

IN THE AGE OF A.I., IS SEEING STILL BELIEVING?

Advances in digital imagery could deepen the fake-news crisis—or help us get out of it.



By Joshua Rothman

November 5, 2018



As synthetic media spreads, even real images will invite skepticism. Illustration by Javier Jaén; photograph by Svetikd / Getty

In 2011, Hany Farid, a photo-forensics expert, received an e-mail from a bereaved father. Three years earlier, the man's son had found himself on the side of the road with a car that wouldn't start. When some strangers offered him a lift, he accepted. A few

minutes later, for unknown reasons, they shot him. A surveillance camera had captured him as he walked toward their car, but the video was of such low quality that key details, such as faces, were impossible to make out. The other car's license plate was visible only as an indecipherable jumble of pixels. The father could see the evidence that pointed to his son's killers—just not clearly enough.

Farid had pioneered the forensic analysis of digital photographs in the late nineteen-nineties, and gained a reputation as a miracle worker. As an expert witness in countless civil and criminal trials, he explained why a disputed digital image or video had to be real or fake. Now, in his lab at Dartmouth, where he was a professor of computer science, he played the father's video over and over, wondering if there was anything he could do. On television, detectives often “enhance” photographs, sharpening the pixelated face of a suspect into a detailed portrait. In real life, this is impossible. As the video had flowed through the surveillance camera's “imaging pipeline”—the lens, the sensor, the compression algorithms—its data had been “downsampled,” and, in the end, very little information remained. Farid told the father that the degradation of the image couldn't be reversed, and the case languished, unsolved.

A few months later, though, Farid had a thought. What if he could use the same surveillance camera to photograph many, many license plates? In that case, patterns might emerge—correspondences between the jumbled pixels and the plates from which they derived. The correspondences would be incredibly subtle: the particular blur of any degraded image would depend not just on the plate numbers but also on the light conditions, the design of the plate, and many other variables. Still, if he had access to enough images—hundreds of thousands, perhaps millions—patterns might emerge.

Such an undertaking seemed impractical, and for a while it was. But a new field, “image synthesis,” was coming into focus, in which computer graphics and A.I. were combined. Progress was accelerating. Researchers were discovering new ways to use neural networks—software systems based, loosely, on the architecture of the brain—to analyze and create images and videos. In the emerging world of “synthetic media,” the work of digital-image creation—once the domain of highly skilled programmers and Hollywood special-effects artists—could be automated by expert systems capable of producing realism on a vast scale.

In a media environment saturated with fake news, such technology has disturbing implications. Last fall, an anonymous Redditor with the username Deepfakes released a software tool kit that allows anyone to make synthetic videos in which a neural network substitutes one person's face for another's, while keeping their expressions consistent. Along with the kit, the user posted pornographic videos, now known as “deepfakes,” that appear to feature various Hollywood actresses. (The software is complex but comprehensible: “Let's say for example we're perverting on some innocent girl named Jessica,” one tutorial reads. “The folders you create would be: *jessica; jessica_faces; porn; porn_faces; model; output.*”) Around the same time, “Synthesizing Obama,” a paper published by a research group at the University of Washington, showed that a neural network could create believable videos in which the former President appeared to be saying words that were really spoken by someone else. In a video voiced by Jordan Peele, Obama seems to say that “President Trump is a total and complete dipshit,” and warns that “how we move forward in the age of information” will determine “whether we become some kind of fucked-up dystopia.”

Not all synthetic media is dystopian. Recent top-grossing movies (“Black Panther,” “Jurassic World”) are saturated with synthesized images that, not long ago, would have been dramatically harder to produce; audiences were delighted by “Star Wars: The Last Jedi” and “Blade Runner 2049,” which featured synthetic versions of Carrie Fisher and Sean Young, respectively. Today's smartphones digitally manipulate even ordinary snapshots, often using neural networks: the iPhone's “portrait mode” simulates

what a photograph would have looked like if it been taken by a more expensive camera. Meanwhile, for researchers in computer vision, A.I., robotics, and other fields, image synthesis makes whole new avenues of investigation accessible.

Farid started by sending his graduate students out on the Dartmouth campus to photograph a few hundred license plates. Then, based on those photographs, he and his team built a “generative model” capable of synthesizing more. In the course of a few weeks, they produced tens of millions of realistic license-plate images, each one unique. Then, by feeding their synthetic license plates through a simulated surveillance camera, they rendered them indecipherable. The aim was to create a Rosetta Stone, connecting pixels to plate numbers.

Next, they began “training” a neural network to interpret those degraded images. Modern neural networks are multilayered, and each layer juggles millions of variables; tracking the flow of information through such a system is like following drops of water through a waterfall. Researchers, unsure of how their creations work, must train them by trial and error. It took Farid’s team several attempts to perfect theirs. Eventually, though, they presented it with a still from the video. “The license plate was like ten pixels of noise,” Farid said. “But there was still a signal there.” Their network was “pretty confident about the last three characters.”

This summer, Farid e-mailed those characters to the detective working the case. Investigators had narrowed their search to a subset of blue Chevy Impalas; the network pinpointed which one. Someone connected to the car turned out to have been involved in another crime. A case that had lain dormant for nearly a decade is now moving again. Farid and his team, meanwhile, published their results in a computer-vision journal. In their paper, they noted that their system was a free upgrade for millions of low-quality surveillance cameras already in use. It was a paradoxical outcome typical of the world of image synthesis, in which unreal images, if they are realistic enough, can lead to the truth.

Farid is in the process of moving from Dartmouth to the University of California, Berkeley, where his wife, the psychologist Emily Cooper, studies human vision and virtual reality. Their modernist house, perched in the hills above the Berkeley campus, is enclosed almost entirely in glass; on a clear day this fall, I could see through the living room to the Golden Gate Bridge. At fifty-two, Farid is gray-haired, energized, and fit. He invited me to join him on the deck. “People have been doing synthesis for a long time, with different tools,” he said. He rattled off various milestones in the history of image manipulation: the transposition, in a famous photograph from the eighteen-sixties, of Abraham Lincoln’s head onto the body of the slavery advocate John C. Calhoun; the mass alteration of photographs in Stalin’s Russia, designed to purge his enemies from the history books; the convenient realignment of the pyramids on the cover of *National Geographic*, in 1982; the composite photograph of John Kerry and Jane Fonda standing together at an anti-Vietnam demonstration, which incensed many voters after the *Times* credulously reprinted it, in 2004, above a story about Kerry’s antiwar activities.

“In the past, anybody could buy Photoshop. But to really use it well you had to be highly skilled,” Farid said. “Now the technology is democratizing.” It used to be safe to assume that ordinary people were incapable of complex image manipulations. Farid recalled a case—a bitter divorce—in which a wife had presented the court with a video of her husband at a café table, his hand reaching out to caress another woman’s. The husband insisted it was fake. “I noticed that there was a reflection of his hand in the surface of the table,” Farid said, “and getting the geometry exactly right would’ve been really hard.” Now convincing synthetic images and videos were becoming easier to make.

Farid speaks with a technologist’s enthusiasm and a lawyer’s wariness. “Why did Stalin airbrush those people out of those photographs?” he asked. “Why go to the trouble? It’s because there is something very, very powerful about the visual image. If you

change the image, you change history. We're incredibly visual beings. We rely on vision—and, historically, it's been very reliable. And so photos and videos still have this incredible resonance." He paused, tilting back into the sun and raising his hands. "How much longer will that be true?"

One of the world's best image-synthesis labs is a seven-minute drive from Farid's house, on the north side of the Berkeley campus. The lab is run by a forty-three-year-old computer scientist named Alexei A. Efros. Efros was born in St. Petersburg; he moved to the United States in 1989, when his father, a winner of the U.S.S.R.'s top prize for theoretical physics, got a job at the University of California, Riverside. Tall, blond, and sweetly genial, he retains a Russian accent and sense of humor. "I got here when I was fourteen, but, really, one year in the Soviet Union counts as two," he told me. "I listened to classical music—everything!"

As a teen-ager, Efros learned to program on a Soviet PC, the Elektronika BK-0010. The system stored its programs on audiocassettes and, every three hours, overheated and reset; since Efros didn't have a tape deck, he learned to code fast. He grew interested in artificial intelligence, and eventually gravitated toward computer vision—a field that allowed him to watch machines think.

In 1998, when Efros arrived at Berkeley for graduate school, he began exploring a problem called "texture synthesis." "Let's say you have a small patch of visual texture and you want to have more of it," he said, as we sat in his windowless office. Perhaps you want a dungeon in a video game to be made of moss-covered stone. Because the human visual system is attuned to repetition, simply "tiling" the walls with a single image of stone won't work. Efros developed a method for intelligently sampling bits of an image and probabilistically recombining them so that a texture could be indefinitely and organically extended. A few years later, a version of the technique became a tool in Adobe Photoshop called "content-aware fill": you can delete someone from a pile of leaves, and new leaves will seamlessly fill in the gap.

From the front row of CS 194-26—Image Manipulation and Computational Photography—I watched as Efros, dressed in a blue shirt, washed jeans, and black boots, explained to about a hundred undergraduates how the concept of "texture" could be applied to media other than still images. Efros started his story in 1948, with the mathematician Claude Shannon, who invented information theory. Shannon had envisioned taking all the books in the English language and analyzing them in order to discover which words tended to follow which other words. He thought that probability tables based on this analysis might enable the construction of realistic English sentences.

"Let's say that we have the words 'we' and 'need,'" Efros said, as the words appeared on a large screen behind him. "What's the likely next word?"

The students murmured until Efros advanced to the next slide, revealing the word "to."

"Now let's say that we move our contextual window," he continued. "We just have 'need' and 'to.' What's next?"

"Sleep!" one student said.

"Eat!" another said.

"Eat" appeared onscreen.

“If our data set were a book about the French Revolution, the next word might be ‘cake,’ ” Efros said, chuckling. “Now, what is this? You guys use it all the time.”

“Autocomplete!” a young man said.

Pacing the stage, Efros explained that the same techniques used to create synthetic stonework or text messages could also be used to create synthetic video. The key was to think of movement—the flickering of a candle flame, the strides of a man on a treadmill, the particular way a face changed as it smiled—as a texture in time. “Zzzzt,” he said, rotating his hands in the air. “Into the time dimension.”

A hush of concentration descended as he walked the students through what this meant mathematically. The frames of a video could be seen as links in a chain—and that chain could be looped and crossed over itself. “You’re going to compute transition probabilities between your frames,” he said. Using these, it would be possible to create user-controllable, natural motion.

The students, their faces illuminated by their laptops, toggled between their notes and their code. Efros, meanwhile, screened a video on “expression-dependent textures,” created by the team behind “Synthesizing Obama.” Onscreen, a synthetic version of Tom Hanks’s face looked left and right and, at the click of a mouse, expressed various emotions: fear, anger, happiness. The researchers had used publicly available images of Hanks to create a three-dimensional model, or “mesh,” of his face onto which they projected his characteristic expressions. For this week’s homework, Efros concluded, each student would construct a similar system. Half the class groaned; the other half grinned.

Afterward, a crowd gathered around Efros with questions. In my row, a young woman turned to her neighbor and said, “Edge detection is sweet!”

Before arriving in Berkeley, I had written to Shiry Ginosar, a graduate student in Efros’s lab, to find out what it would take to create a synthetic version of me. Ginosar had replied with instructions for filming myself. “For us to be able to generate the back of your head, your profile, your arm moving up and down, etc., we need to have seen you in these positions in your video,” she wrote. For around ten minutes, before the watchful eye of an iPhone, I walked back and forth, spun in circles, practiced my lunges, and attempted the Macarena; my performance culminated in downward dog. “You look awesome ;-),” Ginosar wrote, having received my video. She said it would take about two weeks for a network to learn to synthesize me.

When I arrived, its work wasn’t quite done. Ginosar—a serene, hyper-organized woman who, before training neural networks, trained fighter pilots in simulators in the Israel Defense Forces—created an itinerary to keep me occupied while I waited. In addition to CS 194-26, it included lunch at Momo, a Tibetan curry restaurant, where Efros’s graduate students explained how it had come to pass that undergrads could create, as homework, Hollywood-like special effects.

“In 1999, when ‘The Matrix’ came out, the ideas were there, but the computation was very slow,” Deepak Pathak, a Ph.D. candidate, said. “Now computers are really fast. The G.P.U.s”—graphics processing units, designed to power games like Assassin’s Creed—“are very advanced.”

“Also, everything is open-sourced,” said Angjoo Kanazawa, who specializes in “pose detection”—figuring out, from a photo of a person, how her body is arranged in 3-D space.

“And that’s good, because we want our research to be reproducible,” Pathak said. “The result is that it’s easy for someone who’s in high school or college to run the code, because it’s in a library.”

The acceleration of home computing has converged with another trend: the mass uploading of photographs and videos to the Web. Later, when I sat down with Efros in his office, he explained that, even in the early two-thousands, computer graphics had been “data-starved”: although 3-D modellers were capable of creating photorealistic scenes, their cities, interiors, and mountainscapes felt empty and lifeless. True realism, Efros said, requires “data, data, data” about “the gunk, the dirt, the complexity of the world,” which is best gathered by accident, through the recording of ordinary life.

Today, researchers have access to systems like ImageNet, a site run by computer scientists at Stanford and Princeton which brings together fourteen million photographs of ordinary places and objects, most of them casual snapshots posted to Flickr, eBay, and other Web sites. Initially, these images were sorted into categories (carrouseles, subwoofers, paper clips, parking meters, chests of drawers) by tens of thousands of workers hired through Amazon Mechanical Turk. Then, in 2012, researchers at the University of Toronto succeeded in building neural networks capable of categorizing ImageNet’s images automatically; their dramatic success helped set off today’s neural-networking boom. In recent years, YouTube has become an unofficial ImageNet for video. Efros’s lab has overcome the site’s “platform bias”—its preference for cats and pop stars—by developing a neural network that mines, from “life style” videos such as “My Spring Morning Routine” and “My Rustic, Cozy Living Room,” clips of people opening packages, peering into fridges, drying off with towels, brushing their teeth. This vast archive of the uninteresting has made a new level of synthetic realism possible.

On his computer, Efros showed me a photo taken from a bridge in Lyon. A large section of the riverbank—which might have contained cars, trees, people—had been deleted. In 2007, he helped devise a system that rifles through Flickr for similar photos, many of them taken while on vacation, and samples them. He clicked, and the blank was filled in with convincing, synthetic buildings and greenery. “Probably it found photos from a different city,” Efros said. “But, you know, we’re boring. We always build the same kinds of buildings on the same kinds of riverbanks. And then, as we walk over bridges, we all say, along with a thousand other people, ‘Hey, this will look great, let me take a picture,’ and we all put the horizon in the same place.” In 2016, Ira Kemelmacher-Shlizerman, one of the researchers behind “Synthesizing Obama,” applied the same principle to faces. Given your face as input, her system combs the Internet for people who look like you, then combines their features with your own, to show how you’d look if you had curly hair or were a different age.

One of the lessons of image synthesis is that, with enough data, everything becomes texture. Each river and vista has its double, ready to be sampled; there are only so many faces, and your doppelgängers have already uploaded yours. Products are manufactured over and over, and new buildings echo old ones. The idea of texture even extends—“Zzzzt!”—into the social dimension. Your Facebook news feed highlights what “people like you” want to see. In addition to unearthing similarities, social media creates them. Having seen photos that look a certain way, we start taking them that way ourselves, and the regularity of these photos makes it easier for networks to synthesize pictures that look “right” to us. Talking with Efros, I struggled to come up with an image for this looped and layered interconnectedness, in which patterns spread and outputs are recirculated as inputs. I thought of cloverleaf interchanges, subway maps, Möbius strips.

A sign on the door of Efros’s lab at Berkeley reads “Caution: Deep Nets.” Inside, dozens of workstations are arranged in rows, each its own jumble of laptop, keyboard, monitor, mouse, and coffee mug—the texture of workaholism, iterated. In the back, in a lounge with a pool table, Richard Zhang, a recent Ph.D., opened his laptop to explain the newest developments in synthetic-image

generation. Suppose, he said, that you possessed an image of a landscape taken on a sunny day. You might want to know what it would look like in the rain. “The thing is, there’s not just one answer to this problem,” Zhang said. A truly creative network would do more than generate a convincing image. It would be able to synthesize many possibilities—to do for landscapes what Farid’s much simpler system had done for license plates.

Onscreen, Zhang showed me an elaborate flowchart in which neural networks train other networks—an arrangement that researchers call a “generative adversarial network,” or GAN. He pointed to one of the networks: the “generator,” charged with synthesizing, more or less at random, new versions of the landscape. A second network, the “discriminator,” would judge the verisimilitude of those images by comparing them with the “ground truth” of real landscape photographs. The first network riffed; the second disciplined the first. Zhang’s screen showed the system in action. An image of a small town in a valley, on a lake, perhaps in Switzerland, appeared; it was night, and the view was obscured by darkness. Then, image by image, we began to “traverse the latent space.” The sun rose; clouds appeared; the leaves turned; rain descended. The moon shone; fog rolled in; a storm gathered; snow fell. The sun returned. The trees were green, brown, gold, red, white, and bare; the sky was gray, pink, black, white, and blue. “It finds the sources of patterns of variation,” Zhang said. We watched the texture of weather unfold.

In 2016, the Defense Advanced Research Projects Agency (DARPA) launched a program in Media Forensics, or MediFor, focussed on the threat that synthetic media poses to national security. Matt Turek, the program’s manager, ticked off possible manipulations when we spoke: “Objects that are cut and pasted into images. The removal of objects from a scene. Faces that might be swapped. Audio that is inconsistent with the video. Images that appear to be taken at a certain time and place but weren’t.” He went on, “What I think we’ll see, in a couple of years, is the synthesis of events that didn’t happen. Multiple images and videos taken from different perspectives will be constructed in such a way that they look like they come from different cameras. It could be something nation-state driven, trying to sway political or military action. It could come from a small, low-resource group. Potentially, it could come from an individual.”

MediFor has brought together dozens of researchers from universities, tech companies, and government agencies. Collectively, they are creating automated systems based on more than fifty “manipulation indicators.” Their goal is not just to spot fakes but to trace them. “We want to attribute a manipulation to someone, to explain why a manipulation was done,” Turek said. Ideally, such systems would be integrated into YouTube, Facebook, and other social-media platforms, where they could flag synthesized content. The problem is speed. Each day, five hundred and seventy-six thousand hours of video are uploaded to YouTube; MediFor’s systems have a “range of run-times,” Turek said, from less than a second to “tens of seconds” or more. Even after they are sped up, practical questions will remain. How will innocent manipulations be distinguished from malicious ones? Will advertisements be flagged? How much content will turn out to be, to some degree, synthetic?

In his glass-walled living room, Hany Farid and I watched a viral video called “Golden Eagle Snatches Kid,” which appears to show a bird of prey swooping down upon a toddler in a Montreal park. Specialized software, Farid explained, could reveal that the shadows of the eagle and the kid were subtly misaligned. Calling up an image of a grizzly bear, Farid pointed out that, under high magnification, its muzzle was fringed in red and blue. “As light hits the surface of a lens, it bends in proportion to its wavelength, and that’s why you see the fringing,” he explained. These “chromatic aberrations” are smallest at the center of an image and larger toward its edges; when that pattern is broken, it suggests that parts of different photographs have been combined.

There are ways in which digital photographs are more tamper-evident than analog ones. During the manufacturing of a digital camera, Farid explained, its sensor—a complex latticework of photosensitive circuits—is assembled one layer at a time. “You’re laying down loads of material, and it’s not perfectly even,” Farid said; inevitably, wrinkles develop, resulting in a pattern of brighter and dimmer pixels that is unique to each individual camera. “We call it ‘camera ballistics’—it’s like the imperfections in the barrel of a gun,” he said. Modern digital cameras, meanwhile, often achieve higher resolutions by guessing about the light their sensors don’t catch. “Essentially, they cheat,” he said. “Two-thirds of the image isn’t recorded—it’s synthesized!” He laughed. “It’s making shit up, but in a logical way that creates a very specific pattern, and if you edit something the pattern is disturbed.”

Many researchers who study synthesis also study forensics, and vice versa. “I try to be an optimist,” Jacob Huh, a chilled-out grad student in Efros’s lab, told me. He had trained a neural network to spot chromatic aberrations and other signs of manipulation; the network produces “heat maps” highlighting the suspect areas of an image. “The problem is that, if you can spot it, you can fix it,” Huh said. In theory, a forger could integrate his forensic network into a GAN, where—as a discriminator—it could train a generator to synthesize images capable of eluding its detection. For this reason, in an article titled “Digital Forensics in a Post-Truth Age,” published earlier this year in *Forensic Science International*, Farid argued that researchers need to keep their newest techniques secret for a while. The time had come, he wrote, to balance “scientific openness” against the risk of “fueling our adversaries.”

In Farid’s view, the sheer number of distinctive “manipulation indicators” gives forensics experts a technical edge over forgers. Just as counterfeiters must painstakingly address each security feature on a hundred-dollar bill—holograms, raised printing, color-shifting ink, and so on—so must a media manipulator solve myriad technical problems, some of them statistical in nature and invisible to the eye, in order to create an undetectable fake. Training neural networks to do this is a formidable, perhaps impossible task. And yet, Farid said, forgers have the advantage in distribution. Although “Golden Eagle Snatches Kid” has been identified as fake, it’s still been viewed more than thirteen million times. Matt Turek predicts that, when it comes to images and video, we will arrive at a new, lower “trust point.” “‘A picture’s worth a thousand words,’ ‘Seeing is believing’—in the society I grew up in, those were catchphrases that people agreed with,” he said. “I’ve heard people talk about how we might land at a ‘zero trust’ model, where by default you believe nothing. That could be a difficult thing to recover from.”

As with today’s text-based fake news, the problem is double-edged. Having been deceived by a fake video, one begins to wonder whether many real videos are fake. Eventually, skepticism becomes a strategy in itself. In 2016, when the “Access Hollywood” tape surfaced, Donald Trump acknowledged its accuracy while dismissing his statements as “locker-room talk.” Now Trump suggests to associates that “we don’t think that was my voice.”

“The larger danger is plausible deniability,” Farid told me. It’s here that the comparison with counterfeiting breaks down. No cashier opens up the register hoping to find counterfeit bills. In politics, however, it’s often in our interest not to believe what we are seeing.

As alarming as synthetic media may be, it may be more alarming that we arrived at our current crises of misinformation—Russian election hacking; genocidal propaganda in Myanmar; instant-message-driven mob violence in India—without it. Social media was enough to do the job, by turning ordinary people into media manipulators who will say (or share) anything to win an argument. The main effect of synthetic media may be to close off an escape route from the social-media bubble. In 2014, video of the deaths of Michael Brown and Eric Garner helped start the Black Lives Matter movement; footage of the football player Ray Rice assaulting his fiancée catalyzed a reckoning with domestic violence in the National Football League. It seemed as though video evidence, by turning us all into eyewitnesses, might provide a path out of polarization and toward reality. With the advent of

synthetic media, all that changes. Body cameras may still capture what really happened, but the aesthetic of the body camera—its claim to authenticity—is also a vector for misinformation. “Eyewitness video” becomes an oxymoron. The path toward reality begins to wash away.

In the early days of photography, its practitioners had to argue for its objectivity. In courtrooms, experts debated whether photos were reflections of reality or artistic products; legal scholars wondered whether photographs needed to be corroborated by witnesses. It took decades for a consensus to emerge about what made a photograph trustworthy. Some technologists wonder if that consensus could be reestablished on different terms. Perhaps, using modern tools, photography might be rebooted.

Truepic, a startup in San Diego, aims at producing a new kind of photograph—a verifiable digital original. Photographs taken with its smartphone app are uploaded to its servers, where they enter a kind of cryptographic lockbox. “We make sure the image hasn’t been manipulated in transit,” Jeffrey McGregor, the company’s C.E.O., explained. “We look at geolocation data, at the nearby cell towers, at the barometric-pressure sensor on the phone, and verify that everything matches. We run the photo through a bunch of computer-vision tests.” If the image passes muster, it’s entered into the Bitcoin and Ethereum blockchain. From then on, it can be shared on a special Web page that verifies its authenticity. Today, Truepic’s biggest clients are insurance companies, which allow policyholders to take verified photographs of their flooded basements or broken windshields. The software has also been used by N.G.O.s to document human-rights violations, and by workers at a construction company in Kazakhstan, who take “verified selfies” as a means of clocking in and out. “Our goal is to expand into industries where there’s a ‘trust gap,’ ” McGregor said: property rentals, online dating. Eventually, he hopes to integrate his software into camera components, so that “verification can begin the moment photons enter the lens.”

Earlier this year, Danielle Citron and Robert Chesney, law professors at the Universities of Maryland and Texas, respectively, published an article titled “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,” in which they explore the question of whether certain kinds of synthetic media might be made illegal. (One plausible path, Citron told me, is to outlaw synthetic media aimed at inciting violence; another is to adapt the law against impersonating a government official so that it applies to synthetic videos depicting them.) Eventually, Citron and Chesney indulge in a bit of sci-fi speculation. They imagine the “worst-case scenario,” in which deepfakes prove ineradicable and are used for electioneering, blackmail, and other nefarious purposes. In such a world, we might record ourselves constantly, so as to debunk synthetic media when it emerges. “The vendor supplying such a service and maintaining the resulting data would be in an extraordinary position of power,” they write; its database would be a tempting resource for law-enforcement agencies. Still, if it’s a choice between surveillance and synthesis, many people may prefer to be surveilled. Truepic, McGregor told me, had already had discussions with a few political campaigns. “They say, ‘We would use this to just document everything for ourselves, as an insurance policy.’ ”

One evening, Efros and I walked to meet Farid for dinner at a Japanese restaurant near campus. On the way, we talked about the many non-nefarious applications of image synthesis. A robot, by envisioning what it might see around a corner and discovering whether it had guessed right, could learn its way around a building; “pose detection” could allow it to learn motions by observing them. “Prediction is really the hallmark of intelligence,” Efros said, “and we are constantly predicting and hallucinating things that are not actually visible.” In a sense, synthesizing is simply imagining. The apparent paradox of Farid’s license-plate research—that unreal images can help us read real ones—just reflects how thinking works. In this respect, deepfakes were sparks thrown off by the project of building A.I. “When I see a face,” Efros continued, “I don’t know for sure what it looks like from the side. . . .” He paused. “You know what? I think I screwed up.” We had gotten lost.

When we found the restaurant, Farid, who had come on his motorcycle, was waiting for us, wearing a snazzy leather jacket. Efros and Farid—the generator and the discriminator—embraced. They have known each other for a decade.

We took a small table by the window. “What’s really interesting about these technologies is how quickly they went from ‘Whoa, this is really cool’ to ‘Holy crap, this is subverting democracy,’ ” Farid said, over a seaweed salad.

“I think it’s video,” Efros said. “When it was images, nobody cared.”

“Trump is part of the equation, too, right?” Farid asked. “He’s creating an atmosphere where you shouldn’t believe what you read.”

“But Putin—my dear Putin!—his relationship with truth is amazing,” Efros said. “Oliver Stone did a documentary with him, and Putin showed Stone a video of Russian troops attacking ISIS in Syria. Later, it turned out to be footage of Americans in Iraq.” He grimaced, reaching for some sushi. “A lot of it is not faking data—it’s misattribution. On Russian TV, they say, ‘Look, the Ukrainians are bombing Donetsk,’ but actually it’s footage from somewhere else. The pictures are fine. It’s the label that’s wrong.”

Over dinner, Farid and Efros debated the deep roots of the fake-news phenomenon. “A huge part of the solution is dealing with perverse incentives on social media,” Farid said. “The entire business model of these trillion-dollar companies is attention engineering. It’s poison.” Efros wondered if we humans were evolutionarily predisposed to jump to conclusions that confirmed our own views—the epistemic equivalent of content-aware fill.

As another round of beer arrived, Farid told a story. Many years ago, he said, he’d published a paper about a famous photograph of Lee Harvey Oswald. The photograph shows Oswald standing in his back yard, holding the rifle he later used to kill President Kennedy; conspiracy theorists have long claimed that it’s a fake. “It kind of does look fake,” Farid said. The rifle appears unusually long, and Oswald seems to be leaning back into space at an unrealistic angle; in this photograph, but not in others, he has a strangely narrow chin. “We built this 3-D model of the scene,” Farid said, “and it turned out we could explain everything that people thought was wrong—it was just that the light was weird. You’d think people would be, like, ‘Nice job, Hany.’ ”

Efros laughed.

“But no! When it comes to conspiracies, there are the facts that prove our beliefs and the ones that are part of the plot. And so I became part of the conspiracy. At first, it was just me. Then my father sent me an e-mail. He said, ‘Someone sent me a link to an article claiming that you and I are part of a conspiracy together.’ My dad is a research chemist who made his career at Eastman Kodak. Well, it turns out he was at Eastman Kodak at the same time they developed the Zapruder film.”

“Ahhhhh,” Efros said.

For a moment, they were silent. “We’re going to need technological solutions, but I don’t think they’re going to solve the problem,” Farid said. “And I say that as a technologist. I think it’s a societal problem—a human problem.”

On a brisk Friday morning, I walked to Efros’s lab to see my synthetic self. The Berkeley campus was largely empty, and I couldn’t help noticing how much it resembled other campuses—the texture of college is highly consistent. Already, the way I looked at the world was shifting. That morning, on my phone, I’d watched an incredible video in which a cat scaled the outside of an apartment building, reached the tenth floor, then leaped to the ground and scampered away. Automatically, I’d assumed the video was fake. (I Googled; it wasn’t.)

A world saturated with synthesis, I'd begun to think, would evoke contradictory feelings. During my time at Berkeley, the images and videos I saw had come to seem distant and remote, like objects behind glass. Their clarity and perfection looked artificial (as did their gritty realism, when they had it). But I'd also begun to feel, more acutely than usual, the permeability of my own mind. I thought of a famous study in which people saw doctored photographs of themselves. As children, they appeared to be standing in the basket of a hot-air balloon. Later, when asked, some thought they could remember actually taking a balloon ride. It's not just that what we see can't be unseen. It's that, in our memories and imaginations, we keep seeing it.

At a small round table, I sat down with Shiry Ginosar and another graduate student, Tinghui Zhou, a quietly amused man with oblong glasses. They were excited to show me what they had achieved using a GAN that they had developed over the past year and a half, with an undergraduate named Caroline Chan. (Chan is now a graduate student in computer science at M.I.T.)

"O.K.," Ginosar said. On her laptop, she opened a video. In a box in the upper-left corner of the screen, the singer Bruno Mars wore white Nikes, track pants, and an elaborately striped shirt. Below him, a small wireframe figure imitated his posture. "That's our pose detection," she said. The right side of the screen contained a large image of me, also in the same pose: body turned slightly to the side, hips cocked, left arm raised in the air.

Ginosar tapped the space bar. Mars's hit song "That's What I Like" began to play. He started dancing. So did my synthetic self. Our shoulders rocked from left to right. We did a semi-dab, and then a cool, moonwalk-like maneuver with our feet.

"Jump in the Cadillac, girl, let's put some miles on it!" Mars sang, and, on cue, we mimed turning a steering wheel. My synthetic face wore a huge grin.

"This is amazing," I said.

"Look at the shadow!" Zhou said. It undulated realistically beneath my synthetic body. "We didn't tell it to do that—it figured it out." Looking carefully, I noticed a few imperfections. My shirt occasionally sprouted an extra button. My wristwatch appeared and disappeared. But I was transfixed. Had Bruno Mars and I always had such similar hair? Our fingers snapped in unison, on the beat.

Efros arrived. "Oh, very nice!" he said, leaning in close and nodding appreciatively. "It's very good!"

"The generator tries to make it look real, but it can look real in different ways," Ginosar explained.

"The music helps," Efros said. "You don't notice the mistakes as much."

The song continued. "Take a look in that mirror—now tell me who's the fairest," Mars suggested. "Is it you? Is it me? Say it's us and I'll agree!"

"Before Photoshop, did everyone believe that images were real?" Zhou asked, in a wondering tone.

"Yes," Ginosar said. "That's how totalitarian regimes and propaganda worked."

"I think that will happen with video, too," Zhou said. "People will adjust."

"It's like with laser printers," Efros said, picking up a printout from the table. "Before, if you got an official-looking envelope with an official-looking letter, you'd treat it seriously, because it was beautifully typed. Must be the government, right? Now I toss it out."

Everyone laughed.

“But, actually, from the very beginning photography was never objective,” Efros continued. “Whom you photograph, how you frame it—it’s all choices. So we’ve been fooling ourselves. Historically, it will turn out that there was this weird time when people just assumed that photography and videography were *true*. And now that very short little period is fading. Maybe it should’ve faded a long time ago.”

When we’d first spoken on the phone, several weeks earlier, Efros had told me a family story about Soviet media manipulation. In the nineteen-forties and fifties, his grandmother had owned an edition of the Great Soviet Encyclopedia. Every so often, an update would arrive in the mail, containing revised articles and photographs to be pasted over the old ones. “Everyone knew it wasn’t true,” Efros said. “Apparently, that wasn’t the point.”

I mulled this over as I walked out the door, down the stairs, and into the sun. I watched the students pass by, with their identical backpacks, similar haircuts, and computable faces. I took out my phone, found the link to the video, and composed an e-mail to some friends. “This is so great!” I wrote. “Check out my moves!” I hit Send. ♦



WATCH

How to Make a Synthetic Self

Using artificial intelligence, researchers can project the movements of one body onto another’s.

Published in the print edition of the November 12, 2018, issue, with the headline “Afterimage.”



Joshua Rothman, the ideas editor of newyorker.com, has been a writer and an editor at the magazine since 2012.

