personal matter—and contributed nothing to debates about history, science or culture.

How would *Alvarez* apply to deepfakes? None of the Court's opinions addresses whether the First Amendment might

---

[154] *Id*. at 734.
[155] *Id*. at 746 (Alito, J., dissenting)
[156] *Id*. at 740.
[157] *Id*. at 751.

protect deception with video in addition to verbal lying. But, as noted above, the Court has treated the video- and audio-tapes as First Amendment expression, so one might guess that *Alvarez* would presumptively cover that kind of expression as well as verbal lying.

Still, that does not mean that Kennedy and Breyer would necessarily find all restriction of deepfakes to be unconstitutional. First, some deepfakes might be part of, or the cause of, a "legally-cognizable" harm, by being used in fraud, defamation, or false impersonation, for example.[158] For example, where a deepfake defamed a businessman by showing him meeting with a foreign agent it would be subject to civil liability. If someone used a deepfake to impersonate a government official, they would be engaged in the kind of lie that is unprotected. Moreover, although, strict scrutiny is often difficult to meet, courts might find that some deepfake restrictions could so. They might do so, for example, when the deepfake sows panic by depicting a missile strike, a major natural disaster, or a declaration of war by the President. Courts have had little difficulty, after all, answering First Amendment claims by individuals arrested for sending fake anthrax in mailings. In one such case, the Ninth Circuit emphasized that "[f]alse and misleading information indicating an act of terrorism is not a simple lie" and that it "tends to incite a tangible negative response"—wastefully diverting resources of emergency workers and law-enforcement personnel, and creating paralyzing fear among those

---

[158] *See* Chesney and Citron, *supra* note 6 at 1791, 1793-94. *See also* Erwin Chemerinsky, *'Deepfake' Videos Threaten Our Privacy and Politics. Here's How to Guard Against Them*, SACRAMENTO BEE (Jul. 13, 2019), https://www.sacbee.com/opinion/california-forum/article232515577.html#storylink=cpy (stating that "the court has said that speech which is defamatory of public officials and public figures has no First Amendment protection" and arguing this is true of deepfakes used to recklessly portray political candidates taking actions they did not take).

targeted by such threats.[159] And if government could meet Justice Kennedy's strict scrutiny standard in restricting a kind of deepfake—or show that the deepfake presented a legally-cognizable harm—then it could almost certainly meet Justice Breyer's intermediate scrutiny standard as well.

Alvarez therefore presents one framework for letting government address the dangers of deepfakes while protecting their First Amendment uses. First of all, we should extend the First Amendment protection that *Alvarez* provided to lies to all deepfakes—even those designed to deceive—because they too constitute First Amendment speech. Then, the protection would be withdrawn only from those deepfakes that cause harms of a kind traditionally unprotected by the law—or that would enable the government to meet strict or intermediate scrutiny.

There is, however, a problem with this approach to deepfakes. It misunderstands the way in which the deceptive nature of deepfakes, *by itself*, can betray and undermine the reliance we place on video and audio evidence in many different contexts. The threat deepfakes pose doesn't always come in the form of hysteria created by a fake missile attack, riot, or natural disaster, or defamation of a particular individual. The spread of deepfakes in certain contexts can, more generally, erode our ability to trust certain kinds of evidence. Chesney and Citron refer to this aspect of deepfakes as the "liar's dividend;" "As the public becomes more aware of the idea that video and audio can be convincingly faked, some will try to escape accountability for their actions by denouncing authentic video and audio as deep fakes."[160] In addition to the "truth decay" that results when audiences are deceived by

---

[159] United States v. Keyser, 704 F.3d 631, 640 (9th Cir. 2012).
[160] Chesney & Citron, *supra* note 6, at 1785-86.

deepfakes, they argue, there is also "trust decay" that results from the intense skepticism deepfakes will force us to bring even to genuine video, and from the way certain individuals will exploit that erosion of trust.[161]

This is not a harm that flows from any single deepfake. Nor does it arise only from deepfakes that are defamatory, threatening, or falsely depict the occurrence of threats to public safety—such as terrorist or missile attacks. It arises from the *very existence* of deepfakes—and the ease with which people can create them. Even the vivid realism of an otherwise relatively harmless deepfake—a deceptive deepfake showing someone shaking the hand of a United States President they never met or climbing to the peak of a mountain they never visited—will help "prime" a "skeptical public" to doubt other video recordings that society may need to rely upon as evidence.[162] For the most part, Justice Kennedy and Justice Breyer's analysis lack an answer to that threat.

One might respond by arguing that the *Alvarez* plurality can still factor in this concern—by treating the way that deepfakes can undermine trust in our perceptions (and perhaps, in other sources of non-testimonial beliefs) as a kind of "legally-cognizable harm." In fact, the *Alvarez* plurality already did something very much like this in explaining why the First Amendment does not protect lies in which one impersonates a law enforcement officer or other government official. Imagine that someone impersonates a police officer. She might *not,* in doing so, necessarily cause any kind of material or financial harm to others, or any harm to their reputation. But according to Justice Kennedy, the act of impersonating an officer *itself* has caused harm by undermining trust in government:

---

[161] *Id.* at 1786.
[162] *Id.*

The First Amendment leaves government free to punish people who falsely speak in the government's voice—in order to protect the trust individuals place, and often have no choice but to place in government officials (such as law enforcement officers or, for example, officials overseeing responses to public health threats or weather emergencies). It would almost certainly allow punishment of individuals who use deepfakes to make government officials give fictional speeches or announcements—and to portray them as real. If the First Amendment leaves government leeway to restrict lying and deepfakes that undermine trust in government, perhaps it also allows government leeway to restrict uses of deepfakes that undermine our trust in video and other digitally mediated perceptions. Perhaps the latter is also a kind of legally-cognizable harm—or a harm that even strict scrutiny will allow government to restrict when the restriction is careful enough.

The problem with that analysis, however, is that it seems to stretch the meaning of "legally cognizable harm" to the point where it might cover virtually any consequence of disseminating deepfakes—*or* of ordinary lying. The presence of harm in a lie, as Alan Chen and Justin Marceau write, cannot by itself open the door wide to government restriction because "some degree of harm is inextricably wedded to the very act of lying"—"[n]early every lie, if believed, causes some reliance by the listener and produces some combination of benefits and harms."[163] If lies are to receive any meaningful First Amendment protection at all, then, First Amendment protection cannot be withdrawn whenever there is harm (since there always is).

C. An Alternative to Alvarez*: Treating Verbal Lies as*

[163] Alan K. Chen and Justin Marceau, *Developing a Taxonomy of Lies under the First Amendment*, 89 U. COLO. L. REV. 655, 656-59 (2018).

Unprotected

The liar's dividend thus presents a problem for applying *Alvarez* to deepfakes. And one might argue that it presents a problem for *Alvarez more generally*. Seana Shiffrin has already made the same point about another, older kind of liar's dividend and trust decay—the kind that has long arisen from verbal lying itself. "[D]eliberate misrepresentations," she stresses, "undercut the warrants we have to accept each other's testimonial speech."[164] Just as world filled with deepfakes will lead us to distrust even many genuine videos we see, a world filled with verbal lying (that takes place under cover of First Amendment rights), will lead us to distrust many true statements we hear. "When people appreciate that their reasons to accept others' testimony have been diminished, the culture of trust will noticeably deteriorate," and the erosion this culture stems from lying even when identifiable material harms do not.[165] As Shiffrin also notes, the Justices' analyses in *Alvarez* don't adequately account for this deterioration of a "culture of trust" because their arguments focus only on the "*particularized* harms" that certain lies cause "to specific people," and ignore the deeper harm that the social practice of lying causes to the conditions of communication and trust.[166]

Shiffrin then proposes that the solution is to abandon the *Alvarez* framework altogether and reduce First Amendment protection from false statements that are designed to mislead their audience.[167] She rejects the Justices' conclusion in *Alvarez* that

---

[164] SEANA VALENTINE SHIFFRIN, SPEECH MATTERS: ON LYING, MORALITY, AND THE LAW 117 (2014).
[165] *Id.* at 137.
[166] *Id.*
[167] Although she concludes that lies do not have any First Amendment value, she still favors applying "a modified version of intermediate scrutiny" to government

speech restrictions that forbid intentionally false speech are, for that reason alone, content-based. Content-based restrictions on speech typically receive strict scrutiny[168] and this was one reason the plurality applied strict scrutiny to the Stolen Valor Act and found it unconstitutional.[169] Prosecutors can't prove to a judge or jury that a statement like that made by Xavier Alvarez violates the Act's ban on falsely presenting oneself as a Medal of Honor winner unless they can provide evidence about the specific content of Alvarez's statement. But prohibitions on false statements are not content-based in the traditional sense, Shiffrin argues, because they crucially depend not simply on the content of a statement but rather on the speaker's beliefs about it. What the Stolen Valor Act demanded from Xavier Alvarez was not that he or others invariably refrain from making an assertion with specific content (that Alvarez had won a Medal of Honor) but rather that he convey his honest belief, when making any such statement, about the proposition that the statement sets forth.[170] So if Alvarez mistakenly believed his statement to be true, he should (on Shiffrin's view) be fully protected by the First Amendment in claiming to have won a Medal of Honor. If, by contrast, he believed it to be false, the Stolen Valor Act required him to be clear about this belief in its falsity when stating it. What the Act compelled, in other words, wasn't silence, but sincerity. Someone in Alvarez's situation, Shiffrin suggests, might convey that he doesn't believe his own claim to have won a

---

restrictions of lying because of "serious political and structural concerns associated with the regulation of such speech." *Id*. at 154.

[168] *See* Texas v. Johnson, 491 U.S. 397, 411-412 (1989) (finding that because prosecution of defendant for burning the American flag was content-based, the court had to apply "the most exacting scrutiny.").

[169] *Alvarez*, 567 U.S. 709 (stating that "when content-based speech regulation is in question [] exacting scrutiny is required" and as a result "the statutory provisions under which respondent was convicted must be held invalid, and his conviction must be set aside.").

[170] Shiffrin, *supra* note 166, at 125-126.

Medal of Honor "by taking advantage of culturally well-understood mechanisms of disclosure, such as deploying a sarcastic tone, evidently exaggerating in ways that indicate parody or irony, publishing under the rubric of fiction."[171]

In other words, just as a deepfake-creator could conceivably bring what would otherwise be a deceptive deepfake within the First Amendment's coverage by offering contextual clues that it is a work of art or opinion, and not unaltered camera footage[172]—by actually informing the viewer that it is a fiction generated by special effects (like the scene in Forrest Gump) and not genuine camera footage (capturing a real Medal of Honor ceremony)—so a speaker can do so for verbal falsehoods by disclosing their falsity.

Rather than struggle with whether (and how) *Alvarez* applies to deepfakes, one might thus question whether the *Alvarez* framework should continue to apply to any kind of falsity—or whether it is time for a new framework.

It is not only Shiffrin who explores the possibility of replacing the *Alvarez* framework, but also Cass Sunstein. He has recently noted that with the rise of social media disinformation campaigns as well as deepfakes, *Alvarez's* 2012 opinion "seems like a generation ago."[173] To the extent lies could once have been easily debunked by counterspeech or skepticism, as Xavier Alvarez's lie was, perhaps that is not true in an age when there are sophisticated techniques for flooding social media with false information. These techniques include "bots" that will amplify a false claim, making one person's lie seem like it is embraced by thousands of others (who are really just computer programs "voicing" the same

---

[171] *Id.* at 133.
[172] *See supra* Part I.B.
[173] Cass R. Sunstein, *Falsehoods and the First Amendment,* 33 HARV. J.L. & TECH. 387, 388 (2020).

falsehood multiple times); teams of users who seek to spread an outrageous claim not because they have verified it, but because they support it and wish to help spread it; social media algorithms that reward certain falsehoods because they will likely draw more interest.

In some ways, the framework Sunstein proposes as an alternative to *Alvarez* is less radical than that of Shiffrin's. It still insists that government restrictions of false expression be subject to something like heightened scrutiny: Such restrictions should be constitutional, he writes, only when officials "can show that [the lies] threaten to cause serious harm that cannot be avoided through a more speech-protective route."[174] But "serious harm," in his view, isn't limited to "legally-cognizable harm."[175] It can encompass damage to our epistemic practices. In discussing deepfakes' potential harm, for example, Sunstein notes that deepfakes may be more likely to cause harm because they "are more credible than merely verbal representations" and "[i]n a sense . . . self-authenticating."[176] It is thus "plausible to say that deepfakes (and doctored videos) are properly the objects of regulatory attention even if statements that embody their propositional content are not."[177] This framework could also explain why the First Amendment may leave government with greater power to regulate certain kinds of technologically-enhanced verbal lies—like those amplified by bots or disconfirmation campaign on social media—than to regulate isolated oral falsehoods.

---

[174] *See* Sunstein, *supra* note 67, at 4.
[175] More specifically, Sunstein suggests that to analyze the harm caused by falsity, courts should look at (1) what the speaker's state of mind was in promulgating the falsity, and (2) the magnitude, (3) likelihood, and (4) probability of the harm that might result from the falsity. *Id.* at 12-14.
[176] *Id.* at 119.
[177] *Id.*

III. DEEPFAKES AS NON-TESTIMONIAL FALSEHOODS

A. Testimonial and Non-Testimonial Sources of Knowledge (and False Belief)

In this part, I argue that while there may be reasons for courts to adhere to *Alvarez*'s framework for old-fashioned verbal lies—if not for the amplification of falsehoods that occurs on social media—Shiffrin's critique *does* provide a better model for how the First Amendment should treat deepfake audio and videos. More specifically, even if Shiffrin's First Amendment framework isn't the appropriate one for verbal lies, or false *testimony*, something like it should apply to what I will call "non-testimonial" falsehoods.Such non-testimonial falsehoods, moreover, are the more likely to be proper "objects of regulatory attention," as Sunstein argues, but not only because they can cause greater harm for the reasons he explains, but also because in some cases, intentionally deceptive uses of deepfakes shouldn't come within the First Amendment's coverage. Their non-testimonial nature makes them more like the non-expressive counterfeit realities in Part I than the artistic or other fictions discussed there.

It is first helpful to understand the contrast between testimonial and non-testimonial sources of beliefs. In the courtroom, testimonial evidence consists of witness statements.[178] Non-testimonial evidence encompasses all evidence that comes from other sources: "Real evidence" such as a murder weapon, a bag

---

[178] *See* Jay E. Grenig, Rocco M. Scanza, *Understanding Evidence (Part I)*, 70 DISP. RESOL. J. 85, 91 (2015) ("Testimonial evidence comes from the sworn testimony of witnesses.").

of drugs, a bloody shirt, or other objects that were a part of the underlying action [179] or the "documentary evidence" one finds in records of the underlying action, such as police reports written before the trial or video recordings of the events.[180] "Demonstrative evidence" is often treated as yet another category of non-testimonial evidence—even though it is more like testimony than real and documentary evidence, since it consists of maps, graphs, computer animations, or other visual illustrations that a party *creates for the trial* to clarify their testimony.[181]

For purposes of this article, a more helpful framing of the distinction between testimonial and non-testimonial sources of knowledge is the one that comes from epistemology. In the sense used by philosophers, "[t]estimony," as Christopher Green writes "need not be formal testimony in a courtroom but happens whenever one person tells something to someone else."[182] If someone tells us, in casual conversation, that she has won a Congressional Medal of Honor, and we accept her claim as true, then we have formed a testimonially-based belief. More specifically, a testimonial belief is one that I form when I accept as true the *assertions* presented by another person. Green states, "[w]hen someone tells us *p*, where *p* is some statement, and we accept it, then we are forming a testimonially-based belief that *p*."[183] Jonathan Adler similarly writes that a speaker who offers testimony makes an "[a]ssertion put[ting] forth a proposition that the speaker represents as true."[184]

---

[179] *See* Ashley S. Lipson, *"Real" Real Evidence*, 19 LITIGATION 29, 29 (Fall 1992).
[180] *Id.* at 30.
[181] *See* People v. Palacios, 419 P.3d 1014, 1018 (Colo. App. 2018).
[182] Christopher R. Green, *Epistemology of Testimony*, INT. ENCYCLOPEDIA PHIL., https://www.iep.utm.edu/ep-testi/.
[183] *Id.*
[184] Jonathan Adler, *Epistemological Problems of Testimony*, STAN. ENCYCLOPEDIA PHIL., https://plato.stanford.edu/entries/testimony-episprob/.

Jennifer Lackey states that testimony occurs where a speaker makes a communication in which the speaker "reasonably intends to convey information that *p*" or where that communication "is reasonably taken as conveying information that *p*."[185] These definitions are broad enough to include communications that are intentionally false. Although some accounts of testimony narrowly define it to include only communications of the above kind that are an "epistemically good source of belief," the above accounts include in the definition of "testimony" communications that lead people to false beliefs—including a statement like that of Xavier Alvarez about his Congressional Medal of Honor.[186]

Our beliefs have a "non-testimonial" basis, by contrast, when we form them not on the basis of what others tell us (that is, their testimony), but rather on some other grounds—like what we see with our own eyes or what we learn from a measuring device, like a GPS location reader. When my plane has landed in Chicago's O'Hare airport, I might believe I am in Chicago not because the pilot announces that this is so just after landing, but also (or rather) because of what I have seen through the plane window just prior to landing (the John Hancock building, Grant Park, Northerly Island, and other structures or sights that identify the place below as Chicago). My evidence that I am in Chicago comes from my perceptions—and my memory of what Chicago looks like—not just from the pilot's "testimony." The same is true if I learn I am in Chicago, or perhaps at a certain location within the city, from the GPS-generated location data that places me there on a phone mapping application.

---

[185] JENNIFER LACKEY, *Introduction, in* Jennifer Lackey & Ernest Sosa, THE EPISTEMOLOGY OF TESTIMONY 3 (2006).

[186] *See* JENNIFER LACKEY, LEARNING FROM WORDS: TESTIMONY AS A SOURCE OF KNOWLEDGE 15-19 (2008).

Video can similarly be non-testimonial. I might know an event occurred in a certain location not because the narrator of the video tells me that, but because I can *see* the event occurring in that location on the video. Say, for example, that friends of mine wish to support a boast that they were present for the Chicago Cubs' 2016 World Series victory, then they might share a time-stamped video showing them at Game 7 of that World Series or a GPS record placing them at Progressive Field in Cleveland, when the game ended at 12:30 am on Nov. 3, 2016. This boast itself ("we were at Game 7 of the 2016 World Series") is testimonial. But the time-stamped video is *non*-testimonial. To the extent they are sharing the video as proof that they were at that baseball game, they are not simply offering testimony in a different form. The video is not an assertion. It is a camera-generated *record* that *supports* their assertions.

In fact, such video evidence is non-testimonial in two different senses. First, it is like records generated by a host of other machines: GPS readings to determine location, odometer readings to tell how far our car has traveled, thermometer measurements of temperature, blood pressure or pulse rate measurements to understand certain features of our biological functioning, or a clock that provides precise information about the time.[187] Just as a computer-connected thermometer might construct and archive a series of temperature readings as it measures the temperature of the surrounding environment, cameras create video footage records of events occurring in front of them by capturing light that entered the camera. Second, watching a video often feels like a proxy for perceiving the events in it directly. It may well make an audience

[187] *See* Sosa, *supra* note 61, at 121-123 (analyzing "instrumental" knowledge as a kind of non-testimonial knowledge distinct from our "trust in sensory sources").

feel that they are seeing what they would have seen if they had been standing where the camera was. A video recording thus seems to audiences more like a perceptual source of knowledge than do other kinds of machine-generated records.

The above dichotomy between testimonial and non-testimonial sources of knowledge has not had much significance in First Amendment law. It has played a role in two other areas of constitutional law. The Fifth Amendment's privilege against self-incrimination, the Court has said, only bars government from compelling a criminal defendant to provide testimonial evidence.[188] Government may not compel her to make assertions the contents of which can incriminate her, but it may compel her to provide a blood or DNA sample,[189] or a voice sample to compare with a voice recording,[190] or other physical evidence.[191] The Sixth Amendment's Confrontation Clause similarly applies only to testimonial evidence. A criminal defendant must be given the opportunity to "confront" and cross-examine any witness providing testimony against her.[192] But she does not have a similar opportunity to challenge, for

---

[188] *See* Schmerber v. California, 384 U.S. 757, 761 (1966) ("[T]he privilege protects an accused only from being compelled to testify against himself, or otherwise provide the State with evidence of a testimonial or communicative nature, and that the withdrawal of blood and use of the analysis in question in this case did not involve compulsion to these ends.").

[189] *See, e.g.*, United States v. Hook, 471 F.3d 766, 773 (7th Cir. 200) ("taking of blood samples or fingerprints is not testimonial evidence and as such is not protected by the Fifth Amendment," and therefore DNA sample was not protected either).

[190] *See, e.g.*, United States v. Dionisio, 410 U.S. 1, 7 (1973) ("The voice recordings were to be used solely to measure the physical properties of the witnesses' voices, not for the testimonial or communicative content of what was to be said.").

[191] *See, e.g.*, *Hook*, 471 F.3d at 773.

[192] *See* Crawford v. Washington, 541 U.S. 38, 53-54 (2004) ("[T]he Framers would not have allowed admission of testimonial statements of a witness who did not appear at trial unless he was unavailable to testify.").

example, GPS data used to place her in a certain location.[193] In both cases, this focus on testimonial evidence stems from the use of the word "witness" in each amendment's text. The Fifth Amendment bars government from forcing a person "to be a *witness* against" herself.[194] The Sixth Amendment gives criminal defendants the right to be "to be confronted with the *witnesses* against" them.[195]

In free speech law, by contrast, courts haven't felt the need to ask whether information is testimonial or non-testimonial. The key question courts raise when they wish to determine if the First Amendment applies to a given type of conduct is rather whether conduct is expressive or not—not whether it is testimonial. But I argue here that non-testimonial falsehoods, like fabricated videos, merit different First Amendment treatment than does false testimony for two reasons.

B. Speaker Autonomy

First, it may be the case, perhaps, that the First Amendment has to allow—and shield—many verbal lies as a concession to speakers' *autonomy*. Our words, on this view, generally have to remain within our control—even if this allows us to fill them with false content. This causes some damage to our testimonial practices, but this is a price that we may have to pay to keep government out of our decisions about how we describe ourselves. That government regulation of lying might gravely threaten individual autonomy is a common refrain in arguments that the First Amendment should shield even intentional falsehoods. The plurality in *Alvarez* seemed to have such a concern in mind when stating that "a broad censorial power" over lies would cast "a chill the First Amendment cannot

---

[193] *See* United States v. Brooks, 715 F.3d 1069, 1080 (8th Cir. 2013) (finding "GPS reports were non-testimonial" under the Sixth Amendment).
[194] U.S. CONST, amend. V.
[195] U.S. CONST, amend. VI.

permit if free speech, thought, and discourse are to remain a foundation of our freedom."[196] Jonathan Varat likewise argues a government that could ban lying would have more control than we can afford to let it have over the communicative practices that are at the core of our "rights to personal and political self-rule."[197] As noted earlier, David Han has stressed that a core element of such self-rule includes a "self-definition interest" that "by its very nature assumes some element of deception" because "we constantly craft different personas to present to different audiences."[198]

But that does not mean that a person's interest in "*self*-definition" or "*self*-rule" requires that she be able to falsify not only the content of the words that others view as *her* words, but also any evidence that has traditionally been independent of her. There are, after all, other kinds of evidence—apart from her testimony—that others might potentially use to obtain information about her. Imagine, for example, that someone claims she has won a Purple Heart for bravery in battle fighting in a particular Army unit. Apart from relying on that person's testimony to understand if that claim is true, her audience might also talk to other individuals they can find who have fought in that Army unit. Or they might consult government records about that battle (and about military award winners). Or they might view video evidence of the battle or of award ceremonies at which Purple Hearts were awarded.

This evidence all pertains to the speaker's history. But that doesn't mean that her autonomy interests give her a First Amendment right to control the content of all of it. As important as autonomy, or "self-definition" and "self-rule," may be, First

---

[196] *Alvarez*, 567 U.S. at 723.
[197] Jonathan D. Varat, *Deception and the First Amendment: A Central, Complex, and Somewhat Curious Relationship*, 53 UCLA L. REV. 1107, 1109 (2006).
[198] *See* Han, *supra* note 135, at 99.

Amendment law doesn't give force to a speaker's autonomy interests by endowing her with unlimited control over *all* of the sources of evidence others can draw upon to learn about her. Rather, it carves out a certain realm—the realm of *her* expression—where she has extraordinary control and where the government is presumptively excluded from controlling the content she decides to place in this realm. The stories we tell or claims about ourselves are quite clearly (at least most of the time) squarely within this zone of autonomy. Deepfakes, I am arguing here, may not be.

Moreover, whether they are or are not depends not only on their value for a speaker (as the speaker engages in self-definition), but also on the costs that society must sustain if a speaker's authorship extends to video evidence or other evidence that has traditionally been (at least to some extent) viewed as external to that speaker. This brings us to a second reason that deepfakes and other non-testimonial falsehoods may not merit the First Amendment insulation that *Alvarez* accorded to verbal lies: It is possible that such fabricated evidence is not only less crucial for *speaker autonomy*, but also more potentially harmful to *viewers'* autonomy and reliance interests. Lies, as Shiffrin points out, undermine our "testimonial practices."[199] But—at least prior to the age of social media disinformation campaigns—we have been relatively well-equipped to mitigate these harms. Lying, after all, has long been a familiar feature of social life. Individuals in most societies, and certainly in the contemporary United States, are acutely aware that others—particularly others who are unfamiliar to them—might be speaking dishonestly. It is not an insuperable burden, therefore, for individuals faced with possibly false words to play a role that Justice Robert Jackson insisted they must play in a society governed by

[199] Shiffrin, *supra* note 166, at 117.

First Amendment freedom—that is, act as one's *own* "watchman for truth,"[200] and sort out truth and falsity for themselves.

By contrast, machine-based distortions and deceptions are less familiar and it is more plausible to insist that the First Amendment leaves individuals with room to recruit the help of government to defend against those deceptions. So even if Shiffrin's proposed First Amendment framework isn't suitable for verbal lies, it may be suitable for deepfake video and audio footage (and Part VI will suggest that something like it is). The First Amendment may stand in the way when the government wants to force liars to be sincere and to reveal when they know their words to be false. But it should not generally stop government from forcing deepfake creators to *reveal* that their video is a *fake*. More generally, even if the First Amendment does protect false testimony (as *Alvarez* held), it should not protect the fabrication or intentional dissemination of *non-testimonial* falsehoods. And when deepfakes are used to deceive their audience, I argue, this what they are.

C. Viewers' Autonomy and Reliance Interests (and "Epistemic Backstops")

As noted in the introduction, one could imagine an alternative version of the facts in *Alvarez*, wherein the defendant doesn't *only* tell a lie that he has won a Medal of Honor. He *backs up* his lie by creating a fake government database of Medal of Honor winners on the World Wide Web and includes his name in it. Perhaps a deepfake's First Amendment status is more like that of this fake government web site than it is like that of a false claim.

The Justices didn't directly address the question of whether such a fake government database would count as protected speech.

---

[200] Thomas v. Collins, 323 U.S. 516, 545 (1945) (Jackson, J., concurring) (citing W. Va. State Bd. of Educ. v. Barnette, 319 U.S. 624 (1943)).

But they said enough to make clear it would not. First, both Justice Kennedy's plurality opinion and Justice Breyer's concurrence said that the First Amendment can protect individuals' lies that they have won a Medal of Honor in part because listeners can check such statements against more reliable sources of information in the world outside of the speaker's control. Both suggested that an authoritative government web site could provide such a check. "A Government-created database [] list[ing] Congressional Medal of Honor winners," wrote Justice Kennedy, would make it "easy to verify and expose false claims," especially if it were "accessible through the Internet."[201] Justice Breyer agreed: An "accurate, publicly available register of military awards," he said, "may well adequately protect the integrity of an award against those who would falsely claim to have earned it."[202]

Of course, these statements by Justices Kennedy and Breyer only make sense if the "database" or "register" they suggest would *itself* be largely insulated against falsification. Such a web site would hardly be an authoritative way to expose a liar's false statements if the web site itself were just as vulnerable to that liar's control. But it isn't as vulnerable—not only because practical realities make it harder for a liar to pose as the government than to make false statements (for example, a private web site creator may have difficulty in getting a .gov address)[203] but also because of First Amendment law: Both Kennedy and Breyer stress that while the First Amendment protects lying in the absence of certain harms, it

---

[201] United States v. Alvarez, 567 U.S. 709, 729 (2012) (plurality opinion).
[202] *Id*. at 738 (Breyer, J., concurring).
[203] *But see* Brian Krebs, *It's Way Too Easy to Get a .Gov Domain Name*, KREBSONSECURITY (Nov. 26, 2019), https://krebsonsecurity.com/2019/11/its-way-too-easy-to-get-a-gov-domain-name (arguing that "trust [in .gov domain names] may be severely misplaced, and that it is relatively straightforward for anyone to obtain their very own .gov domain.").

does *not* give individuals a right impersonate government officials or to fabricate government records.[204]

In their view, this is because when lying involves impersonation of a government official it comes with harms that disqualify it from the First Amendment protection that lying ordinarily receives. In Justice Kennedy's words, impersonation of government officials threatens "the integrity of Government processes, quite apart from merely restricting false speech."[205] This explanation of why the First Amendment leaves impersonators of government unprotected follows from Kennedy's broader analysis of lies' First Amendment status: He insists that lies can be restricted by government not simply when they are false, but when they cause *additional* harm above and beyond the falsity itself. So, if the falsity in impersonating a government official (or creating fake government records) is unprotected, it must be because it is not only deceptive, but harmful in other ways.

But there is something missing in this harm-based account of why imitating a government web site is punishable while lying is not. The problem lies not just in harms that are independent of deception (like financial or reputational harms that result from it)—but in the fool-proof nature of the deception itself.[206] It will be far harder for a listener or viewer to look outside of, and evaluate, liars' deception if these liars can plant such deception not only in their own words, but also in the external evidence the listener or viewer

---

[204] *Alvarez*, 567 U.S. at 721 (plurality opinion) (acknowledging that statutes "prohibit falsely representing that one is speaking on behalf of the Government" and explaining why these statutes are constitutional); *Id*. at 735 (Breyer, J., concurring) (same). Given that Justice Alito and the other dissenting Justices do not believe that Xavier Alvarez's verbal lie is protected by the First Amendment, see text accompanying *supra* notes 156-158, it follows they would likely reach the same conclusion about a fake government web site register or database record inaccurately listing him as a Medal of Honor winner.

[205] *Id*. at 721 (plurality opinion).

[206] *See* Blitz, *supra* note 31, at 67.

relies upon to check those statements. If such external evidence were to be absorbed in the realm of speech that is internal to Xavier Alvarez's First Amendment-shielded expression, and thus, subject to his unfettered alteration or fabrication, then it no longer provides solid epistemic ground for a listener to retreat to when faced with doubts about Alvarez's words.

In fact, Kennedy's and Breyer's opinions make clear that authoritative web registers or databases are valuable not only because—unlike questionable verbal claims—audiences can rely on them. They are also valuable because they make it possible for society to *afford* to leave liars with a First Amendment freedom to lie. As the Justices' arguments make clear, even where verbal lies threaten to undermine the government's "compelling" or "substantial" interest in protecting the integrity of the Medal of Honor, an authoritative web register provides back-up protection for this interest. Thanks to the safeguard provided by a web register or database, Alvarez's lie can do little harm: It can be easily debunked by checking it against an authoritative (and easily accessible) government record. In this sense, the web register imagined by the Justices functions as what Regina Rini calls an "epistemic backstop." It is a check against the deception those statements might cause. Such a backstop, as Rini points out, is also likely to serve as a deterrent to liars, who—once aware of a database's existence—will have to worry that their lie can be exposed as false. [207]

A fake government web site, then, has a First Amendment status quite different than that of a verbal lie. When it comes to verbal lies, as Justice Jackson has said, the First Amendment leaves

---

[207] *See* Rini, *supra* note 33, at 1, 3-4.

individuals to act as their *own* "watchm[e]n for truth."[208] They are supposed to be careful and vigilant explorers in the marketplace of ideas and sort out truth and falsity for themselves. But they are *not* left to their own devices in the same way when relying on an official government database. They can rather rely on its veracity. As Helen Norton has written, this is because there are many circumstances where it is crucial to know that a government message is really coming from the government.[209] Listeners, she observes, may need to place "automatic reliance" in evidence that tells them certain speech is an authoritative government message.[210]

Robert Post provides a more general framework for how First Amendment coverage relates to such a need for automatic reliance. The First Amendment, on his analysis, doesn't give speakers a right to freely shape all sources of information. It only does so where speakers and audiences have a relationship that is "dialogic and independent."[211] This is the case where, under existing social conventions, "audiences autonomously query" the "meaning and authority" of a work of art or other expression.[212] But not all sources of information we encounter fit this description. "Navigation charts," for example, "do not receive First Amendment protection, because we interpret them as speaking monologically to their audience, as inviting their audience to assume a position of dependence and to rely on them."[213] Navigation charts, in other words, are like authoritative government registers: They are

---

[208] Thomas v. Collins, 323 U.S. 516, 545 (1945) (Jackson, J., concurring) (citing W. Va. State Bd. of Educ. v. Barnette, 319 U.S. 624 (1943)).
[209] *See* Helen Norton, *The Measure of Government Speech: Identifying the Expression's Source*, 88 B.U. L. REV. 587, 597 (2008).
[210] *Id.*
[211] Robert Post, *Recuperating First Amendment Doctrine*, 47 STAN. L. REV. 1249, 1254 (1995).
[212] *Id.*
[213] *Id.*

information sources where the "primary legal value" is "protect[ing] the integrity of that reliance" audiences place in them— *not* giving authors or other creators free rein to control their content.[214] Thus, there is no First Amendment right to alter or forge a navigation chart.

It is worth making two observations about such government web registers and navigation charts before understanding how they might be a model for thinking about deepfakes and other non-testimonial falsehoods. First, the logic that removes them from the ambit of a deceiver (or other author's) control does not come from traditional doctrines of First Amendment coverage. The key question for a court is not whether such a government web site or navigation chart has an understandable particularized message that satisfies the Spence test. Nor is it whether a web site or navigation chart like this is "inherently expressive." It is rather whether, given the social conventions surrounding such an an information source, others can have a First Amendment *right* to manipulate it.[215]

Second, there *is* some First Amendment protection for *truthful* sharing of the information in these sources—but not for fabrication or falsification of it. This is particularly clear with respect to the government web register of Medal winners. If such a government web register existed, the First Amendment would likely protect a person's decision to print it out and share it with others. It would also likely protect someone if they shared a navigation chart with someone who wishes to read it. Such truthful sharing of

---

[214] *Id*. at 1254-55.

[215] *See id*. at 1252-55. As Post notes, in considering whether an information source is the kind that we will treat as one which we are expected to autonomously query, it is insufficient to simply inquire—as the Spence test does—into "speaker's intent, a specific message, and an audience's potential reception of that message." It is rather necessary to inquire into the "social conventions and practices shared by speakers and audience" about that information.

information doesn't undercut the reliance interest audiences bring to these sources. In fact, it may even be necessary to assure that reliance can be exercised: A ship pilot who can't obtain a navigation chart can't rely on it for safe guidance.

In fact, there is a First Amendment right for individuals to share such information with each other even when it does *not* amount to testimony because it does not contain an assertion of any kind. At times, a person's expression may, instead of asserting something as true, simply provide data for a listener to evaluate. For example, imagine one scientist decides to share with other scientists a series of temperature readings a thermometer has recorded because she believes it will be valuable for their research. This data does not amount to an assertion that the scientist has any particular belief about the data. It is even possible that she sends it without looking it at its content and thus remains unaware of what the data might indicate. She is, in this case, more a messenger who delivers the information than an author who shapes it. She is delivering *non-testimonial* data to an audience rather presenting it as her own testimony. Such a transfer of information, Jane Bambauer has argued, is nonetheless First Amendment expression.[216] And there are precedents of the Supreme Court and other courts that appear to agree. As already noted, the Second Circuit said in *Universal City Studios v. Corley*, in finding that the First Amendment protected the transfer of computer code, that free speech law protects "[e]ven dry information, devoid of advocacy, political relevance, or artistic expression."[217] The Supreme Court, in *Sorrell v. IMS Health*, cited *Corley* in noting that a data mining company's transfer of information it had collected about doctors' prescription practices

---

[216] Jane Bambauer, *Is Data Speech?*, 66 STAN. L. REV. 57, 61 (2014).
[217] Universal City Studios, Inc. v. Corley, 273 F.3d 429, 446 (2d Cir. 2001)

was likely First Amendment speech. Even unadorned facts, from which a speaker draws no conclusions, "are the beginning point for much of the speech that is most essential to advance human knowledge and to conduct human affairs."[218] Moreover, it is hard to see why the audio footage the Court shielded from state censorship in *Bartnicki v. Vopper* would constitute protected speech unless the First Amendment shielded raw information. The Court's opinion suggested the playing of the audio would be protected even if the radio station playing the recording didn't supplement it with any assertions. It was the sharing of the information in the audio itself, not merely any testimony that might accompany it, that received First Amendment protection.[219]

But that there is a First Amendment right to share accurate versions of such information sources does not mean there is a right to forge or falsify them. It is accurate facts, not false data, that constitutes the "beginning point for much of the speech that is most essential to advance human knowledge."[220] Government is *not* barred by the First Amendment from punishing individuals who provide *inaccurate* navigation or aeronautical charts.[221] Nor is it barred from imposing malpractice liability on doctors who give false medical information, even though is barred from restricting the sharing of honest and accurate medical advice by a doctor with a patient.[222] It may ban misleading advertisements by a company

---

[218] Sorrell v. IMS Health, Inc., 564 U.S. 552, 570 (2011).
[219] Bartnicki v. Vopper, 532 U.S. 514, 526-27 (2001).
[220] *Sorrell*, 564 U.S. at 570.
[221] *See, e.g.*, Saloomey v. Jeppesen & Co., 707 F.2d 671, 672 (2d Cir. 1983) (not raising any First Amendment concerns in permitting that suit for flawed navigation chart). *See also* Frederick Schauer, *The Boundaries of the First Amendment: A Preliminary Exploration of Constitutional Salience*, 11 HARV. L. REV. 1765, 1802 (2004) ("Liability for misleading [] maps . . . is generally (and silently) understood not to raise First Amendment issues.")
[222] *See* Pickup v. Brown, 740 F.3d 1208, 1226, 1231 (9th Cir. 2014); *see also* ROBERT C. POST, DEMOCRACY, EXPERTISE, AND ACADEMIC FREEDOM: A FIRST

even though the First Amendment doesn't allow it to censor or ban the communication of truthful commercial speech (unless it can meet certain intermediate scrutiny requirements).[223]

All of these are examples of an asymmetry that Justice Kennedy takes note of in *United States v. Alvarez*: "Some false speech," he observes, "may be prohibited even if analogous true speech could not be."[224] This is particularly true where the value of the "analogous true speech" is to meet a reliance interest that isn't met, and is in fact betrayed, when the speech is falsified in a way that is hidden from the audience. When someone forges or fabricates a web register or navigation chart, and adds false information to it, they are really engaged in two lies: One adding the false information to the speech, the second clothing their false information in the false garb of a normally reliable source (an authoritative government web site or navigation chart).

Forging a video with deepfake technology might at first seem to be quite different from all of these examples: It doesn't dress evidence up in the clothing of a government record or in that provided by an expert. But it does do something similar. A deepfake also hides the *source* of the information it delivers. It disguises a person's own (and traditionally easy to falsify) story about the world in the garb of a kind of record that has traditionally been much harder to falsify: a camera-captured record of light and

---

AMENDMENT JURISPRUDENCE FOR THE MODERN STATE 47 (2012) (writing that a theory of the First Amendment that "immediately converts every effort to regulate professional practice into a constitutional question is surely suspect [since] professional practices are subject to many regulations, like ordinary malpractice law, that do not" raise First Amendment questions).

[223] Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 563 (1980).

[224] United States v. Alvarez, 567 U.S. 709, 721 (plurality opinion).

sound.[225] To use the terminology I used earlier, it disguises a person's false testimony as a non-testimonial record. It gives a person's untrustworthy say-so the false appearance of proof that is *external* to her testimony. As Rini puts it, it takes testimony and gives it a non-testimonial form. It takes an author's vision of the world and transforms it into something we see not as another person's creation but as the "equivalent of our 'perception'—the equivalent of "see[ing] something with [our] own eyes."[226]

It is plausible then that First Amendment doctrine should allow government to protect the integrity of videos and audio footage—as well as other non-testimonial evidence—just as it allows it to protect the integrity of government-generated records and charts.[227] First, the mediated perception made possible by video

---

[225] As Helen Norton writes, information about the source of speech is especially valuable for listeners because it lets them make crucial judgments about the trustworthiness of speech, or in some cases about the evidentiary value of non-testimonial evidence. Helen Norton, *Robotic Speakers and Human Listeners*, 41 SEATTLE U. L. REV. 1145, 1150 (2018). One could argue that even when individuals are engaged in public discourse, the First Amendment does not protect forgeries in which they take on the guise of *another speaker*—but might allow them to engage in imitation of general *style or medium of communication*. However, our ability to rely on mediated or extended perception provided by video or other technology doesn't just depend on our being able to rely on specific speakers—but on the technology itself. The same is true of other non-testimonial evidence: When we rely on clocks, GPS location devices, or other instruments to be free from First Amendment-protected manipulation, it seems odd to think we do so only if we view the clock, GPS location tracker, or other instrument as carrying knowledge from a particular speaker.

[226] Regina Rini, *Deepfakes are Coming; We Can No Longer Believe What We See*, N.Y. TIMES (Jun. 10, 2019), https://www.nytimes.com/2019/06/10/opinion/deepfake-pelosi-video.html.

[227] Not only are deepfakes in some respects like a falsified navigation chart—a new form of deepfake is being used to generate a very similar kind of fake geographic information. As one article reported, "geographers are concerned about the spread of fake, AI-generated satellite imagery" that "could mislead in a variety of ways"—for example, by "creat[ing] hoaxes about wildfires or floods, or to discredit stories based on real satellite imagery." James Vincent, *Deepfake Satellite Imagery Poses a Non-So Distant Threat*, THE VERGE, Apr. 27, 2021, https://www.theverge.com/2021/4/27/22403741/deepfake-geography-satellite-imagery-ai-generated-fakes-threat. Deepfakes might thus not only hijack (and ultimately weaken) the reliance we place in visual records of events, but also that

and audio evidence can sometimes play the same role for audiences as does an official web site "register" or "database." It can be a source of knowledge that individuals can turn to and rely upon when they cannot fully rely upon testimony. As Rini emphasizes, videos can serve as "epistemic backstops" for testimony.[228] Footage from cell phones and various kinds of surveillance cameras often provide such a backstop: They can, as Rini points out, adjudicate "conflicting or confused testimony."[229] Video or audio- recordings provide a firm foundation for reaching confident conclusions where bias and imperfect memory may distort each witness's claims, and lead to conflict between them. This is one of the reasons that policymakers have strongly advocated that police wear body cameras so that footage is available to adjudicate claims of wrongful police behavior.[230]

In fact, video chats and audio calls provide a commonly used backstop for phishing attempts. Cybersecurity experts often advise individuals who receive an e-mail purporting to be from an employer, a friend, or a relative to exhibit some skepticism rather than assuming it is really from that person. For example, one article advises that, when employees encounter a suspicious e-mail, they might "reach out to the sender directly" to confirm they are really the one sending it or perhaps "reach out to other trustworthy people like coworkers and supervisors to confirm the content of the suspicious email."[231] In this case, direct contact with an e-mail

---

which we place in maps and other geographic records we rely upon to understand our environment and chart a path through it.

[228] Rini, *supra* note 33, at 1.

[229] *Id*. at 3.

[230] *See* Seth W. Stoughton, *Police Body-Worn Cameras*, 96 N.C. L. REV. 1363, 1365 (2018).

[231] Tom Kelly, *How Hackers Are Using COVID-19 to Find New Phishing Victims*, SECURITY (Jun. 23, 2020), https://www.securitymagazine.com/articles/92666-how-hackers-are-using-covid-19-to-find-new-phishing-victims.

sender—in person, in a phone call or via a video chat—acts as a way to verify that they are who they say they are. It is an epistemic territory we can retreat to that is safer—that is, harder to falsify—than the name or address of an e-mail we receive. Deepfakes, however, can bring an imposter even into this epistemic refuge. At least when our contact takes place over an audio or video call (rather than in person), deepfakes might allow strangers to take on the digital guise of people we know and trust. According to one report on deepfakes, hackers are already attempting to use deepfakes in this way.[232]

More generally, Rini points out, such developments might lead to growing distrust of deepfakes that "will gradually eliminate the epistemic credentials of *all* recordings, to an extent that" our loss of confidence in video and audio generate a "[b]ackstop cris[is]."[233] If a deceiver has the same First Amendment right to falsify a video as she does to falsify her words, video ceases to serve this function. Thus, it is conceivable that for video as for a web register, a false version of a video—a deceptive deepfake—might be unprotected by the First Amendment even if sharing an authentic camera-generated record of that video is protected.

IV. THE CONSTITUTIONAL CHALLENGES OF TRANSFORMATIVE TECHNOLOGIES

A. Deepfakes and Shifting Constitutional Boundary Lines

Part III argued that, unlike a deepfake that clearly embodies an author's artistic vision, political opinion, or testimony, a deepfake *lacks* the status of First Amendment expression when it clothes itself

---

[232] *See* DEEPTRACE LABS, THE STATE OF DEEPFAKES: LANDSCAPE, THREAT AND IMPACTS 13-14 **(**2019), https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

[233] Regina Rini, *supra* note 33, at 8. As noted earlier, this is what Chesney and Citron describe as "trust decay." *See* Chesney & Citron, *supra* note 6, at 1785-86. *See also* Mary Anne Franks & Ari Ezra Waldman, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, 78 MD. L. REV. 892, 895 (2019) ("deep fakes erode the trust that is necessary for social relationships").

in the garb of a non-testimonial record. Such a deepfake is a work of authorship—but it is not experienced that way by its audience. Just as the fake corpse described in Part I will predictably be perceived by its audience as a real corpse (and not as an artist's creation), so a deepfake security camera sequence, lifelogging video, or video feed on an Internet chat or meeting will—at least in the current day and age—predictably deceive its audience into seeing it as an unaltered camera recording. It smuggles authorship into a setting where an audience doesn't expect it and likely won't detect it. This kind of hijacking of our trust occurs not only when someone dresses testimony (or a collection of false data) in the garb of a non-testimonial record but, more generally, whenever they recast their own claims or artistry as any kind of record that speaks to us "monologically," and on which we thus will likely wrongly rely upon, believing it is an authoritative source of accurate information (and not only their false speech in disguise).

But there is a question and challenge that Part III left unanswered. Is it possible that, in emulating non-testimonial evidence, deepfakes don't simply emulate it but *transform* it into something else? What if, in an age of deepfakes, non-testimonial evidence *becomes* a setting where authorship *is* expected. Consider again the variation on the facts of *United States v. Alvarez* in which the defendant creates a deepfake video that shows President Reagan speaking of his bravery in battle and presenting him with a Medal of Honor. In the past, perhaps, a viewer of such a video would be justified it in assuming that such a video must be a record of an actual event. While Alvarez could insert false content into his own words, he could not without special effects technology available only to major movie studios, create a video of an event that never occurred. But in the current age of deepfakes, one might argue, a deepfake video does *not* betray a viewer's justified reliance on the

veracity of what a video shows because such reliance isn't justified in the first place.

Perhaps, one might thus argue, Part III is thus mistaken in saying that deepfakes must *remain* outside of the scope of the First Amendment when they emulate the type of non-testimonial video that we have seen, in the past, as speaking to us "monologically." Perhaps the rise of deepfakes instead moves such video to the "dialogic" side of Robert Post's dichotomy—so that when we see a video of a Medal of Honor ceremony we should not immediately assume it depicts an actual event captured by a camera but should instead examine it for ourselves, "autonomously query" it, and form our own judgment about whether to believe it or not. We should, in other words, assume we have to serve as our own "watchmen for truth," not only when we hear other's testimony, but when we see video footage or hear audio recordings. Just as we must question the content of the words they convey to us, perhaps we must, going forward, question whether any authoritative speech source (whether it is a video recording or other previously-authoritative record) is really what it appears to be. The remainder of this Part, however, argues that courts should resist easily surrendering the terrain of non-testimonial evidence, or other information to which we bring a reliance interest, to the realm of First Amendment-protected authorship.

B. Equilibrium Adjustment Theory, the Fourth Amendment and the First Amendment

Recent Fourth Amendment law provides a model for how First Amendment law can respond to transformative technologies, such as deepfakes. Fourth Amendment law protects individuals against "unreasonable searches" by shielding them—in the words of

the Court—against a "too permeating police surveillance."[234] The Fourth Amendment, for example, does not let police enter and search a person's home unless they have obtained a warrant based upon probable cause. But if courts adhered woodenly to a rule requiring a warrant only when police *physically* enter a home, they would let advances in technology erode that privacy—by letting police find new ways to gather information from a home without physically entering it. As Anthony Amsterdam has stressed, individuals should not have to crowd themselves into a smaller and smaller corner of private space to escape enhanced state surveillance: "[A]nyone," he notes, "can protect himself against surveillance by retiring to the cellar, cloaking all the windows with thick caulking, turning off the lights and remaining absolutely quiet."[235] But "this much withdrawal" should "not [be] required" given the kind of "freedom and privacy" the Fourth Amendment is meant to secure for us against state surveillance—especially in our homes.[236]

The Supreme Court put the force of its caselaw behind this point in *Kyllo v. United States* in 2001.[237] Home life is normally blocked from outsiders' view by the walls. But thermal imagers can construct a detailed picture of what lies behind such walls.[238] As the Court wrote, technologies such as thermal imagers thus confront the court with the question of "what limits there are upon this power of technology to shrink the realm of guaranteed privacy."[239] The Court's answer was that Fourth Amendment law does not "permit

---

[234] United States v. Di Re, 332 U.S. 581, 595 (1948).
[235] Anthony G. Amsterdam, *Perspectives on the Fourth Amendment*, 58 MINN. L. REV. 349, 402 (1974).
[236] *Id.*
[237] Kyllo v. United States, 533 U.S. 27, 40 (2001).
[238] *Id.*
[239] *Id.* at 34.

police technology to erode" the privacy we have traditionally found in the home.[240] Rather than restructure our lives so that we reveal less to technologically-empowered police observers, it is rather the police who have to find a way to use thermal imagers that preserves the in-home privacy that existed prior to the emergence of this technology.

Riley v. California is another example of how the Court adjusts Fourth Amendment law to new technologies:[241] Prior to 2014, Fourth Amendment doctrine allowed police to search all objects they found on the person of an arrestee—like a cigarette box, a wallet, or a notebook—and to do so without first obtaining a warrant.[242] But when the item police find on an arrestee's person is a Smartphone, this allows them to make a far greater warrantless intrusion into the arrested individual's privacy than they could before the age of Smartphones.[243] What they are searching then is akin to a massive library of data about the person that, if it was available at all in past times, was likely to available only in a person's home files and personal computers—places they cannot generally search without a warrant.[244] So the Court held that police now need such a warrant to search an arrestee's cell phone.[245]

These Fourth Amendment holdings are examples of what Orin Kerr calls "equilibrium adjustment." Kerr argues that the Supreme Court has often implicitly used such an approach to assure

---

[240] *Id.*
[241] Riley v. California, 573 U.S. 373, 403 (2014).
[242] *See* United States v. Robinson, 414 U.S. 218, 235 (1973) (holding in "a lawful custodial arrest a full search of the person is [] an exception to the warrant requirement of the Fourth Amendment" and a "reasonable" search).
[243] *Riley*, 573 U.S. at 386.
[244] *See Riley*, 572 U.S. at 393 ("One of the most notable distinguishing features of modern cell phones is their immense storage capacity. Before cell phones, a search of a person was limited by physical realities and tended as a general matter to constitute only a narrow intrusion on privacy.").
[245] *Id.* at 403.

the Fourth Amendment continues to preserve an existing constitutional balance between the public's need for effective crime investigation on the one hand and individuals' need for freedom from excessive surveillance on the other.[246] It is, says Kerr, "a judicial response to changing technology and social practice. When new tools and new practices threaten to expand or contract police power in a significant way, courts adjust the level of Fourth Amendment protection to try to restore the prior equilibrium."[247]

As Kerr's language makes clear, courts adjust the scope of Fourth Amendment protection not only when technological developments "shrink the realm of guaranteed privacy,"[248] but also when they expand it in ways that leave too little space for law enforcement to effectively investigate crime. For example, as Kerr points out, if thermal imagers cut deeply into the privacy of the home by letting law enforcement officers see what is within its walls, the building of fences and walls around the property surrounding a home does the opposite: Such a barrier doesn't simply block government officials from observing the inside of a home (including a lamp of the kind Danny Kyllo used for growing marijuana in *Kyllo*)—it blocks them from seeing an illegal marijuana growing operation *outside* of the home in a home owner's backyard.[249] Unlike the space inside the home, the "curtilage" outside the home is an area law enforcement have long been able to observe from public vantage points.[250] The Court in *California v. Ciraolo* therefore made it clear that when individuals use high fences expand the realm of activity that is insulated against

---

[246] Orin S. Kerr, *An Equilibrium-Adjustment Theory of the Fourth Amendment*, 125 HARV. L. REV. 476, 480 (2011).
[247] *Id.*
[248] *Id.* at 497.
[249] *Id.* at 524.
[250] *Id.*

government observation—and thus, make it a potential safe-zone for illegal activity—the Fourth Amendment does not block police from warrantlessly taking countermeasures, such as flying over a property with airplanes (where their view is no longer blocked by a fence).[251] As Chief Justice Burger emphasized in *Ciraolo*, government needs *somewhere* to look to collect the evidence they will need to obtain a warrant—and that somewhere is a realm they can observe from public space.[252] The Fourth Amendment safeguards a zone of privacy in the home and other private spaces, but will not let this zone swallow up all of public space, and thus, leave government with nowhere to begin an investigation.

First Amendment doctrine on deepfakes should draw on the same type of equilibrium adjustment. To some extent, First Amendment law already does. In *Universal City Studios v Corley*, the Second Circuit Court of Appeals issued a ruling on encryption code that provided, in First Amendment law, a kind of equilibrium adjustment akin to that which *Ciraolo* undertook in Fourth Amendment law. More specifically, the Second Circuit had to decide in *Corley* whether government had violated the First Amendment when it relied upon the Digital Millennium Copyright Act (DMCA) to prosecute publishers of a hacking magazine for disseminating the code in a decryption program which, when run, would let individuals circumvent the technology on digital video disks (DVDs) that blocked users from unauthorized viewing or copying of the movies or other content on these DVDs.[253] The publisher of the hacking magazine insisted that, in publishing the code, it was engaged in First Amendment speech—and the Second

---

[251] California v. Ciraolo, 476 U.S. 207, 209, 213-14.
[252] *See id.* at 213 (noting that observation from a public vantage point is "what a judicial officer needs to a provide a basis for a warrant").
[253] Universal City Studios v. Corley, 273 F.3d 429, 435-36 (2d Cir. 2001).

Circuit agreed: Publishing the code, it said, was an integral part of the way that the magazine communicated to its computer-savvy readers about various computer programs.[254] But computer programs have a double-character. Not only can they be read and understood by a human reader, they also can be executed by a computer to produce *non-speech* action, like cutting through a digital fence. Distributing computer code is thus not only a form of speech, but also a sharing of powerful tools for non-speech conduct. Software, for example, can now give people the means to conduct cyberattacks on power generators, hospitals, automobiles and other computer-connected devices in the "Internet of Things."

Responding to this development therefore required a form of equilibrium adjustment: Even though the Second Circuit treated the dissemination of programs *qua reading materials* as First Amendment expression, it treated dissemination of programs *qua cutting tools* as *non*-expressive, and let the government regulate this latter dimension of computer programs—just as it has always been able to regulate, without First Amendment constraints, the sharing of "skeleton keys to open door locks."[255] In other words, the fact that a variant of such a key had now taken the form of computer code—which can be read by an audience and not simply used by a copyright violator—didn't mean the government would henceforth be disabled from combatting this threat to property rights.

This is analogous to what the Supreme Court did in *Ciraolo*. In that case, it found that individuals could not stretch the super-strong privacy shielding the Fourth Amendment provided for the home so that it cloaked, not only the home's interior, but also a garden of marijuana plants outside. Similarly, in *Corley*, the Second

---

[254] *Id.* at 449.
[255] *Id.* at 452.

Circuit found that individuals could not use the expressive dimension of software code to stretch the First Amendment shield covering it so that it embraced not just the communication carried out with such code, but the *non-speech* action generated by the same code (the theft or threat of theft it makes possible). The digitization of property has forced the government's property-protection methods to move into a space—the realm of software—where it is in close quarters with First Amendment expression that occurs through, or about, software, but that doesn't mean such government property-protection methods are now barred by the First Amendment's free speech clause.

The use of deepfakes to fabricate non-testimonial records, this article suggests, demands a similar response. Modern life increasingly requires individuals to rely on digitally-mediated perceptions rather than seeing items in their immediate physical environment. To the extent that such digitally-mediated perception has to occur through video or audio records (and live feeds), it will share space with a medium of communication and knowledge-transmission that doesn't only extend our perception, but also serves as raw material for others' artistic and other expression. But that shouldn't mean that we have to abandon these modes of knowledge-transmission to those who would use them to manipulate us. Tools for theft don't morph from non-speech conduct into protected First Amendment speech when they take the form of software (even though that software also be used in expression). Manipulation of our perceptual knowledge likewise doesn't become fully protected speech as soon as it occurs through alteration of video footage or video feeds (even those such video can also be raw material for art or political speech).

One might resist this analogy between deepfakes and the decryption code regulated in *Corley* by highlighting a difference.

When a would-be intellectual property thief uses decryption code as the digital equivalent of "a skeleton key that can open a locked door, a combination that can open a safe, or a device that can neutralize the security device attached to a store's products,"[256] they are quite clearly using code for an act that is non-expressive. It isn't First Amendment speech to unlock doors or overcome the security that protects a safe or a store's products. By contrast, one might argue, one *is* in the realm of First Amendment expression when one adds to someone's knowledge or otherwise influences someone's beliefs, whether by sending them a text or e-mail, *or by showing them security camera footage to make them understand what happened* at a certain place and time. Unlocking a security measure with a digital skeleton key, in other words, lies just as clearly outside of the scope of the First Amendment as unlocking it with a physical key. But deepfaking security camera footage posted on web site or tweet, or sending it to others in a text message, does *not* lie as clearly outside the First Amendment's scope, so it doesn't make sense to demand that equilibrium adjustment keep it there when it is not there in the first place.

But as Part III argued, it is too simple to classify all information sources—even when falsified—as a kind of First Amendment-protected expression. It is not only physical action (like unlocking a barrier) that might place an activity outside the First Amendment's shield for expression but also our *reliance* interests in the integrity or accuracy of certain information. Deception, in other words, might become fair game for regulators not only when it threatens us with physical or financial harm—but also when it undercuts the epistemic backstops we rely upon to make sense of, and learn about, the world around us. Consider again the most

[256] *Id*. at 453.

plausible reason that the First Amendment would let government punish a deceiver's manipulation of an official database of military award winners. It isn't because such falsification would necessarily cause financial or physical harm. It is because without such a database, we might have no easy way of verifying others' testimony about these issues. The Fourth Amendment establishes and safeguards an equilibrium between our privacy interests and the public's security interests. It recognizes spaces—the home and other private areas—where we can have extraordinary insulation against government surveillance. But it also recognize other territory—including most public and observable activity—where government can vigorously investigate. Similarly, the First Amendment should be understood as establishing and safeguarding an equilibrium between speakers' interests in autonomy and audiences' reliance interests. It recognizes certain activities, such as speakers' testimony and artistic expression, where speakers get extraordinary insulation against government restriction of their authorship and self-definition. By contrast, it treats other information sources—such as government records or non-testimonial records—as realms where authorship can be limited to protect an audience's expectations that the information is accurate.[257]

V. DEEPFAKE DECEPTION, PUBLIC DISCOURSE, AND ARTISTIC EXPRESSION

A. Deepfake Deceptions and Public Discourse

It is not only technological shifts—like the emergence of

[257] Helen Norton has proposed other ways in which First Amendment law might appropriately balance speakers' interests in autonomy and audiences' reliance interests. *See, e.g.*, *Powerful Speakers and Their Listeners*, 90 U. COLO. L. REV. 441, 453 (2019) (discussing how "[a] listener-centered approach thus understands the First Amendment to permit the government to require comparatively knowledgeable and powerful speakers to make accurate disclosures about certain matters, even if those speakers resist their discussion.").

deepfakes—that can blur the First Amendment boundary lines between information reserved for author's control (such as their art or the verbal claims they make) and information in which the integrity of audiences' reliance interests is the primary value. Such blurring can *also* occur when speakers themselves take an information source from one context and repackage or repurpose it.

For example, imagine that speakers take an information source that *usually* speaks to us monologically and move it to a realm, such as that of public discourse or art, where we are usually expected to question what we see or judge it for ourselves. One might argue that, once shifted to this new realm, the information is *no longer* the kind an audience can rely upon. If we want a medical report that we can rely on to make important decisions about our health, for example, we can find it in the private communications we have with our physician, in the context of a professional physician-patient relationship.[258] If we want to know how to use an electrical product safely, we can rely on the guidance that the product manufacturer is duty-bound to provide us. If the navigator of a ship or plane wants to rely on a navigation chart, they can use one provided by the Federal Aviation Administration or a commercial mapmaker that owes them a duty of accuracy.

By contrast, one might argue, none of these information sources can be trusted in the same way when we instead receive them in the rough-and-tumble of public debate. If we see medical advice on a social media site, even that of a licensed physician, we cannot rely on it the way we would rely on the advice given to us by our own physician in the context of a doctor-patient relationship.[259]

---

[258] *See* POST, *supra* note 225, at 12-13, 43-45 (2012).

[259] Jane Bambauer has argued that experts' statements should in some cases be left, by the First Amendment, subject to government restriction even when they occur in public discourse, if they occur in circumstances where an audience will

The same is true of advice on how to safely use a product or a navigation chart we receive not from a government agency or professional map-maker but rather on a web site where it has been posted by some anonymous source. In other words, one might argue that navigation charts and other sources of information no longer "spea[k] monologically to" us and credibly "invite" us to "assume a position of dependence and to rely on them,"[260] where they have been torn out of the fiduciary or commercial relationships that justify such reliance—and placed in the realm of public discourse. They have now, one might argue, been transformed into a part-and-parcel of someone's highly questionable expression.

One might argue that the same is true of security camera footage, lifelogging, or other video recordings. We can trust footage that comes from a camera that we own, or that we receive from a security company that vouches for its accuracy. What we cannot do, one might argue, is extend the *same* trust to security camera footage that we find on a social media site, because *that* camera footage is a part of public discourse, and thus, fair game for speakers to edit (or fabricate) in any way they like. When we encounter video or other non-testimonial evidence in public discourse, in other words, the First Amendment requires we serve as our own "watchmen for truth." It is only *outside of public discourse* that we can insist others have a duty to provide us with information (verbal, visual, or in another form) that is accurate—in sources of information that are protected against hacking or other alteration by our property rights in that information (of that of a company or organization that commits to keep it secure), or in certain fiduciary relationships.

---

likely and justifiably rely on those statement. Jane R. Bambauer, *Snake Oil Speech*, 93 WASH. L. REV. 73, 76 (2018).

[260] Post, *supra* note 214, at 1254.

B. Deepfake Deceptions and Artistic Expression

A similar blurring of First Amendment boundaries can occur when an artist commandeers an audience's trust in a non-testimonial source of knowledge—or other authoritative record—and makes it a part of an artistic performance. Earlier in the article, I discussed how some merging of artistic fiction and other practices (like contacting someone on Zoom chats) generate a risk of deception. But in some art or gameplay, deception of an audience isn't simply a worrisome risk—it is one of the *goal*s of the artist or game designer. That is, a kind of artwork or virtual game can't succeed unless there is at least temporary deception.[261]

Perhaps the most famous example of this is Orson Welles's 1938 War of the Worlds radio broadcast. In that broadcast, he famously terrified thousands of radio listeners by reading a passage from H.G. Wells's alien invasion story, *The War of the Worlds*, as though he was a reporter conveying breaking news.[262] That deception worked not just because Welles took on the false role of a news reporter at the site of an alien landing, but also because he read this fictional passage in a communication channel (news radio) that many listeners treated as a source of trusted factual reporting.[263] The effect on his audiences was similar to that which someone today (or in the near future) might cause by creating a deepfake of the United States President appearing on television to gravely warn of an alien

---

[261] It is arguably central to the work of illusionists, such as David Copperfield, who made the Statue of Liberty seem to disappear in a 1983 show—although one might argue that the audiences for such shows realize that the illusionist's talent is to make them see what has not actually occurred. *Cf. David Copperfield: Statue Of Liberty Explained: How Did He Do It?*, REBELMAGIC, https://rebelmagic.com/david-copperfield-statue-of-liberty (last visited Sept. 26, 2017).

[262] A. Brad Schwarz, *The Infamous "War of the Worlds" Radio Broadcast Was a Magnificent Fluke*, SMITHSONIAN MAGAZINE (May 6, 2015), https://www.smithsonianmag.com/history/infamous-war-worlds-radio-broadcast-was-magnificent-fluke-180955180.

[263] *Id.*

invasion in progress, or a deepfake that actually shows vivid video footage of the aliens landing, emerging from spaceships, and attacking pedestrians.

Wells is far from the only artist who made deception an integral part of artistic work or performance. The 1999 horror film, The Blair Witch Project, became a cultural phenomenon in large part because the film itself—and the advertising campaign surrounding it—temporarily succeeded in convincing many viewers that it was a documentary exploring the unsettling disappearance of three film students searching for a legendary witch.[264] The movie-makers did so by clothing their fictional movie (in its form, and in its marketing) with the feel and conventions of a documentary.[265]

Other performance artists and actors go beyond simply telling autobiographical lies with words. Their crafting of personas to present others occurs not merely through language—but with disguises, alteration of appearance, and skillful acting. The actor, Joaquin Phoenix, for example, spent eighteen months pretending to be a fictional version of himself: a taciturn, jaded, and erratic individual who had left acting behind to begin a career in hip hop music.[266] He ultimately revealed this new identity was an act created as an "experiment" and to provide material for a documentary by Casey Affleck.[267] Comedians such as Andy Kaufman and Sacha Baron Cohen have similarly created false

---

[264] *See* Joe Berkowitz, *Blair Witch: The Challenge of Following Up the Most Effective Marketing Campaign Ever*, FAST COMPANY MAGAZINE (Sept. 21, 2016), https://www.fastcompany.com/3063621/the-challenge-of-making-a-sequel-to-the-most-effective-movie-marketing-campaign; THE BLAIR WITCH PROJECT (Haxan 1999).
[265] *See* Berkowitz, *supra* note 267.
[266] *See* Ben Child, *Joaquin Phoenix 'Documentary' I'm Still Here is a Fake, Casey Affleck Admits*, GUARDIAN (Sept. 17, 2010), https://www.theguardian.com/film/2010/sep/17/im-still-here-fake-affleck.
[267] *Id.*

personas.[268] Candid Camera used trickery in the late twentieth-century by confronting unsuspecting individuals with staged scenarios.[269]

One can also imagine other immersive or interactive art experiences where individuals *invite* deception. Those who play the "Turing test" version of the imitation game, for example, usually do so knowing that the very point of this game is for a remote computer to try to deceive them into believing it is a human interlocutor.[270] And they welcome the challenge. Virtual reality gameplayers or explorers may similarly recruit the aid of a designer of VR—or other digital environments—to try to deceive them with fake environments (digital or physical), with artificially-intelligent "bots" that imitate human speakers, or with deepfake videos. The experience of being deceived, in other words, might have value for many who seek out videos or other images not simply as a source of truth, but as part of a practice where brief confusion about the truth is integral to learning, or being entertained, or appreciating a kind of performance art. [271] In some cases, like the trickery of Joaquin

---

[268] *See* Aja Romano, *Sacha Baron Cohen's Political Provocations are Exhausting and Dangerous*, VOX (Jul. 13, 2018), https://www.vox.com/culture/2018/7/13/17568448/sacha-baron-cohen-roy-moore-sarah-palin-who-is-america-fake-news; Roberta Smith, *A Comedian as Artist*, N.Y. TIMES (Feb. 8, 2013), https://www.nytimes.com/2013/02/09/arts/design/creating-reality-by-andy-kaufman-at-maccarone.html (describing an art exhibition, "On Creating Reality, by Andy Kaufman," on Andy Kaufman's creation of personas and other conceptual art); Jordan Zakarin, *How Andy Kaufman Invented Half of Modern Day Comedy*, BUZZFEED (July 19, 2013), https://www.buzzfeed.com/jordanzakarin/andy-kaufman-influence-on-comedy.

[269] *Candid Camera, Television Academy Foundation:* INTERVIEWS, https://interviews.televisionacademy.com/shows/candid-camera.

[270] *See* Graham Oppy & David Dowe, *Turing Test*, STAN. ENCYCLOPEDIA PHIL. (Apr. 9, 2003) (revised Feb. 8, 2016).

[271] In fact, the First Amendment's protection for freedom of thought might conceivably give individuals a right to deceive themselves with virtual reality, deepfakes, or other technologies—even if *others* don't have a right to manipulate them into such false perceptions or beliefs. *See* Marc Jonathan Blitz, *The Right to*

Phoenix or Sacha Baron Cohen, it may be part of the way these artists convey a cultural or political message—by emphasizing how much of what passes for "reality TV" might be unreal (in Phoenix's case) or to critique the simplicity of certain political views (as in Baron Cohen's). Alan Chen has written that false verbal statements—including those in fake news—may have value in part because of the emotional satisfaction they can provide to listeners and viewers, and that may be another reason some audiences may seek deepfakes or other deception.[272]

The larger question here is whether such artists or other speakers have a First Amendment right to borrow cues—or other appearances—that viewers traditionally rely upon as signals that invite reliance, but then use these cues for deceptive artistic or political speech that has a very different value for audiences.

Deepfakes and other technological advances complicate this question because they make it increasingly easy for speakers to borrow (or fabricate) fake versions of these cues or appearances. As I have written in previous scholarship on the First Amendment status of forgery, "[n]ow that newspapers are on websites, the code or design of which can be easily copied by digital means, creating fake versions of established newspapers is far simpler than it was when newspaper production relied on possessing and using a powerful printing press."[273] It is far easier now than it once was to clothe their own false claims in the appearance of a Washington Post, Miami Herald, or Chicago Tribune article.

Filming tricks and editing tools have long given filmmakers

---

an Artificial Reality?: Freedom of Thought and the Fiction of Philip K. Dick, 27 MICH. TECH. L. REV. 377 (2021).

[272] *See* Alan Chen, *Free Speech, Rational Deliberation, and Some Truths About Lies*, 62 WM. & MARY L. REV. 357, 395 (2020).

[273] *See* Blitz, *supra* note 31, at 114.

the power to mask their shaping of a film—making it seem like unedited footage. But deepfakes and other artificial-intelligence-generated animation greatly enhances this power. In the past, even when we see a video that clearly has an author, such as a documentary, news report, or social media post, we have generally assumed that not *every* aspect of the video can be authored or edited: Watching a video captured from a plane flying over the Andes mountains, for example, such as that in the television documentary, *Magical Andes*[274]or *Kingdoms of the Sky*,[275] we assume that the video-maker has determined which shots to take and how to edit the footage, and perhaps has shaped some of the conversations through his interactions with the people or environment he captures—but not that he was the author, in any respect, of the Andes. Rather than shaping each and every aspect of the scene ex nihilo, the camera footage is constrained by the features of a material world it captures.

Deepfake and similar computer technology frees a video creator from the limits of this material reality. Rather than editing within the constraints of the reflected light that our cameras have actually captured from the Andes, we can instead create camera images without relying on a camera, or the external light it captures, designing—from our minds—the mountain range we once relied on external reality to provide. The mountain range that the filmmaker once saw with her own eyes, and could choose to capture on film, she can now sculpt into any form she wishes it to have in her film. She is an artist who designs the mountain range rather than a visual chronicler who captures a record of its appearance.

C. Safe Zones, Authentication, and Shelters from Deepfake Deceptions

---

[274] *See* ANDES MAGICOS (Trailer Film 2019).
[275] *See* KINGDOMS OF THE SKY: HIMALAYA, ROCKIES, ANDES (PBS 2018).

One might argue that, in a world where the appearance of an authoritative video or other record is free for the taking, our reliance interests have to retreat, in a sense, back to safer territory—where reliance is justified not just by our perception of visual cues or other appearances, but by more robust, harder-to-falsify authentication measures, or other safeguards.

First, one might argue, we can rely on property rights that continue to protect—from fabrication or alteration—speech that remains firmly under our control. The First Amendment leaves government leeway to protect—against hackers—the integrity of security camera footage, lifelog videos, or private video chats. Sabotaging such records often involves violating another's property rights—for example, by breaking into a camera or a computer server owned by someone else. The reason a person can typically trust camera footage in these cases is not simply that it is non-testimonial evidence, but that it is secured—technologically and with legal property protections—against outside interference. *Alvarez* protects lying, and perhaps the visual equivalent with a deepfake. But it doesn't protect a speaker's violation of others' property rights. The First Amendment may give me a right to create my own deepfake showing a fictional event on Fifth Avenue in New York. But it doesn't give me a right to insert that deepfake into the surveillance camera archives kept by a local store owner or body camera footage archives kept by a police station—because those aren't for me to access. The federal Computer Fraud and Abuse Act (CFAA) doesn't violate the First Amendment when it makes it a criminal offense for a hacker to access a U.S. government computer—or another computer that isn't his—and swap a file on it with another one he has created. Moreover, as noted earlier, under traditional tests for First Amendment coverage, hacking into and modifying camera footage or other stored records likely wouldn't count as

"speech."[276]

Second, private actors can create other kinds of refuge from deception. Chesney and Citron imagine something like this when they envision a video lifelogging service of the future that might serve as a backstop against deepfake technology itself. Armed with "immutable lifelogs or authentication trails," "a victim of a deep fake" could "produce a certified alibi credibly proving that he or she did not do or say the thing depicted."[277] Of course, just as web site database of military award winners couldn't provide us with a way to check and expose liars if the web site might itself be a lie, so lifelogging footage couldn't protect against deepfakes if it was itself vulnerable to being faked. Aware of this concern, Chesney and Citron emphasize that, in order for lifelogging to serve as an effective defense against deepfakes, it cannot be the product solely of an individual recording her own experience—it must rather be the product of a more complicated process, likely presided over by companies with trusted authentication mechanisms, perhaps built upon blockchain-based recording keeping, A lifelogging service, they write, could only serve as an antidote to deepfakes if it "earn[ed] a strong reputation for the immutability and comprehensiveness of its data; the service otherwise would not have the desired effect when called upon in the face of an otherwise-devastating deep fake."[278] In short, because videorecording can no longer be trusted in an age of deepfakes, an effective lifelogging system has to have *another* epistemic layer which (unlike the video itself) is better fortified against a deceiver's capacity to fake. Rather than let government attempt to roll back all deepfake fabrication of

---

[276] *See supra* text accompanying note 138.
[277] Chesney & Citron, *supra* note 6, at 1814.
[278] *Id.*

video, then, the First Amendment could instead leave government with the narrower power to prevent deceivers from posing as the lifelogging service itself. This type of reaction to the rise of deepfakes is not unlike the reaction that Jonathan Zittrain writes has arisen in response to the cybersecurity challenges of an open Internet on which anyone is free to share the content they wish to share. As Zittrain observes, although the open Internet provides the benefit of giving computer users tools to create and share the programs and other content of their choice, it also leaves individual computer users vulnerable to a host of threats—such as computer viruses, worms, and other cybersecurity attacks.[279] This has moved technology companies to offer and consumers to adopt—"locked down" systems like those one finds in iPhones and other smartphones—environments which trade away the freedom that came with an unmonitored connection between one's computer and the Internet for a well-protected path to it that it is controlled and closely-monitored by Apple, Google and other companies and can better keep cyber-criminals and other threats out. Deepfakes threaten to do the same thing to our epistemic sources that cybersecurity threats have already done to our Internet activities.[280] They might drive us into "safe zones," closely managed by actors in the government or private industry, that can assure that the videos or other non-testimonial evidence we rely upon is authentic. In fact, as noted below, some legislators and authors have urged Facebook, Twitter, YouTube and other prominent social media sites to establish such a deepfake-free zone on their own sites—and many of these social media companies have answered this call (as part of a

---

[279] *See* JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET—AND HOW TO STOP IT 3-5 (2008).
[280] *Id.*

larger push to police and exclude what they identify as misinformation).[281]

Such a safe-zone-oriented solution arguably has the advantage also of leaving political and artistic expression unfettered outside of these zones—where speakers can still engage in unconstrained public discourse free of government restriction and have First Amendment rights to shape this discourse not only with the testimony of their choice, but also with non-testimonial sources of information (or fabrications of it).

D. First Amendment Space for Regulating Forgeries and Fabrications

This article does not entirely reject this stance on deepfakes. When the deceptive power of a deepfake is intertwined with artistic, political, or other protected expression, it cannot be an excuse for the government to simply suppress that expression. But reiterating the argument made in the previous Part, on equilibrium adjustment, this article argues there is a better First Amendment response than leaving audience reliance interests with no alternative but to retreat into safe zones marked out by property-rights and secured by technology, or to rely on authentication methods that might require audiences to suspend belief where they could previously exhibit automatic reliance.

Rather, I argue here, the First Amendment should leave government with a broader range of tools to counteract deepfake deception (and other kinds of non-testimonial falsehoods) even when they occur in public discourse, or merge with art. It should give government leeway to impose certain kinds of disclosure requirements, and more generally address certain deepfake deceptions in ways that comport with intermediate scrutiny and

---

[281] *See infra* text accompanying notes 303-307.

viewpoint neutrality requirements.

A part of this may involve helping to enable and protect authentication mechanisms developed by the government and private actors. The United States Army has recently announced a technology called "DefakeHop" to detect and expose deepfakes.[282] Computer scientists at the University of Buffalo have developed a system for detecting deepfakes of people by analyzing the reflections in the eyes of the individuals pictured in the deepfake.[283] Microsoft and Facebook have also recently made progress in developing deepfake detection technology.[284] Perhaps then the law will not need to restrict deceptive deepfakes if audiences refrain from trusting them in the absence of some new, secure method of technological authentication.

To use Regina Rini's terminology, we might use technology to provide a new "epistemic backstop" (like life-logging

---

[282] *See Breakthrough Army Technology is a Game Changer for Deepfake Detection*, United States Army, Apr. 29, 2021. https://www.army.mil/article/245728/breakthrough_army_technology_is_a_game_changer_for_deepfake_detection

[283] *See Scientists Developed a Clever Way to Detect Deepfakes by Analyzing Light Reflections in the Eyes*, NEURAL, Mar. 11, 2021. https://thenextweb.com/news/ai-detects-deepfakes-analyzing-light-reflections-in-the-cornea-eyes-gans-thispersondoesnotexist

[284] On September 1, 2020, Microsoft announced that it was releasing "Microsoft Video Authenticator," a program that "can analyze a still photo or video to provide a percentage chance, or confidence score, that the media is artificially manipulated" and "[i]n the case of a video . . . can provide this percentage in real-time on each frame as the video plays." *See* Tom Burt, *New Steps to Combat Disinformation*, MICROSOFT ON THE ISSUES, (Sept. 1, 2020), https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/. Facebook had earlier held a "Deepfake Detection Challenge" to "spur creation of innovative new technologies to detect deepfakes and manipulated media." *See Deepfake Detection Challenge Results: An Open Initiative to Advance AI*, FACEBOOK AI (June 12, 2020), https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/.

[284] *See, e.g.*, Occupy Columbia v. Haley, 866 F. Supp. 2d 545, 560 (D.S.C. 2011) (finding that protestors were likely to succeed in challenging a policy preventing them from being on the South Carolina State House grounds after 6:00 pm where "there [was] no evidence that the 6:00 p.m. policy has been applied consistently to all organizations and groups seeking to use the State House grounds")

authentication which uses blockchain technology) to *replace* the one that deepfake technology has eroded. However, video authentication will not successfully restore this equilibrium if it leaves audiences with significant burdens in verifying information they could once trust automatically. Again, as Norton pointed out with respect to government data, audiences often need to place "automatic reliance" on certain data sources without having to conduct an in-depth investigation of their veracity.[285] This kind of automatic reliance takes place, for example, when we assume that "seeing is believing" and trust the evidence before our own eyes. If that is replaced by an approach that "verifying is believing," the verification isn't really a sufficient replacement for our now-lost ability to trust our perceptions unless it can take place without substantially more burden than we used to face in trusting what videos show.

Fourth Amendment analogies can once again be instructive here: As Anthony Amsterdam has argued, when the government invades the privacy of the home, we can't plausibly say that this lost privacy has been restored if it can only be reclaimed by retreating into a small corner of our homes and by "cloaking all the windows with thick caulking, turning off the lights and remaining absolutely quiet."[286] Similarly, we can't plausibly say that audiences will have regained the ability to rely on camera records (or other false records) if the reliance they once placed automatically in these records can now be confidently provided only *after* a careful and somewhat laborious investigation. They will only do so if they allow us to navigate the world as effectively as our now-suspect perceptual knowledge once did.

---

[285] *See* Norton, supra note 212, at 597.
[286] Amsterdam, *supra* note 238, at 402.

Moreover, as Hany Farid points out, some of the same artificial intelligence improvements that enable detection of deepfakes might also allow deceivers to make them harder to detect[287]—and Rini, drawing upon this observation, emphasizes that measures for detecting deepfakes might thus just be the "first move in a machine learning arms race, where fakers continually change strategies to stay a bit ahead of detection."[288] The equilibrium restored by technology, in other words, might quickly be undermined by it again.

One might also argue that, even when we *lack* methods of identifying non-testimonial evidence as a fabrication, it should still count as protected speech when used by someone to support their testimony. Consider how the First Amendment should apply to what we might describe as *"show and tell" speech*. I might display an astronomical instrument like an astrolabe, for example, in explaining how medieval navigators used it to calculate the time of day or their position, or how it is constructed. Or I could hold up a navigation chart to support my argument that a certain route to a destination is superior to another one. Or I could share a GPS location record to support my claim that I was in a certain place on a certain day. Or show video footage (perhaps from a security camera) to do so. In each of these cases, the non-testimonial evidence I share isn't simply a part of my testimony, It is rather evidence an audience is likely to view as external to my speech—as something I obtained rather than created. But on the argument I am considering here, non-testimonial evidence *becomes* First Amendment speech as soon as it is *folded* into public discourse—

---

[287] Hany Farid, *Digital Forensics in a Post-Truth Age*, 289 FORENSIC SCIENCE INTERNATIONAL 268 (2018).
[288] Rini, *supra* note 33, at 7.

and thus fair game for the speaker to surreptitiously alter and falsify. (It is only when a fiduciary duty of some kind rules out this deception that the speaker loses the First Amendment right normally possessed by speakers to manipulate the content of their expression).

But there are reasons to doubt that such an argument for extending First Amendment protection to fabricated evidence. First, it certainly is not always true that someone's use of non-testimonial evidence becomes First Amendment speech as soon as they use it to support an argument: Imagine that, to support my claim that a crime has occurred, I direct my audience's attention to a fake crime scene with a fabricated corpse of the kind described in Part 1-A. That I am showing my audience a part of the environment to support my claim doesn't give me a First Amendment right to manipulate that environment.

One might argue that when videos are used by someone to support their claim, they have a much greater claim to count as First Amendment expression than do many other props. To a greater extent than many other forms of non-testimonial evidence, video footage is *designed* to be shaped and reshaped with widely-available editing tools. Deepfakes, one might argue, only add to already existing tools for editing videos—and thus audiences should be expected to question the accuracy of video more carefully when such editing is possible than when they obtain security camera footage from a system that is well-insulated against outsiders' alteration of recordings.

But matters are different, Part III has suggested, where components of the video serve a function that isn't consistent with authorship—and which it can't serve unless we can rely on the video as a *record* of what actually occurred. Courts have not found that we have a First Amendment right to forge the records we use to prove to others that certain things as true about us or about our

history. In *United States v. O'Brien*, for example, the Court found the First Amendment permitted Congress to enact a federal law punishing any individual who "forges, alters, knowingly destroys, knowingly mutilates, or in any manner changes" a certain kind of government record—namely, a Selective Service Registration Certificate.[289] The Court's focus was on David O'Brien's burning (and destruction of) his draft card. But the Court also stressed that a legal prohibition on "forgery" or other "deceptive misuse of certificates" was "clearly valid."[290] In short, First Amendment law appears to leave government with leeway to protect the integrity of the record itself—however it is shared with an audience, and no matter who does the sharing—whether in private communications or public discourse. If this is true for government records, it also seems likely to be true for business records or marks of authenticity that private actors use to identify an item is genuine (such as the authentications of baseball cards or signatures done by Professional Sports Authenticator (PSA)).[291] Courts have faced (and disagreed about) similar questions involving the deceptive display of military medals or badges. They have disagreed about how *Alvarez* applies to situations where individuals falsely portray themselves as military award winners not with words (as Alvarez did) but by wearing military medals they didn't receive (in violation of a now-repealed ban on such medal wearing in 18 U.S.C. §704(a)). In *United States v. Swisher*, the Ninth Circuit found that bans on 18 U.S.C. 704(a)'s

---

[289] United States v. O'Brien, 391 U.S. 367, 379–80 (1968).
[290] *Id.*
[291] *See* Ottiano v. Prof'l Sports Authenticator, No. CV09-00025-PHX-MHM, 2009 WL 3722996, at *1, *1 (D. Ariz. Nov. 4, 2009) ("PSA is an independent sports and trading card authentication and grading service that is located in Santa Ana, California. . . . its mark and grade is relied upon by collectors as proof that a trading card is genuine and that it has not been altered in any way from its original issue except by the ravages of time."); *Professional Sports Authenticator: Services Offered*, PSA, https://www.psacard.com/services/tradingcardgrading/.

restriction of deceptive wearing of medals violated the First Amendment for the same reason that 704(b)'s ban on false claims about such medals did so: It barred a speaker's attempt to convey a message—a message that he was entitled to the respect owed winners of such medals.[292] Medal-wearing is expression, and false expression is protected. The Fourth Circuit—in *United States v. Hamilton*—reached the contrary result, upholding § 704's ban on unauthorized medal wearing.[293] It distinguished *Alvarez*, finding that the deception perpetrated by a medal wearer was harder to expose as false: "Although speech [that is true] may effectively counter other matters that a person hears," the court said, "speech may not effectively counter that which a person sees."[294]

Essentially, for the Ninth Circuit, the wearing of the medal is like testimony in a different form: It is a message an audience can question and reject. For the Fourth Circuit, the visible medal has a different impact on its audience: It will be harder for the audience to question the proof of an award viewers see with their own eyes than to question someone's verbal self-characterization.

The foregoing discussion in this article provides more support for the Fourth Circuit's position. Imagine the government gave military heroes government-issued identification cards that only they could possess. If it can protect the integrity of such ID cards, as *O'Brien* suggests it can, then why can it not protect the integrity of medals that serve the same function? Of course, wearing a medal one hasn't received is not exactly the same as forging or altering one. Federal law continues to forbid unauthorized

---

[292] United States v. Swisher, 811 F.3d 299, 305 (9th Cir. 2016).
[293] United States v. Hamilton, 699 F.3d 356, 373(4th Cir. 2012).
[294] *Id.*

manufacture of Medals of Honor.[295] But this too cuts against finding that deceptive medal-wearing is speech: If it were, one would expect the First Amendment to protect not only the expression that individuals engage in when they wear medals they haven't earned, but also the medal-creation that makes this deceptive expression possible.[296] The larger point here, and in the previous part, is that where art and public discourse become intertwined with other informational practices that government has an interest in regulating—to protect audiences' reliance or other interests—this intertwining doesn't force government to simply surrender its interest in protecting the public. Even where commercial speakers recruit skilled film directors,

---

[295] *See* 18 U.S.C.A. § 704(a) (West) ("Whoever knowingly. . . manufactures . . . any decoration or medal authorized by Congress for the armed forces of the United States, or any of the service medals or badges awarded to the members of such forces, or the ribbon, button, or rosette of any such badge, decoration or medal, or any colorable imitation thereof, except when authorized under regulations made pursuant to law, shall be fined under this title or imprisoned not more than six months, or both.").

[296] As Ashutosh Bhagwat explains, First Amendment expression would be insecure if the Constitution protected only the act of communication—and offered no protection to "antecedent conduct necessary to produce a desired communication." Ashutosh Bhagwat, *Producing Speech*, 56 WM. & MARY L. REV. 1029, 1034-1036 (2015). Thus, as explained earlier, where videos are used to convey information, it is not merely disseminating a video, but also recording and creating it, that is protected. *See supra* text accompanying notes 121-122.

This logic should also extend to wearing a fake medal. Assume, for the sake of argument, the Ninth Circuit was correct—that wearing a fake medal to falsely portray oneself as the recipient of a military award is an act of First Amendment-protected expression. One would then expect that the antecedent conduct necessary to engage in such expression should *also* receive at least some First Amendment protection against Congressional regulation. Under the framework Bhagwat proposes, in fact, this protection would be quite strong where the government is restricting the antecedent conduct *for the purpose of* restricting the expression it makes possible. *Id.* at 1061, 1063-1064. And it is hard to understand Congress's ban on producing unauthorized medals, or selling them, as unconnected to such a purpose. I make these points not to argue that Congress's ban on production of fake medals, or on sale of medals, is unconstitutional. On the contrary, it is to argue that the intuition that Congress may restrict the production of this deceptive evidence about someone's history of military heroism supports the Fourth Circuit's conclusion that this kind of deception (through wearing a medal that one hasn't received) generally lacks the First Amendment status accorded to verbal claims like those made by *Alvarez*.

cinematographers, and animators to help craft visually striking commercials, they can still be bound by laws that restrict misleading advertising. Even when a threat of violence is bound up with political expression, the First Amendment allows government to address the threat. In the next Part, I explain more fully how such a framework can apply to deepfakes.

VI. DEEPFAKES, DISCLOSURE, AND DOCTRINES FOR FIRST AMENDMENT MIDDLE GROUNDS

A. Borderline Cases and First Amendment "Middle Grounds"

If the First Amendment doctrine set forth in *Alvarez* isn't sufficient to protect backstops—and answer the need for equilibrium adjustment in the face of deepfakes—what First Amendment doctrine will answer this need? This Part argues that the key doctrinal answers to this threat—whether it comes from deepfakes or other, similar technological transformation of non-testimonial evidence—lie in a set of doctrines that exist for what we might call "First Amendment middle grounds," that is situations where the activity targeted by a regulation is neither fully on the expressive or non-expressive side of the line between speech and conduct, or where its character is ambiguous. Or where the activity being regulated lies on a different kind of borderline: Not that between entirely protected and unprotected speech, but rather between speech that is fully protected (like discussions about art or politics) and speech that receives less protection (such as commercial speech). In some cases, as Frederick Schauer points out, such borderline challenges arise simply because certain types of conduct lie at the boundary line between what the First Amendment speech clause covers, and activity it doesn't: "Borderline cases of First Amendment coverage," he points out, "will display attributes of

coverage and noncoverage, just as borderline cases of almost anything will display attributes lying on both sides of the border."[297]

Such borderline cases sometimes arise because of the kind of unsettling of equilibria discussed in the previous Part. In the Fourth Amendment context, for example, the action police officers take when they look at the exterior of a house from a vantage point on a public street has long been something that is entirely unconstrained by the Fourth Amendment's ban on unreasonable searches. It's not a "search" at all. Just as any passer-by is free to look at a person's house from a public street, so too is a police officer. But when such previously unproblematic observation involves using thermal imagers, then observation from the outside starts to take on some of the qualities of observation from inside the home. Police don't enter the home, but—armed with thermal imagers—they can now see aspects of in-home life they couldn't previously see without entering. So, this high-tech observation from the outside takes on an ambiguous status, with some features of a traditional non-search (it takes place from a public street) and some features of a traditional search (it allows observation of in-home details). In *Kyllo*, the Court viewed its task as resolving that ambiguity by clarifying which side of the Fourth Amendment line (between searches or non-searches) the law should place the activity on.

In the First Amendment context, courts sometimes take a similar approach to ambiguous cases. When artistic expression merges with non-artistic "functional activity," like the sale of merchandise on sidewalks, courts sometimes take it upon themselves to explore whether the artistic or functional elements "predominate"—and then, based upon this determination—classify

---

[297] Frederick Schauer, *Out of Range: On Patently Uncovered Speech*, 128 HARV. L. REV. F. 346 (2015).

the activity as expressive (where it's primarily art) or non-speech conduct (where it's primarily commercial or functional).[298]

But courts also take a different approach to such First Amendment middle grounds and that is the one this Part suggests will typically be more suitable for addressing hard cases raised by deepfakes. Rather than simply classifying an ambiguous case as "speech" or "conduct," they live with the reality that it has elements of both—and try to develop doctrine that can simultaneously protect the expressive components, while leaving government leeway to regulate the unprotected or less expressive dimensions of the conduct. Below, I will examine three doctrinal tools courts have developed for addressing First Amendment middle grounds, and how they might apply to deepfakes: (1) disclosure rules, (2) viewpoint-neutrality requirements, and (3) intermediate scrutiny.

B. Disclosure Requirements

As I noted in Part I, deepfakes threaten to confuse—and unsettle the line between authored video and recorded video records— because they blur the line we have in the past been able to draw between (1) videos that are clearly and wholly the product of artistry, like painting, drawing, and computer graphics, and the author's imagination and (2) video that is largely or wholly a record of light captured by a camera or sound captured by a microphone. Perhaps the clearest way to counter such confusion is to redraw the bright dividing line that deepfakes have obscured—by assuring the fake camera footage is clearly marked as fake. Government might

---

[298] *See, e.g.*, Mastrovincenzo v. New York, 435 F.3d 78, 82 (2d Cir. 2006) (finding that "the sale of plaintiff's clothing nonetheless has a predominantly expressive purpose and therefore merits First Amendment protection."); Kleinman v. San Marcos, 597 F.3d 323, 327 (5th Cir. 2010) (finding that the non-expressive qualities in a junked car used a sculpture "objectively dominate any expressive component of its exterior painting," but nonetheless applying intermediate scrutiny as an alternative ground to uphold a city's ordinance forbidding junked cars in open areas.).

do so requiring that those who create or knowingly disseminate deepfakes identify them as deepfakes. Providing room for such disclosure requirements, then, represents one possible method of equilibrium adjustment: In the face of deepfakes' threats to our capacity to distinguish records of fact and fiction, it should leave room in First Amendment law for government to help bolster this capacity.

In fact, some government officials have already proposed precisely this kind of legislation. A bill proposed in the House of Representatives in 2019, The DEEPFAKE Accountability Act, for example, would require anyone creating a deepfake video to identify it as fake, using an "embedded digital watermark," as well as textual descriptions.[299] Moreover, such disclosure requirements are not unlike the stance that Seana Shiffrin proposes First Amendment doctrine should take toward false verbal statements: Give First Amendment protection to those who wish to make such statements, but only if they were willing to use some "culturally well-understood mechanisms of disclosure" to make it clear that the false assertion is fiction.[300]

Some argue that the responsibility for such disclosures should lie not with anybody who creates or knowingly disseminates a deepfake, but rather with large social media platforms that often host such videos: YouTube, Facebook, Twitter and other sites with large customer bases. Facebook has already heeded such calls, announcing a ban on deepfakes in January, 2020.[301] Twitter

---

[299] *See* Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong.
[300] Shiffrin, *supra* note 166, at 133;*see also supra* text accompanying note 173.
[301] *See* Monica Bicket, *Enforcing Against Manipulated Media*, FACEBOOK (Jan. 6, 2020) (announcing in a Facebook statement that Facebook "will remove misleading manipulated media" that meets certain criteria, including use of manipulation that "is the product of artificial intelligence or machine learning that

announced in February 2020 that it would identify manipulated video[302] (and has even put that label on a video disseminated by the Trump campaign).[303] Because social media companies like Facebook and Twitter are not themselves constrained by First Amendment limits—as private companies, they can make their own rules about what user speech is permissible—some proposed laws have encouraged them to take the lead in fighting deepfakes, and impose limits that government cannot itself impose by law.[304]

Richard Hasen has proposed a disclosure requirement that would apply specifically to these kinds of social media companies. He argues that the First Amendment allows them to be subject to a "truth in labeling" requirement imposed by law.[305] Existing free speech doctrine, he argues, allows government to require that "websites and social media platforms with large numbers of users

---

merges, replaces or superimposes content onto a video, making it appear to be authentic."); Sam Shead, *Facebook to Ban "Deepfakes*," BBC (Jan. 7, 2020).

[302] *See* Yoel Roth & Ashita Achuthan, *Building Rules in Public: Our Approach to Synthetic & Manipulated Media*, TWITTER BLOG (Feb. 4, 2020) (announcing Twitter's plan to label manipulated media); *Synthetic and Manipulated Media Policy*, TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/manipulated-media.

[303] *See* Russell Brandom, *Twitter Labels Trump Video as 'Manipulated Media,'* VERGE (June 18, 2020), https://www.theverge.com/2020/6/18/21296518/twitter-trump-video-manipulated-media-deepfake-carpe-donktum.

[304] The bill proposed in the Senate by Senator Sasse, for example, would not only punish deepfakes with characteristics that, under *Alvarez*, would even allow punishment of lies (e.g., their use in "tortious conduct"). It would also shield the actions that social media companies or any other "a provider of an interactive computer service" might take "in good faith to restrict access to or availability of deep fakes" or "to enable or make available to information content providers or other persons the technical means to restrict access to deep fakes." *See* Malicious Deep Fake Prohibition Act of 2018, S. 3805, 115th Cong. Disclosure can also be built into deepfake-creation technology itself: The SmartPhone application, Avatarify—one of several new applications allowing consumers to create deepfakes—recently announced that it would be automatically adding a watermark to each fake video it generates. *See* Mike Butcher, *Deepfake Video App Avatarify, Which Processes On-Phone, Plans Digital Watermark for Videos*, TECHCRUNCH, Apr. 14, 2021, https://techcrunch.com/2021/04/14/deep-fake-video-app-avatarify-which-process-on-phone-plans-digital-watermark-for-videos/.

[305] Richard L. Hasen, *Deep Fakes, Bots, and Siloed Justices: American Election Law in A "Post-Truth" World*, 64 ST. LOUIS U.L.J. 535, 552 (2020).

label altered videos as 'altered.'"[306]

One potential First Amendment barrier to any such a disclosure requirement is that it compels speakers to do something the First Amendment normally protects them against: It lets government essentially commandeer people's speech and force such speakers to give voice to government's mandated expression rather than their own. The Supreme Court has warned that "measures compelling speech are at least as threatening" as measures restricting it.[307] As Caroline Corbin points out, speech compulsions threaten a number of key First Amendment interests: They chill speech. They distort the speaker's expression by altering its content. And they infringe speaker autonomy.[308]

But there are types of disclosure requirements for which the First Amendment allows more room. Most notably, in *Zauderer v. Office of Disciplinary Counsel*, the Court left government with significant room to compel advertisers and other commercial speakers to make certain disclosures about their products or services. Government could, said the Court, require commercial speakers to disclose "purely factual and uncontroversial information" about its products or service, where doing so is "reasonably related to the State's interest in preventing deception of consumers."[309] As Hasen points out, a deepfake disclosure requirement would almost certainly satisfy the first part of this requirement. A video's creator has either used deepfake technology or she hasn't—and whether she has or not is likely to be factual and (once established) uncontroversial. Where hidden use of deepfake

---

[306] *Id*. at 551.
[307] Janus v. Am. Fed'n of State, Cty., & Mun. Employees, Council 138 S. Ct. 2448, 2464 (2018).
[308] *See* Caroline Mala Corbin, *Compelled Disclosures*, 65 ALA. L. REV. 1277, 1280 (2014).
[309] Zauderer v. Office of Disciplinary Counsel of Ohio, 471 U.S. 626 (1985).

technology is likely to confuse its viewers, then making them aware of it is clearly related to "preventing consumer and voter deception" of those viewers.[310]

The major problem with relying on *Zauderer* to provide a foundation for deepfake disclosure requirements is that many of the deepfakes that commentators and policymakers worry about won't count as commercial speech. While the Court has been inconsistent in its pronouncements about how to define commercial speech, its best-known answers define it as speech "proposing a commercial transaction,"[311] or speech that can be characterized as "commercial" based on the "combination" of its character as an advertisement, its reference to specific products and services, and/or the economic motivation for the speech.[312] Deepfakes that sow panic by portraying missile attacks likely don't fit this description. Nor will deepfakes that seek to undermine political candidates or other public figures by making them voice words they never said or seek to celebrate them by displaying valor they never showed. Hasen suggests that even these and other non-commercial deepfakes might be subject to the kind of "truth in labelling" requirement *Zauderer* allows when the social media companies that make them available do so as part of these companies' commercial activity. Even when individuals create and disseminate political speech, the social media company's hosting of the speech gives it a commercial character.[313]

But such a stance seems to require a significant expansion of what can count as commercial speech under *Zauderer*. A social media company's status as a corporation doesn't likely transform all

---

[310] *See* Hasen, *supra* note 308, at 551-552.
[311] *See* Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n of New York, 447 U.S. 557, 562 (1980).
[312] *See* Bolger v. Youngs Drug Prod. Corp., 463 U.S. 60, 66–67 (1983).
[313] Hasen, *supra* note 308, at 551 (describing regulation aimed at large social media companies as mandating disclosure "in a commercial context.").

of the expression it hosts into commercial speech. To understand why, imagine that Congress, after its Stolen Valor Act was struck down in *Alvarez*, responded by enacting a new legal mandate to social media companies: Every time someone claims—in a social media post or video—that she has received a Congressional Medal of Honor or other military award, the social media company hosting this expression must check a government database of award winners and then, if it cannot locate the speaker's name in the database, add a note classifying the claim as "inconsistent with government records." It's unlikely that courts would consider this disclosure requirement a commercial speech requirement subject to *Zauderer*—because the expression modified by the compelled disclosure (a speaker's claim to have won a military award) is no more commercial than Xavier Alvarez's lie in *Alvarez* itself. It might be disseminated on Facebook, Twitter, or YouTube—but it is not a claim about these social media companies' commercial services (The Supreme Court has recently refused to apply *Zauderer* to speech content that is unrelated to "the services that" the speaker provides).[314] If a deepfake of a Congressional Medal of Honor ceremony is analogous to Alvarez's lie (for First Amendment purposes), then government could no more require social media companies to label it a deepfake under existing First Amendment law than they could force the same companies to label Alvarez's statement a lie.

However, as Part III has argued, the lie and the deepfake are *not* analogous.[315] Where a lie is believed, it is the basis of a testimonial belief that a listener forms because she accepts what

---

[314] Nat'l Inst. of Family & Life Advocates v. Becerra, 138 S. Ct. 2361, 2372 (2018).
[315] *See supra* Part III.

Alvarez says. Where the deepfake is believed, it is generally on the basis of a non-testimonial belief that a viewer forms because she thinks the deepfake is a camera-generated record that allows her to see for herself who received a Medal. As explained in the next section, it is that difference between lies and deepfakes that explains why many deepfake disclosure requirements (and perhaps some other restrictions on deepfakes) should likely be constitutional.

C. Viewpoint Neutrality and Intermediate Scrutiny

We should see the non-testimonial dimension of deepfake or other evidence as akin to *non-expressive* conduct that becomes intertwined with speech. This isn't new to First Amendment law. There are two contexts where courts have dealt with it before.

First, when people express ideas they don't *only* express ideas. The expression is almost always combined with some non-expressive conduct. This is certainly true when individuals express themselves not with words—but by taking action that has symbolic significance such as burning a Selective Service Registration card to protest a war.[316] But it also occurs even when people speak with words rather than symbolic action: Anti-war protestors might express their anti-war sentiments by chanting and displaying written messages on signs rather than burning a draft card. But if they march into an intersection as they chant and hold signs, then they aren't simply expressing themselves. They are engaging in physical conduct—namely, moving through an intersection and perhaps blocking traffic as they do so.[317]

In situations like this, the First Amendment raises an almost

---

[316] United States v. O'Brien, 391 U.S. 367, 369, 370, 376-77 (1968).

[317] *See, e.g.*, Akinnagbe v. New York, 128 F. Supp. 3d 539, 549 (E.D.N.Y. 2015) (finding that while strict scrutiny would apply to police measures targeting protestors' speech, only intermediate scrutiny applies to measures protecting public safety in public spaces and assuring the free flow of traffic).

insurmountable barrier against any government attempts to silence the ideas that speakers express. In the above examples, government may generally *not* ban or punish the expression of anti-war messages. It could only do so in the highly unlikely event that it can overcome the nearly-insuperable hurdle of strict scrutiny.[318] But although government is generally blocked by the First Amendment from restricting expression, it is left with more freedom to regulate the physical conduct that comes packaged with expression. It can regulate the threat to safety or government property generated by the burning of draft cards (or other objects) in expressive conduct.[319] It can likewise regulate the "time, place, and manner [of speech]" to prevent the disruption to traffic that occurs when people march in an intersection.[320] Even here, government does not have an entirely free hand. As noted below, it must still satisfy intermediate scrutiny.[321] But as a general matter, the First Amendment gives government leeway—so long as it doesn't target the content of the expression but rather enacts a content-neutral regulation that aims at the non-communicative components of the speech rather than the ideas in it.[322]

At times, moreover, the same kind of problem arises not because protected expressive content becomes intertwined with non-

---

[318] *See supra* text accompany notes 23-24, 152.
[319] *See O'Brien*, 391 U.S. at 377-378, 380-386.
[320] *Akinnagbe*, 128 F. Supp. 3d at 548-549.
[321] *See infra* Section VI.C.2.
[322] *See* Texas v. Johnson 491 U.S. 397, 407, 412 (stating that the Court has limited "the applicability of O'Brien's relatively lenient standard to those cases in which the governmental interest is unrelated to the suppression of free expression, and otherwise subjects such restriction to the "most exacting scrutiny"); Turner Broad. Sys., Inc. v. F.C.C., 520 U.S. 180, 185 (1997) (noting the requirements of O'Brien constitute a kind of "intermediate scrutiny" under the First Amendment and that this scrutiny requires that the government not restrict substantially more speech than necessary to achieve an important government interest); Wilson v. Lynch, 835 F.3d 1083, 1095 (9th Cir. 2016) ("intermediate scrutiny applies when a law is directed at the non-communicative portion of conduct that contains both communicative and non-communicative.").

expressive conduct (like burning draft cards or blocking traffic), but rather because it becomes intertwined with *unprotected expression*—such as threats of violence. The Supreme Court has made clear that such threats generally do not count as protected speech: The First Amendment, it has said, does not protect "true threats."[323] It does not protect speakers when they "communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals."[324] But unprotected threats sometimes come bound together with protected expression of particular political or other views. Imagine, for example, that an anti-war protestor calls a company that manufactures drones—and threatens its headquarters with "a deadly attack" of the kind the caller says is frequently carried by the companies' drones. Here, an instance of First Amendment-protected anti-war expression (opposition to drone warfare and the manufacture of drones used in it) is conveyed through an *unprotected* threat of violence.

As the Court made clear in *R.A.V. v. St. Paul*, the doctrine for such true threats is analogous to that which the Court applies to expressive conduct and time, place, and manner regulation.[325] In those cases, government generally has leeway under the First Amendment to protect the integrity of government records or to protect the flow of traffic. But it is not permitted to misuse that leeway by using it to target *only* the expression it disfavors. Government would violate the First Amendment if it punished destruction of records or traffic disruption *only* when it was carried out by anti-war protestors while leaving war supporters, for

---

[323] Virginia v. Black, 538 U.S. 343, 359 (2003).
[324] *Id.*
[325] *See* R.A.V. v. City of St. Paul, 505 U.S. 377, 386 (1992).

example, free to burn items or disrupt traffic.[326] Similarly, government is generally left free by the First Amendment to punish threats of violence, even when these threats are combined with political messages. But it violates the First Amendment when it misuses this power to target *only* threats with anti-war content or other messages it disfavors (leaving pro-war threat-makers unpunished). A true-threat ban can't be fully content-neutral, because it applies only to communications with threatening content.[327] But the Court has suggested it must be "neutral" with respect to all *other* content besides that which makes it a threat: It can't favor certain threats and disfavor others because of other expressive content in the threat (such as its antiwar message).[328]

1. Viewpoint Neutrality

This shared First Amendment doctrine on "content-neutral" laws and regulation of true threats or other "low value" speech can't be summarized as imposing one specific level of scrutiny. Courts apply intermediate scrutiny to regulations of symbolic conduct or the time, place and manner of speech, but apply only rational basis review or "minimal scrutiny" to regulations of true threats.[329] What

---

[326] *See* Pahls v. Thomas, 718 F.3d 1210, 1229 (10th Cir. 2013) (noting that "[a]t the core of the First Amendment is the idea that 'government has no power to restrict expression because of its message, its ideas, its subject matter, or its content,'" and finding that relocation protestors based on their viewpoint would violate this, but finding that, in this case, officers had qualified immunity).

[327] Black, 538 U.S. at 361-62.

[328] *See* Sorrell v. IMS Health, Inc., 131 S. Ct. 2653, 2680 (2011) (noting that even though "content-based restrictions on protected expression are sometimes permissible," for example, to categories of content that receive less protection, such as commercial speech, that is only true where there is a "neutral" justification that focuses on the harms created by the speech, such as the "risk of fraud," not a government interest in suppressing the views in that speech); *see also Black*, 538 U.S. at 362 (noting that "it would be constitutional to ban only a particular type of threat," but not simply because of the ideas in that threat but rather because aspects of it make the intimidation it causes more serious).

[329] *See* Christopher P. Guzelian, *False Speech: Quagmire?,* 51 SAN DIEGO L. REV. 19, 55 (2014) (noting that unprotected categories of speech can be punished if "the sanctioning law survives rational basis review").

unites all of these First Amendment cases is that courts actually impose a type of variable scrutiny: Government gets significant leeway—in the form of intermediate or minimal scrutiny—to regulate speech when it targets some "proscribable" element in it, according to the Court, but this leeway is not a blank check, and disappears when government misuses it to target ideas.[330] The consequence is that the Court's level of scrutiny varies according to what the government does. The clearest and most common-form this variable scrutiny takes focuses on a requirement of viewpoint neutrality (or a prohibition on viewpoint discrimination): Government misuses the leeway it is given to protect traffic, regulate the physical effects of symbolic conduct, or address the fear-generating capacity of threats, when—instead of focusing on those goals—it targets protestors or threats on the basis of their ideological *views* rather than their effects.

This discussion of viewpoint neutrality is instructive because it provides a starting point for how the First Amendment might apply to deepfakes and other ways that speakers might manipulate the digital evidence we use to make sense of the world. When government restricts deepfake videos, the First Amendment might confront it with the virtually insuperable hurdle of strict scrutiny *only* when the government is targeting the author's views or aesthetic choices, or other embodiments of an author's ideas. When government instead tries to stop individuals from disguising deepfakes as genuine camera footage, it should receive more leeway. In other words, the First Amendment might allow government to protect the integrity of the medium of audiovisual recording without letting government control the message or other

---

[330] *R.A.V.*, 505 U.S. at 385 ("a particular instance of speech can be proscribable on the basis of one feature . . . but not on the basis of another.").

informational content conveyed over that medium (or extend this protection of accurate recording to artistic contexts where audiences don't expect or need video to play that informational function).

To be sure, the Court's Justices have sometimes worried that such a viewpoint neutrality requirement isn't always up to the task of protecting the First Amendment interests at stake. Justice Alito, for example, expressed such a worry even as he argued, in his *Alvarez* dissent, that the First Amendment should leave government with the extensive power to restrict autobiographical lies (like that of Xavier Alvarez).[331] There were other types of lies, Alito stressed, that it would be far more dangerous to let the government restrict—specifically lies about history, philosophy, science or other matters of public concern. It might be theoretically possible, Alito acknowledged, for courts to apply *R.A.V. v. St. Paul* here—and leave the government with leeway to regulate such lies *only on the basis of their falsity*, and *not* on the basis of their political or cultural content.[332] But Alito didn't feel this framework could reliably keep government control over public debate in check.[333] The risk was too high that those with power to restrict lies would target false content not solely because it was false, but because they disagreed with it.[334]

Chesney and Citron raise a similar concern about deepfake restrictions: "Dislike of minority or unpopular viewpoints, combined with ambiguity surrounding a deep-fake creator's intent, might result in politicized enforcement."[335] But deepfakes might be different. Alito stressed that for government to restrict lies on subjects like

---

[331] United States v. Alvarez, 567 U.S. 709, 739, 751-752 (Alito, J., dissenting).
[332] *Id.* at 752.
[333] *Id.*
[334] *Id*. *See also* Sunstein *supra* note 67, at 44-46 (noting that "[e]ven when it is regulating falsehoods, the government needs to take the right kind of "slice" at the problem" and explaining how such restrictions might be bound by viewpoint- or other content-neutrality requirements).
[335] *See* Chesney & Citron, *supra* note 6, at 1789.

history and science, it would have to become an "arbiter" of what is true and false in such fields—and that it is generally ill-equipped to play that role well or with neutrality.[336] Particular government officials may each confidently embrace different views, for example, of how much various sectors of the economy improved under President Obama and President Trump—and may disagree markedly over which accounts of these economic developments count as false or misleading. By contrast, as noted earlier, there are emerging technologies that might allow even observers of very different political persuasions to agree that a certain video sequence was generated by AI rather than captured by a camera.[337] Just as observers with very different political or cultural views might often be able to come to an agreement about whether a Picasso painting is a forgery, so they could presumably come to an agreement about whether a video is a fake. They would not have to make the same controversial judgment about history, philosophy or science—even when a deepfake is about these topics—than they would have to make as an arbiter of false statements about the same topics. In fact, for deepfake disclosure laws to work effectively, there has to be *some* way for regulators and private entities to identify what a deepfake is—and to do so independently of the deepfake's political or cultural content.

Controversies over other types of edited videos may raise doubts about the possibility of such neutral judgments. Someone who edits a political video, for example, might be accused of shortening a political opponent's speech—and, in the process, omitting certain parts of it that mislead her audience. But the video creator might claim to have done so for the very innocent purpose of

---

[336] *Alvarez*, 567 U.S. at 751-52 (Alito, J., dissenting).
[337] *See supra* text accompanying notes 285-287.

making the video short enough to be viewed by busy social media users (and that social media users are likely aware that a two-minute video clip of a longer speech is the product of conscious editing choices). In deciding whether such a video edit is misleading, a government regulator or social media company—one might worry—will be influenced by its own views about the candidate in the video. A decision-maker who supports President Biden may be more likely than one who opposes him to classify as misleading any edits that seem to portray him in a harsher light than the longer video.

Deepfake creators might likewise claim that their alterations are meant not to deceive their audiences but rather to draw their attention to certain truths conveyed by the video. Imagine, for example, that a candidate for office (or a group of supporters) wishes to highlight and criticize a Tweet or other social media post made by an opposing candidate—and does so with a deepfake that shows the opponent passionately *saying* words in the tweet—words that the opponent wrote but never actually said. One might argue, drawing on examples of this kind, that First Amendment law should not leave government with leeway to make judgments about when to restrict such a dramatic visual translation of a person's written words (whether with deepfake creation tools or with other technology).

Still, as Hasen writes, it is simpler to impose certain kinds of disclosure rules on edited videos, including deepfakes,than on false statements. Whatever purpose a person may have had in cutting out segments of a video, or altering it with a deepfake, a "truth in labelling" or other disclosure requirement can still require that they disclose simply that they *have* altered the video (regardless of what

dispute may exist over their reasons for doing so, or the extent to which their alteration distorts the meaning of what it depicts).[338]

Prosecutors and others charged with enforcing the law could still, of course, try to apply neutral criteria for identifying (and mandating disclosures of) deepfakes only against those with ideological views they disfavor. But this could potentially occur in the enforcement of *any* regulation that is supposed to be a content-neutral regulation (such as a law against burning draft cards or blocking intersections) or a regulation of true threats. The Court has not taken back the leeway it gives the government in making such regulations (through intermediate or a viewpoint neutrality requirement) simply on the basis that the potential for abuse exists. It does so only when there is evidence such selective ideological enforcement has occurred, or where judges applying the holding of *R.A.V.* or other First Amendment doctrines conclude (as the *Alvarez* dissenters did with respect to regulation of lies on matters of public concern) that doing so cannot effectively reveal and deter such ideological enforcement.[339]

2. Intermediate Scrutiny

Even when government's regulation of deepfake is *not* a cover for ideologically driven suppression of speech, even when government genuinely wants to protect viewers' reliance on video records rather than push them towards its own preferred ideology,

---

[338] Hasen, *supra* note 308, at 553-54 ("[T]he deep fakes problem is surprisingly *easier* to solve (once the technology is in place) than the problem of low-tech false information. When it comes to whether video or audio has been manipulated, there is an objective truth of the matter: a scientific comparison of original content with content posted online.").

[339] *See, e.g.*, Occupy Columbia v. Haley, 866 F. Supp. 2d 545, 560 (D.S.C. 2011) (finding that protestors were likely to succeed in challenging a policy preventing them from being on the South Carolina State House grounds after 6:00 pm where "there [was] no evidence that the 6:00 p.m. policy has been applied consistently to all organizations and groups seeking to use the State House grounds"). *See also supra* text accompanying notes 334-37.

there's still a First Amendment problem. Sometimes the informational function we rely upon video to serve is inextricably bound up with its artistic or other expressive uses of the video. It is for this dilemma that First Amendment doctrine has relied on what courts call "intermediate scrutiny." Consider again what the Second Circuit did in *Universal City Studios v. Corley* when it explored whether government could restrict the dissemination of decryption code.[340] On the one hand, distributing this code is an act of expression: Sharing code is one of the ways that programmers communicate with each other about programming.[341] On the other hand, the *same* code also has a *non-expressive* use: When run by a computer, it functions as a kind of digital hacksaw that cuts through a movie studio's electronic safeguards.[342]

The way it addressed this merging of expression and non-expressive conduct in *Corley* was the way it had done so in *United States v. O'Brien* and other cases on content-neutral regulation of speech: It applied intermediate scrutiny.[343] Because of the First Amendment interests at stake in these cases, this standard requires government to show—as it does when courts apply strict scrutiny—(1) that its interest in restricting expression is weighty enough to justify the damage it does to our speech and (2) that its restriction on expression is "narrowly-tailored" to the problem and thus, isn't imposing more First Amendment damage than it needs to.[344] But because government needs *some* leeway to protect the public from burning of draft cards, obstruction of traffic, or dissemination of decryption code, courts make the hurdle easier for government to

---

[340] Universal City Studios v. Corley, 273 F.3d 429 (2d Cir. 2001).
[341] *Id.* at 446-49.
[342] *Id.* at 453-54.
[343] *Id.* at 442; *see also* United States v. O'Brien, 391 U.S. 367, 377 (1968).
[344] *Id.*

meet than that presented by strict scrutiny. Government needs only an interest that is "important," "significant," or "substantial"—not the kind of extraordinary "compelling" interest necessary in strict scrutiny.[345] And while its measure needs to be narrowly-tailored, it need not (as in strict scrutiny) be the least speech-restrictive measure imaginable: As long as it doesn't restrict substantially more speech than necessary, it will meet intermediate scrutiny.[346]

Applying intermediate scrutiny to deepfake laws may seem, at first, to be a recipe for constitutional uncertainty. When the Court applied intermediate scrutiny under the Equal Protection clause, Justice Rehnquist worried (in a dissent) that courts had no criteria for determining "what objectives are important."[347] Nor, he said, would they be able to assess whether the fit between the means and ends was close enough where the rule says only that the relationship must be "substantial."[348]

Some such uncertainty is likely unavoidable when one moves from strict scrutiny (or fully unprotected speech) to a First Amendment middle ground. However, courts can—and sometimes do—translate intermediate scrutiny into a less amorphous and more structured inquiry.[349] First, the significant interest prong often

---

[345] *See* Turner Broad. Sys., Inc. v. F.C.C., 520 U.S. 180, 189 (1997).

[346] *Id*. Relying on this framework, the Second Circuit in *Corley* found that the DMCA's limits on dissemination of decryption software were constitutional. The government had a significant interest in protecting copyright owners from widespread distribution of tools that could override copyright protection measures on DVDs, and there was little it could do to assure such protection without forbidding the sharing of such tools on the Internet. Universal City Studios v. Corley, 273 F.3d 429, 450-58 (2d Cir. 2001).
Thus, the Second Circuit found, applying the DMCA to dissemination of decryption software met the requirements of intermediate scrutiny. *id*.

[347] Craig v. Boren, 429 U.S. 190, 221 (1976) (Rehnquist, J., dissenting).

[348] *Id*.

[349] Alexander Tsesis says more about how intermediate scrutiny—including Justice Breyer's use of it in *Alvarez*—can be a form of "systematic balancing" that proceeds in a more predictable way than unconstrained free-form balancing. *See* FREE SPEECH IN THE BALANCE 40 (2020).

presents a relatively low hurdle for government—perhaps because courts are ill-equipped to challenge government officials' claims to serve safety, health, or other interests they invoke when imposing content-neutral restrictions of speech. As a general matter, government will have a powerful interest in combatting the effects that Chesney and Citron describe respectively as "truth decay" (where a deepfake is mistaken for genuine video) and "trust decay" (where deepfakes lead people to doubt the accuracy of videos in general).[350]

Since the significant interest prong is often easily met, the permissibility of a deepfake law is instead likely to turn on the narrow tailoring prong—which will effectively push the government towards less speech restrictive alternatives. In deepfake regulation, for example, courts might look least favorably on (1) prohibitions on deepfake creation or dissemination, and strongly prefer (2) disclosure mandates that let individuals create or distribute deepfakes, as long as they reveal their nature. Even government disclosure mandates might fail intermediate scrutiny, however, when (3) there are numerous self-help measures available that individuals can use to detect deepfakes on their own (or with the help of private actors, like social media companies), and thus don't need disclosure mandates to reveal for them what they can fairly easily discover themselves.[351] To the extent government help is necessary then, it might permissibly come in the form of bolstering

---

[350] Chesney & Citron, *supra* note 6, at 1786. There may be exceptions: Certain types of deepfakes—or deepfakes in certain contexts—might be less likely fool viewers, and less likely to undercut the trust they place in other types of videos. *See supra* text accompanying notes 67-68, raising questions about how the form deepfakes take, or frequency with which we encounter them, might affect how likely we are to be deceived by them. The strength of the government's interest in restricting a particular deepfake or type of deepfake may also depend in part on the harm it is likely to cause. See Sunstein, *supra* note 67, at 12-14, 119.

[351] *See supra* text accompanying notes 285-287.

video authentication methods that private actors have created—and protecting private companies' marks of authentication (rather than video itself) against forgery.[352] (Of course, there might be specific variations on each of these alternatives and I will examine some of these below).

Finally, there is one more consideration that is likely to play an important role as courts analyze the effectiveness of deepfake prohibitions, disclosure rules, or self-help mechanisms—and that is the cost that an instance of each of these would have to speakers' autonomy and listeners' autonomy interest (or to democratic discourse) in particular situations. As noted in Part III, where falsity takes the form of non-testimonial evidence, a restriction of it does not generally undermine speaker autonomy in the same way as a restriction on the speaker's control of her own words.[353] But a ban or disclosure requirement for videos could do so, for example, when it applies to artistic expression on the ground that it includes or carries a high risk of deception. In doing so, it could simultaneously undercut listeners' or viewers' right to receive such artistic expression, and democratic discourse that benefits from it. Where this is true, courts applying intermediate scrutiny should hesitate to impose a ban or compelled speech requirement.

How would this framework apply to some of the situations I have discussed earlier in which artistic and political dimensions of deepfakes are intertwined with deception? Consider some of the cases where potentially deceptive deepfakes might also count as art

---

[352] *See* Blitz, *supra* note 31, at 115-116. (stating that "what the First Amendment may have to allow room for" to protect detection of forged evidence "is similar to the role played by the Digital Millennium Copyright Act (DMCA) in the realm of copyright law" which is to "let private parties create [protective] technology themselves, and then legally shield the technology from those who would circumvent it.").

[353] *See supra* Part III.B.

or creative expression. As noted above, in the near future, individuals might use deepfake filters on Zoom other video chats to appear in the guise of someone else.[354] Should the First Amendment, then, let government require disclosure of their use, or otherwise restrict it? In certain circumstances, it already does. Most states have "false personation" or "false impersonation" laws that prohibit individuals for impersonating others in any way that obtains gain for themselves or causes harm to someone else.[355] But such false impersonation laws tend to punish not falsity alone—but only the kind of falsity that the *Alvarez* plurality said was left open to punishment by the First Amendment—namely, falsity that is part of or accompanied by "legally cognizable harm."[356] Under the intermediate scrutiny regime proposed here, by contrast, government may well be permitted to impose a disclosure requirement to prevent successful use of a hyper-realistic false appearance or voice even in certain circumstances when it can't point to any specific material harm caused by, or advantaged derived from, the ruse.[357]

Deepfake disclosure requirements for videos disseminated on social media may present a more challenging case—since it is

---

[354] *See supra* text accompanying notes 112-113.

[355] California's penal code, for example, makes it a crime to "falsely personat[e] . . . and in that assumed character . . Does any other act whereby, if done by the person falsely personated, he might, in any event, become liable to any suit or prosecution, or to pay any sum of money, or to incur any charge, forfeiture, or penalty, or whereby any benefit might accrue to the party personating, or to any other person." CAL. PENAL CODE § 529 (West 2011). It also bars other uses of false identity unlikely to arise in video chats.

[356] United States v. Alvarez, 567 U.S. 709, 719 (2012).

[357] Even though Justice Breyer does not squarely address deepfakes or similar technology in *Alvarez*, in his own argument for applying intermediate scrutiny even to verbal lies, he takes an approach quite similar to the approach taken here. He notes that "[s]tatutes forbidding impersonation of a public official typically focus on *acts* of impersonation, not mere speech," *Alvarez*, 567 U.S. at 735 (Breyer, J., concurring), suggesting that deception that is carried out by action that creates a false appearance may not be expression at all.

becoming quite common for users to encounter doctored video of all kinds on social media (just as they have long encountered doctored photographs) and one can plausibly argue that anyone who treats a YouTube, Twitter, or Facebook video as non-testimonial evidence would be unjustified in doing so. But to the extent individuals *do* treat such videos as sources for information about election-related information, laws that require disclosure of deepfake technology (or even prohibitions on it) may survive intermediate scrutiny. If they do, it is quite likely that such disclosure requirements could apply not only to the social media companies that carry deepfake videos,[358] but to those that create and post them as well.

Certain kinds of performance and avant-garde art raise more difficult questions in cases, like those above where the artist intentionally deceives an audience as part of the performance. Courts might hesitate to exclude deepfake deception—or indeed, intentional deception of any kind—from the ambit First Amendment when it is part of an artistic practice. In some art or gameplay, as noted earlier, deception of an audience isn't simply a worrisome risk—it is one of the *goal*s of the artist or game designer. That is, a kind of artwork or virtual game can't succeed unless there is at least temporary deception. How should the First Amendment apply to the deception involved in these artistic projects? Should they also be subject to intermediate scrutiny? In the first place, free speech law should not, except in unusual cases, let government protect individuals from *inviting* deception as part of an artistic performance they know includes it. In certain physical and virtual environments—like theaters, museums, immersive art exhibits, and virtual reality games—viewers know that what looks real may not be. The couple described in Part I, for example, would likely *not*

[358] *See supra* text accompanying note 316.

treat as real a fake corpse that they encounter—not in their hotel room—but rather in a wax museum, an amusement park haunted house, or art exhibit on movie special effects. The same would likely be true in a variant of the imitation game that challenges viewers to distinguish between real video and deepfakes.

The First Amendment analysis is more difficult when artistic use of deception is uninvited and entirely unexpected. In such cases, viewpoint neutrality requirements and intermediate scrutiny should apply. That is, the First Amendment should generally leave government free to protect individuals from surreptitious manipulation of their perceptual sources of knowledge, or other non-testimonial evidence. It should, on the other hand, block government from suppressing artistic expression on the basis of the ideas in it. Where it can't protect users from deception without simultaneously depriving them of artistic expression, it should let government act to prevent the deception only where there are significant government interests at stake, and where its restriction doesn't cause grave and unnecessary harm to the First Amendment autonomy interests of the artist. A very brief deception that is quickly corrected before a listener or viewer can rely on it in any significant way, for example, may not be the kind of deception that government has a strong interest in regulating. By contrast, where viewers do predictably rely on deepfake deceptions, weaving such a deception into an artistic project of some kind shouldn't disable the government from protecting viewers' reliance interests—any more than integrating distribution of decryption code into an immersive art exhibit or performance would disable the government from regulating it in the way the Second Circuit found permissible in *Corley*.

The intermediate scrutiny approach set out here may seem, on the surface, to bring us a First Amendment framework that could

be reached by a less circuitous route—by treating deepfakes as analogous to lies and then subjecting them to the intermediate scrutiny that Justice Breyer uses in his *Alvarez* concurrence.[359]

However, the framework proposed here is different. In two ways, it would likely give government more room to regulate deceptive deepfakes. First, where government acts *solely* against the use of deepfakes to fabricate non-testimonial evidence and does so without impinging on the use of deepfakes for art or to give visual form to testimony, then the First Amendment may not apply at all, except in so far as it imposes a viewpoint neutrality requirement on such a government measure. Just as it doesn't typically run afoul of the First Amendment when it bans the destruction or forgery of a government ID card, or the generation of fake GPS data, so it won't if it restricts attempts to pass off a deepfake as genuine security camera footage or an unaltered camera feed. This is true as long as the government is really targeting non-testimonial falsehoods evenhandedly. If it is targeting only that non-testimonial evidence which supports beliefs it disfavors, then this does implicate the First Amendment and triggers strict scrutiny.

Second, Justice Breyer would not apply intermediate scrutiny to all laws on lying: When government restricts "false statements about philosophy, religion, history, the social sciences, the arts, and the like" then Breyer would apply strict scrutiny.[360] But the same is not true for deepfakes under the approaches proposed here: Even if a deepfake depicts events or people that are matters of public interest, that doesn't mean government suddenly faces strict scrutiny. Just as the government has authority to regulate the forgery of government or business records that deal

---

[359] *See supra* text accompanying notes 153-155.
[360] *Alvarez*, 567 U.S. at 731 (Breyer, J., concurring).

with matters of public interest, so it should have some leeway to apply disclosure requirements to, or otherwise regulate, the fabrication of video or audio records on matters of public interest. Intermediate scrutiny insulates such fabrication—at least, to some extent—from government restriction, but this is not because the deepfake deals with certain topics, but rather for a different reason. While this framework is similar to that of Justice Breyer's *Alvarez* concurrence in applying intermediate scrutiny, its underlying logic resembles that of Justice Alito's *Alvarez* dissent. The reason the Court's prior First Amendment protects some intentional false statements, said Alito, is not because they have First Amendment value, but to provide "breathing space" for similar truthful speech that does.[361] Such an approach may ultimately be a poor fit for lies, because it allows the government to intrude too deeply into a speaker's control over their own words.[362] But it is, the article has argued, a good fit for deepfakes and other non-testimonial falsehoods: The reason even intentionally deceptive deepfakes should be protected by intermediate scrutiny is not because the hijacking of our perceptions or other non-testimonial sources of information itself has First Amendment value, but rather because individuals may need "breathing space" for the artistic or other expressive uses of deepfake that do.

CONCLUSION

The Supreme Court ruled in *Alvarez* that individuals are presumptively protected by the First Amendment when they deceive

---

[361] *Alvarez*, 567 U.S. at 750 (Alito, J., dissenting) (quoting Gertz v. Welch, 418 U.S. 323, 342). *See also* Shiffrin, *supra* note 166, at 154 (arguing that lies have "no free speech value" but advocating some First Amendment protection—with a "modified version of intermediate scrutiny"—to protect against possible ways government might abuse a power to restrict falsehoods).

[362] *See supra* text accompanying notes 199-201.

other people by making false statements.[363] But when and to what extent are they likewise protected by the First Amendment when they carry out the same kind of deception in other ways? This is the question that this article has addressed, focusing on the example of deepfakes.

After all, individuals can deceive each other not only with words– but also with non-verbal conduct that is calculated to mislead others about their plans. Consider the example offered by the philosopher Immanuel Kant in his *Lectures on Ethics*: If someone wants to create the false impression that she is about to go on a long journey, she might do so not only with a lie ("I'm leaving for a long trip") but also by packing a large suitcase in view of the person he's trying to deceive.[364]

As noted in the article, a modern deceiver might create such a false impression not simply by engaging in misleading deeds, but also by creating misleading evidence: She might create a fake airplane ticket and leave it lying on a desk—or a fake e-mail from an airline company confirming the purchase of such a ticket. In a more elaborate digital ruse, she might send the target of his deception a link allowing that person to track a ride to the airport that doesn't actually occur. Or a fake GPS reading that shows her inside of the airport, at a place she wouldn't be permitted without a purchased airline ticket. Or, perhaps, a time-stamped deepfake video showing her waiting to board the plane.

Is all such deception protected by the First Amendment? If not, does it at least protect the deepfake video on the ground that video is now a recognized medium of expression? My argument in

---

[363] *Alvarez*, 567 U.S. at 718, 727-730.

[364] IMMANUEL KANT, *Lectures on Ethics*, *in* Peter Heath and J.B. Schneewind, LECTURES ON ETHICS: THE CAMBRIDGE EDITION WORKS OF IMMANUEL KANT 202 (1997).

this article is that it does not. Video is, of course, in many circumstances, a medium of artistic expression, and deepfake technology can play a role in such artistic expression. Not only it is a tool for professional filmmakers to tell fictional stories. It is a means by which authors can visually illustrate or embellish their arguments. But my argument here has been that video doesn't always serve as a vessel for an author's ideas. It has long served as a record of what a camera captured rather than as a picture and storyteller or argument-maker wishes to show us. First Amendment law should leave government with room to preserve this *non-testimonial* function of video.

That doesn't mean that those who disseminate deepfakes that emulate such videos should be entirely without First Amendment protection: The same deepfake that deceives an audience in one context, after all, can educate and entertain it in another. The same deepfake that is viewed as evidence external to a speaker might, at another time, be seen as a vessel for a speaker's expression. Deepfakes are thus in a First Amendment middle ground—one where courts should seek to protect them when and to the extent they are expressive, but let government expose them as deepfakes when they pose as genuine camera footage.