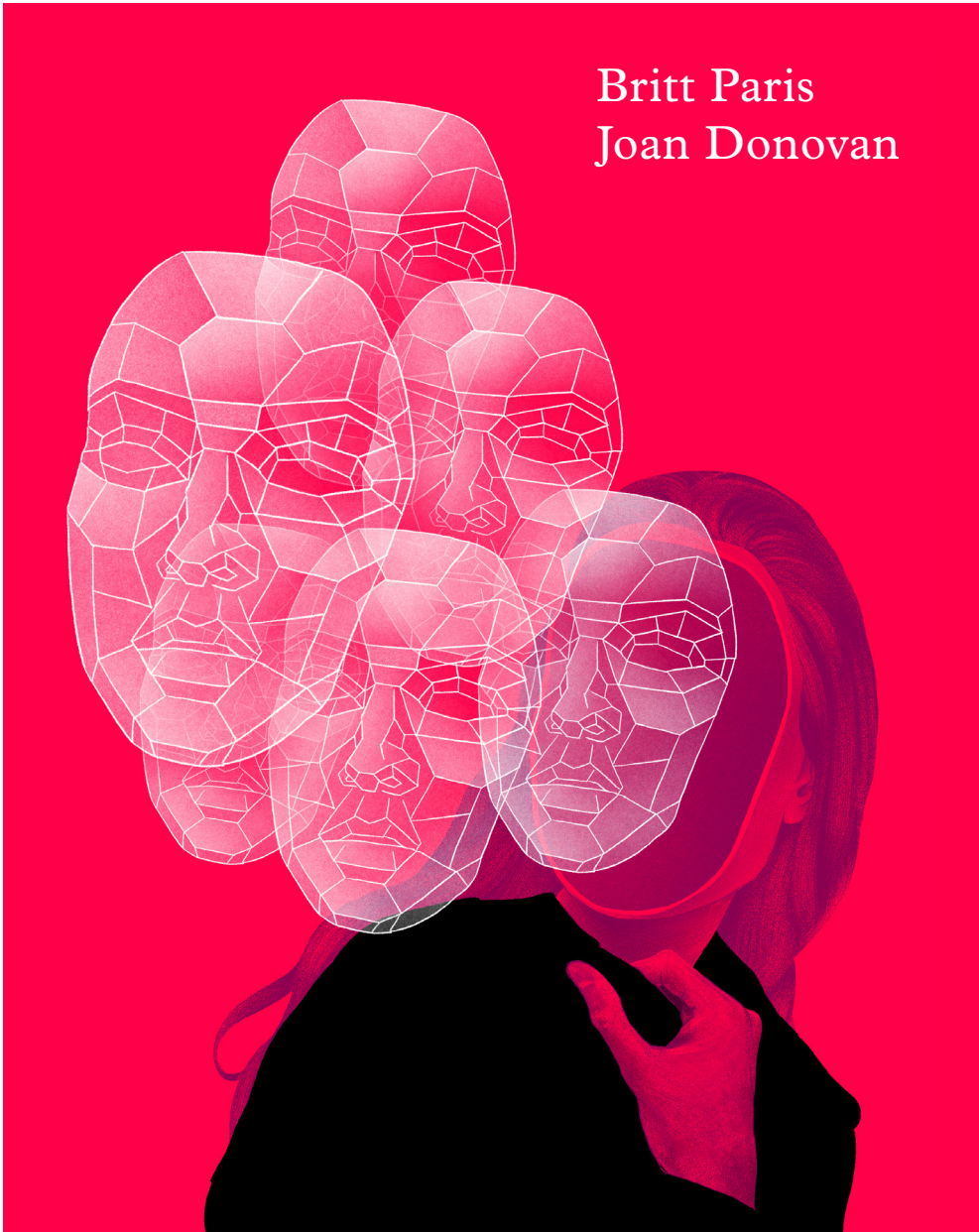
DATA&
SOCIETY

DEEPFAKES AND CHEAP FAKES

THE MANIPULATION OF AUDIO AND VISUAL EVIDENCE

Britt Paris
Joan Donovan



CONTENTS

Author: Britt Paris, assistant professor of Library and Information Science, Rutgers University; PhD, 2018,Information Studies, University of California, Los Angeles.

Author: Joan Donovan, director of the Technology and Social Change Research Project, Harvard Kennedy School; PhD, 2015, Sociology and Science Studies, University of California San Diego.

This report is published under Data & Society's Media Manipulation research initiative; for more information on the initiative, including focus areas, researchers, and funders, please visit https://datasociety.net/research/media-manipulation

EXECUTIVE SUMMARY

Do deepfakes signal an information apocalypse? Are they the end of evidence as we know it? **The answers to these questions require us to understand what is truly new about contemporary AV manipulation and what is simply an old struggle for power in a new guise.**

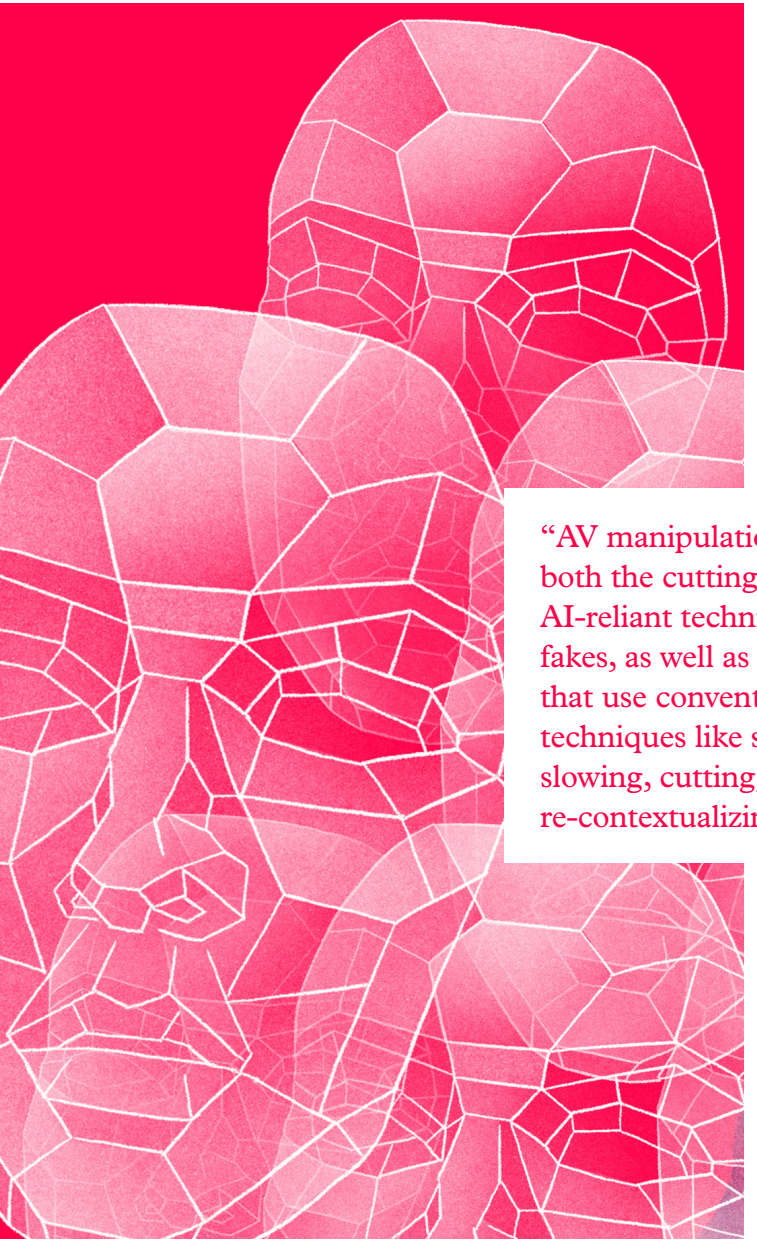
The first widely-known examples of amateur, AI-manipulated, face swap videos appeared in November 2017. Since then, the news media, and therefore the general public, have begun to use the term "deepfakes" to refer to this larger genre of videos—videos that use some form of deep or machine learning to hybridize or generate human bodies and faces. News coverage claims that deepfakes are poised to assault commonly-held standards of evidence, that they are the harbingers of a coming "information apocalypse." **But what coverage of this deepfake phenomenon often misses is that the "truth" of audiovisual content has never been stable—truth is socially, politically, and culturally determined.**

Deepfakes which rely on experimental machine learning represent one end of a spectrum of audio-visual AV manipulation. The deepfake process is both the most computationally-reliant and also the least publicly accessible means of creating deceptive media. **Other forms of AV manipulation – "cheap fakes" – rely on cheap, accessible software, or no software at all.** Both deepfakes

and cheap fakes are capable of blurring the line between expression and evidence. **Both can be used to influence the politics of evidence: how evidence changes and is changed by its existence in cultural, social, and political structures.**

Locating deepfakes and cheap fakes in the longer history of the politics of evidence allows us to see:

- decisions over what counts as "evidence" have historically been a crucial tool in defending the privilege of the already powerful;
- the violence of AV manipulation can only be addressed by a combination of technical and social solutions;
- public agency over these technologies cannot be realized without addressing structural inequality; and
- the violence of AV manipulation will not be curtailed unless those groups most vulnerable to that violence are able to influence public media systems.



"AV manipulation includes both the cutting edge, AI-reliant techniques of deep-fakes, as well as "cheap fakes" that use conventional techniques like speeding, slowing, cutting, re-staging, or re-contextualizing footage."

INTRODUCTION

In June 2019, artists Bill Posters and Daniel Howe posted a fake video of Mark Zuckerberg to Instagram.[1] Using a proprietary video dialogue replacement model (VDR) made by Canny, an Israeli advertising company, Posters and Howe produced a video of Zuckerberg talking about amassing power and control through a nefarious organization called Spectre. The video was created using a suite of machine learning and artificial intelligence techniques that enable the sophisticated manipulation of visual data, specifically the movement of bodies. While some are fascinated by the expressive possibilities of these technologies, others see dire consequences in the ability to put words and actions in the mouths and bodies of others. If video can no longer be trusted as proof that someone has done something, what happens to evidence, to truth?

The first widely known examples of amateur, AI-manipulated, face-swap videos appeared in November 2017, when a Reddit user with the username "deepfakes" uploaded a series of videos with the faces of famous female actors, including Gal Gadot and Scarlett Johansson, grafted onto other actors' bodies in pornography. Since then, the news media, and therefore the general public, have begun to use the term "deepfakes" to refer to this genre of videos that use some form of "deep" or machine learning to hybridize or generate human bodies and faces.

Deepfakes, however, are just one component of a larger field of audiovisual (AV) manipulation. AV manipulation includes any sociotechnical means for influencing the interpretation

1 https://www.instagram.com/p/ByaVigGFP2U/

of media. AV manipulation includes both the cutting edge, AI-reliant techniques of deepfakes, as well as "cheap fakes" that use conventional techniques like speeding, slowing, cutting, re-staging, or re-contextualizing footage.

How will deepfakes complicate the larger field of AV manipulation? Many of the journalists who cover deepfakes have declared them the harbingers of a coming "information apocalypse."[2] Journalists, politicians, and others allege deepfakes' ability to destroy democracy: to tamper with elections, compromise national security, or foment widespread violence.[3] News coverage claims that deepfakes are poised to destroy video's claim to truth by permanently blurring the line between evidentiary and expressive video. But what coverage of this deepfake phenomenon often misses is that the "truth" of AV content has never been stable—truth is socially, politically, and culturally determined. And people are able to manipulate truth with deepfakes and cheap fakes alike.

2 Charlie Warzel, "He Predicted the 2016 Fake News Crisis. Now He's Worried About An Information Apocalypse.," BuzzFeed News (blog), February 11, 2018, https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news; Franklin Foer, "The Era of Fake Video Begins," The Atlantic, April 8, 2018, https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/; Samantha Cole, "AI-Assisted Fake Porn Is Here and We're All Fucked," Motherboard (blog), December 11, 2017, https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn; Joshua Rothman, "In the Age of A.I., Is Seeing Still Believing?," The New Yorker, November 5, 2018, https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing; Jennifer Finney Boylan, "Will Deep-Fake Technology Destroy Democracy?," The New York Times, October 19, 2018, sec. Opinion, https://www.nytimes.com/2018/10/17/opinion/deep-fake-technology-democracy.html.

3 Warzel, "He Predicted the 2016 Fake News Crisis. Now He's Worried About An Information Apocalypse"; James Vincent, "US Lawmakers Say AI Deepfakes 'Have the Potential to Disrupt Every Facet of Our Society,'" The Verge, September 14, 2018, https://www.theverge.com/2018/9/14/17859188/ai-deepfakes-national-security-threat-lawmakers-letter-intelligence-community; Daniel Funke, "A Potential New Marketing Strategy for Political Campaigns: Deepfake Videos," Poynter, June 6, 2018, https://www.poynter.org/news/potential-new-marketing-strategy-political-campaigns-deepfake-videos; Boylan, "Will Deep-Fake Technology Destroy Democracy?"

We contextualize the phenomenon of deepfakes with the history of the politics of AV evidence to show that the ability to interpret the truth of evidence has been the work of institutions—journalism, the courts, the academy, museums, and other cultural organizations.[4] Every time a new AV medium has been decentralized, people have spread content at new speeds and scales; traditional controls over evidence have been upset until trusted knowledge institutions weighed in with a mode of dictating truth. In many cases, panic around a new medium has generated an opening for experts to gain juridical, economic, or discursive power.

However, there are two related phenomena that are truly new today. First, deepfakes are being applied to the digitization of bodies, including one's voice and likeness in increasingly routine ways. These are techniques that require training data of only a few hundred images. With thousands of images of many of us online, in the cloud, and on our devices, anyone with a public social media profile is fair game to be faked. And we are already seeing that the unbounded use of tools

4 Jean Baudrillard et al., In the Shadow of the Silent Majorities, trans. Paul Foss et al. 1979. (Los Angeles: Cambridge, Mass: Semiotext, 2007); Jean Baudrillard, Simulacra and Simulation. 1980, trans. Sheila Faria Glaser, 14th Printing edition (Ann Arbor: University of Michigan Press, 1994); John Fiske and Kevin Glynn, "Trials of the Postmodern," Cultural Studies 9, no. 3 (October 1, 1995): 505–21, https://doi.org/10.1080/09502389500490541; Ian Hacking, "Prussian Numbers 1860-1882," in The Probabilistic Revolution, Volume 1 (Cambridge, MA: MIT Press, 1987), 377–94; Herbert Schiller, Information Inequality, 1 edition (New York, NY: Routledge, 1995); Jean Baudrillard, The Gulf War Did Not Take Place. 1993. (Indiana University Press, 1995); Tal Golan, Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America (Cambridge, MA; London: Harvard University Press, 2007); Sarah E. Igo, The Averaged American: Surveys, Citizens, and the Making of a Mass Public (Cambridge, MA: Harvard University Press, 2008); Randall C. Jimerson, Archives Power: Memory, Accountability, and Social Justice (Society of American Archivists, 2009); Saloni Mathur, "Social Thought & Commentary: Museums Globalization," Anthropological Quarterly 78, no. 3 (2005): 697–708.trans. Sheila Faria Glaser, 14th Printing edition (Ann Arbor: University of Michigan Press, 1994

for AV manipulation is most likely to have a negative impact on women, people of color, and those questioning powerful systems.[5]

Second, AV manipulations of any type, deepfakes or cheap, can be transmitted at the speed and scale of today's online platforms. This increase in distribution provides a real challenge to traditional counters to manipulation and makes efforts like moderation or fact-checking harder to accomplish. Further, in some applications, like WhatsApp, encrypted messages are circulated along private connections,[6] achieving a type of hidden virality in which these fake videos can spread to a wide network of viewers while avoiding content moderation or mainstream media coverage.

Currently, technologists. policymakers, and journalists are responding to deepfakes with calls for what scholars call technical and legal closures[7]—that is, regulations, design features, and cultural norms that will determine the role of this technology. Venture capitalists, technologists, and

5 Jessie Daniels, Cyber Racism: White Supremacy Online and the New Attack on Civil Rights (Lanham, MD: Rowman & Littlefield Publishers, 2009); Safiya Umoja Noble, Algorithms of Oppression: How Search Engines Reinforce Racism, 1 edition (New York: NYU Press, 2018); Mary Anne Franks, "Unwilling Avatars: Idealism and Discrimination in Cyberspace," Columbia Journal of Gender and Law 20, no. (2011) (May 9, 2019), https://cjgl.cdrs.columbia.edu/article/unwilling-avatars-idealism-and-discrimination-in-cyberspace?article=unwilling-avatars-idealism-and-discrimination-in-cyberspace&post_type=article&name=unwilling-avatars-idealism-and-discrimination-in-cyberspace; Franks, "The Desert of the Unreal: Inequality in Virtual and Augmented Reality," U.C.D. L. Rev., January 1, 2017, 499; Danielle Citron, "Addressing Cyber Harassment: An Overview of Hate Crimes in Cyberspace," Journal of Law, Technology, & the Internet 6, no. 1 (January 1, 2015): 1.

6 Recently Facebook's Zuckerberg made a statement that Facebook's suite of platforms, of which WhatsApp is one, should follow WhatsApp's mode of encryption and private messaging.

7 Trevor J. Pinch and Wiebe E. Bijker, "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other," Social Studies of Science 14, no. 3 (1984): 399–441.

entrepreneurs in particular have called for new forms of technical redress, from the automated identification of fakes to the mandatory registration of content creators.[8] But such strategies do not address the thornier, and more important, issues of social, cultural, and historical context. There is a risk that these technical and legal closures will be directed by those who already hold economic and political power in ways that, following the political history of evidence, further consolidate truth for the status quo.

8 Antonio García Martínez, "The Blockchain Solution to Our Deepfake Problems," Wired, March 26, 2018, https://www.wired.com/story/the-blockchain-solution-to-our-deepfake-problems/; Siwei Lyu, "The Best Defense against Deepfake AI Might Be… Blinking," Fast Company, August 31, 2018, https://www.fastcompany.com/90230076/the-best-defense-against-deepfakes-ai-might-be-blinking; Tristan Greene, "Researchers Developed an AI to Detect DeepFakes," The Next Web, June 15, 2018, https://thenextweb.com/artificial-intelligence/2018/06/15/researchers-developed-an-ai-to-detect-deepfakes/; Facebook, "Expanding Fact-Checking to Photos and Videos," Facebook Newsroom, September 13, 2018, https://newsroom.fb.com/news/2018/09/expanding-fact-checking/; TruePic, "Photos and Videos You Can Trust," accessed December 10, 2018, https://truepic.com/about/; Sam Gregory and Eric French, "OSINT Digital Forensics," WITNESS Media Lab (blog), accessed June 20, 2019, https://lab.witness.org/projects/osint-digital-forensics/; Matt Turek, "Media Forensics," Defense Advanced Research Projects MediFor, 2018, https://www.darpa.mil/program/media-forensics.

THE DEEPFAKES/ CHEAP FAKES SPECTRUM

This spectrum charts specific examples of audiovisual (AV) manipulation that illustrate how deepfakes and cheap fakes differ in technical sophistication, barriers to entry, and techniques. From left to right, the technical sophistication of the production of fakes decreases, and the wider public's ability to produce fakes increases. Deepfakes—which rely on experimental machine learning—are at one end of this spectrum.

The deepfake process is both the most computationally reliant and also the least publicly accessible means of manipulating media. Other forms of AV manipulation rely on different software, some of which is cheap to run, free to download, and easy to use. Still other techniques rely on far simpler methods, like mislabeling footage or using lookalike stand-ins.

TECHNOLOGIES

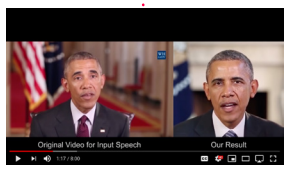
- **Recurrent Neural Network (RNN); Hidden Markov Models (HMM) and Long Short Term Memory Models (LTSM)**
- **Generative Adversarial Networks (GANs)**
- **Video Dialogue Replacement (VDR) model**
- **FakeApp / After Effects**
- **After Effects, Adobe Premiere Pro**
- **Sony Vegas Pro**
- **Free real-time filter applications**
- **Free speed alteration applications**
- **In-camera effects**
- **Relabeling/ Reuse of extant video**

DEEPFAKES More expertise and technical resources required — Less expertise and fewer technical resources required **CHEAP FAKES**

TECHNIQUES

- Virtual performances *(page 35)*
- Virtual performances
- Voice synthesis *(page 38)*
- Face swapping *(page 35)*
- Lip-synching *(page 38)*
- Face swapping: Rotoscope
- Speeding and slowing *(page 30)*
- Face altering/ swapping
- Speeding and slowing
- Lookalikes *(page 27)*
- Recontextualizing *(page 28)*

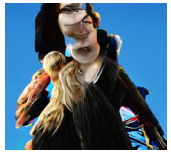
EXAMPLES



Suwajanajorn et al. Face2Face: Synthesizing Obama

Posters and Howe's: Mark Zuckerberg

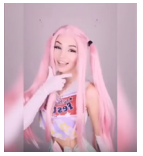
Deepfakes: Gal Gadot (not pictured because of image content)



Mario Klingemann: AI Art

Jordan Peele and BuzzFeed: Obama PSA

Paul Joseph Watson: Acosta Video



Belle Delphine: Hit or Miss Choreography

Rana Ayyub (not pictured because of image content)

Huw Parkinson: Uncivil War

SnapChat: Amsterdam Fashion Institute

Unknown: BBC NATO newscast

THE CHEAP FAKES-DEEPFAKES SPECTRUM

The upper end of the spectrum (*page 10*) contains experimental methods of computationally intensive image production, exclusive to a small set of graphics researchers and professional media producers. The deepfake process was initially developed by those working in the graphics processing industry.[9] NVIDIA, the market leader in the production of graphics processing units (GPUs),[10] has been at the forefront of "deep learning"—a type of machine learning that uses layers of algorithms called "neural networks" to sort through visual data to make predictions.[11] In recent years, these treatments have included weather prediction, cancer detection,

9 The graphical processing unit (GPUs) are found in chips in a computer's motherboard, with the CPU, or as a plug-in unit. It renders images, animations and video for the computer's screen. A GPU can perform parallel processing and pixel decoding much faster than a CPU. The primary benefit is that it decodes pixels to render smooth 3D animations and video.

10 Jensen Huang, "Accelerating AI with GPUs: A New Computing Model," The Official NVIDIA Blog, January 12, 2016, https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/; Moor Insights and Strategy, "A Machine Learning Landscape: Where AMD, Intel, NVIDIA, Qualcomm And Xilinx AI Engines Live," Forbes, accessed June 30, 2018, https://www.forbes.com/sites/moorinsights/2017/03/03/a-machine-learning-landscape-where-amd-intel-nvidia-qualcomm-and-xilinx-ai-engines-live/; "Intel, Nvidia Trade Shots Over AI, Deep Learning," eWEEK, accessed June 30, 2018, https://www.eweek.com/servers/intel-nvidia-trade-shots-over-ai-deep-learning.

11 Machine learning is a set of complicated algorithms used to train machines to perform a task with data, such as sorting a list of names in alphabetical order. The first few times, the lists would be mostly correct, but with a few ordering errors. A human would correct these errors and feed that data back into the system. As it sorts through different hypothetical lists of names, the machine does it faster with fewer errors. Machine learning requires a human for error correction. Deep learning, on the other hand, is a subset of machine learning that layers these algorithms, called a neural network, to correct one another. It still uses complicated algorithms to train machines to sort through lists of names, but instead of a human correcting errors, the neural network does it.

and responsive navigation systems for self-driving vehicles.[12] Computer scientists are also experimenting with ever more sophisticated strains of deep learning neural networks such as recurrent neural networks (RNNs) or generative adversarial networks (GANs), to turn audio[13] or audiovisual clips[14] into realistic, but completely fake, lip-synced videos.[15] With these techniques, media producers are able to create entirely virtual performances by recognizable figures, as in the cases of synthesizing various performances of Obama in prototypes produced by the University of Washington,[16] Stanford,[17] and at the State University of New York at Albany.[18] It is these spectacular, high-profile experiments

12 Economist Special Report, "From Not Working to Neural Networking," The Economist, June 25, 2016, https://www.economist.com/special-report/2016/06/25/from-not-working-to-neural-networking.; Jon Markman, "Deep Learning, Cloud Power Nvidia," Forbes, November 22, 2016, https://www.forbes.com/sites/jonmarkman/2016/11/22/deep-learning-cloud-power-nvidia-future-growth/; Mario Casu et al., "UWB Microwave Imaging for Breast Cancer Detection: Many-Core, GPU, or FPGA?," ACM Trans. Embed. Comput. Syst. 13, no. 3s (March 2014): 109:1–109:22, https://doi.org/10.1145/2530534.

13 Supasorn Suwajanakorn, Steven Seitz, and Ira Kemelmacher-Shlizerman, "Synthesizing Obama: Learning Lip Sync from Audio," in ACM Transactions on Graphics, vol. 36. 4, Article 95 (SIGGRAPH 2017, New York, NY: Association for Computing Machinery, 2017), http://grail.cs.washington.edu/projects/AudioToObama/.

14 Aayush Bansal et al., "Recycle-GAN: Unsupervised Video Retargeting," ArXiv: 1808.05174 [Cs], August 15, 2018, http://arxiv.org/abs/1808.05174.

15 In 2011, NVIDIA began partnering with deep learning computer scientists at leading research universities to develop image and pattern recognition technology. https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/

16 Suwajanakorn, et al., "Synthesizing Obama: Learning Lip Sync from Audio."

17 Justus Thies, Michael Zollhfer, Marc Stamminger, Christian Theobalt, and Matthias Neissner, "Face2Face: Real-Time Face Capture and Reenactment of RGB Videos," accessed June 27, 2019, https://cacm.acm.org/magazines/2019/1/233531-face2face/fulltext.

18 Gregory Barber, "Deepfakes Are Getting Better, But They're Still Easy to Spot," Wired, May 26, 2019, https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/.

that have driven much of the press coverage of deepfakes.[19]

Computer scientists are not the only ones with access to deepfake techniques. Consumer-grade animation software like Adobe After Effects enables anyone with average computing power and a little time to create similar audiovisual fakes through machine learning. Individual programmers have also created open source projects for deepfake creation—projects like FakeApp,[20] FaceSwap,[21] and DeepFace Lab.[22] Most of these are hosted in public repositories like GitHub and can produce output ranging from face-mapped performance videos to artificial lip-synchs.

Even without deep and machine learning, some video producers use a technique called digital rotoscoping to produce similar effects to deepfake face projection. This technique requires creators to manually outline target areas in every frame of a video and is therefore both difficult and time consuming. For instance, Australian artist Huw Parkinson used this technique to create an Avengers-themed political parody, manually mapping politicians' faces onto superheroes' bodies. Many other video producers manipulate media with simple methods of slowing, speeding, or cutting. In

19 Artists have also used deep learning methods to produce abstract images that sell for great amounts of money and exhibit in the most renowned galleries across the world, for example, new media artist Mario Klingeman's work shown in the typology above has shown at MoMA Le Centre Pompidou. Rama Allen, "AI Will Be the Art Movement of the 21st Century," Quartz, March 5, 2018, https://qz.com/1023493/ai-will-be-the-art-movement-of-the-21st-century/; "Artificial Intelligence and the Art of Mario Klingemann," Sothebys.com, February 8, 2019, https://www.sothebys.com/en/articles/artificial-intelligence-and-the-art-of-mario-klingemann.

20 "FakeApp 2.2.0 - Download for PC Free," Malavida, accessed June 1, 2018, https://www.malavida.com/en/soft/fakeapp/.

21 deepfakes, Deepfakes/Faceswap, Python, 2019, https://github.com/deepfakes/faceswap.

22 iperov, DeepFaceLab Is a Tool That Utilizes Machine Learning to Replace Faces in Videos. Includes Prebuilt Ready to Work Standalone Windows 7,8,10 Binary (Look Readme.Md), Iperov/DeepFaceLab, Python, 2019, https://github.com/iperov/DeepFaceLab.

"By using lookalike stand-ins, or relabeling footage of one event as another, media creators can easily manipulate an audience's interpretations."

2018 and 2019, respectively, videos of CNN reporter Jim Acosta and Speaker of the House of Representatives Nancy Pelosi were the subject of widespread media coverage; both of these had been altered by speeding or slowing using consumer editing software.

Countless free mobile apps now offer limited versions of the above techniques: speeding and slowing, but also forms of facial tracking and manipulation. Apps like SnapChat and TikTok offer dramatically lowered computational and expertise requirements that allow users to generate various manipulations in real time. SnapChat offers elements that mimic rotoscoping simply by clicking on filters. TikTok has a time filter that allows people to speed and slow video both as they capture it and afterward.

Finally, the most accessible forms of AV manipulation are not technical but contextual. By using lookalike stand-ins, or relabeling footage of one event as another, media creators can easily manipulate an audience's interpretations. One such example is a video emailed to Indian journalist Rana Ayyub as a form of blackmail. According to Ayyub, the manipulators had staged a pornographic video with a body double in an attempt to silence her critiques of the Indian

government.[23] Similarly accessible techniques were used in another Indian disinformation effort in 2018. Manipulators circulated a video via WhatsApp that showed footage of a nerve gas attack in Syria from 2013. By misidentifying the time and events of the video, manipulators used it to support false claims of child kidnapping and motivated violence in rural India.[24] These types of staging and re-contextualizing are possible for nearly anyone to reproduce, and technical forensic differences are even harder to detect, because there are no pixels out of order.[25]

Regardless of where a particular example rests on the spectrum of techniques, it is critical to attend to the ways in which these examples work as both expression and evidence. Both technically sophisticated and exceedingly simple techniques can be used in works of art or fiction. But problems arise when these techniques are used to create works that are interpreted as evidence. When that happens, despite the techniques or intentions behind production, manipulated media becomes a tool for changing or maintaining the distribution of power.

23 Rana Ayyub, "I Was the Victim of a Deepfake Porn Plot Intended to Silence Me," HuffPost UK, November 21, 2018, https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316; Rana Ayyub, "In India, Journalists Face Slut-Shaming and Rape Threats," The New York Times, May 22, 2018, https://www.nytimes.com/2018/05/22/opinion/india-journalists-slut-shaming-rape.html.

24 Timothy McLaughlin, "How WhatsApp Fuels Fake News and Violence in India," Wired, December 12, 2018, https://www.wired.com/story/how-whatsapp-fuels-fake-news-and-violence-in-india/.

25 Gregory and French, "OSINT Digital Forensics."

THE POLITICS OF EVIDENCE

In the past year, much news coverage has claimed that AI-generated video signals an incoming "information apocalypse."[26] For instance, the MIT Technology Review declared that "AI could set us back 100 years when it comes to how we consume news."[27] The Atlantic characterized the advent of AI-video production capacity as the "collapse of reality."[28] And The New Yorker questioned "In the age of A.I., is seeing still believing?"[29] Across these panicked summations of AV manipulation there is an underlying assumption: that video and photography, before our current moment, worked as objective, accurate evidence.

However, reviewing the history of visual evidence, it becomes clear that the relationship between media and truth has never been stable. There has always been a politics of evidence around audiovisual media—how evidence changes and is changed by its existence in cultural, social, and political structures.

The treatment of visual media as an objective documentation of truth is a 19th century legal construct. Science studies scholar Tal Golan[30] has documented how photographic

26 Warzel, "He Predicted the 2016 Fake News Crisis. Now He's Worried About an Information Apocalypse."

27 Jackie Snow, "AI Could Send Us Back 100 Years When It Comes to How We Consume News," MIT Technology Review, November 7, 2017, https://www.technologyreview.com/s/609358/ai-could-send-us-back-100-years-when-it-comes-to-how-we-consume-news/.

28 Foer, "The Era of Fake Video Begins."

29 Joshua Rothman, "In the Age of A.I., Is Seeing Still Believing?," The New Yorker, November 5, 2018, https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing.

30 Tal Golan, Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America (Cambridge, MA: Harvard University Press, 2007).

evidence has been made admissible in courts on a case-by-case basis since the 1850s. In the 19th century, witness testimony had long been the gold standard for courts, despite the fact that personal accounts were not always reliable.[31] Similarly, written historical records were also taken as fact in courts and by historians,[32] because they were seen as less mysterious than newer technologies, such as photography.[33]

> "There has always been a politics of evidence around audiovisual media—how evidence changes and is changed by its existence in cultural, social, and political structures."

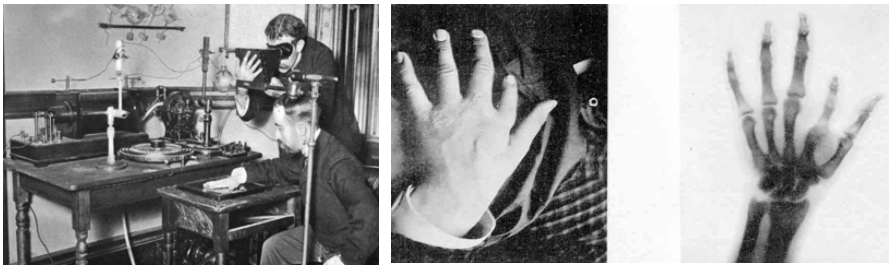
Golan describes that, as a result of the widespread distrust in image-capture technology, careful pains had to be taken to provide juries and judges with transparency into the process of image capture. For example, submitting an X-ray into evidence in court in the 1890s required the testimony from an expert technician to explain how a machine could see inside a body, that certain lines were evidence of a broken bone and not an artifact of developing the image. Op-ed columnists reacting to the introduction of X-rays

31 John Durham Peters, "Witnessing," Media, Culture & Society 23, no. 6 (November 1, 2001): 707–23, https://doi.org/10.1177/016344301023006002.

32 Sarah E. Igo, The Averaged American: Surveys, Citizens, and the Making of a Mass Public (Cambridge, MA: Harvard University Press, 2008); Matthew S. Hull, Government of Paper: The Materiality of Bureaucracy in Urban Pakistan, First edition (Berkeley: University of California Press, 2012); Durba Ghosh, "Optimism and Political History: A Perspective from India," Perspectives on History 49, no. 5 (2011): 25–27; Donald Francis McKenzie, Oral Culture, Literacy & Print in Early New Zealand: The Treaty of Waitangi (Victoria University Press, 1985); Harold Innis, Empire and Communications (Toronto, Canada: Dundurn Press Limited, 1948); Marshall McLuhan, The Gutenberg Galaxy, 1962. Centennial Edition (Toronto: University of Toronto Press, Scholarly Publishing Division, 2011); Walter J. Ong, Orality and Literacy, 1980. 30th Anniversary Edition (London ; New York: Routledge, 2012).

33 Golan, Laws of Men and Laws of Nature.

expressed fear that all doctors might have to become skilled X-ray technicians, to carry X-ray equipment with them on house visits.



At left, taking an X-ray image, late 1800s, published in the medical journal "Nouvelle Iconographie de la Salpetrière." Center, a hand deformity; at right, the same hand seen using X-ray technology. (Wikimedia Commons)

The use of expertise to determine what media is evidence of still happens today. Even simple, unsophisticated manipulations can have dire impact on how media is interpreted as evidence. The mutability of video in constructing legal evidence was clearly demonstrated in 1991, during the trial of four Los Angeles police officers for the beating of Rodney King. The trial hinged on a central piece of video evidence – captured by onlooker George Holliday – that shows King being brutally beaten by the group of officers.[34] During the defense, the officers' lawyers played the Holliday video, slowed to a fraction of its normal speed. As they did, they asked the officers, "Was he complying here with your order to stay down?..."[35] Slowed to a crawl, the video made King's involuntary physical reactions to blows appear as if he were attempting to get up.

34 Charles Goodwin, "Professional Vision," American Anthropologist, 1994.

35 John Fiske and Kevin Glynn, "Trials of the Postmodern," Cultural Studies 9, no. 3 (October 1, 1995): 505–21, https://doi.org/10.1080/09502389500490541.

The jurors decided the police officers were not guilty of using excessive force on King and said that the slow video in the courtroom "made all the difference."[36] The slowed video captured the events of March 3, 1991, but its manipulation allowed new interpretations, a new evidentiary function—one that maintained police sovereignty to brutalize members of the public.[37]

Visual evidence is not only interpreted in formal courtroom settings. Within journalism, visual evidence plays a key role in the construction of public opinion and the arrangement of political power. Journalists also serve as experts in the politics of evidence, deciding how to frame media as representative of truth to the public.

Media theorist Jean Baudrillard used the term "simulacra" to describe how broadcast journalism coverage of the First Gulf War defined how the West perceived the conflict – turning the daily news into around-the-clock reality TV.[38] Audiences tuned in to the spectacle of night-vision missile launches, in retaliation to less powerful but equally visually stunning Scud missiles launched by Iraqis, which provided evidence that a war was raging.[39] News media turned the Gulf War into fight between evenly matched opponents,

36 Brian Stonehill, "Real Justice Exists Only in Real Time: Video: Slow-Motion Images Should Be Barred from Trials.," Los Angeles Times, May 18, 1992, http://articles.latimes.com/1992-05-18/local/me-15_1_real-time.

37 Meredith D. Clark, Dorothy Bland, Jo Ann Livingston, "Lessons from #McKinney: Social Media and the Interactive Construction of Police Brutality," (McKinney, 2017), https://digital.library.unt.edu/ark:/67531/metadc991008/; Kimberlé Crenshaw, "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color," Stanford Law Review 43, no. 6 (July 1991): 1241–99. Both these pieces are foundational to the thinking of the first author in her work on how evidence is wielded by law enforcement and by groups pushing for police accountability and frame how evidence from official statistics to images online are used to reify structural inequality.

38 Jean Baudrillard, The Gulf War Did Not Take Place. 1993. (Indiana University Press, 1995).

39 Baudrillard, 42–45.

but did not mention or show the uneven damage and death toll: 20,000 dead Iraqi soldiers and an estimated 75,000 civilian casualties, compared with 286 deaths of American soldiers.[40] According to Baudrillard, US media mythologized the conflict through the stylized misrepresentation of events, justifying a colonization effort by a country with a strong and well-resourced military. [41] These images were real images. What was manipulative was how they were contextualized, interpreted, and broadcast around the clock on cable television.

Gulf War coverage on CNN in January 1991 featured infrared night time reporting of missile fire, heightening the urgency of US operations in the Persian Gulf. (Wikimedia Commons)

Simulacra describes how those who are already privileged in the shaping of social structures make media into evidence. From these examples, we see that evidentiary media's effect on society is not only produced by the technical specifics of its ability to capture the world. Instead, the

40 Baudrillard, 71-73; Robert Fisk, The Great War for Civilisation: The Conquest of the Middle East, Reprint edition (New York: Vintage, 2007)."DCAS Reports - Persian Gulf War Casualties - Desert Storm," accessed April 9, 2019, https://dcas.dmdc.osd.mil/dcas/pages/report_gulf_storm.xhtml.

41 Baudrillard, 78–83.

expert interpreter's expression – their creation of a frame – shapes the evidence and its effect on society.[42] When that expert interpreter works in service of the powerful, power is consolidated, regardless of what "reality" is captured in media.

This history shows that evidence does not speak for itself. That is, it is not simply the representational fidelity of media that causes it to function as evidence. Instead, media requires social work for it to be considered as evidence. It requires people to be designated as expert interpreters. It requires explicit negotiations around who has the status to interpret media as evidence. And because evidence serves such a large role in society – it justifies incarcerations, wars, and laws – economically, politically, and socially powerful actors are invested in controlling those expert interpretations. Put another way, new media technologies do not inherently change how evidence works in society. What they do is provide new opportunities for the negotiation of expertise, and therefore power.

It is critical to understand the current panic around deepfakes and the drive toward technical solutions as part of this historical trajectory. While many understand the evidence they commonly use to inform their daily decisions as neutral, it has never been so. Supposed evidentiary neutrality has often worked to uphold and justify the privilege of those who benefit from social inequality. Moving forward, it is important for us to question whether those who claim or construct expertise around deepfakes will do so in a way that reinforces economic, political, and social privilege.

42 Pinch and Bijker, "The Social Construction of Facts and Artefacts."

"When that expert interpreter works in service of the powerful, power is consolidated, regardless of what "reality" is captured in media."

CHEAP FAKES ON SOCIAL MEDIA

Today, social media is experiencing a sped-up version of the cycles of hype, panic, and closure that still and moving images went through in the last 200 years. The camera function became a competitive feature of mobile phones in the mid-2000s.[43] At the same time, social media platforms, like Facebook, began amassing popularity. The combination of mobile cameras and social media radically expanded the possibilities of who could distribute photographic media, to whom, and at what speeds.

Cultural scholar Henry Jenkins has responded to one aspect of the activity around social media by identifying what he calls "Photoshop for democracy." He analyzes a number of online communities that used image manipulation software and social media to make "participatory culture [become] participatory government."[44] Jenkins illustrates this with a set of images that criticize political figures. He refers to an image with faces of Howard Dean, John Kerry, and John

43 Nathan Jurgenson, The Social Photo: On Photography and Social Media (New York: Verso, 2019).

44 Henry Jenkins, "Photoshop for Democracy," MIT Technology Review, June 4, 2004, https://www.technologyreview.com/s/402820/photoshop-for-democracy/.

Edwards grafted onto the Three Stooges. Another depicts the Stooges as George W. Bush, Colin Powell, and Donald Rumsfeld. Jenkins argues these uses of Photoshop are a meaningful measure taken by the public to hold elected officials accountable. Indeed, Jenkins put into language an important component of the history of evidence we have constructed thus far: While the already-dominant shapers of discourse centralize power through expert interpretation, members of the public can redistribute that power when they have the ability to spread expressive messages at larger scales, with less expert oversight than ever before.

But while Jenkins helps us understand what happens to the horizon of political expression with increased scale and speed, evidentiary media carries a different valence. Clearly, while manipulated, the "Stooges" images make no evidentiary claim. They are, like political cartoons or protest signs, an expression of political beliefs. Therefore, to Jenkins' excitement for participatory democracy, we must add that Photoshop and other image manipulation tools can have real consequences for individuals when expression is not clearly distinguishable from evidence.

The remainder of this section discusses techniques of contemporary cheap fakes: photoshopping, lookalikes, recontextualization, and speeding and slowing moving images. Each technical tool set described was previously only available to experts, but in the context of technological advancement and widespread social media use, these technical tool sets are more accessible to amateurs and their outputs reach larger scales at higher speeds. These examples show how the structural power of manipulators and subjects determine how media can be read as evidence that justifies physical, sexual, and political violence.

Photoshopping

Photoshop and similar image manipulation software can be used to manipulate photographic media in countless ways, but the stakes of such manipulations are most clear in cases of faked pornography. People have manipulated pornographic photos and films since the very inventions of those technologies.[45] The distribution of faked pornographic images has become more widespread since mid-2000s[46] as Photoshop and related digital editing tools became more accessible, and aggregated websites dedicated to pornography became more common. Today, pornographic cheap fakes are often played off as expression even as they work as evidence in image-based violence against women and other vulnerable groups.[47] Feminist legal scholars Claire McGlynn, Erica Rackley, and Ruth Hoffman refer to the act of superimposing images of an individual's head or body part into pornographic content, without their consent, to make it appear as if an individual is engaged in sexual activity, as "sexualized photoshopping."[48] McGlynn et al. call this type of gendered, sexualized abuse of images "image-based sexual abuse." The authors suggest that if a photo is originally shared publicly, it becomes a private image when it has been edited to depict the person in the image in a sexualized way. When the image is manipulated, the individual

45 Richard Abel, Encyclopedia of Early Cinema (Taylor & Francis, 2004); John Pultz, The Body and the Lens: Photography 1839 to the Present, First edition (New York: Harry N Abrams Inc., 1995).

46 Clare McGlynn, Erika Rackley, and Ruth Houghton, "Beyond 'Revenge Porn': The Continuum of Image-Based Sexual Abuse," Feminist Legal Studies 25, no. 1 (April 1, 2017): 25–46, https://doi.org/10.1007/s10691-017-9343-2.

47 McGlynn, Rackley, and Houghton, "Beyond 'Revenge Porn'"; Danielle Keats Citron and Mary Anne Franks, "Criminalizing Revenge Porn," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, May 19, 2014), https://papers.ssrn.com/abstract=2368946.

48 McGlynn, Rackley, and Houghton, "Beyond 'Revenge Porn,'" 33.

pictured loses agency over their own likeness and the "portrayal of their sexual self."[49]

For instance, in 2016 a 17-year-old Australian woman, Noelle Martin, found her face photoshopped onto pornographic images circulating on the internet, which caused her anguish over diminished job prospects.[50] Martin is hardly alone. Countless examples of image-based violence online link pornographic expression to existing power relations in society.[51] The internet at once facilitates access to pornographic materials made with and without consent, and provides multiple unchecked and unregulated environs that "normalize misogyny, hurtful sexuality, racism, and even seeking revenge on female ex-partners."[52] The debate and legislation over this phenomenon continues today and sets the background for most legal discussion around deepfake pornography.

Lookalikes

Image-based sexual violence can be used to inflict other types of harm, such as the suppression of the press, civil society,

49 McGlynn, Rackley, and Houghton, "Beyond 'Revenge Porn,'" 33.

50 "Noelle Martin Fights to Have Harmless Selfie Removed from 'parasite' Porn Sites - ABC News (Australian Broadcasting Corporation)," accessed February 18, 2019, https://www.abc.net.au/news/2016-10-12/womans-fight-to-have-harmless-selfie-removed-from-porn-site/7924948.

51 McGlynn, Rackley, and Houghton, "Beyond 'Revenge Porn'"; Citron and Franks, "Criminalizing Revenge Porn"; Feona Attwood, "No Money Shot? Commerce, Pornography and New Sex Taste Cultures," Sexualities 10, no. 4 (October 1, 2007): 441–56, https://doi.org/10.1177/1363460707080982.

52 Walter S. DeKeseredy and Martin D. Schwartz, "Thinking Sociologically About Image-Based Sexual Abuse: The Contribution of Male Peer Support Theory," Sexualization, Media, & Society 2, no. 4 (October 20, 2016): 237462381668469, https://doi.org/10.1177/2374623816684692.p 6.

or political opposition. Many of the most public examples of this have been directed toward female politicians or activists, often simply using similar-looking actors to depict them in sexualized or pornographic footage. This strategy is a type of in-camera editing that requires no technical skill, simply the desire and ability to find a lookalike for the target.

In the Philippines in 2016, the legal counsel of President Rodrigo Duterte used a faked sex video of Senator Leila De Lima to justify her imprisonment, in seeming retribution for her critique of his authoritarian rule.[53] Similar tactics have been used with female journalists who call out abuses of power as in the case of Rana Ayyub, mentioned earlier, who critiqued the nationalist Bharatiya Janata Party (BJP) that ran the Indian government in Gujarat responsible for extrajudicial murders following the Gujarat riots in 2002.[54] Since her book was originally published in 2010, and as her reporting continues to follow abuses of power, she has experienced ongoing harassment. In 2018, she discovered that a pornographic video of "herself" circulated through WhatsApp had been posted onto a BJP fan page and circulated widely in order to ruin her credibility and force her into silence.[55] In reality, the pornographic fake simply featured an actor who resembled her.

53 "De Lima on Sex Video: It Is Not Me," philstar.com, accessed March 29, 2019, https://www.philstar.com/headlines/2016/10/06/1630927/de-lima-sex-video-it-not-me.

54 Rana Ayyub, Gujarat Files: Anatomy of a Cover Up (India: CreateSpace Independent Publishing Platform, 2016).

55 Ayyub, "I Was The Victim of a Deepfake Porn Plot Intended To Silence Me."

Re-contextualizing

Perhaps the easiest mode of producing a cheap fake is simply cutting together existing footage and spreading it under false pretenses. For example, in April 2018, a falsified BBC report began circulating on WhatsApp, presenting a false story of nuclear escalation between NATO and Russia.[56] The shared clip was four minutes long, and featured nuclear mushroom clouds, the Queen's evacuation from Buckingham Palace, a Russian ship firing missiles, and NATO aircraft being launched in retaliation. As the clip spread, alarmed viewers began sending messages to the BBC, which issued a statement[57] explaining that the clip was cut together from YouTube footage uploaded in 2016 by the Benchmarking Assessment Group, an Irish marketing company. The original footage was a 30-minute video designed as a psychometric test for reactions during a disaster scenario. Outside of the focus group, the video was never intended to be taken seriously. But, a still-unknown manipulator was able to re-contextualize and disseminate it through WhatsApp as an imminent threat.

WhatsApp's secure encryption model gives the service an air of authenticity—if one is only receiving information from verified contacts, then everything should be trustworthy. This feeling of authenticity is precisely what gets exploited by manipulators. Once a video like BBC NATO does make

56 Martin Coulter, "BBC Issues Warning after Fake News Clips Claiming NATO and Russia at War Spread through Africa and Asia," London Evening Standard, accessed April 25, 2018, https://www.standard.co.uk/news/uk/bbc-issues-warning-after-fake-news-clips-claiming-nato-and-russia-at-war-spread-through-africa-and-a3818466.html; "BBC Forced to Deny Outbreak of Nuclear War after Fake News Clip Goes Viral," The Independent, April 20, 2018, https://www.independent.co.uk/news/media/bbc-forced-deny-fake-news-nuclear-war-viral-video-russia-nato-a8313896.html.

57 Chris Bell, "No, the BBC Is Not Reporting the End of the World," April 19, 2018, https://www.bbc.com/news/blogs-trending-43822718.

"Platforms are not equipped to handle fake videos even when these videos are cheap fakes—not generated with sophisticated artificial intelligence."

it onto a node of one of these hidden social networks, new recipients can feel a political, emotional, or social duty to share information received from their family or friend groups, regardless what judgment they may have about the factual content of the information.[58]

This and other more prominent recent examples where WhatsApp has been implicated in spreading disinformation illustrate that in the contemporary climate, platforms struggle to moderate problematic content traveling at large scales.[59] Platforms are not equipped to handle fake videos even when these videos are cheap fakes—not generated with sophisticated artificial intelligence. One method platforms have used to attempt to address this is to create encrypted social media networks like WhatsApp that allow sharing

58 Daniel Kreiss, "The Fragmenting of the Civil Sphere: How Partisan Identity Shapes the Moral Evaluation of Candidates and Epistemology," American Journal of Cultural Sociology 5, no. 3 (October 2017): 443–59, https://doi.org/10.1057/s41290-017-0039-5; Santanu Chakrabarti, Lucile Stengel, and Sapna Solanki, "Fake News and the Ordinary Citizen in India," Beyond Fake News (London: British Broadcasting Corp., November 12, 2018), http://downloads.bbc.co.uk/mediacentre/duty-identity-credibility.pdf.

59 Jason Koebler, Derek Mead, and Joseph Cox, "Here's How Facebook Is Trying to Moderate Its Two Billion Users," Motherboard (blog), August 23, 2018, https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works; Sarah T. Roberts, Behind the Screen, Yale University Press (New Haven, CT: Yale University Press, 2019), https://yalebooks.yale.edu/book/9780300235883/behind-screen.

between small groups.[60] They claim this will promote authentic communication between trusted parties, and thus this technical solution would discourage the spread of disinformation. However, this has also enabled disinformation to enjoy the cover of hidden virality in which trusted parties in private networks link quickly to wider networks of users while remaining impenetrable to outside oversight.

Speeding and Slowing

Recalling how the video of Rodney King's beating was slowed in court to attempt to sow doubt in jurors' minds, we can see that the speed of a video has a large effect on its interpretation. We were recently reminded of this in May 2019, when a faked video appeared to depict House Speaker Nancy Pelosi drunkenly talking about Donald Trump[61] caused waves of public outrage. It was amplified by Trump and other prominent pundits.[62] Placing the "drunk Pelosi" clip side by side with the original C-SPAN footage[63] make the manipulation clear: a roughly 75% decrease in speed, which could have easily been accomplished with any consumer video editing program, from Adobe Premiere to iMovie.

60 Mark Zuckerberg, "A Privacy-Focused Vision for Social Networking," Facebook (blog), March 6, 2019, https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/.

61 The Guardian, "Real v Fake: Debunking the 'drunk' Nancy Pelosi Footage," The Guardian, May 24, 2019, https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video.

62 Contemptor, Team Trump Goes All In on Pushing Doctored "Drunk" Pelosi Videos, accessed June 7, 2019, https://www.youtube.com/watch?time_continue=43&v=FGBc3PJ4OoM.

63 The New York Times, "Edited Pelosi Video vs. the Original: A Side-by-Side Comparison," The New York Times, May 24, 2019, https://www.nytimes.com/video/us/politics/100000006525055/pelosi-video-doctored.html.

Not so long ago, on November 7, 2018, another widely publicized struggle over evidentiary media came from the social media account of the White House press secretary. Following a heated exchange between Donald Trump and CNN reporter Jim Acosta, the White House announced that it would be revoking Acosta's press credentials after he struck an intern while reaching for a microphone. Along with the announcement, White House press secretary Sarah Sanders posted an altered version of the original C-SPAN footage that showed the interaction between Acosta and the intern.[64]

Within hours social media users identified the edited version of the Acosta video[65] as one created by Paul Joseph Watson,[66] editor and writer for conspiracy site Infowars. The video Watson produced was taken from the original C-SPAN footage, but zoomed in on the moment of contact, apparently spliced in with a low-frame rate GIF.[67] To many viewers and experts, the resulting interaction between Acosta

64 Sarah Sanders, "We Stand by Our Decision to Revoke This Individual's Hard Pass. We Will Not Tolerate the Inappropriate Behavior Clearly Documented in This Video," Twitter, accessed November 8, 2019, https://twitter.com/PressSec/status/1060374680991883265.

65 Mike Stuchbery, "Oh Lord, in His Fevered Attempts to Claim He Didn't Edit the Acosta Video, @PrisonPlanet Posted a Picture of His Vegas Timeline — That Shows He Added Frames...," Twitter, accessed November 8, 2019, https://twitter.com/MikeStuchbery_/status/1060535993525264384; Rafael Shimunov, "1) Took @PressSec Sarah Sanders'' Video of Briefing 2) Tinted Red and Made Transparent over CSPAN Video 3) Red Motion Is When They Doctored Video Speed 4) Sped up to Make Jim Acosta's Motion Look like a Chop 5) I've Edited Video for 15+ Years 6) The White House Doctored It," Twitter, accessed November 8, 2019, https://twitter.com/rafaelshimunov/status/1060450557817708544.

66 Paul Joseph Watson, "Sarah Sanders Claims Jim Acosta 'Placed His Hands on a Woman.' Jim Acosta Says This Is a 'Lie'. Here's the Video. You Be the Judge," Twitter, accessed November 8, 2019, https://twitter.com/PrisonPlanet/status/1060344443616800768.

67 Paris Martineau, "How an InfoWars Video Became a White House Tweet," Wired, November 8, 2018, https://www.wired.com/story/infowars-video-white-house-cnn-jim-acosta-tweet/.

and the intern appeared more forceful in the Watson/Sanders version.

All of the examples above – whether photoshopping, stand-ins, or conventional editing – show that politically dangerous AV manipulation in the form of cheap fakes already exists. Society is already struggling to mitigate the harm of "traditional" techniques: changes in speed, cutting apart or together, circulating in new, vulnerable contexts. It is into this landscape that so-called deepfakes emerge.

Deepfakes Present and Future

Much of the current coverage of deepfakes strikes the same tone as Franklin Foer's assessment of the phenomenon in early 2018: "We'll shortly live in a world where our eyes routinely deceive us … we're not so far from the collapse of reality."[68] While video has always been susceptible to manipulation, deepfakes have produced a refreshed panic. But it is important to not let this general panic shape the discourse around AV manipulation. Instead, we should pay close attention to the full range of such manipulations, to understand how these technologies are actually being interpreted and acted upon.

The range of these newly widespread, technically advanced modes of AV manipulation include virtual performances, face swapping, and lip-synching or voice synthesis. Some instances may include all of these techniques, and others may only use one. While at present amateurs may certainly participate in the production of video content, the technical

68 Foer, "The Era of Fake Video Begins."

barriers to entry are still high, as is the expertise needed to produce the faked videos.

During our survey of audiovisual manipulation online,[69] and background interviews with producers of these videos,[70] we found evidence of a range of communities of practice[71] represented among those producing and discussing these deepfake videos. There are professional effects artists who use sophisticated and expensive equipment to generate material for Hollywood movies and other big budget media spectacles. Then there is a loose collection of hobbyist and tech enthusiasts who organize online to reproduce these expensive AI techniques with free and consumer-grade technology. Finally, a subset of this non-professional group produces and consumes problematic content—the most prominent of which is manipulated pornography.

The lines between these communities, however, are fuzzy at best. And online, a wide variety of media labeled as AI-generated is not actually made with AI but is presented alongside other videos that are in YouTube playlists or Reddit threads. In our interviews with producers across the range of communities of practice, we identified a driving,

69 From January 2018 to June 2019, we collected and analyzed over 70 examples of deepfake videos and images from YouTube, Instagram, Twitter, TikTok, PornHub, Arxiv, and Mr. Deepfakes.

70 We completed eleven semi-structured interviews between September and December 2018.

71 This term is drawn from sociological and anthropological literature on knowledge production refer to groups of individuals organize both on and off-line to experiment, learn, and professionalize around practice-based activities. Jean Lave and Etienne Wenger, Situated Learning: Legitimate Peripheral Participation (Cambridge University Press, 1991); Line Dubé, Anne Bourhis, and Réal Jacob, "The Impact of Structuring Characteristics on the Launching of Virtual Communities of Practice," Journal of Organizational Change Management, April 1, 2005, https://doi.org/10.1108/09534810510589570.

optimistic belief that technological advancement is necessarily good,[72] which minimizes the potential harm brought by new technologies. What remains important to articulate is where to draw a line between experimental, expressive play and new forms of image-based abuse, whether sexual, political, or otherwise.

Virtual Performances

Initially, the public reception of AI-generated manipulations was shaped by a small number of prominent examples in major Hollywood films. Before deepfakes, Hollywood effects artists were using computer-generated imagery that faked performances. For instance, Tom Hanks' titular character Forrest Gump interacting with long-dead individuals by inserting Gump into already existing footage using digital compositing.

Recent Hollywood movies are able to create even more ambitious virtual performances, not just combining two sources of footage, but generating entirely new representations of face, voice, and body. Most notable among these are recent films in the Star Wars franchise. The effects artists at Industrial Light and Magic created a younger version of actor Carrie Fischer for 2017's Rogue One and a posthumous version of her for 2018's The Last Jedi. Indeed, even in amateur circles, Star Wars content is particularly popular. A popular Reddit user with the name "derpfakes," also created their own version of Fischer's scenes in Rogue

72 This observation from our interviews is backed up in interviews from news coverage, for example, in Megan Farokhmanesh, "Is It Legal to Swap Someone's Face into Porn without Consent?" The Verge, January 30, 2018, https://www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal.

One – highlighting just how close free and open source tools can get to the "expensive and labor intensive" ones used by Hollywood.[73]

Face Swapping

Much of the activity of deepfake-focused communities has involved specific technologies for face swapping. Original software like FakeApp, FaceSwap, and DeepFace Lab allow users to map one actor's face onto another actor's performance. Several software users have used this technique to reproduce or comment on various Hollywood versions, such as an October 2018 video that mapped a young Harrison Ford's face onto Alden Ehrenreich's body in Solo: A Star Wars Story.

The most publicly visible examples of open source face swapping, however, have been a series of videos made by projecting female actors' faces onto the bodies of performers in pornography, using software like FakeApp and Adobe After Effects. In November of 2017, for instance, one user uploaded such a video using Gal Gadot's face to Reddit. In April 2018, pornographic deepfakes were banned from Reddit, but the videos remain available on sites like Mr. Deepfakes and Pornhub. In an interview with Motherboard, the anonymous creator of the Gal Gadot video said, "Every technology can be used with bad motivations, and it's impossible to stop that.... The main difference is how easy

73 Insight gleaned from exchange with derpfakes on November 14, 2018, and derpfake video comments on YouTube and Reddit.

[it is] to do that by everyone. I don't think it's a bad thing for more average people [to] engage in machine learning research."[74] Another Redditor posted, "[T]he work that we create here in this community is not with malicious intent. Quite the opposite. We are painting with revolutionary, experimental technology, one that could quite possibly shape the future of media and creative design."[75] In this example, this new "revolutionary technology" is firmly situated in a misogynist environment in which its practitioners mobilize the bodily identity of women, without their consent, as an unproblematic creative practice.[76]

Indeed, it is not a bad thing for people to engage in machine learning research. It is not inherently wrong to make, circulate, and consume pornography. However, problems arise when experimenting with machine learning is used in the service of non-consensual objectification and harassment. Regardless of the intention of those who create videos, the effects wrought by technological play matters. AI-generated pornographic videos highlight a set of interconnected

74 Samantha Cole, "AI-Assisted Fake Porn Is Here and We're All Fucked," Motherboard, December 11, 2017, https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn.

75 Quote from Reddit user Quiet Horse, now deleted from Reddit, in Megan Farokhmanesh, "Is It Legal to Swap Someone's Face into Porn without Consent?," The Verge, January 30, 2018, https://www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal.

76 McGlynn, Rackley, and Houghton, "Beyond 'Revenge Porn'"; Franks, "Unwilling Avatars"; Citron, "Addressing Cyber Harassment"; Citron and Franks, "Criminalizing Revenge Porn." While this is generally true in online porn communities, pornographic practice is by no means monolithic. The production of pornography by and for women, LGBTIQ individuals, and people of color highlights that these groups have agency in producing and enjoying their own counter-hegemonic pornographic content. The higher degree of visibility and access to sexual culture for these groups might be described as liberatory, or more radically democratic. In this vein, Feona Attwood suggests that altporn in the form of SuicideGirls and Nerve.com represent "participatory cultures which serve corporate and community needs" which can be understood as radically democratic. (Feona Attwood, "No Money Shot? Commerce, Pornography and New Sex Taste Cultures," Sexualities 10, no. 4 (October 1, 2007): 441–56, https://doi.org/10.1177/1363460707080982.)

"Regardless of the intention of those who create videos, the effects wrought by technological play matters."

problems regarding the "democratization" of a technical skill set. That the new technology's first notable amateur output depicts one woman engaged in a sex act without her consent, while also erasing the original female actor in the original video highlights how the expressive use of technology can be wielded as harassment or harm individuals who do not command public attention, and cannot command the resources necessary to refute falsehoods.

Lip-synching and Voice Synthesis

Another strand of deepfake production focuses not on dynamic virtual performances of face swapping but on the presentation of new speech content for recognizable figures. In November 2017, Stanford researchers published a paper and model for Face 2 Face,[77] a RNN-based video production model that allows third parties the ability to "put words in the mouths" of public figures like Barack Obama – in real time. The software captures the third party's facial expressions as they talk into a webcam and then melds those movements with the person's face in the original video.

77 Thies et al, "Face2Face."

In February 2018, Jordan Peele teamed with BuzzFeed to produce a "deepfake" video in which Obama appears to provide a public service announcement about the dangers of the new technology.[78] Like the derpfakes example mentioned above, the video was produced using Adobe After Effects, FakeApp, and a team of creative professionals to swap in Peele's voice as he spoke in a voice that many know as his Obama impression.[79] It took hours for the FakeApp application, trained by a human, to render a video in which Obama's mouth moves along with the Peele's voice.

Versions of these techniques are part of the proprietary software that was used to create the Zuckerberg fake from our introduction.[80] Their bespoke video-faking tool used audio and video files from Zuckerberg's Senate testimony from April 2018 to create a new virtual performance. This proprietary model was in some senses more sophisticated than the GAN or RNN-generated video production generated by computer science labs, as it incorporated more audio and video training data. The artists who built it maintain that the VDR model but was built on the work featured at 2017 SIGGRAPH[81] conference coming from a computer science lab at the University of Washington on a project led by Supasorn Suwajanakorn and the Face2Face tool.[82]

78 Craig Silverman, "How To Spot a DeepFake Like The Barack Obama-Jordan Peele Video," accessed April 18, 2018, https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed.

79 James Vincent, "Watch Jordan Peele Use AI to Make Barack Obama Deliver a PSA about Fake News," The Verge, April 17, 2018, https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed.

80 Cade Metz, "A Fake Zuckerberg Video Challenges Facebook's Rules," The New York Times, June 11, 2019, sec. Technology, https://www.nytimes.com/2019/06/11/technology/fake-zuckerberg-video-facebook.html.

81 The Association for Computing Machinery is a professional organization that supports SIGGRAPH, a special interest group on computer graphics and interactivity.

82 "Canny AI: Imagine World Leaders Singing" Fxguide, accessed June 24, 2019, https://www.fxguide.com/featured/canny-ai-imagine-world-leaders-singing/.

"But as these techniques become more accessible, the risk of harm to private figures will increase, especially those who are politically, socially, or economically vulnerable."

CONCLUSION

The problems wrought by audiovisual manipulation are many and difficult to remedy, but we can work to reduce risk and mitigate harm. This report has shown that individual women, journalists, and others who are antagonistic to those who hold economic and political power are going to be the first to confront the politics of evidence in a "post-truth" world. In these scenarios, questions of evidence are key: Who should we trust, and on what basis?

The high-profile deepfakes cases discussed above are just a fraction of the examples that exist. And more are being created every day. So far, news coverage and public discussion has paid the most attention to those cases that involve the imitation – whether virtual performance, face swap, or voice synthesis – of celebrities, politicians, and public figures. But as these techniques become more accessible, the risk of harm to private figures will increase, especially those who are politically, socially, or economically vulnerable.

At present, anyone with a social media profile is fair game to be faked. Social media relies on forms of soft consent, buried in the terms of service that users habitually click through. That soft consent is now being used to justify the collection of data far beyond web surfing or ad clicking habits.

It involves data that captures our faces and bodies[83]—a condition that demands greater transparency and true consent.[84] Social media functions as vast troves of images of faces and bodies, and amateur programmers have already built bespoke scrapers to extract these images as training data.[85]

There already have been numerous cases of AV manipulation techniques used against private individuals: to settle personal vendettas,[86] exact blackmail,[87] and trick people into participating in personalized financial scams.[88] People are already faking their classmates and co-workers in amateur AI-generated

83 Hamza Shaban, "A Google App That Matches Your Face to Artwork Is Wildly Popular. It's Also Raising Privacy Concerns.," Washington Post. January 17, 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/01/16/google-app-that-matches-your-face-to-artwork-is-wildly-popular-its-also-raising-privacy-concerns/; Cole and Maiberg, "People Are Using AI to Create Fake Porn of Their Friends and Classmates"; Natasha Singer, "Facebook's Push for Facial Recognition Prompts Privacy Alarms," The New York Times, July 11, 2018, sec. Technology, https://www.nytimes.com/2018/07/09/technology/facebook-facial-recognition-privacy.html.

84 Pavel Korshunov and Sebastien Marcel, "DeepFakes: A New Threat to Face Recognition? Assessment and Detection," ArXiv:1812.08685 [Cs], December 20, 2018, http://arxiv.org/abs/1812.08685; Owen Hughes, "Is FaceApp Safe? Don't Be so Quick to Share Your Face Online, Warn Security Experts," International Business Times UK, April 27, 2017, https://www.ibtimes.co.uk/faceapp-safe-dont-be-so-quick-share-your-face-online-warn-security-experts-1618975; Robert Chesney and Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, July 14, 2018), https://papers.ssrn.com/abstract=3213954; M. C. Elish and danah boyd, "Situating Methods in the Magic of Big Data and AI," Communication Monographs 85, no. 1 (January 2, 2018): 57–80, https://doi.org/10.1080/03637751.2017.1375130.

85 Instagram Scraper and the Chrome extension DownAlbum are among many open source tools that allow anyone to easily download images from publicly available Facebook or Instagram accounts to do many things, including to generate training data for fakes.

86 Cole and Maiberg, "People Are Using AI to Create Fake Porn of Their Friends and Classmates."

87 "Bella Thorne Steals Hacker's Thunder, Publishes Nude Photos Herself," Naked Security (blog), June 18, 2019, https://nakedsecurity.sophos.com/2019/06/18/bella-thorne-steals-hackers-thunder-publishes-nude-photos-herself/; "Beware the Fake Facebook Sirens That Flirt You into Sextortion," Naked Security (blog), March 23, 2018, https://nakedsecurity.sophos.com/2018/03/23/beware-the-fake-facebook-sirens-that-flirt-you-into-sextortion/.

88 BBC News, "Fake Voices 'Help Cyber-Crooks Steal Cash,'" BBC, July 8, 2019, https://www.bbc.com/news/technology-48908736; Stacey Colino, "Don't Fall Victim to the Grandparents Scam," AARP, April 18, 2018, http://www.aarp.org/money/scams-fraud/info-2018/grandparent-scam-scenarios.html.

pornography in online porn sites.[89] Faked videos featuring private individuals carry great consequence for the ways we conceive of the evidentiary relationship between bodies and audiovisual content, expertise, and structural inequalities.

The question of how to effectively understand and address these images is non-trivial. Existing legal and platform policy cannot effectively address the harm that has and will arise from AV manipulation techniques. Even in the US there are many legal paths one could take to stop fake pornographic or otherwise harmful videos from being disseminated,[90] but there is no law to undo the unique damage these images unleash.[91] Moreover, the misuse of one's likeness is expensive to address through the legal system. In the US, laws like Section 230 of Title 47 of the 1996 Communications Decency Act are intended to remedy instances where courts fall short by allowing platforms to respond to users' requests to remove harmful content. But even Section 230 has proven incapable of fully addressing the range of problems that have been thrown up by massive platform use.[92]

The stakes are high for accurately understanding AV manipulation. The fear of a hyperreality spawned by deepfakes can be used to foreclose other possible interventions.

89 Drew Hartwell, "Fake-Porn Videos Are Being Weaponized to Harass and Humiliate Women: 'Everybody Is a Potential Target,'" Washington Post, accessed April 5, 2019, https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/; Samantha Cole and Emanuel Maiberg, "People Are Using AI to Create Fake Porn of Their Friends and Classmates," Motherboard (blog), January 26, 2018, https://motherboard.vice.com/en_us/article/ev5eba/ai-fake-porn-of-friends-deepfakes.

90 "47 U.S. Code § 230 - Protection for Private Blocking and Screening of Offensive Material," Legal Information Institute, accessed March 1, 2019, https://www.law.cornell.edu/uscode/text/47/230.

91 Franks, "The Desert of the Unreal"; Citron and Franks, "Criminalizing Revenge Porn."

92 47 U.S. Code § 230 - Protection for Private Blocking and Screening of Offensive Material,"; "Section 230 Protections," Electronic Frontier Foundation, August 25, 2011, https://www.eff.org/issues/bloggers/legal/liability/230.; Danielle Citron and Benjamin Wittes, "The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity," Fordham Law Review 86, no. 2 (November 1, 2017): 401; Citron and Franks, "Criminalizing Revenge Porn."

Sociologist and science studies scholar Donald Mackenzie has written extensively about the consequences of creating technology without critically assessing its potential for harm. His research led him to ask a peculiar question: "Can we uninvent the bomb?"[93] No matter how much we may want to uninvent the bomb, we cannot. Once new technology is generated, it must be reckoned with as a social and cultural phenomenon.

Nonetheless, many are focused on uninventing the "bomb" they see in AI-generated video production. In the wake of alarmist news coverage, multiple popular press articles have proposed programs of improved literacy,[94] arguing that people must take steps to equip themselves to detect fakes. Technologists are rushing to undo the coming "end of the world" through a range of new technical solutions.[95] Two of the leading technological approaches are: automated fake detection and distributed verification technologies to verify all online and even offline interaction. Facebook has been one of the loudest voices advocating automated systems to detect AI-generated manipulations.[96] The start-up TruePic[97] uses blockchain technology to verify press-captured images and videos. The US Defense Advanced Research Projects Agency (DARPA) funded MediFor to run diagnostics of

93 Donald Mackenzie, "Uninventing the Bomb?" Medicine and War 12, no. 3 (October 22, 2007): 202–11, https://doi.org/10.1080/13623699608409285.

94 Natasha Bernal, "Journalists at Wall Street Journal to Be Taught to Identify 'Deepfakes,'" The Telegraph, November 15, 2018, https://www.telegraph.co.uk/technology/2018/11/15/journalists-wall-street-journal-taught-identify-deepfakes/; Suzanne Sunne, "What to Watch for in the Coming Wave of "Deep Fake" Videos," Global Investigative Journalism Network, May 28, 2018, https://gijn.org/2018/05/28/what-to-watch-for-in-the-coming-wave-of-deep-fake-videos/; Siwei Lyu, "The Best Defense against Deepfake AI Might Be … Blinking," Fast Company, August 31, 2018, https://www.fastcompany.com/90230076/the-best-defense-against-deepfakes-ai-might-be-blinking.

95 Martínez, "The Blockchain Solution to Our Deepfake Problems"; Lyu, "The Best Defense against Deepfake AI Might Be… Blinking"; Greene, "Researchers Developed an AI to Detect DeepFakes"; "Exposing Fake Videos."

96 Facebook, "Expanding Fact-Checking to Photos and Videos | Facebook Newsroom," Facebook, September 13, 2018, https://newsroom.fb.com/news/2018/09/expanding-fact-checking/.

97 TruePic, "Photos and Videos You Can Trust."

videos to determine differences in pixels within videos that signal fakery.[98] Other such verification proposals involve doing away with net neutrality in the interest of marking the provenance of all packets carrying video data so they are easily surveilled to determine real-life identities of creators.[99] At present, there are four legislative bills in nascent stages of consideration to force individuals to label content that they have manipulated, and to allow the federal government to impose massive fines against individuals who have manipulated content deemed harmful.[100]

Those forwarding these totalizing technical security steps hold up the alleged uniqueness of deepfakes as justification. Locating deepfakes in the longer history of the politics of evidence, however, allows us to see that: a) decisions over what counts as "evidence" have historically been a crucial

98 Matt Turek, "Media Forensics," Defense Advanced Research Projects MediFor, 2018, https://www.darpa.mil/program/media-forensics.

99 H. R. Hasan and K. Salah, "Combating Deepfake Videos Using Blockchain and Smart Contracts," IEEE Access 7 (2019): 41596–606, https://doi.org/10.1109/ACCESS.2019.2905689.

100 Yvette Clarke, "H.R.3230 Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019," webpage, June 24, 2019, https://www.congress.gov/bill/116th-congress/house-bill/3230; Ben Sasse, "S.3805 Malicious Deep Fake Prohibition Act of 2018," webpage, December 21, 2018, https://www.congress.gov/bill/115th-congress/senate-bill/3805/text; Tim Grayson "AB-1280 Crimes: Deceptive Recordings.," webpage, April 22, 2019, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB1280. Bryan Hughes "SB 751," webpage, February 11. 2019. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB1280; In addition, Virginia has expanded existing revenge porn law to cover fake nude images. Marcus Simon, "HB 2678 Amendment" webpage, February 11, 2019. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?-bill_id=201920200AB1280

tool in defending the privilege of the already powerful;[101] b) the violence of AV manipulation can only be addressed by a combination of technical and social solutions; c) public agency over these technologies cannot be realized without addressing structural inequality; and d) the violence of AV manipulation will not be curtailed unless groups most vulnerable to that violence are able to influence public media systems.

Scholars, journalists, and activists are already considering how we might reconfigure technical systems, their ownership, and regulatory bodies to consider the politics of evidence. Beyond the "social" policy solutions mentioned above that penalize individuals for harmful behavior, more encompassing solutions might be to enact federal measures on corporations to encourage them to more meaningfully address the fallout from their massive gains in the past 15 years. Information scholar Sarah T. Roberts has suggested that contrary to the practice as it exists at present, the work of content moderators should be valued and supported by platform companies as one of many necessary steps to build community that fosters pro-social values that go beyond

101 Literature in both science and technology studies (STS) and media studies discuss how evidence in the form of formalized and informal data and discourses around these types of evidence are shaped by social, political and cultural structures that reflect and reify the status quo. At the same time, evidence can be mobilized by less powerful entities to make justice-based arguments about the distribution of harms and power in society. But even then, community-based counter-data is always judged according to rules set by dominant discourses around truth that exist to protect the status quo of privilege. The authors have published collaborative work on the politics of evidence as they relate to community-based counter-data practices. Britt S. Paris et al., "Pursuing a Toxic Agenda," September 2017, https://100days.envirodatagov.org/pursuing-toxic-agenda/; Morgan Currie, Joan Donovan, and Britt S. Paris, "Preserving for a More Just Future: Tactics of Activist Data Archiving," in Data Science Landscape: Towards Research Standards and Protocols, ed. Usha Mujoo Munshi and Neeta Verma, Studies in Big Data 38 (Singapore: Springer, 2018), 67–78; Britt S. Paris and Jennifer Pierre, "Naming Experience: Registering Resistance and Mobilizing Change with Qualitative Tools," InterActions: Journal of Education and Information Studies 13, no. 1 (2017), https://escholarship.org/uc/item/02d9w4qd; Britt S. Paris and Jennifer Pierre, "Bad Data — Cultural Anthropology," Cultural Anthropology Field Insights, April 28, 2017, https://culanth.org/fieldsights/1107-bad-data.\\uc0\\u8221{} in {\\i{}Data Science Landscape: Towards Research Standards and Protocols}, ed. Usha Mujoo Munshi and Neeta Verma, Studies in Big Data 38 (Singapore: Springer, 2018

profit motives.[102] Feminist legal scholar Danielle Citron has suggested tech companies could foster digital citizenship by building in thoughtful mechanisms to discourage user-generated content that is false or misleading, that harasses, or negatively targets specific groups by developing more rigorously pro-social terms of service (TOS) for taking down and labeling nefarious content, and following these TOS in a way that is transparent and open to public oversight.[103] Finally, Citron and legal scholar Mary Ann Franks have suggested the necessity of reconsidering how to hold platform intermediaries legally accountable for spreading tortious content in ways that would promote more effective cross-platform injunctions of harmful content.[104] These solutions taken together begin to show a long and difficult way forward, but one that is, nonetheless, possible.

Regardless of the difficulty, it is crucial to consider these larger problems that contribute to complications of AI-generated media as they enter an already troubled media landscape. While we cannot start again, Mackenzie's statement of the atomic bomb is apt when thinking about what to do about audiovisual manipulation: "We do not know, and should not pretend to know, at this point in time, the solution to all

102 Sarah T. Roberts, Behind the Screen (New Haven, CT: Yale University Press, 2019), https://yalebooks.yale.edu/book/9780300235883/behind-screen.

103 Danielle Keats Citron, Hate Crimes in Cyberspace, Reprint edition (Cambridge, MA: Harvard University Press, 2016); Danielle Keats Citron and Helen Norton, "Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age," Boston University Law Review 91 (2011): 1435.Reprint edition (Place of publication not identified: Harvard University Press, 2016

104 Danielle Citron and Benjamin Wittes, "The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity," Fordham Law Review 86, no. 2 (November 1, 2017): 401; Mary Anne Franks and Danielle Citron, "Criminalizing Revenge Porn," Wake Forest Law Review 49 (2014): 345–92; Mary Anne Franks, "Drafting an Effective 'Revenge Porn' Law: A Guide for Legislators" (Cyber Civil Rights Initiative, 2016), https://www.cybercivilrights.org/guide-to-legislation/; Christopher Zara, "The Most Important Law in Tech Has a Problem | Backchannel," Wired, January 3, 2017, https://www.wired.com/2017/01/the-most-important-law-in-tech-has-a-problem/.

the problems it may throw up."[105] What we do know is that limiting the harm of AV manipulation will require an understanding of the history of evidence, and the social processes that produce truth, in order to avoid new consolidations of power for those who can claim an exclusive expertise.

105 Donald Mackenzie, "Uninventing the Bomb?" Medicine and War 12, no. 3 (October 22, 2007): 202–11, https://doi.org/10.1080/13623699608409285, 120

ACKNOWLEDGMENTS

We are grateful to everyone at Data & Society Research Institute for their generous support, guidance, and helpful feedback through the duration of this project. In particular, we'd like to thank Patrick Davison for his thoughtful editing and revisions. We are forever thankful for the collaboration and ongoing discussions with members of Data & Society's Media Manipulation Initiative, past and present. Thanks to interviewees for graciously taking time to provide background information on this work. Special thanks to external reviewers Fenwick McKelvey, Tiziana Terranova, Aayush Bansal, and Samuel Gursky for their generous feedback on the draft. A big thank you to our friends, family, and partners for supporting us in so many ways through this project and during everything else.

DATA & SOCIETY

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net
@datasociety
Illustration by Jim Cooke

DATA&
SOCIETY

www.datasociety.net

@datasociety