Disinformed

Adam Garfinkle

There have been fakes as long as there have been frauds, and that is a very long time; but *deep*fakes are new fakes, and having initially loitered along the margins of general awareness, they are now occupied in haunting it. Tens of thousands of deepfakes have already been created. The technical means of fiddling with images is hardly new. Standing beside Joseph Stalin in one photograph taken along the newly completed White Sea Canal, Nikolai Yezhov disappeared from the very same photograph some months later, as he, in fact, had disappeared from life. The fakery is fine, but it is no better than that, the ensuing photograph visually unbalanced by a lot of gray canal water where Yezhov had once stood. It is thanks to a technology invented in 2014 that deepfakery is capable of taking verisimilitude to a new level.

The ability to produce ever more persuasive deepfakes has been made possible by a recent form of machine learning called generative adversarial networks—or GANs. A GAN operator pits a generator (G) against a discriminator (D) in a gamelike environment in which G tries to fool D into incorrectly discriminating between fake and real data. The technology works by means of a series of incremental but rapid adjustments that allows D to discriminate data while G tries to fool it.

How fast are these adjustments? Very fast. A computer can play 24 trillion games of Texas Hold'em every second. To beat human opponents, a computer does not need to assess their strategies. It relies on the patterns it picks out, and assumes only that human strategy is limited to a few flexible tactics. DeepMind beat human players at 99.8% of StarCraft II games, a game subtler and more abstract than Texas Hold'em.

GAN technology is not particularly exotic; the software is available commercially, and anyone who can write code can figure out how to use it. If simply using it is open admission, what about using it to change the 2020 election? That, David Doermann argues, "would take a massive amount of computing power." Rogue actors, he adds, are too small to do much. "A nation state is required."[1]

What about an organized group scaled somewhere between a rogue actor and a rogue state?

It is too late to ban GANs. But it is possible to criminalize certain uses, and efforts are afoot to do so. Beyond the ambit of domestic law, legal remedies are less likely to be effective. GANs have any number of applications. Some are pure as the driven snow. GANs can reconstruct three-dimensional images from two-dimensional photographs. They can be used to visualize industrial design, improve astronomical images by filling in statistically what real cameras cannot capture, and generate showers of imaginary particles for high-energy physics experiments. GANs can also be used to visualize motion in static environments, which could help find people lost or hiding in forests or jungles. In 2016, GAN technology was used to generate new molecules for a variety of protein targets in cells implicated in fibrosis, inflammation, and cancer.[2]

So much for Dr. Jekyll. Mr. Hyde now follows. What makes GANs frightening is their power to produce photographic images of people who do not exist, or to generate video from voice recordings, or to doctor images of people who do exist to make them seem to be someone else, or to say things they never did or would say. GANs can be used to create pornography by using an image without the subject's knowledge or consent. According to the company Sensity, formerly Deeptrace, of the 15,000 online deepfakes detected by September 2019, 96% were pornographic.[3]

GAN technology is *intended* to deceive.

And the technology is flexible. Those who mean mischief favor the adversarial neural network; those who do not, the discriminators. This allows authorities to better detect deepfake attacks; but it also makes them adept at offense if they themselves go rogue. Any formula that helps the defense can be used to improve an attack. In planting false positives, clever operators tag real videos as fakes. Ambiguity infects the entire informational domain.

Pornography aside, many other nefarious uses of deepfakery are obvious. In August 2019, the *Wall Street Journal* reported on the first big-money case of identity fraud.[4] Scammers used voice-changing technology to impersonate a chief executive. The money is gone; *they* have not been caught. Business leaders or banking lenders can be made to say things that dupe investors and markets, the ensuing herd yielding millions for those in the know.

Political uses carry enormous potential. If Russian efforts got Donald Trump elected, as former Director of National Intelligence General James Clapper suggested, one could hardly think of a more ominous threat.[5] A GAN-generated deepfake already exists of Nancy Pelosi sounding drunk and saying things she never said.[6] It is primitive, and thus easy to detect, but that did not prevent both President Trump and Rudy Giuliani from retweeting it. The same principle was at work in recent deepfakes of political figures in Gabon and Malaysia.[7] Social context critically determines the effect of technological tomfoolery. The fake alone need only go so far, and by the time a fake is found out, it may be too late to prevent an incensed or excited mob from violence.

Terrorism is often an attempt to lure a target into reacting in a way that undermines its own principles and sources of political legitimacy.[8] ISIS and al-Qaeda both proved more technologically adept than was at first thought. These organizations could easily use GANs to assign to various national leaders speeches or sentiments that might incite riots from Karachi to Fez, as when Crown Prince Mohammed bin Salman, technologically refreshed and so reborn, claims that Saudi military forces, having secretly obtained three nuclear weapons from Pakistan, are preparing to bomb Tehran, Qom, and Bandar Abbas.

Will regional publics and governments believe it real? If not, what will they do?

Disinformation has been a part of espionage for centuries. What is new about deepfakery, then? For one thing, an illusion of reality more convincing than any produced in the past. For another, entirely new social and cultural contexts. The result is a vamping up of venerable means to satisfy modern goals. The Roman coliseum was useful for stirring up a mob by means of leather-lunged orators. Useful but limited. The advent of movable type? Better. Yet only a small minority ever gained literacy. The radio? *Much* better. Radio provided even the linen-lunged the power to reach mass audiences. Benito Mussolini's regime was a pioneer in the 1920s. The Nazis soon followed, combining the use of radio with old-fashioned spectacle such as the Nuremberg rallies. Father Charles Coughlin used radio to dangerously good effect. The creation of the Federal Communications Commission in 1934 testified to the concern of US democratic elites.

Sound is one thing; sight is another. People believe their eyes before their ears. In 1984, something like artificial intelligence beamed Max Headroom to American viewers through their television sets. The technology was primitive, and Max was actually a man in facial prosthetics and a plastic suit. If his original purpose was entertainment, he was, at once, hijacked for political advocacy. On November 22, 1987, two Chicago television stations had their signal taken over by unknown individuals, one of whom wore a Max Headroom look-alike costume. The fake Max rambled on for about ninety seconds contemning the real Max's commercial endorsements, and concluding with a pair of exposed buttocks being whacked by a fly swatter before normal programming resumed.

The culprits, it is gratifying to recount, were never apprehended, still less identified.

Graphic capabilities have now progressed from Max Headroom to computer-aided anime and CGI technologies—child's play compared to GANs. The new technology requires sophisticated techniques all its own. Twitter works as well as it does by promoting an obvious sense of both immediacy and intimacy. There it is: the naked thought, shorn of layers, lawyers, fillers, or filters. To communicate to the American people, Trump prefers tweets to press conferences. Intimacy of this sort requires many individual technical platforms to be linked together. More than five billion people now have mobile devices, over half of them smartphones.[9] The iPhone came on the market in June 2007 and took a decade to reach initial market saturation. This is a fast-moving development, and one with radical effects. If the Arab Spring was driven by young people, it was made possible by social media.[10] A platform as anodyne as Facebook was sufficient to deepen, if not cause, ethnic cleansing in Myanmar.[11] The country had recently emerged from a military dictatorship, and as the internet was relatively new, those incited to violence were not able to distinguish real information from false. They were not about to take any chances.

Technologically advanced countries have some chance of using legal and regulatory means to deter and limit the damage. That makes countries like Mali, Malaysia, and Egypt more vulnerable to attacks than countries like Germany, South Korea, and Israel. But *between* countries, and particularly between unfriendly regimes, the potential for legal and regulatory efforts is much smaller. The landscape for destructive deepfakery is more international than national, even if national and personal uses are more frequent.

Under US law, the present basis for protection and remediation goes back to the 1996 Communications Decency Act, Section 230. Websites are defined as platforms, not publications, so as not to encumber free speech. Thus the recent difference of opinion between Twitter and Facebook about the appropriateness of political advertisements. The legal regime remains so loose that issues are decided on a case-by-case basis. Legislation introduced by Senator Ben Sasse and Representative Yvette Clarke—the DEEPFAKES Accountability Act, referred to the House Subcommittee on Crime, Terrorism, and Homeland Security last year—would ban deepfakes from creating pornography and punish deepfakers for election interference.[12] This act, assuming it becomes law, is almost certain to generate constitutional challenges. In such a situation, proposals to make major tech-communications platforms into regulated utilities would have a robust future.

If these platforms represent a critical infrastructure, why not treat them as every other critical infrastructure? The fact that this infrastructure is inextricably international, as opposed to more or less hermetically national like an electricity grid, enormously complicates matters.

It does not alter the principle.

Power distributions among states used to be overwhelmingly a matter of mass, brawn, and physical resources. Those qualities are not altogether obsolete. But human capital, social trust, and institutional coherence mean far more now than they did a century ago. Relatively small polities that can protect themselves from malicious attack now punch above their weight class, and brawnier actors that cannot are at greater risk.

Philosophers have long distinguished five sources that together generate our knowledge of the facts: *empirical*—from the senses; *rational*—from reason; *introspective*—from self-knowledge; *memorial*—from memories; and *testimonial*—from what others tell us. "Only a very small part of my knowledge of the world," Alfred Schütz observed, "originates within my personal experience. The greater part is socially derived, handed down to me by my friends, my parents, my teachers and the teachers of my teachers."[13]

Getting things right is a process obviously vulnerable to disruption. If there are ways of getting things right, there must be ways of getting them wrong. And there are. We may be mistaken, deluded, entirely in error, misled, or deceived. Those who understand how these vulnerabilities work can manipulate them deliberately. That is what cutting-edge marketing does. The fact that it works has led one observer to describe human beings as moist robots—creatures easily duped.[14] Thanks to Citizens United, who sued the Federal Election Commission, corporations are entitled to a political voice as if they were individual citizens. As a result, corporations that create and use virtual influencers can be sheltered under the free speech protections of the First Amendment.

It is a privileged position if one's goal is to influence moist robots.

A spectacle is an attention-arresting display. I am not referring to special effects that the audience knows are fabricated, such as the Death Star explosion in *Return of the Jedi*. I have in mind what used to make circus freak shows so captivating. The reality-TV genre is an example. Is it real, or fake, or some of both? It is not fully scripted, and so designed to evoke uncertainty. The same applies to World Federation Wrestling. It is fake, and yet sufficiently ambiguous that those who wish may suspend their disbelief. Spectacle evokes what some psychologists refer to as an astounding complex. Astounding complexes are ubiquitous as technical events in television and movies, most often in the form of rapid scene cuts, where the screen takes our senses places where our bodies cannot go. Rapid scene cuts are more used in commercials than in regular programming because, though they cost marginally more to make, their effects are worth it to advertisers. Viewers are made more alert through the multiplication of scene shifts and are more likely to remember and hence to buy the product. Large numbers of people living in technologically advanced environments see more mediated visual images than real ones. We have gotten so used to mediated visual events, many of them freak-show-like spectacles, that what is real and what is not have become blurred by the time most people become young adults.

In the case of deepfakes, we are being astounded, and fooled. Thirst for the astounding is a characteristic of time and place. When big-screen movie houses first opened in American cities in the early 1920s, medical teams had to be called to the scene on occasion because silent versions of *Frankenstein*, for example, were so frightening that they caused some patrons to faint or even have heart attacks.

A poignant description of how much the sensitivity- and shock-bars of culture can change over time comes from, not entirely surprisingly, a Michael Crichton novel:

> What is the dominant mode of experience at the end of the twentieth century? How do people see things, how do they expect to see things? The answer is simple. In every field, from business to politics to marketing to education, the dominant mode has become entertainment. ...
>
> Today, everybody expects to be entertained, and they expect to be entertained all the time. Business meetings must be snappy, with bullet lists and animated graphics, so executives aren't bored. Malls and stores must be engaging, so they amuse as well as sell us. Politicians must have pleasing video personalities and tell us only what we want to hear. Schools must be careful not to bore young minds that expect the speed and complexity of television. Students must be amused—everyone must be amused, or they will switch: switch brands, switch channels, switch parties, switch loyalties. This is the intellectual reality of Western society at the end of the century.
>
> In other centuries, human beings wanted to be saved, or improved, or freed, or educated. But in our century, they want to be entertained.[15]

A deepfake may have a vastly greater chance of working because it is delivered in a manner that has become completely seamless.

Toward the end of Book VIII of Plato's *Republic*, Socrates tells us that "the tyrant [is surrounded by] his spectacular, iridescent, numerous, and

ever-changing bodyguard." The people who designed the Nazi rallies understood Plato. They took the measure of those mesmerized by the tyrant's spectacular bodyguard, and they were familiar with the cave.

In Book VII, Plato suggests that humanity lives in a world of shadows and illusions. Unaware of the fully lighted world outside, a man's vision is dim, so that he is drawn to and becomes mesmerized by whatever light there is. The luster of gold coins will spellbind him unless and until his eyes adjust to the brighter light outside the cave that enables him to see the divine gold of wisdom. Once he does, *if* he does, he has no further use for gold. It follows that the more dim-sighted citizens are, the more easily manipulated; the more ruled by appetite, the more dim-sighted. If we expect no more from citizens than that they gratify their appetites, they will be enticed by spectacle and easily taken in by lies. Disinformation will thrive.

As it is thriving now.

Adam Garfinkle is a non-resident Distinguished Fellow at the S. Rajaratnam School of International Studies at the Nanyang Technological University in Singapore.

1. David Doermann, quoted in Siddharth Venkataramakrishnan, "Can You Believe Your Eyes? How Deepfakes Are Coming for Politics," *FT Weekend*, October 24, 2019.
2. Meg King, "It's Meg King... Or Maybe It's a Deepfake," Wilson Center podcast, September 24, 2019..
3. Henry Ajder et al., "The State of Deepfakes 2019: Landscape, Threats, and Impact," *Deeptrace Labs*, September 2019.
4. Catherine Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case," *Wall Street Journal*, August 30, 2019.
5. See James R. Clapper, *Facts and Fears* (Penguin USA, 2019).
6. Donald Trump, (@realDonaldTrump), "PELOSI STAMMERS THROUGH NEWS CONFERENCE," Twitter, May 24, 2019, 3:09 a.m.
7. Ali Breland, "The Bizarre and Terrifying Case of the 'Deepfake' Video that Helped Bring an African Nation to the Brink," *Mother Jones*, March 15, 2019. Philip Golingai, "Is it Azmin or a Deepfake?" *The Star*, June 15, 2019.
8. See the classic essay, still revelatory to some after all these years, by David Fromkin, "The Strategy of Terrorism," *Foreign Affairs*, July 1975.
9. Laura Silver, "Smartphone Ownership Is Growing Rapidly around the World, but Not Always Equally," Pew Research Center, February 5, 2019.
10. Jessi Hempel, "Social Media Made the Arab Spring but Couldn't Save It," *Wired*, January 26, 2016.
11. Alexandra Stevenson, "Facebook Admits It Was Used to Incite Violence in Myanmar," *New York Times*, November 6, 2018.
12. Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R.3230, 116th Congress, introduced in House June 12, 2019.
13. Alfred Schütz, *Collected Papers* (The Hague: Martinus Nijhoff, 1967), 13. Kenneth Burke similarly distinguished the mass of what we know from our bio-sensory bit of reality, what he called "the paper-thin line of our own particular lives." Kenneth Burke, *Language as Symbolic Action* (Berkeley: University of California Press, 1966), 5.
14. I refer to Scott Adams, the creator of the *Dilbert* comic strip, in *Win Bigly: Persuasion in a World Where Facts Don't Matter* (New York: Portfolio, 2017).
15. Michael Crichton, *Timeline* (New York: Ballentine, 1999), 442–43.

Published on September 28, 2020

https://inference-review.com/article/disinformed