

Pyrite or Panic?

Deepfakes, Knowledge, and the Institutional Backstop

Gavin Wilde
February 2025

Abstract: Breakthroughs in machine learning over the past decade have made synthetically created audio-visual content – so-called “deepfakes” – both easier to produce, and less discernible from genuine recordings. These advances have prompted concerns about a coming epistemic crisis: If media can no longer be relied upon as a “backstop” against human perceptual frailty and gullibility, deepfakes are said to imperil collective sensemaking in a range of institutions. This paper argues, however, that such arguments tend to oversimplify the evidentiary role that audio-visual media play, and to conflate identity politics and entertainment with the formulation of knowledge. To the extent that deepfakes do pose an epistemological threat, this paper draws on historical parallels and contemporary scholarship to argue for improved institutional – rather than digital – “backstops.” Deepfake fears stem from a recent overinvestment in the latter, while truth and expertise are primarily the byproducts of *social* processes – which have gone relatively underserved.

About the Author: Gavin Wilde is a non-resident fellow at the Carnegie Endowment for International Peace, a senior researcher at the Latvian Institute of International Affairs, and an adjunct lecturer at Johns Hopkins University and Riga Stradiņš University. He previously served in leadership and analytic roles in the U.S. intelligence community for fifteen years, including as a co-author of assessments of foreign interference in elections from 2016-2020. He holds a BA in Russian Studies from the University of Utah, and an MS in National Security Strategy from the National War College in Washington, DC.

Introduction

Thanks to advances in machine learning – a subdiscipline of the technology now marketed as “artificial intelligence” – it is now easier than ever to create and share audio and visual content that is virtually indistinguishable from genuine recordings.¹ The rise of these synthetic materials – commonly known as “deepfakes” – have been cause for concern among policymakers and researchers for years now.² Among the most frequently cited fears: an explosion of targeted propaganda and a slow collapse in trust toward media of any kind. A gradual, widespread loss of confidence in recordings and their correspondence to reality conceivably puts the shared wisdom of societies at catastrophic risk.

However, such alarm tends to overlook the complexity of two key relationships: that between any audio-visual medium and the reality it presumably represents; and, by extension, that between media content and the processes involved in human knowledge formulation. The historical backdrop of media innovations and the academic literature exploring human-media interaction both lend necessary nuance to research and policy discussions about deepfakes. In the so-called “post-truth” era, these conversations have tended to foreground the epistemological implications of new media technologies at the expense of the cultural ones – which are often more salient.

Historical Parallels

The advent of any new technology – particularly audio-visual media – has typically been accompanied by both outsized hype and panic about their prospects.³ Such panics tend to pit the new medium against a monolithic mass audience, and assume the opening of a new direct pathway to the worst conceivable outcomes.⁴ If history is any guide, however, the novelty eventually dissipates, and societies gradually adapt.⁵

¹ For a short explainer on the technology underpinning deepfakes, see Michael Copeland, “What’s the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning?”, *NVIDIA Blog*, July 29, 2016, <https://blogs.nvidia.com/blog/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai>.

² Manipulated media that use less sophisticated methods are often referred to as “cheapfakes.” See Sijia Qian, Cuihua Shen, and Jingwen Zhang, “Fighting Cheapfakes: Using a Digital Media Literacy Intervention to Motivate Reverse Search of out-of-Context Visual Misinformation,” *Journal of Computer-Mediated Communication* 28, no. 1 (January 1, 2023), <https://doi.org/10.1093/jcmc/zmac024>.

³ Tom Wheeler, *From Gutenberg to Google: The History of Our Future* (Brookings Institution Press, 2019), 42-47.

⁴ Amy Orben, “The Sisyphean Cycle of Technology Panics,” *Perspectives on Psychological Science* 15, no. 5 (September 1, 2020): 1143–57, <https://doi.org/10.1177/1745691620919372>.

⁵ Todd C. Helmus and Bilva Chandra, “Generative Artificial Intelligence Threats to Information Integrity and Potential Policy Responses” (RAND Corporation, April 16, 2024), 7, <https://www.rand.org/pubs/perspectives/PEA3089-1.html>; Jessica Poon, “GANs Gone Wild: Public Perceptions of Deepfake Technologies on YouTube” (Czechia, Univerzita Karlova, 2021), <https://dspace.cuni.cz/handle/20.500.11956/150489>.

From this historical perspective, audio recordings, photographs, and motion pictures are still relatively young, only dating from two centuries ago or less.⁶ The birth of each new medium over that period was usually attended by predictions about the demise of a newly outmoded one: Painters feared photographs spelled the end of their artistry.⁷ Wider circulation of major newspapers meant the book was dead.⁸ The phonograph was said to mark the end of type, and the telephone would wipe out regional dialects.⁹ Motion pictures would replace textbooks.¹⁰ Live radio broadcasts would make static recordings obsolete.¹¹ History is replete with such “fickle fantasies and dashed expectations.”¹²

However technically sophisticated a new invention may be, the cultural context into which it is born also plays a major role in how it is ultimately perceived and used. For instance, light projection in the mid-1600s was intended to demystify the science of reflection and optics. Instead, audiences were captivated by the sense of illusion. Scientific demonstrations and magic shows would go on to become both wildly popular and largely indistinguishable from one another.¹³ Similar scenarios recurred as “scientific instrument[s] devised to demonstrate one concept of vision became a popular device whose interpretation and habitual use asserted another,” from the stereoscope to the camera.¹⁴ Thomas Edison likewise envisioned sound recording primarily for business dictation, attracting major investors in the late 1870s. Within two decades, the phonograph had instead ushered in a mass market for music, revolutionizing radio broadcasting in the process.¹⁵ It hardly occurred to early marketers of telephones in the early 1900s to tout the device as a tool for idle socializing – which became a major selling point for residential customers in subsequent decades.¹⁶

⁶ Regina Rini, “Deepfakes and the Epistemic Backstop,” *Philosopher’s Imprint* 20, no. 24 (August 2020), 4, <https://philpapers.org/archive/RINDAT.pdf>.

⁷ Amanda Prusila, “Truth, Lies and ‘Deepfakes’: The Epistemology of Photographic Depictions” (Ottawa, Ontario, Carleton University, 2022), 41, <https://doi.org/10.22215/ctd/2022-14911>.

⁸ Megan Garber, “How Americans Put Reality on Life Support,” *The Atlantic*, December 1, 2016, <https://www.theatlantic.com/entertainment/archive/2016/12/the-image-in-the-age-of-pseudo-reality/509135>.

⁹ Claude S. Fischer, *America Calling: A Social History of the Telephone to 1940* (University of California Press, 1992), 24-27.

¹⁰ Jeffrey Westling, “Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction to Deep Fakes,” SSRN Scholarly Paper (Rochester, NY, July 24, 2019), <https://doi.org/10.2139/ssrn.3426174>.

¹¹ Jonathan Sterne, *The Audible Past, Cultural Origins of Sound Reproduction* (Duke University Press, 2003), 179–214.

¹² Nicholas Carr, *The Shallows: What The Internet Is Doing To Our Brains* (W. W. Norton & Company, 2010), 109-10; John Keane, “Communicative Abundance,” in *Democracy and Media Decadence*, 1st ed. (Cambridge University Press, 2013), 9, <https://doi.org/10.1017/CBO97811073300767>.

¹³ Lisa Bode, “Deepfaking Keanu: YouTube Deepfakes, Platform Visual Effects, and the Complexity of Reception,” *Convergence* 27, no. 4 (August 1, 2021): 919–34, <https://doi.org/10.1177/13548565211030454>; Brian Merchant, “AI Really Is Smoke and Mirrors,” *Blood in the Machine*, May 21, 2024, <https://www.bloodinthemachine.com/p/ai-really-is-smoke-and-mirrors>.

¹⁴ Laura Burd Schiavo, “From Phantom Image to Perfect Vision: Physiological Optics, Commercial Photography, and the Popularization of the Stereoscope,” in *New Media 1740-1915*, eds. Lisa Gitelman and Geoffrey B. Pingree (Cambridge: MIT Press, 2003), 131.

¹⁵ Lisa Gitelman, “How Users Define New Media: A History of the Amusement Phonograph,” in *Rethinking Media Change*, ed. David Thorburn and Henry Jenkins (Cambridge: MIT Press, 2003).

¹⁶ Fischer, *America Calling: A Social History of the Telephone to 1940*, 68, 75-84.

From these developments, a common theme emerges – one which now applies equally to deepfakes. American historian and philosopher Lewis Mumford put it best: “The camera, the phonograph, and the moving picture... all these forms of permanent record were first employed chiefly for amusement;” the artistic interest behind them often outpacing the practical.¹⁷ For instance, despite prevailing modern notions about photography’s documentary qualities, the medium served throughout the 19th century as an artistic extension of the painter’s canvas. It was not until well into the 1900s that the public came to expect these renderings to function as evidence.¹⁸

To assess the long-term impacts of deepfakes, the distinction and interplay between a medium’s technical and social applications must likewise be considered.¹⁹

The presumption that any medium necessarily captures an unvarnished, unbiased snapshot of reality has always been somewhat misguided; subjectivity on the part of the creator and interpretation on the part of the observer are both unavoidable.²⁰ The old adage that “seeing is believing” actually only captures half the story; the original quote from 1732 closes with “...but feeling is the truth.”²¹ The line between *expression* and *evidence* in any audio-visual medium – in other words, whether it *depicts* or *detects* its subject – has always been blurry.

In either case, fakery has been an indelible part of audio-visual media since their inception – whether to maliciously distort or faithfully depict events.²² For instance, early cameras captured landscapes poorly, leading photographers to either add clouds by hand or superimpose them from the negatives of other shots. Throughout the nineteenth century, photo negatives were widely treated as “first drafts, to be corrected, reordered, or overwritten as needed.” By the late 1800s, major newspapers routinely doctored the images accompanying stories.²³ In similar fashion, early sound recordings often featured staged recreations of famous historical speeches.²⁴

¹⁷ Lewis Mumford, *Technics And Civilization* (New York, NY: Harcourt, Brace, 1934), 242.

¹⁸ Graham Clarke, *The Photograph*, (Oxford: Oxford Univ. Press, 1997), 41, 47.

¹⁹ Britt Paris and Joan Donovan, “Deepfakes and Cheap Fakes” (Data & Society Research Institute, September 18, 2019), https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf; David Campany, *Exposures: Photography and Cinema* (London, UK: Reaktion Books, 2008), 11; Sterne, *The Audible Past, Cultural Origins of Sound Reproduction*, 215–86; Prusila “Truth, Lies and ‘Deepfakes’: The Epistemology of Photographic Depictions,” 32-40.

²⁰ Rob Horning, “The Science of the Concrete,” *Internal Exile*, December 29, 2023, <https://robhorning.substack.com/p/the-science-of-the-concrete>.

²¹ Matthew Groh et al., “Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video” (arXiv, January 15, 2024), <http://arxiv.org/abs/2202.12883>.

²² Dylan Love, “The World Before Photoshop: Fake And Doctored Pictures Throughout History,” *Business Insider*, August 4, 2011, <https://www.businessinsider.com/fake-photos-history-2011-8>; Warren Neidich, “American History Reinvented,” 1989, <https://americanhistoryreinvented.com>; Marvin, *When Old Technologies Were New*, 204-206.

²³ Daniel Immerwahr, “What the Doomsayers Get Wrong About Deepfakes,” *The New Yorker*, November 13, 2023, <https://www.newyorker.com/magazine/2023/11/20/a-history-of-fake-things-on-the-internet-walter-j-scheirer-book-review>; Walter Scheirer, *A History of Fake Things on the Internet* (Stanford University Press, 2023), 60-62; Parker Settecase, “Parker’s Pensées,” Ep. 249, “Deepfakes and the Epistemic Apocalypse,” <https://podcasts.apple.com/us/podcast/parkers-pens%C3%A9es/id1527286862>.

²⁴ Sterne, *The Audible Past, Cultural Origins of Sound Reproduction*, 242.

Even so, societies gradually came to rely on such recordings as indisputable proof that a given thing took place.²⁵ This reliance was generally an improvement over a hand-drawn sketch or an oral recounting of the same scene or event. It ultimately came to serve as what philosopher Regina Rini calls an epistemic “backstop” for the comparatively flawed – or potentially deceptive – testimony of others and our own perceptions.²⁶ The wider adoption of newfangled devices like microscopes and X-rays meanwhile promised insights otherwise unavailable to humans, eventually mooting so much expert and eyewitness testimony. As legal scholars debated where illustration ended and proof began, “let nature speak for itself” became the guiding philosophy.²⁷

This approach may have been largely unproblematic, but was never entirely foolproof. In the modern era, digital processes have largely replaced analog optical and auditory ones, further blurring the lines between *capturing* and *manipulating* a sound or image.²⁸ In the space of a single generation, the transfer of light and sound waves onto a physical material like cellulose film or magnetic tape transformed into a fully computerized process from start to finish.²⁹ By the late 1980s, the average person could tinker with recorded sound or imagery on their home computer – culminating in the nearly ubiquitous use of Photoshop editing software roughly a decade later.³⁰ Most audio-visual media nowadays are primarily machine-captured and machine-readable, requiring minimal human involvement, if at all.³¹ The underlying files are only briefly available to our eyes and ears through interfaces – themselves also digital – putting even more conceptual distance between people and the events a medium purports to represent.³²

To grapple with this evolution, a new discipline arose in the 1990s: media forensics. Alarm about the forthcoming flood of manipulated content spurred some scholars to devise technically sophisticated means of detecting fabrications – particularly in high-stakes contexts like the criminal justice system and electoral politics. Such fraught cases have turned out to be sparse, however. Most simply did not require so much sleuthing. For instance, in an early “viral” hoax, a photo purportedly showing the body of Princess Diana

²⁵ Nicholas Carr, “Beautiful Lies: The Art of the Deep Fake,” *Los Angeles Review of Books*, December 6, 2021, <https://lareviewofbooks.org/article/beautiful-lies-the-art-of-the-deep-fake>.

²⁶ Rini, “Deepfakes and the Epistemic Backstop.”

²⁷ Tal Golan, “The Authority of Shadows: The Law and X-Rays,” in *Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America* (Harvard University Press, 2004), 176–210, <https://doi.org/10.2307/j.ctv1smjnhv.8>; Westling, “Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction to Deep Fakes.”

²⁸ Prusila, “Truth, Lies and ‘Deepfakes’: The Epistemology of Photographic Depictions,” 24.

²⁹ Kris Paulsen, “The Index and the Interface,” *Representations* 122, no. 1 (May 1, 2013): 83–109, <https://doi.org/10.1525/rep.2013.122.1.83>.

³⁰ Jenn Shreve, “Photoshop: It’s All the Rage,” *Wired*, November 19, 2001, <https://www.wired.com/2001/11/photoshop-its-all-the-rage>; Scheirer, *A History of Fake Things on the Internet*, 71.

³¹ Gideon Jacobs, “A.I. Is the Future of Photography. Does That Mean Photography Is Dead?,” *The New York Times*, December 26, 2023, <https://www.nytimes.com/2023/12/26/opinion/ai-future-photography.html>; Andrew Deck, “Errol Morris on Whether You Should Be Afraid of Generative AI in Documentaries,” *Neiman Lab*, July 10, 2024, <https://www.niemanlab.org/2024/07/errol-morris-on-whether-you-should-be-afraid-of-generative-ai-in-documentaries>.

³² Trevor Paglen, “Invisible Images (Your Pictures Are Looking at You),” *The New Inquiry*, December 8, 2016, <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you>.

being extracted from her fatal car crash circulated widely in 1997. The fake was readily dismissed by French authorities who noted disparities with paramedics' uniforms and the vehicles involved.³³ A Hungarian teenager joking with his friends in the aftermath of 9/11 would lead to a similar viral episode.³⁴ As had been true for new media over the previous two centuries – and contrary to prevalent worries that Photoshop would spell the end of trust in photographs – most manipulation turned out to be for “taking private jokes to a larger audience.”³⁵

Amid the curiosity and speculation that new technological advances generate, it can be easy to overlook these cultural contexts, as well as the institutional norms and disciplinary practices underpinning the media people routinely encounter.³⁶ When considering the threats posed by deepfakes to society's collective sensemaking, commentators often cultivate fear of losing some unmitigated access to reality that has never truly existed.³⁷ Insofar as deepfakes do imperil such access, the challenge may be as much to correct an historical *overreliance* on technological “backstops” as it is to identify new ones upon which to rely.³⁸ Society has never depended – and arguably never should – on singular recorded images or audio to formulate knowledge or understand events. Audio-visual media have always required “context and narration...to serve as evidence or to testify to a particular event or action,” says historian Kris Paulsen, “[they] are, by nature, always open to interpretation and doubt.”³⁹

“Info-Pocalypse” Now

The algorithmic technology underlying deepfakes made its greatest advances since 2015 – incidentally the period when the world became seized by concerns about the spread of false information and propaganda online.⁴⁰ By 2017, social media platforms and online message boards began seeing the first video deepfakes – short clips featuring uncannily

³³ Scheirer, *A History of Fake Things on the Internet*, 112-114, 138.

³⁴ Leo Hickman, “Tracking down the Tourist of Death,” *The Guardian*, November 30, 2001, <https://www.theguardian.com/technology/2001/nov/30/september112001.ethicalliving>.

³⁵ Shreve, “Photoshop: It’s All the Rage”; Immerwahr, “What the Doomsayers Get Wrong About Deepfakes.”

³⁶ Zsolt Ziegler, “Michael Polányi’s Fiduciary Program against Fake News and Deepfake in the Digital Age,” *AI & Society* 38, no. 5 (October 1, 2023): 1949–57, <https://doi.org/10.1007/s00146-021-01217-w>.

³⁷ Ignas Kalpokas and Julija Kalpokiene, “On Alarmism: Between Infodemic and Epistemic Anarchy,” in *Deepfakes: A Realistic Assessment of Potentials, Risks, and Policy Regulation*, ed. Ignas Kalpokas and Julija Kalpokiene (Cham: Springer International Publishing, 2022), 41–53, https://doi.org/10.1007/978-3-030-93802-4_5.

³⁸ Benjamin N. Jacobsen and Jill Simpson, “The Tensions of Deepfakes,” *Information, Communication & Society* 27, no. 6 (April 25, 2024): 1095–1109, <https://doi.org/10.1080/1369118X.2023.2234980>; Katja de Vries, “You Never Fake Alone. Creative AI in Action,” *Information, Communication & Society* 23, no. 14 (December 5, 2020): 2110–27, <https://doi.org/10.1080/1369118X.2020.1754877>.

³⁹ Paulsen, “The Index and the Interface,” 89, 105.

⁴⁰ Joseph Bernstein, “Bad News: Selling the Story of Disinformation,” *Harper’s Magazine*, August 9, 2021, <https://harpers.org/archive/2021/09/bad-news-selling-the-story-of-disinformation>.

realistic portrayals of prominent politicians and celebrities.⁴¹ The increasing ease of creating such material, its potential to go viral on social media, and the lack of any reliable technical or policy countermeasures all prompted an outpouring of speculation about what might lay ahead – particularly for democracy and the knowledge economy.

In a seminal article, legal scholars Robert Chesney and Danielle Citron posited that the problem deepfakes pose is two-fold: they might flood the information ecosystem with falsehoods, consequently leading to a widescale distrust of *all* audio-visual content. This crisis could, in turn, undermine elections, the criminal justice system, and other democratic institutions, as an erosion of confidence “makes it easier for liars to avoid accountability for things that are in fact true.”⁴² This so-called Liar’s Dividend thus threatens to gradually diminish individual and collective capacity to ascertain truth at a societal scale, argue Chesney and Citron.⁴³ Academics, journalists, and civil society researchers have since expanded on the prospects for this looming “information apocalypse.” Their concerns generally pivot upon how realistic and easily personalized content can be.⁴⁴

These catastrophic outcomes are eminently conceivable, particularly in the aftermath of recent national traumas in which propaganda campaigns and the spread of false

⁴¹ Maximilian Schreiner, “Deepfakes: How It All Began - and Where It Could Lead Us,” *The Decoder*, April 28, 2022, <https://the-decoder.com/history-of-deepfakes>; Jennifer Kavanagh and Michael D. Rich, “Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life” (RAND Corporation, January 16, 2018), https://www.rand.org/pubs/research_reports/RR2314.html.

⁴² Robert Chesney and Danielle Keats Citron, “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,” SSRN Scholarly Paper (Rochester, NY, July 14, 2018), <https://doi.org/10.2139/ssrn.3213954>. Also see Rebecca Delfino, “The Deepfake Defense—Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers,” SSRN Scholarly Paper (Rochester, NY, February 12, 2023), <https://doi.org/10.2139/ssrn.4355140>.

⁴³ For instance, a politician might more readily claim that a genuine recording of their own illicit behavior is “fake.” See Kaylyn Jackson Schiff, Daniel S. Schiff, and Natália S. Bueno, “The Liar’s Dividend: Can Politicians Claim Misinformation to Evade Accountability?,” *American Political Science Review*, February 20, 2024, 1–20, <https://doi.org/10.1017/S0003055423001454>.

⁴⁴ For example, see Editorial Board, “A Reason to Despair about the Digital Future: Deepfakes,” *Washington Post*, January 7, 2019, https://www.washingtonpost.com/opinions/a-reason-to-despair-about-the-digital-future-deepfakes/2019/01/06/7c5e82ea-0ed2-11e9-831f-3aa2c2be4cbd_story.html; Simon Parkin, “The Rise of the Deepfake and the Threat to Democracy,” *The Guardian*, June 22, 2019, <http://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy>; Madeleine Kearns, “Deepfakes Are Deeply Worrying,” *National Review*, August 27, 2019, <https://www.nationalreview.com/corner/deepfakes-deeply-worrying>; Joe Littel, “Don’t Believe Your Eyes (or Ears): The Weaponization of Artificial Intelligence, Machine Learning, and Deepfakes,” *War on the Rocks*, October 7, 2019, <https://warontherocks.com/2019/10/dont-believe-your-eyes-or-ears-the-weaponization-of-artificial-intelligence-machine-learning-and-deepfakes>; Cristian Vaccari and Andrew Chadwick, “Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News,” *Social Media + Society* 6, no. 1 (January 1, 2020), <https://doi.org/10.1177/2056305120903408>; Don Fallis, “The Epistemic Threat of Deepfakes,” *Philosophy & Technology* 34, no. 4 (December 1, 2021): 623–43, <https://doi.org/10.1007/s13347-020-00419-2>; Jeffrey T. Hancock and Jeremy N. Bailenson, “The Social Impact of Deepfakes,” *Cyberpsychology, Behavior, and Social Networking* 24, no. 3 (March 2021): 149–52, <https://doi.org/10.1089/cyber.2021.29208.jth>; Erik Hoel, “The Semantic Apocalypse,” *The Intrinsic Perspective*, February 27, 2024, <https://www.theintrinsicperspective.com/p/the-semantic-apocalypse>; John Fletcher, “Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance,” *Theatre Journal* 70, no. 4 (2018): 455–71; Matthew Burtell and Thomas Woodside, “Artificial Influence: An Analysis Of AI-Driven Persuasion” (arXiv, March 15, 2023), <http://arxiv.org/abs/2303.08721>; Paris and Donovan, “Deepfakes and Cheap Fakes”; James Greig, “Is AI Really a Threat to Democracy?,” *Dazed Digital*, February 29, 2024, <https://www.dazeddigital.com/life-culture/article/62043/1/is-ai-going-to-destroy-democracy-us-elections-deepfakes-voter-manipulation>; Mateusz Łabuz and Christopher Nehring, “Information Apocalypse or Overblown Fears—What AI Mis- and Disinformation Is All about? Shifting Away from Technology toward Human Reactions,” *Politics & Policy*, 2024, 1–18, <https://doi.org/10.1111/polp.12617>; Groh et al., “Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video;” Rini, “Deepfakes and the Epistemic Backstop.”

information featured prominently – for example, Russia’s shutdown of flight MH-17 over Ukraine, Moscow’s interference in Western elections, and the COVID-19 pandemic. In this regard, such scenarios may be overly biased toward recent experience than necessarily predictive about the future.⁴⁵ A range of literature, from social and cognitive sciences to media and communication studies, suggests the need to temper prevalent threat-perceptions about deepfake-induced, wholesale knowledge collapse.

Realism and Epistemic Value

The sophistication of the technology underlying deepfakes, the perceived “realness” of the content, or the mechanics of how deepfakes spread online often get the most attention from researchers and policymakers. The processes audiences undergo when they encounter such media are far less understood.⁴⁶ When considering how humans accumulate knowledge and establish truth, it is worth noting two differing kinds of reasoning that come into play: the curious, deliberative, slow kind that helps us model and grapple with reality; and the intuitive, reflexive, quick kind that is often guided more by emotion than calculation.⁴⁷

By extension, these different modes of cognition also undergird two different types of beliefs. Factual ones – like the law of gravity or the need to avoid predators – guide much of our behavior, since failure to adopt them poses real-world hazards. Meanwhile, symbolic beliefs – for instance, religious or partisan ones – mostly service social needs, signaling membership in a given group. We often adhere to these, even in the face of contradictory evidence or privately held skepticism.⁴⁸ As a result, we are used to juggling two versions of reality: “one that feels true and that we publicly advocate, and...another that we use to effectively interact with the world,” says anthropologist Manvir Singh.⁴⁹ Both modes serve useful functions in daily life, and can even be mutually complementary.⁵⁰

When media comes into play, the processes of *interpreting* it for social purposes and *evaluating* it for accuracy are also distinct.⁵¹ Failure to account for this difference likely lies

⁴⁵ Anna Broinowski, “Deepfake Nightmares, Synthetic Dreams: A Review of Dystopian and Utopian Discourses Around Deepfakes, and Why the Collapse of Reality May Not Be Imminent—Yet,” *Journal of Asia-Pacific Pop Culture* 7, no. 1 (May 1, 2022): 109–39, <https://doi.org/10.5325/jasiapacipopcult.7.1.0109>; Karin Wahl-Jorgensen and Matt Carlson, “Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes,” *Journalism Practice* 15, no. 6 (July 3, 2021): 803–20, <https://doi.org/10.1080/17512786.2021.1908838>.

⁴⁶ Lucas Christian Staab, “Making Sense of Deepfakes” (University of Twente, 2023), <https://essay.utwente.nl/97795>.

⁴⁷ Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).

⁴⁸ Sperber, “Intuitive and Reflective Beliefs.”

⁴⁹ Manvir Singh, “Don’t Believe What They’re Telling You About Misinformation,” *The New Yorker*, April 22, 2024, <https://www.newyorker.com/magazine/2024/04/22/dont-believe-what-theyre-telling-you-about-misinformation>.

⁵⁰ Anne Strainchamps, “Stop Worrying About Deepfakes,” *Nautilus*, December 20, 2023, <https://nautil.us/stop-worrying-about-deepfakes-470212>.

⁵¹ Becca Rothfeld, “Yes, People Lie Online. But It May Matter Less than We Fear.,” *Washington Post*, December 2, 2023, <https://www.washingtonpost.com/books/2023/12/02/history-fake-things-internet-walter-scheirer-review/>; Staab, “Making Sense of Deepfakes”; Sperber et al., “Epistemic Vigilance,” 10.

at the root of so much alarmism around media falsehood in general.⁵² Cognitive scientists warn against concluding that passionate commitment to certain ideas necessarily operates the same way as knowledge formulation.⁵³ Political scientists likewise caution that people insincerely make or accept certain claims as a way to publicly signal allegiance in a community, even while harboring quiet doubts about their veracity.⁵⁴

In practice, this means that while a politically-charged deepfake video may be realistic and evocative enough to share with friends, this resonance should not necessarily be conflated with wholesale acceptance as – or indifference toward – epistemic truth.

Catastrophic scenarios around deepfakes often tend to cast the public as a unified entity, easily duped, for whom all synthetic content will be of equal relevance.⁵⁵ But this pessimism conflates society's quest for Truth (in a grand, philosophical sense) with its ability to find the truth about very specific questions. When examining the media's role in knowledge formation, as in related fields, such oversimplified "folk theories" are often used as the baseline for analysis.⁵⁶ However, people rarely approach anything in media (or in nature, for that matter) as a binary choice between inherently truthful or calculatingly deceptive. Nor do they assess content with the foremost goal of confirming or refuting a given belief or argument.⁵⁷ Doing so would require far more mental and social energy than people can typically expend.⁵⁸ Humans tend to either ignore or forget media content – real or otherwise – which is not immediately, deeply relevant.⁵⁹

This does not rule out an individual or group being duped by a deepfake. But doing so in one context is unlikely to fuel distrust toward media in *every* context – as each has unique variables and degrees of significance to understanding the world.⁶⁰ Just as a photo-finish at

⁵² Groh et al., "Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video"; Carolin-Theresa Ziemer and Tobias Rothmund, "Psychological Underpinnings of Misinformation Countermeasures: A Systematic Scoping Review," *Journal of Media Psychology*, January 23, 2024, <https://doi.org/10.1027/1864-1105/a000407>.

⁵³ Dan Sperber et al., "Epistemic Vigilance," *Mind & Language* 25, no. 4 (2010): 359–93, <https://doi.org/10.1111/j.1468-0017.2010.01394.x>; Anderson and Johnston, *From Information Literacy to Social Epistemology: Insights from Psychology*, 134; Jean Claude Kwitonda and Symone Campbell, "'The Medium Is the Massage/Message': Functions of Synthetic Media in Sense-Making Conditions," in *Black Communication in the Age of Disinformation: DeepFakes and Synthetic Media*, ed. Kehbama Langmia (Cham: Springer International Publishing, 2023), 131–46, https://doi.org/10.1007/978-3-031-27696-5_7.

⁵⁴ Michael Hannon, "Disagreement or Badmouthing? The Role of Expressive Discourse in Politics," in *Political Epistemology*, ed. Elizabeth Edenberg and Michael Hannon (Oxford University Press, 2021), 297–317, <https://doi.org/10.1093/oso/9780192893338.003.0017>.

⁵⁵ Westling, "Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction to Deep Fakes."

⁵⁶ Celia Paris, "Reconciling Democratic Theories and Realities: A Review of Christopher H. Achen and Larry M. Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government*," *The Good Society* 25, no. 1 (2017): 119–27, <https://doi.org/10.5325/goodsociety.25.1.0119>.

⁵⁷ Prusila, "Truth, Lies and 'Deepfakes': The Epistemology of Photographic Depictions," 59.

⁵⁸ Daniel Williams, "The Marketplace of Rationalizations," *Economics & Philosophy*, March 3, 2022, 1–25, <https://doi.org/10.1017/S0266267121000389>.

⁵⁹ Dan Sperber and Deirdre Wilson, *Relevance: Communication and Cognition* (Malden: Blackwell Publishing, 1995); Hugo Mercier, *Not Born Yesterday: The Science of Who We Trust and What We Believe* (Princeton, NJ: Princeton University Press, 2020); de Vries, "You Never Fake Alone. Creative AI in Action."

⁶⁰ Keith Raymond Harris, "Real Fakes: The Epistemology of Online Misinformation," *Philosophy & Technology* 35, no. 3 (August 31, 2022): 83, <https://doi.org/10.1007/s13347-022-00581-9>.

the Olympics and an action sequence in a movie might rely on the exact same instruments to produce, the vastly different contexts must be taken into account to assign significance to what they represent.⁶¹ To overlook these complexities is to conflate what is technologically plausible with what is sociologically probable.⁶² In terms of how realistic deepfakes might be, the question of persuasiveness requires a great deal of nuance – particularly to distinguish what might be sufficient to go viral online from what might be sufficient to fool a judge or a jury.⁶³ The degree of relevance and the gravity of these contexts should not be confused with the mere propensity to draw attention.

Meanwhile, the spaces currently most flooded with deepfake content are those that are *already* least relevant to collective sensemaking, where attention-grabbing fakery is already deeply embedded: advertising, political memes, pornography, and entertainment. Such content is “most often (but not always) created, indexed, and received as a trick for the eye rather than with the intent to deceive the mind.”⁶⁴ In this regard, deepfakes may simply prove to be the latest in a long trend of cultural kayfabe – like professional wrestling or stage magic. In music, for instance, the “vaporwave” genre relies on splicing, stretching, and manipulating samples of existing works in an unapologetically derivative, ironic, and uncanny way – designed less as a demonstration of musicality or composition than of *aesthetics*.⁶⁵ These winking forms make few, if any, formal epistemic claims, deliberately leaving it to audiences to project their own meaning onto, or infer it from, the performance.⁶⁶

Paradoxically, a preoccupation with AI-generated content as a uniquely threatening kind of fakery – rather than a novel form of political cartoon, meme, kayfabe, aesthetic, or

⁶¹ Rob Horning, “Truth and Consequences,” *Internal Exile*, March 8, 2024, <https://robhorning.substack.com/p/truth-and-consequences>; Prusila, “Truth, Lies, and ‘Deepfakes’: The Epistemology of Photographic Depictions,” 40.

⁶² Gavin Wilde, “From Panic to Policy: The Limits of Foreign Propaganda and the Foundations of an Effective Response,” *Texas National Security Review*, March 26, 2024, <https://tnsr.org/2024/03/from-panic-to-policy-the-limits-of-foreign-propaganda-and-the-foundations-of-an-effective-response>; Ciaran Martin, “Deepfakes Are Here and Can Be Dangerous, but Ignore the Alarmists – They Won’t Harm Our Elections,” *The Guardian*, June 11, 2024, <https://www.theguardian.com/commentisfree/article/2024/jun/11/deepfakes-ignore-alarmists-elections>.

⁶³ Shannon Bond, “People Are Trying to Claim Real Videos Are Deepfakes. The Courts Are Not Amused,” *NPR*, May 8, 2023, <https://www.npr.org/2023/05/08/1174132413/people-are-trying-to-claim-real-videos-are-deepfakes-the-courts-are-not-amused>.

⁶⁴ Rob Horning, “Reciprocal Human Investment,” *Overland Literary Journal*, February 7, 2024, <https://overland.org.au/2024/02/reciprocal-human-investment>; Bode, “Deepfaking Keanu: YouTube Deepfakes, Platform Visual Effects, and the Complexity of Reception”; Harris, “Real Fakes: The Epistemology of Online Misinformation,” 11.

⁶⁵ Lucas Cantinelli, “Anemoia and the Vaporwave Phenomenon: The ‘New’ Aesthetic of an Imagined Nostalgia,” *Imaginaires*, no. 26 (June 30, 2024): 132–51, <https://doi.org/10.34929/imaginaires.vi26.57>.

⁶⁶ “Kayfabe” can be defined as “the contemporary concept wherein we know we’re being deceived but we play along, sometimes because we must and sometimes because it is actually pleasurable.” See Eero Laine, “Professional Wrestling Scholarship: Legitimacy and Kayfabe,” *Popular Culture Studies Journal* 6, no. 1 (2018): 82–96, https://www.mpaaca.org/files/ugd/5a6d69_463fd96f910e4d5b8b9402683fa7f266.pdf; Abraham Josephine Riesman, “The Best Way to Explain the G.O.P. Is Found in the W.W.E.,” *The New York Times*, February 26, 2023, <https://www.nytimes.com/2023/02/26/opinion/vince-mcmahon-wwe-trump-kayfabe.html>; and Nadisha-Marie Aliman, Leon Kester, “AI-Related Misdirection Awareness in AIVR,” *PhilArchive*, June 1, 2024, <https://philarchive.org/rec/ALIAMA-3>; Sterne, *The Audible Past, Cultural Origins of Sound Reproduction*, 244.

Frankfurtian bullshit⁶⁷ – may lend more epistemic credence to these time-worn, familiar phenomena than any of them have ever warranted. The vast majority of deepfakes may turn out to be more phony than fraudulent, designed with little purpose beyond spectacle, much less any calculated deceit.⁶⁸ Notably, researchers at Notre Dame University who set out to catalog inauthentic content online found that the majority of it was “cultural expression, not the traditional political propaganda...more satirical than diabolical...it was memes all the way down.”⁶⁹

More recently, with more than 40 national elections slated to take place last year, 2024 threatened to become the year that deepfakes posed the greatest danger to democratic processes worldwide.⁷⁰ Instead, researchers and platforms like YouTube similarly saw “a whole lot of parody and little actual malice...the fakery [was] the point.”⁷¹ Most of the inauthentic media targeting voters was easily debunked; even for those who may have been duped, it isn’t clear how their views would have been meaningfully impacted.⁷² “We see limited effective use of AI...[it] isn’t necessarily boosting productivity for threat actors, nor is it being used more than any other analog or existing content generation tools,” says Olga Belogolova, Director of the Emerging Technologies Initiative at the Johns Hopkins School of Advanced International Studies.⁷³

Concurrent advances in machine-learning and automation, paired with pervasive data-collection online, now make audiences more easily identifiable – and their preferences more narrowly targetable – than ever before.⁷⁴ This has given rise to concern about the “personalization” of deepfakes, and what it might mean for individual or group-level

⁶⁷ Harry G. Frankfurt, *On Bullshit* (Princeton NJ: Princeton University Press, 2005), <https://doi.org/10.2307/j.ctt7t4wr>; Łabuz and Nehring, “On the Way to Deep Fake Democracy? Deep Fakes in Election Campaigns in 2023”; ⁶⁷ Joe Slater Hicks, James Humphries, Michael Townsen, “ChatGPT Isn’t ‘Hallucinating’—It’s Bullshitting!,” *Scientific American*, July 17, 2024, <https://www.scientificamerican.com/article/chatgpt-isnt-hallucinating-its-bullshitting/>; Rob Horning, “Truth and Consequences.”

⁶⁸ Henry Ajder and Joshua Glick, “Just Joking! Deepfakes, satire, and the politics of synthetic media,” *WITNESS and MIT*, December 2012, <https://cocreationstudio.mit.edu/just-joking>.

⁶⁹ Scheier, “*A History of Fake Things on the Internet*,” xvi.

⁷⁰ Pranshu Verma and Cat Zakrzewski, “AI Deepfakes Threaten to Upend Global Elections. No One Can Stop Them.,” *Washington Post*, April 23, 2024, <https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india>.

⁷¹ From an in-person Q&A with a YouTube official on June 6, 2024. Also see Russell Brandom, “India’s Election Wasn’t the Deepfake Doomsday Many Feared,” *Rest of World*, May 30, 2024, <https://restofworld.org/2024/exporter-india-deepfake-trolls>.

⁷² Lluís de Nadal, Peter Jančárik, “Beyond the Deepfake Hype: AI, Democracy, and ‘The Slovak Case,’” *HKS Misinformation Review*, 6, no. 1 (August, 2024), <https://doi.org/10.37016/mr-2020-153>; Sara Fischer, Megan Morrone, “The AI-Generated Disinformation Dystopia that Wasn’t,” *Axios*, July 22, 2024, <https://www.axios.com/2024/07/22/ai-deepfakes-trump-news-misinformation/>; Louis Anslow, “Deepfakes Haven’t Destroyed Democracy—Yet,” *Daily Beast*, July 25, 2024, <https://www.thedailybeast.com/deepfakes-havent-destroyed-democracy-yet/>; Will Oremus, “What a Viral, Fake Image of Rafah Tells Us about AI Propaganda,” *Washington Post*, May 30, 2024, <https://www.washingtonpost.com/politics/2024/05/30/all-eyes-on-rafah-instagram-ai-policies/>; Mark Scott, “We Need to Calm down about AI and Elections,” *Politico*, April 18, 2024, <https://www.politico.eu/newsletter/digital-bridge/we-need-to-calm-down-about-ai-and-elections>.

⁷³ Interview with author, August 16, 2024. See also Kevin Collier, “Russian Propagandists are Still Targeting Americans in Influence Operations, Meta Says,” *NBC News*, August 15, 2024, <https://www.nbcnews.com/tech/security/russian-propagandists-are-still-targeting-americans-influence-operatio-rena166648>; and Teresa Weikmann and Sophie Lecheler, “Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network,” *Digital Journalism*, 8, <https://doi.org/10.1080/21670811.2023.2194665>.

⁷⁴ Josh A. Goldstein and Girish Sastry, “The Coming Age of AI-Powered Propaganda,” *Foreign Affairs*, April 7, 2023, <https://www.foreignaffairs.com/united-states/coming-age-ai-powered-propaganda>.

attempts at manipulation. However, there is no previously unmet market for material that deepfakes now somehow satisfy to support a given worldview. Would-be propagandists have always been limited more by audience *demand* than lack of content *supply*.⁷⁵ Particularly online, audiences are already well-attuned to the potential for fakery, says media theorist Rob Horning: “they know when the stakes are low enough to suspend skepticism and derive whatever enjoyment they can from it.”⁷⁶ Beyond their novelty, it is unclear why a targeted campaign to mislead the public would necessarily favor or require deepfakes over other kinds of content.⁷⁷

Meanwhile, casting the pitfalls of personalized synthetic media through the (still largely speculative) prism of politics and news tends to overshadow their most prevalent illicit uses-cases.⁷⁸ AI-generated graphic imagery is now routinely used both to harass and portray non-consenting victims – particularly women and minors – while voice-cloning targets vulnerable groups like the elderly and undermines identity verification (for instance, by financial institutions).⁷⁹ In terms of societal-level harms from personalized deepfakes, rampant misogyny and fraud constitute actual agonies, from which “wishful worries” like mass political manipulation or epistemic crisis can sometimes serve as convenient distractions.⁸⁰

Perils of Pollution

In a world reliant on the Internet for both the connection and the curation necessary to locate facts and establish knowledge, a deluge of synthetically produced content might ultimately drown out whatever quality information exists online, leading publics to feel helpless and apathetic about their prospects of ever finding it at all.⁸¹ Such a scenario looks increasingly plausible, particularly as search engines and digital archives – resources upon which several generations have come to depend for seeking out credible information – have

⁷⁵ Westling, “Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction to Deep Fakes,” 19.

⁷⁶ Rob Horning, “Have You Heard the Word,” *Internal Exile*, December 21, 2023, <https://robhorning.substack.com/p/have-you-heard-the-word>; de Vries, “You Never Fake Alone. Creative AI in Action.”

⁷⁷ Tim Hwang, “Deepfakes: A Grounded Threat Assessment,” *Center for Security and Emerging Technology*, July 2020, <https://cset.georgetown.edu/publication/deepfakes-a-grounded-threat-assessment>.

⁷⁸ Graham Meikle, *Deepfakes* (John Wiley & Sons, 2022), 4; Chandell Gosse and Jacquelyn Burkell, “Politics and Porn: How News Media Characterizes Problems Presented by Deepfakes,” *Critical Studies in Media Communication* 37, no. 5 (October 19, 2020): 497–511, <https://doi.org/10.1080/15295036.2020.1832697>.

⁷⁹ Sophie Compton, “More Women are Facing the Reality of Deepfakes, and They’re Ruining Lives,” *Vogue*, March 15, 2021, <https://www.vogue.co.uk/news/article/stop-deepfakes-campaign>; Joe Hernandez, “That Panicky Call from a Relative? It Could Be a Thief Using a Voice Clone, FTC Warns,” *NPR*, March 22, 2023, <https://www.npr.org/2023/03/22/1165448073/voice-clones-ai-scams-ftc>; Katie McQue, “AI Is Overpowering Efforts to Catch Child Predators, Experts Warn,” *The Guardian*, July 18, 2024, <https://www.theguardian.com/technology/article/2024/jul/18/ai-generated-images-child-predators>.

⁸⁰ David Brock, “Our Censors, Ourselves: Commercial Content Moderation,” *Los Angeles Review of Books*, July 25, 2019, <https://lareviewofbooks.org/article/our-censors-ourselves-commercial-content-moderation>.

⁸¹ Alison MacKenzie, Jennifer Rose, and Ibrar Bhatt, “Dupery by Design: The Epistemology of Deceit in a Postdigital Era,” *Postdigital Science and Education* 3, no. 3 (October 1, 2021): 693–99, <https://doi.org/10.1007/s42438-020-00114-7>; Rini, “Deepfakes and the Epistemic Backstop.”

recently shifted their business models away from indexing and curation towards something else entirely: synthesis and summarization.⁸² The impact of this shift on common knowledge has thus far been mixed, at best.⁸³ False, misleading, or wholly nonsensical search results now dominate some of the digital spaces that could previously be relied upon – even as “just good enough” – to service our routine queries, or at least to chart a path toward answers to weightier questions.⁸⁴

As the financial incentives for both online content creation and curation disappear, some analysts fear the best days of the Internet as a source of knowledge production have now passed.⁸⁵ For instance, at least one third of all websites that existed at some point over the last decade no longer do.⁸⁶ Reams of academic research have disappeared into the ether as publishers gradually stop maintaining their online holdings.⁸⁷ Indeed, the dream of the Internet as a commonly accessible library for the world’s knowledge now seems naively optimistic – beset at first by copyright disputes, then by the “built-in obsolescence” of digital storage, now finally by a flood of generated, nonsensical “slop.”⁸⁸ This shifting ecosystem also threatens to upend digital literacy best-practices – themselves the product of years of research – which assumed the Internet would perpetually be populated by human-produced (if often low-quality) content.⁸⁹

This looming reconfiguration of how people gain knowledge online holds potential, however, to serve as a long-overdue course-correction away from automation, back toward the historical norms of iteration and social interaction.⁹⁰ What is frequently lamented as the

⁸² Chirag Shah and Emily M. Bender, “Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web?,” *ACM Trans. Web* 18, no. 3 (April 15, 2024): 33:1-33:24, <https://doi.org/10.1145/3649468>.

⁸³ Erik Hoel, “Here Lies the Internet, Murdered by Generative AI,” *The Intrinsic Perspective*, March 16, 2022, <https://www.theintrinsicperspective.com/p/here-lies-the-internet-murdered-by>.

⁸⁴ Mike Caulfield, “Truth Is in the Network,” *Project Information Literacy*, June 4, 2019, <https://projectinfo.org/smart-talk-interviews/truth-is-in-the-network>; Adam Rogers, “The New ChatGPT Clones from Google and Microsoft Are Going to Destroy Online Search,” *Business Insider*, February 9, 2023, <https://www.businessinsider.com/ai-chatbots-chatgpt-google-bard-microsoft-bing-break-internet-search-2023-2>.

⁸⁵ Judith Donath and Bruce Schneier, “It’s the End of the Web as We Know It,” *The Atlantic*, April 22, 2024, <https://www.theatlantic.com/technology/archive/2024/04/generative-ai-search-llmo/678154>.

⁸⁶ Gonzalo Rivero et al., “When Online Content Disappears,” *Pew Research Center*, May 17, 2024, <https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears>.

⁸⁷ Jeffrey Brainard, “Dozens of Scientific Journals Have Vanished from the Internet, and No One Preserved Them,” *Science*, September 8, 2020, <https://www.science.org/content/article/dozens-scientific-journals-have-vanished-internet-and-no-one-preserved-them>.

⁸⁸ Keane, “Communicative Abundance,” 29-30; Hoel, “A.I.-Generated Garbage Is Polluting Our Culture”; Ashley Belanger, “Internet Archive Forced to Remove 500,000 Books after Publishers’ Court Win,” *Ars Technica*, June 21, 2024, <https://arstechnica.com/tech-policy/2024/06/internet-archive-forced-to-remove-500000-books-after-publishers-court-win>; Jason Koebler, “Where Facebook’s AI Slop Comes From,” *404 Media*, August 6, 2024, <https://www.404media.co/where-facebooks-ai-slop-comes-from>.

⁸⁹ Joel Blechinger, “Reflections on information literacy in the ChatGPT era,” *Pathfinder: A Canadian Journal for Information Science Students and Early Career Professionals*, 4, no. 1 (2023): 163-172. <https://doi.org/10.29173/pathfinder78>; Elise Silva, “AI-Powered Search and the Rise of Google’s ‘Concierge Wikipedia,’” *Tech Policy Press*, May 28, 2024, <https://www.techpolicy.press/ai-powered-search-and-the-rise-of-googles-concierge-wikipedia>.

⁹⁰ Chirag Shah and Emily M. Bender, “Situating Search,” in *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR ’22: ACM SIGIR Conference on Human Information Interaction and Retrieval*, Regensburg Germany: ACM, 2022), 221–32, <https://doi.org/10.1145/3498366.3505816>.

current “post-truth” era might be more accurately dubbed a “post-speed” one.⁹¹ “The idea that you can just quickly Google something and know what’s fact and what’s fiction – I don’t know think it works like that anymore,” says Oxford Internet Institute’s Sandra Wachter.⁹² It likely never did, writes media literacy guru Mike Caulfield: “Search is often seen as a marketplace, but it’s more like a rummage sale.”⁹³ After all, knowledge was never rooted solely in the recording, indexing, and transmission of content, nor by casting it as binary, zero-sum conceptual battle between fact and fiction, with little in between.⁹⁴

In a bid to demystify the new flood of content, industry-led collaboratives aim to standardize and simplify digital forensics.⁹⁵ Fortunately, some of these “deepfake detection” methods apply a holistic approach – analyzing content “together with all correlated evidence” like metadata, eyewitness testimony, near-identical versions of the same materials, creators claiming credit (or lack thereof), and their bona fides and affiliations.⁹⁶ Given the “black box” nature of the technology underpinning deepfakes, the *absence* of such evidence may be among the best indicators of inauthenticity.⁹⁷ Even so, deepfake generation and detection are likely to be locked in a long-term arms race.⁹⁸ The latter, even if perfected, is still likely insufficient to curb the extant fraud and harassment harms of deepfakes, much less to stave off the (still mostly speculative) worst-case epistemological scenarios.⁹⁹

⁹¹ Michael Hannon, “The Politics of Post-Truth,” *Critical Review* 35, no. 1–2 (April 3, 2023): 40–62, <https://doi.org/10.1080/08913811.2023.2194109>.

⁹² Melissa Heikkila, “An AI Startup Made a Hyperrealistic Deepfake of Me That’s so Good It’s Scary,” *MIT Technology Review*, April 25, 2024, <https://www.technologyreview.com/2024/04/25/1091772/new-generative-ai-avatar-deepfake-synthesis>.

⁹³ Mike Caulfield, “Truth Is in the Network,” *Project Information Literacy*, June 4, 2019, <https://projectinfoilit.org/smart-talk-interviews/truth-is-in-the-network>. Also see Lincoln Michel, “The Library of Blather,” *Counter Craft*, March 17, 2023, <https://countercraft.substack.com/p/the-library-of-blather>.

⁹⁴ Noortje Marres, “Why We Can’t Have Our Facts Back,” *Engaging Science, Technology, and Society* 4 (July 24, 2018): 423–43, <https://doi.org/10.17351/ests2018.188>; Janet Vertesi, “The Encyclopedia Project, or How to Know in the Age of AI,” *Public Books*, June 13, 2024, <https://www.publicbooks.org/the-encyclopedia-project-or-how-to-know-in-the-age-of-ai>.

⁹⁵ See, for instance, the Coalition for Content Provenance and Authenticity (C2PA) at <https://c2pa.org>.

⁹⁶ Luisa Verdoliva, “Media Forensics and DeepFakes: An Overview,” *IEEE Journal of Selected Topics in Signal Processing* 14, no. 5 (August 2020): 910–32, <https://doi.org/10.1109/JSTSP.2020.3002101>; Scheier, *A History of Fake Things on the Internet*, 116, 131; Amanda Morris, “Deepfake-Detection System Is Now Live,” *Northwestern Engineering*, July 9, 2024, <https://www.mccormick.northwestern.edu/news/articles/2024/07/deepfake-detection-system-is-now-live>; also see the U.S. Defense Advanced Research Projects Agency’s (DARPA) Semantic Forensics (SemaFor) program at <https://www.semanticforensics.com>.

⁹⁷ Sanford C. Goldberg, “Epistemically Engineered Environments,” *Synthese* 197, no. 7 (July 2020): 2783–2802, <https://doi.org/10.1007/s11229-017-1413-0>.

⁹⁸ Fakhar Abbas and Araz Taeihagh, “Unmasking Deepfakes: A Systematic Review of Deepfake Detection and Generation Techniques Using Artificial Intelligence,” *Expert Systems with Applications* 252 (October 15, 2024): 124260, <https://doi.org/10.1016/j.eswa.2024.124260>.

⁹⁹ Kelley Saylor and Laurie Harris, “Deepfakes and National Security,” *Congressional Research Service*, April 17, 2023, <https://crsreports.congress.gov/product/pdf/IF/IF11333>.

The Institutional “Backstop”

Beneath widely held fears about the proliferation of deepfakes and the pollution of the Internet with machine-made content lies a faulty sense of inevitability: that a search engine was fated to be the lynchpin of human understanding, or that artificial intelligence is destined to play a new central role in every audio-visual medium.¹⁰⁰ But technological innovation is but one factor driving the future – institutions, norms, values, and cultural choices play just as decisive a role.¹⁰¹ To assume otherwise benefits and solidifies one set of institutions – namely, the owners and producers of these technologies – while distracting from others that have endured, and must continue to, for much longer.¹⁰²

Media theorist Neil Postman once wrote that truth is “a product of a conversation man has with himself about and through the techniques of communication he has invented.”¹⁰³ Despite the technological promise that such techniques might spare us from the inherent frailties of that conversation, humanity’s search for knowledge is an inescapably imperfect and social process.¹⁰⁴ To talk about the authenticity and authoritativeness of a picture, a book, a recording, or a film is to refer to something far deeper than the medium or the content themselves. It is to implicate the scaffolding of social rituals and institutions that went into their production.¹⁰⁵

True, the advent of new technical instruments enabled society to rely on a snapshot or soundbite as a piece of evidence. But in equal part, this was due to the technicians and techniques, organizational credentials, norms governing operation and maintenance, and bona fides surrounding publication and dissemination that emerged in tandem – built on communally developed expertise and best practice, over time.¹⁰⁶ However complex, opaque, or outdated as these might seem, they still carry a great deal of weight. For example, a written signature or an oath in court are archaic and easily falsifiable, but still universally signify one’s commitment to a range of legal, political, and ethical

¹⁰⁰ Jascha Bareis and Christian Katzenbach, “Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics,” *Science, Technology, & Human Values* 47, no. 5 (September 1, 2022): 855–81, <https://doi.org/10.1177/01622439211030007>.

¹⁰¹ L. M. Sacasas, “Is Resistance Futile? The Myth of Tech Inevitability,” *Local Futures*, April 26, 2021, <https://www.localfutures.org/is-resistance-futile-the-myth-of-tech-inevitability>; Schiavo, ““From Phantom Image to Perfect Vision: Physiological Optics, Commercial Photography, and the Popularization of the Stereoscope,” 114.

¹⁰² Bernstein, “Bad News: Selling the Story of Disinformation”; John Horgan, “Premature Freak-Outs about Techno-Enhancement,” *Scientific American*, January 23, 2021, <https://www.scientificamerican.com/article/premature-freak-outs-about-techno-enhancement>.

¹⁰³ Neil Postman, *Amusing Ourselves to Death: Public Discourse in the Age of Show Business* (Penguin, 2005), 11, 24.

¹⁰⁴ Dan Cavedon-Taylor, “Photographically Based Knowledge,” *Episteme* 10, no. 3 (September 2013): 283–97, <https://doi.org/10.1017/epi.2013.21>. For more on knowledge production as a social process, see Mandi Astola and Mark Alfano, “Social Epistemology,” *Open Encyclopedia of Cognitive Science*, July 24, 2024, <https://doi.org/10.21428/e2759450.e29a27cc>.

¹⁰⁵ Anthony Anderson and Bill Johnston, *From Information Literacy to Social Epistemology: Insights from Psychology* (Chandos Publishing, 2016); Sandy Goldberg, “The Division of Epistemic Labor,” *Episteme* 8, no. 1 (February 2011): 112–25, <https://doi.org/10.3366/epi.2011.0010>; Sterne, *The Audible Past, Cultural Origins of Sound Reproduction*, 219..

¹⁰⁶ Rob Horning, “Ex Noumena,” *Internal Exile*, May 3, 2024, <https://robhorning.substack.com/p/ex-noumena>; Habgood-Coote, “Deepfakes and the Epistemic Apocalypse,” 103.

responsibilities.¹⁰⁷ Film may have long become obsolete in newsrooms or forensic labs, yet we continue to trust in photography because experts and practitioners of the discipline still prize authenticity and penalize fraud in highly relevant contexts.¹⁰⁸ Ultimately, “the ability to interpret the truth of evidence has been the work of institutions—journalism, the courts, the academy, museums, and other cultural organizations” – less breakthroughs in technology or individual savvy, says library scientist Britt Paris.¹⁰⁹

These factors may make wholesale epistemic collapse an unlikely prospect, but they also suggest that an institutional – rather than digital – “backstop” is the optimal strategy, in any case. The urge to mitigate the harms of deepfakes through primarily digital methods – like automated detection or voluntary watermarking – only further perpetuates a mythical, unattainable standard of technologically-derived objectivity that might help us conclusively discern between real and artificial.¹¹⁰ This approach risks doubling down on an already bad bet, on the very overreliance on technology that sparked such profound fears about deepfakes in the first place.¹¹¹ Moreover, techno-centric solutions are likely to further solidify power in the hands of the already powerful, to the exclusion of those being truly victimized by deepfake-related harms.¹¹²

Insofar as deepfakes might threaten our collective grip on reality, research and policy focus is best concentrated on reinforcing the social and institutional “backstops” that truly underpin common knowledge: libraries and library science,¹¹³ local journalism,¹¹⁴ publicly accessible digital resources and archives,¹¹⁵ standards-setting and credentialing

¹⁰⁷ Catherine Kerner and Mathias Risse, “Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds,” *Moral Philosophy and Politics* 8, no. 1 (April 1, 2021): 81–108, <https://doi.org/10.1515/mopp-2020-0024>.

¹⁰⁸ Joshua Habgood-Coote, “Deepfakes and the Epistemic Apocalypse,” *Synthese* 201, no. 3 (March 9, 2023), <https://doi.org/10.1007/s11229-023-04097-3>; Westling, “Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction to Deep Fakes.”

¹⁰⁹ Paris and Donovan, “Deepfakes and Cheap Fakes.”

¹¹⁰ Jacobsen and Simpson, “The Tensions of Deepfakes.”

¹¹¹ Andrew Buzzell and Regina Rini, “Doing Your Own Research and Other Impossible Acts of Epistemic Superheroism,” *Philosophical Psychology* 36, no. 5 (July 4, 2023): 906–30, <https://doi.org/10.1080/09515089.2022.2138019>; Miller and Record, “Justified Belief in a Digital Age: On the Epistemic Implications of Secret Internet Technologies.”

¹¹² Britt Paris, “Seeing Through the Fog of War: Assessing Epistemic Burden Around Cheapfakes and Deepfakes of Geopolitical Crisis,” in *Re-Thinking Mediations of Post-Truth Politics and Trust* (Routledge, 2023), <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003388975-8/seeing-fog-war-britt-paris>; Habgood-Coote, “Deepfakes and the Epistemic Apocalypse,” 102–104.

¹¹³ Shannon Mattern, “Library as Infrastructure,” *Places Journal*, June 9, 2014, <https://doi.org/10.22269/140609>.

¹¹⁴ Michael Koliska, Erin Moroney, and David Beavers, “Trust Through Relationships in Journalism,” *Journalism Studies* (May 2023): 1–18, <https://doi.org/10.1080/1461670X.2023.2209807>.

¹¹⁵ Maria Bustillos, “The Internet Archive Is a Library,” *Flaming Hydra*, June 13, 2024, <https://flaminghydra.com/the-internet-archive-is-a-library/>; Katharine Schwab, “Inside Starling Lab, a Moonshot Project to Preserve the World’s Most Important Information,” *Fast Company*, March 30, 2022, <https://www.fastcompany.com/90731729/inside-starling-lab-a-moonshot-project-to-preserve-the-worlds-most-important-information>.

organizations,¹¹⁶ and related civil society efforts.¹¹⁷ These collectives and practices have long underpinned our common knowledge, and must continue to. Many of them, however, suffer chronic underfunding and politically motivated attacks. Salvaging and bolstering them in the current era is indeed a lofty goal – but arguably no loftier than the quest for a digital shortcut to epistemic purity.

The Pyrite Premium

Deepfakes could be usefully compared to pyrite – commonly known as “fool’s gold.” Some early prospectors were indeed duped by the mineral, but soon thereafter it became a novelty item, and even proved useful in its own right.¹¹⁸ Its relative abundance did not diminish the value of real gold, nor did it diminish the odds of striking it. Over time, fool’s gold merely placed a greater premium on the processes and methods of discovering and validating the real thing.¹¹⁹ The worth of pure gold is, ultimately, a social construct – based not only on its relative scarcity, but on the customs and disciplines involved in mining and assaying it. The so-called “gold standard” originated not as a claim of perfection, but as a way to curb counterfeits through standardized practices.¹²⁰ Similarly, the emergence of deepfakes should spur less fear of an information apocalypse, more focus and investment in the institutional “backstops” that have long enabled societies to build knowledge and establish truth.¹²¹

¹¹⁶ Jessica Peake, “Challenges of Using Digital Evidence for War Crimes Prosecutions: Availability, Reliability, Admissibility,” *American Journal of International Law*, 118 (January 2024): 57–61, <https://doi.org/10.1017/aju.2024.5>; Hayley Evans and Mahir Hazim, “Digital Evidence Collection at the Int’l Criminal Court: Promises and Pitfalls,” *Just Security*, July 5, 2023, <https://www.justsecurity.org/87149/digital-evidence-collection-at-the-intl-criminal-court-promises-and-pitfalls/>; Benj Edwards, “Deepfakes in the Courtroom: US Judicial Panel Debates New AI Evidence Rules,” *Ars Technica*, April 24, 2024, <https://arstechnica.com/information-technology/2024/04/deepfakes-in-the-courtroom-us-judicial-panel-debates-new-ai-evidence-rules>.

¹¹⁷ Shannon Mattern, “Library as Infrastructure,” *Places Journal*, June 9, 2014, <https://doi.org/10.22269/140609>; A. W. Geiger, “Most Americans – Especially Millennials – Say Libraries Can Help Them Find Reliable, Trustworthy Information,” *Pew Research Center*, August 30, 2017, <https://www.pewresearch.org/short-reads/2017/08/30/most-americans-especially-millennials-say-libraries-can-help-them-find-reliable-trustworthy-information/>; Lee Rainie, “The Information Needs of Citizens: Where Libraries Fit In,” *Pew Research Center*, April 9, 2018, <https://www.pewresearch.org/internet/2018/04/09/the-information-needs-of-citizens-where-libraries-fit-in/>; Tracie Hall, “Attacks on Libraries Are Attacks on Democracy,” *Time Magazine*, September 20, 2023, <https://time.com/collection/time100-voices/6315724/banned-books-library-access>.

¹¹⁸ David Rickard, *Pyrite: A Natural History of Fool’s Gold* (Oxford, New York: Oxford University Press, 2015), 17-20.

¹¹⁹ Jessica Silbey and Woodrow Hartzog, “The Upside of Deep Fakes,” *Maryland Law Review* 78, no. 4 (January 1, 2019): 960-66.

¹²⁰ Stanley L. Brodsky and Bronwen Lichtenstein, “The Gold Standard and the Pyrite Principle: Toward a Supplemental Frame of Reference,” *Frontiers in Psychology* 11 (March 31, 2020): 562, <https://doi.org/10.3389/fpsyg.2020.00562>.

¹²¹ Andreas Jungherr, “Artificial Intelligence and Democracy: A Conceptual Framework,” *Social Media + Society* 9, no. 3 (July 1, 2023), <https://doi.org/10.1177/20563051231186353>; Rob Horning, “The Egg of Experience,” *Internal Exile*, September 29, 2023, <https://robhorning.substack.com/p/the-egg-of-experience>.