



Deepfake real value

Flipping the script on
deepfake technologies



It's early 2020, and digital artist Denis Shiryaev has just uploaded a new video to YouTube. It seems relatively simple, just a short video of a train pulling into a station stop. But something's odd. The clip is from "The Arrival of a Train at La Ciotat Station," a film recorded in 1895 in 35mm, but it's appearing in 4K resolution at 60 frames per second.

It's March 2019, and an executive at an energy firm based in the United Kingdom gets a call from his boss, asking him to quickly transfer €220,000 to a Hungarian vendor. He transfers the money to the bank account specified by his boss—who turns out not, in fact, to be his boss. The voice he heard was a sophisticated fake of his real boss's voice, convincing enough to con the company out of a significant amount of real money.¹

It's 2021, and someone is playing *Grand Theft Auto V*—but not the original version. The player is navigating a version of the open world game generated by a neural network as he plays, with the network first trained using playthroughs of the real game.² The movement of the car, the shadows it creates, the shifting background as the player navigates—everything is generated automatically.

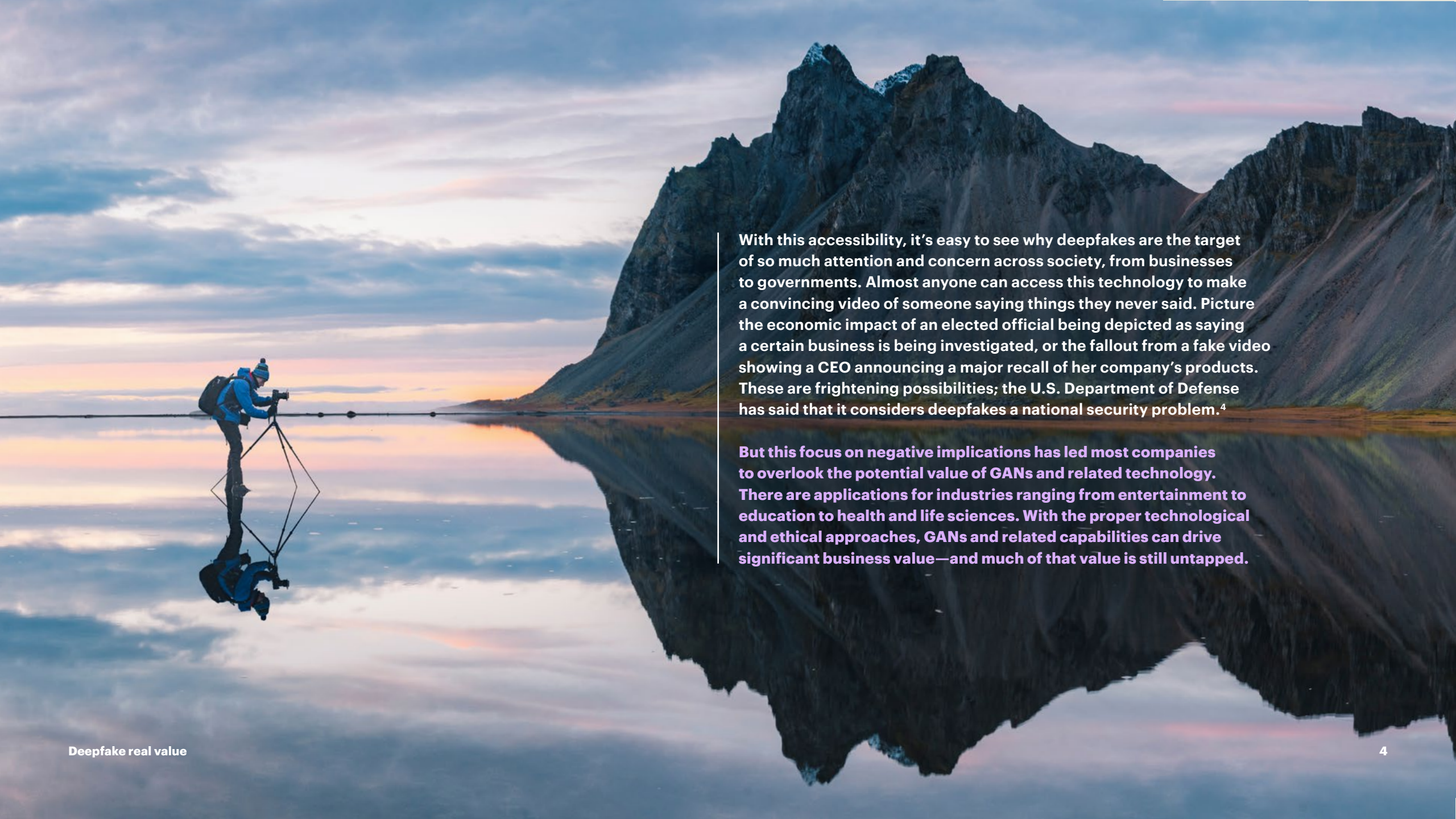
All of these real efforts were powered by the same underlying technology: generative adversarial networks (GANs), a machine learning framework. (The researchers who worked with *Grand Theft Auto* call their version GAN Theft Auto.)

GANs are called adversarial because two neural networks are set up to essentially try to beat each other at a game. A “generator” neural net creates content that goes to the second, “evaluator” neural net. The first tries to generate data that fools the other, and the other tests the first's output for realism. If the evaluator can tell the difference between a real example and a fake one, the generator tries again. This continues until it becomes difficult for the evaluator to correctly identify the generated cases.

If you've seen GANs mentioned in the media, it's probably been in the context of their use in deepfakes. GANs have famously been used to generate fake videos of real people saying or doing things they haven't done. In one well known video, an anonymous YouTube creator altered a video of a David Letterman interview with comedian Bill Hader. The original interview shows Hader

doing impressions of actors Seth Rogen and Tom Cruise. In the altered video, created using cloud computing platform Paperspace, Hader's face morphs from his own into Rogen's and Cruise's as he does impressions of them, while his voice remains the same. It's done so smoothly you could miss it if you weren't watching carefully.

That application was merely a novelty. But it demonstrates how easy “DIY deepfakes” have become. In an interview with *The Guardian*, the creator explained that his work with deepfakes was entirely self-taught.³ It requires a computer with sufficient processing power, but ultimately, he used an existing machine learning platform that's available to everyone.



With this accessibility, it's easy to see why deepfakes are the target of so much attention and concern across society, from businesses to governments. Almost anyone can access this technology to make a convincing video of someone saying things they never said. Picture the economic impact of an elected official being depicted as saying a certain business is being investigated, or the fallout from a fake video showing a CEO announcing a major recall of her company's products. These are frightening possibilities; the U.S. Department of Defense has said that it considers deepfakes a national security problem.⁴

But this focus on negative implications has led most companies to overlook the potential value of GANs and related technology. There are applications for industries ranging from entertainment to education to health and life sciences. With the proper technological and ethical approaches, GANs and related capabilities can drive significant business value—and much of that value is still untapped.

Generating value

At its core, generative network technology is about creating realistic synthetic data. In the case of deepfakes, the goal is to create audio or video that can fool human viewers. But synthetic data can generate value in many enterprise applications, across many industries.

Product development

In the consumer packaged goods industry, testing new product formulations is time-consuming and costly. Changing regulations, shifting consumer preferences, and broken supply chains all create pressure to change the ingredients used in products, but each new possible “recipe” requires substantial testing. Experts must select, process and combine ingredients to deliver specific properties, functionality and performance, and then physically test each combination. A product like toothpaste can be based on dozens of ingredients selected from thousands of potential components.

Labs has used synthetic data to speed up this process while exploring more possible formulations than before. Just as GANs can generate deepfake videos that are

“believable”—where the required metric is **realistic** data—generative models can uncover surprising new combinations that are likely to meet performance requirements for a product. With this approach, generative models can help guide product developers toward new and surprising formulations that may have been previously overlooked, or exist in areas that have yet to be explored. Just like with deepfakes, the models are trained on real observations and learn to generate synthetic, but realistic, data.

Better training for AI systems

This type of GAN-generated synthetic data can also be used to improve AI systems for a wide range of applications. As an example, take voice or image recognition. Machine learning systems like those used to understand human speech or recognize faces in photographs need data to train on: recordings of speech, or photographs that contain different faces. Ideally, that data will also be labeled. For example, a recording of a particular word might be labeled as being that word.

Sometimes, though, there simply aren't enough data available, or the data aren't labeled. With languages, it's easy to see why this might be the case. For common languages like English there are many large (and labeled) datasets available to researchers. But less common languages, like Swahili, don't have such datasets available. And if you think about use cases for voice recognition technology—like enabling conversation with those who

speak different languages when human translators aren't available—being able to understand those less common languages becomes quite important. So how can we train machine learning systems to perform well when presented with them, or with other limited or unlabeled datasets?

GANs offer a solution here as well. Facebook AI's [Wav2vec-U method](#) uses a GAN to train speech recognition models with unlabeled data. The generator network tries to predict distinct units of sound for each audio segment, and the evaluator network looks at whether the predicted sequences look realistic. The two networks go back and forth and, even with limited, unlabeled data, the system's transcriptions become impressively accurate over time. They're still not as good as those created by systems that use labeled data, but they show that GANs offer a path to useful machine learning capabilities even when data is limited.

GANs have also been used extensively to generate synthetic data for the autonomous driving space. Self-driving vehicles need to be trained on millions of scenarios to be able to operate safely. Some scenarios would come up frequently in real test drives, like pedestrians in a crosswalk. But other scenarios—like something falling off a truck in front of the vehicle while it drives on the highway—are so rare that a vehicle might never “see” them live, even after thousands of miles on the road. Synthetic data has become a key element of training such systems so that they're ready to respond to “edge case” scenarios. It can also be used to augment real road test data to reflect varied conditions. Rather than collecting different versions for different seasons, lighting conditions, traffic conditions, weather, and so on, real data can be augmented to train systems for the different possibilities.

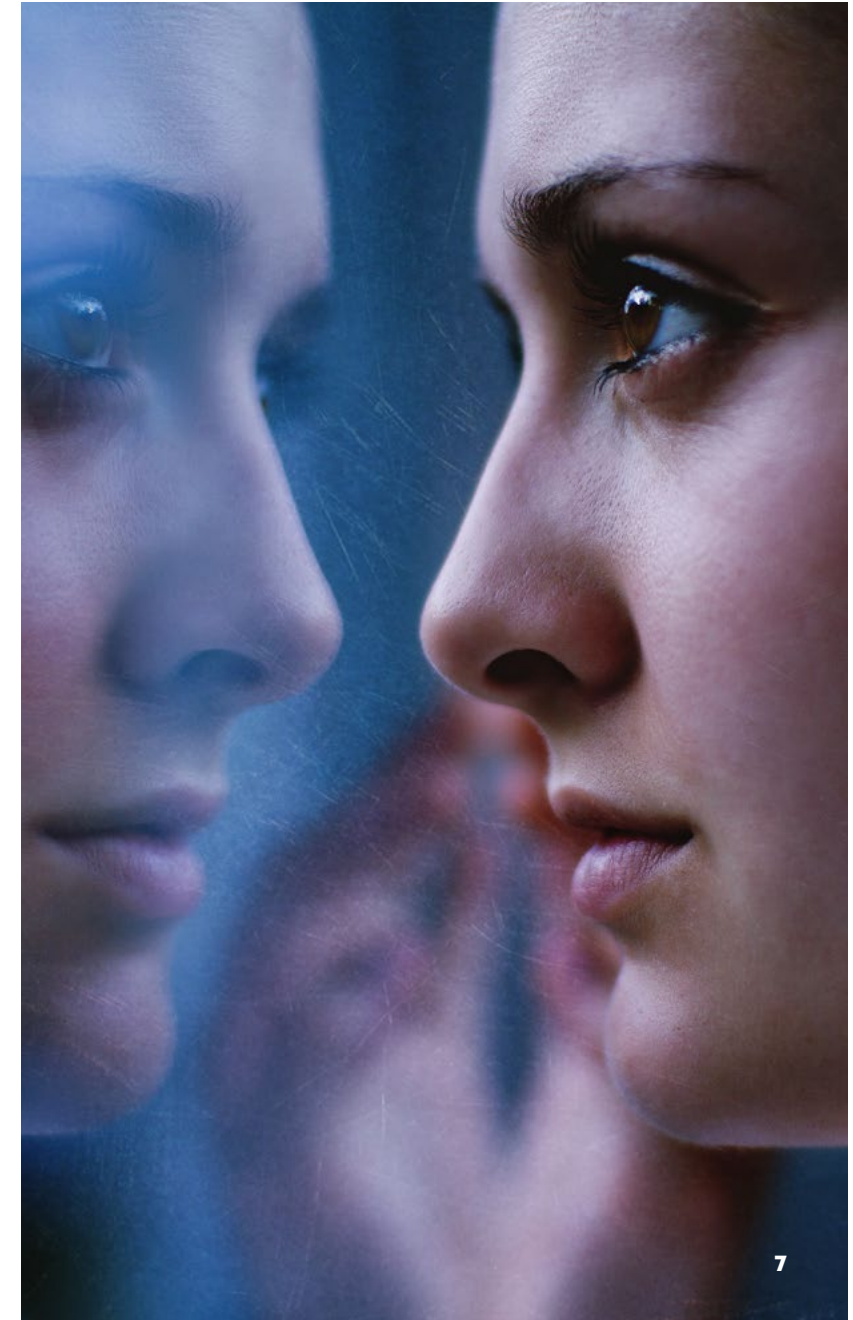
Synthetic data, real privacy

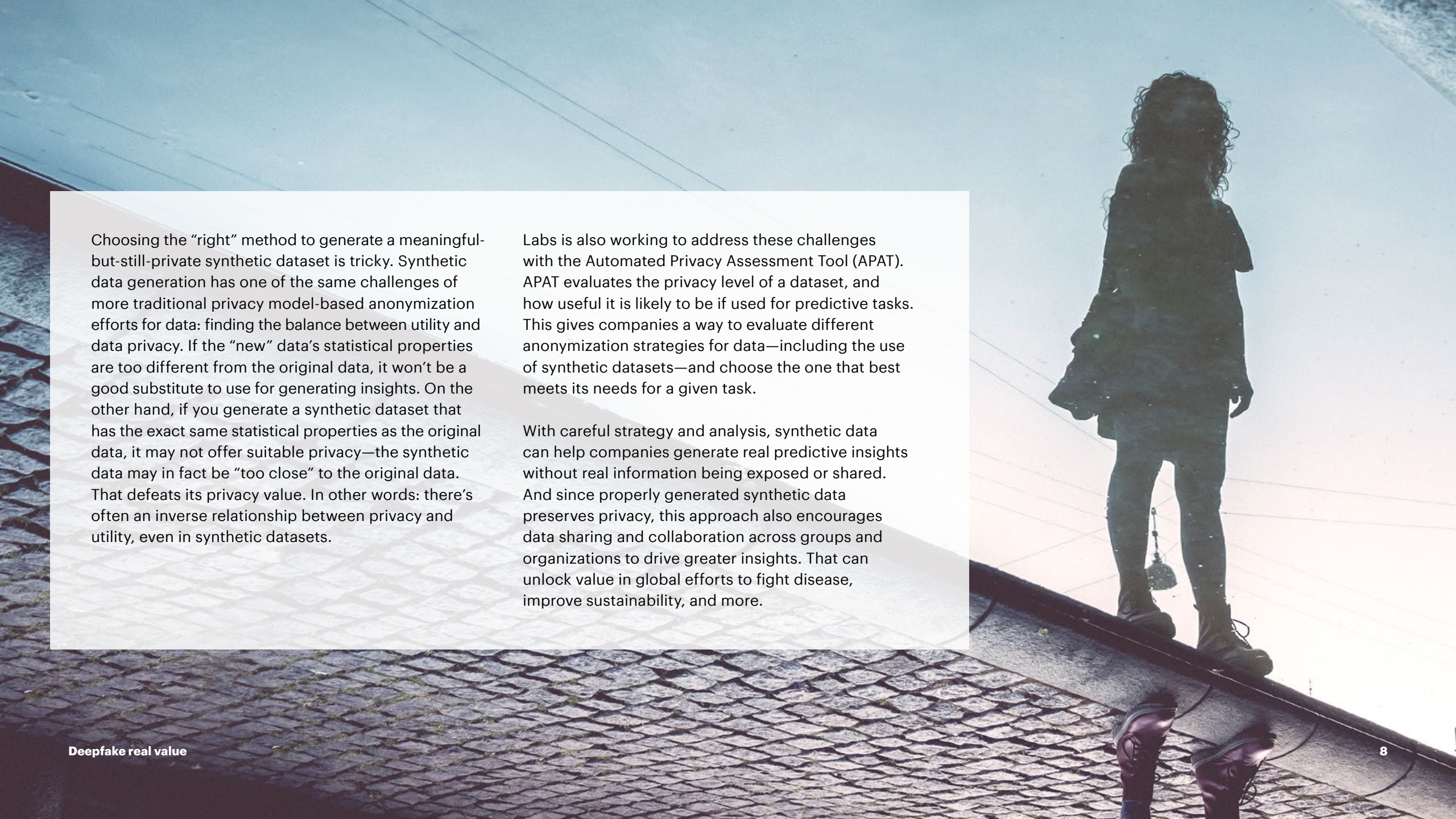
When it comes to data, sometimes it's not a question of scarcity but practicality. There are many scenarios where valuable data is available to businesses, but where using it is problematic.

The healthcare space may be the most prominent of these, where companies must comply with regulatory requirements around healthcare information (like HIPAA), along with data privacy regulations like GDPR.

Enter synthetic data. It's possible to generate a larger synthetic dataset that retains the valuable statistical information from the real dataset but doesn't contain the original data. When synthetic datasets are properly generated, organizations can use them for modeling and analysis and be confident of the results—and, critically, maintain trust with those whose data they are able to access.

There are several different methods available to generate synthetic data from existing datasets. They range from more general methods to those specifically focused on a given domain (e.g., the medGAN and ehrGAN frameworks are focused on generating or augmenting medical records). Each of these methods has advantages and disadvantages for use in different situations. Accenture Labs has explored many of them to compare machine learning performance using datasets they've generated, based on accuracy, precision, sensitivity, and specificity.





Choosing the “right” method to generate a meaningful-but-still-private synthetic dataset is tricky. Synthetic data generation has one of the same challenges of more traditional privacy model-based anonymization efforts for data: finding the balance between utility and data privacy. If the “new” data’s statistical properties are too different from the original data, it won’t be a good substitute to use for generating insights. On the other hand, if you generate a synthetic dataset that has the exact same statistical properties as the original data, it may not offer suitable privacy—the synthetic data may in fact be “too close” to the original data. That defeats its privacy value. In other words: there’s often an inverse relationship between privacy and utility, even in synthetic datasets.

Labs is also working to address these challenges with the Automated Privacy Assessment Tool (APAT). APAT evaluates the privacy level of a dataset, and how useful it is likely to be if used for predictive tasks. This gives companies a way to evaluate different anonymization strategies for data—including the use of synthetic datasets—and choose the one that best meets its needs for a given task.

With careful strategy and analysis, synthetic data can help companies generate real predictive insights without real information being exposed or shared. And since properly generated synthetic data preserves privacy, this approach also encourages data sharing and collaboration across groups and organizations to drive greater insights. That can unlock value in global efforts to fight disease, improve sustainability, and more.

Artistic enhancement

It's safe to say that GANs open up new possibilities in quantitative research. But they also have a lot to offer in the creative spaces.

In an Accenture artificial intelligence lab, visitors can enjoy a demo of GANs' artistic capabilities. Starting from a photo, a GAN-driven system can create a caricature in line with a particular artistic style. It can also “convert” a known painting to a reimagined render in another style—for example, a “Picasso-ized Van Gogh.”

These novelties are quick, fun ways to demonstrate the power of GANs. In fact, though, they offer a hint at the larger value companies can unlock.

Moving from the physical to the digital space, GANs offer new opportunities in content creation. An anti-malaria PSA with David Beckham, part of the “[Malaria Must Die](#)” campaign, provided an early glimpse of the possibilities.

In a video from the campaign, [Beckham seems to speak nine languages](#), though he was only filmed speaking English. The eight other clips were generated using deepfake technology, allowing the same base message to reach a much wider audience. In a later phase of the same campaign, production company Digital Domain used machine learning to digitally age Beckham.⁵ They blended real footage of him with an older stand-in actor

to generate “footage” of Beckham as an old man, giving a speech on the day that the world has put an end to malaria. This didn't require any extensive data capture; just standard video of Beckham (principal photography).

The possibilities are enormous. Imagine the extended reach that's available if you could shoot a film in just one language but release it in many—without dubbing or subtitles, but with the real actors appearing to speak the language, eventually in their own voices.

Companies in any industry could use the same approach to develop truly global ad campaigns, or for internal purposes, training materials. This would allow for more inclusive communication and wider distribution of content customized for specific geographical and cultural needs.

For organizations whose focus is on content creation, this could offer a path far beyond just “restoring” old films. The technology may eventually support full-on synthetic content creation. In time, we may see aging Hollywood action stars appearing in stunt scenes without having to film them or use a stand-in. Or, ultimately, new Hollywood stars that don’t exist in real life—entirely synthetic actors and actresses.

GANs also offer the opportunity to interact with digital versions of real people in new ways. Perhaps we’ll see individual viewers becoming the stars of the movies they’re watching. Or, as in the Museum of Art and

Photography (MAP) in Bangalore, people can interact with the digital avatar of someone who is no longer with us.

MAP had emphasized the importance of creating a museum-going culture that engaged younger generations. The museum cited that as one of the motivations behind its work with Accenture Labs on a conversational digital persona of deceased artist M. F. Husain. With this exhibit, the museum experience reaches far beyond reading from a pamphlet or following an audio tour. Through a combination of GANs, natural language processing, and emotion detection,

the interactive experience allows visitors to ask questions to “Husain” in their own words. The digital persona responds based on real information from Husain’s life.

Of course, now that we’re talking about digital versions of real people, we have to address some of the same issues that companies already face with deepfakes: ethical, privacy, and security concerns.



A series of illuminated arches at night, reflecting in water. The arches are lit with blue and red lights, creating a vibrant, futuristic scene. The water in the foreground is calm, mirroring the lights and the structure above. The sky is a deep purple, suggesting dusk or dawn.

Preparing for larger impact

Like every technology, generative approaches create risks for business and society. When companies apply them to address business challenges and opportunities, they must do so responsibly. Likewise, they must prepare for the reality that others could use these technologies maliciously.

Responsibility by design

Deepfake audio and videos are already here and becoming more common, as well as easier and cheaper to create. For better or worse, the line between reality and realistic-but-not-real-content will continue to blur.

Will society begin to question whether it can ever trust what it can't see in person? Or worse, whether it can trust anything at all? Like it or not, companies exist in the same reality that's currently being undermined by malicious deepfakes. Organizations must use care, then, when exploring these underlying technologies, even if their own goals in using GANs have nothing to do with deception or misinformation. To maintain trust with consumers, regulators, and the public, companies must incorporate ethical considerations into their decision-making process from the start.

As part of its work in responsible innovation, Accenture Labs teams with Carnegie Mellon University. CMU professor David Danks points out the importance of the “deepfake” terminology itself: “I think the word ‘fake’ is really important. What the AI is being used to do is not just create some whole fictional world,” says Danks, “But something that is deliberately supposed to be presenting a fake image, a fake idea, a fake speech as though it were real, which is part of where all the challenging social implications arise.”⁶

For companies, then, a first question might be: what is the purpose of using a GAN for a particular task? If the goal is in fact to convince viewers or listeners of something that is not real—a true deepfake—there are rampant ethical concerns. If the goal, instead, is to generate synthetic data to preserve the privacy of someone who originally contributed the underlying data, it is perhaps less ethically fraught—but regulatory compliance will still be important.

Most enterprise applications would likely fall somewhere in between these two “ends” of the ethical spectrum. And, in many cases, there won't be a straightforward answer to the question of how best to handle ethical concerns.

Consider the new territory we might find ourselves in as GANs grow ever more sophisticated in their capabilities. If a company “creates” a spokesperson using GANs—a lifelike persona that can speak for the company in a way that's convincingly real—do they need to include a disclosure indicating that it's not a real person? Or is it only valuable as an asset if someone believes it **is** a real person? (European Union regulations proposed in 2021—though not yet adopted as of this writing—would require creators of deepfakes to disclose that the content was artificially generated or manipulated.⁷)

What about GAN-driven “avatars” of real people? David Beckham would have been given the chance to review the GAN-created “versions” of himself delivering an anti-malaria PSA before they were aired. But what about deepfakes that “speak” for people who are no longer with us, like the *Dalí Lives* exhibit at the Dalí Museum in Florida?

Dalí died in 1989, but visitors to the museum could “hear” from his avatar in 2019, and “Dalí” would even offer to snap a selfie with them before they left.⁸ The museum's exhibition was created and launched with permission from the Dalí Foundation; the sole heir named in Dalí's will was the Spanish Kingdom, and he has no living family. But what if he did? Can family members reliably determine what a deceased person would have agreed to “say” or “do”? Who can give consent for the dead to be brought to life in digital form?

Technologically, questions remain even when ethical concerns are addressed. Certification or provenance techniques may become critical when these technologies are used for societal applications, to verify the “authenticity” of content (or at least the veracity of its claims). This isn't a problem that's solved yet, but one that's clearly important to consider.

Dealing with bad actors

Companies can and must use care when applying GANs to drive value. But, as with any technology, bad actors can take advantage of these capabilities as well.

We've already seen deepfake technology used to trick an employee into sending money to a scammer. In some ways that's just the logical progression of long-running phishing scams. Rather than an email claiming to be from your boss, asking you to buy a bunch of gift cards and send him the numbers, it's a voicemail from your boss.

We're conditioned to recognize people's voices and acknowledge them as evidence that we know who we're talking to; some financial organizations even use your voice to verify your identity when you call. We trust

voices. It's logical, then, that scammers would embrace a technology that lets them spoof a person's voice for fraudulent purposes. And given our growing reliance on video calling and even video messaging, it's not hard to imagine that fake video messages could be next—whether for targeted phishing or scam attempts, or simply to spread misinformation.

Many organizations are working on deepfake detection techniques to help mitigate these and other threats. Not surprisingly, researchers have turned to other AI systems to try to detect deepfaked content, but with mixed success. Accenture's Cyber Lab developed an ensemble of AI models—a mix of previously well-known methods and novel approaches—to analyze content. Each of the models

learn different features of the content being analyzed, which minimizes the chances that the results will focus on irrelevant features of the video being scrutinized. Once each model has done its job, the solution calculates the likelihood that the content being examined is the result of deepfake technology.

Organizations can apply techniques like these to monitor for deepfake content on distribution channels under their control, like email or collaboration platforms. Of course, a significant amount of misinformation also makes its way into the public eye via external platforms. That's why companies like Facebook are deeply interested in preventing the spread of deepfakes as well.

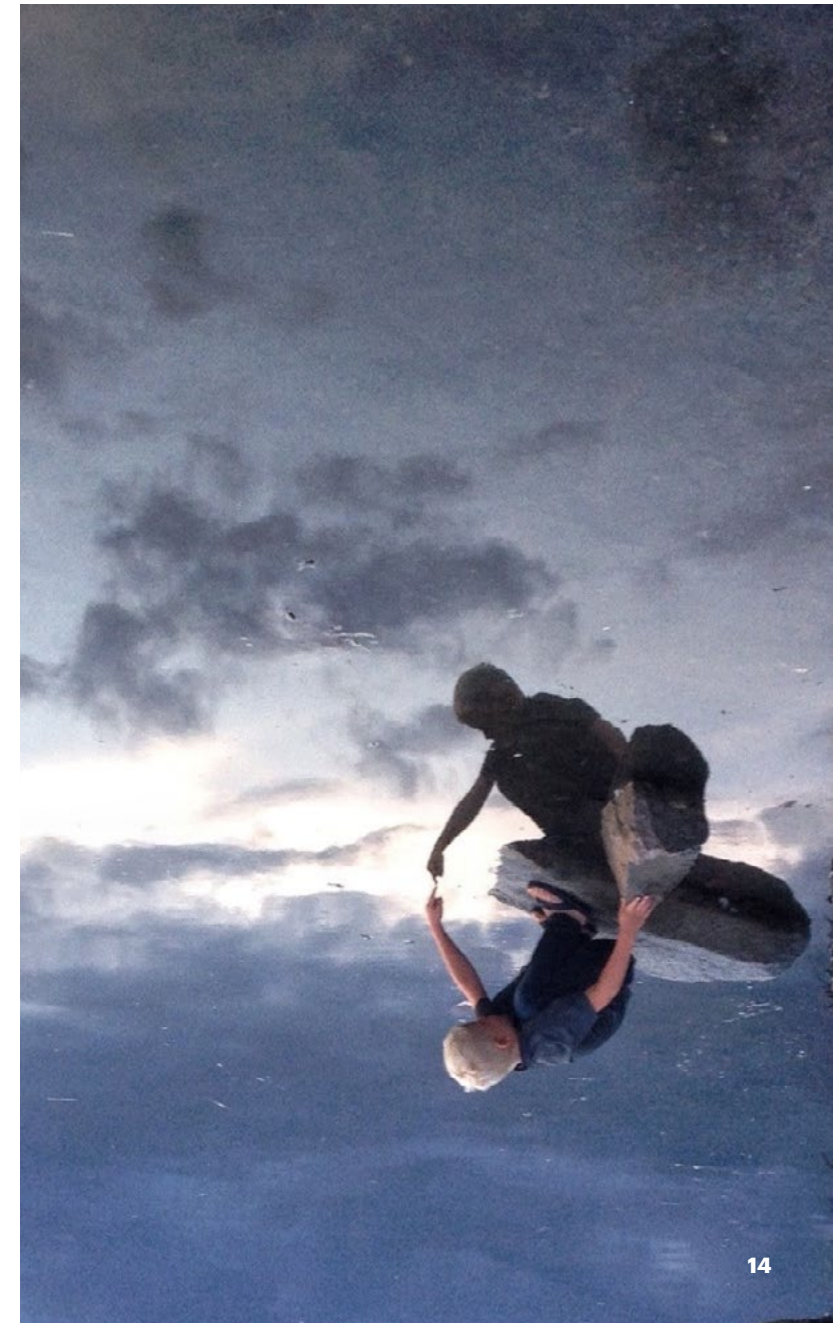
The company recently collaborated with researchers at Michigan State University in an effort that reaches beyond detecting deepfaked images and into identifying their source.⁹ Prior research had been able to identify which of a list of known AI models had generated a particular deepfaked image. In this latest effort, however, the researchers from MSU and Facebook were also able to pinpoint markers associated with as-yet unknown AI models—identifying what’s essentially a digital fingerprint left on the deepfaked video by the model that created it. This approach is not only useful in detecting a deepfake in the first place, it gives the company a forensic method to work backward and find the original source of faked content.

That’s important, because as CMU’s Danks says, there is no one perfect solution to the challenge of identifying or preventing negative impact from deepfakes. It will take multiple efforts from stakeholders across industries

and geographies working together to address these challenges—and progress will be incremental.

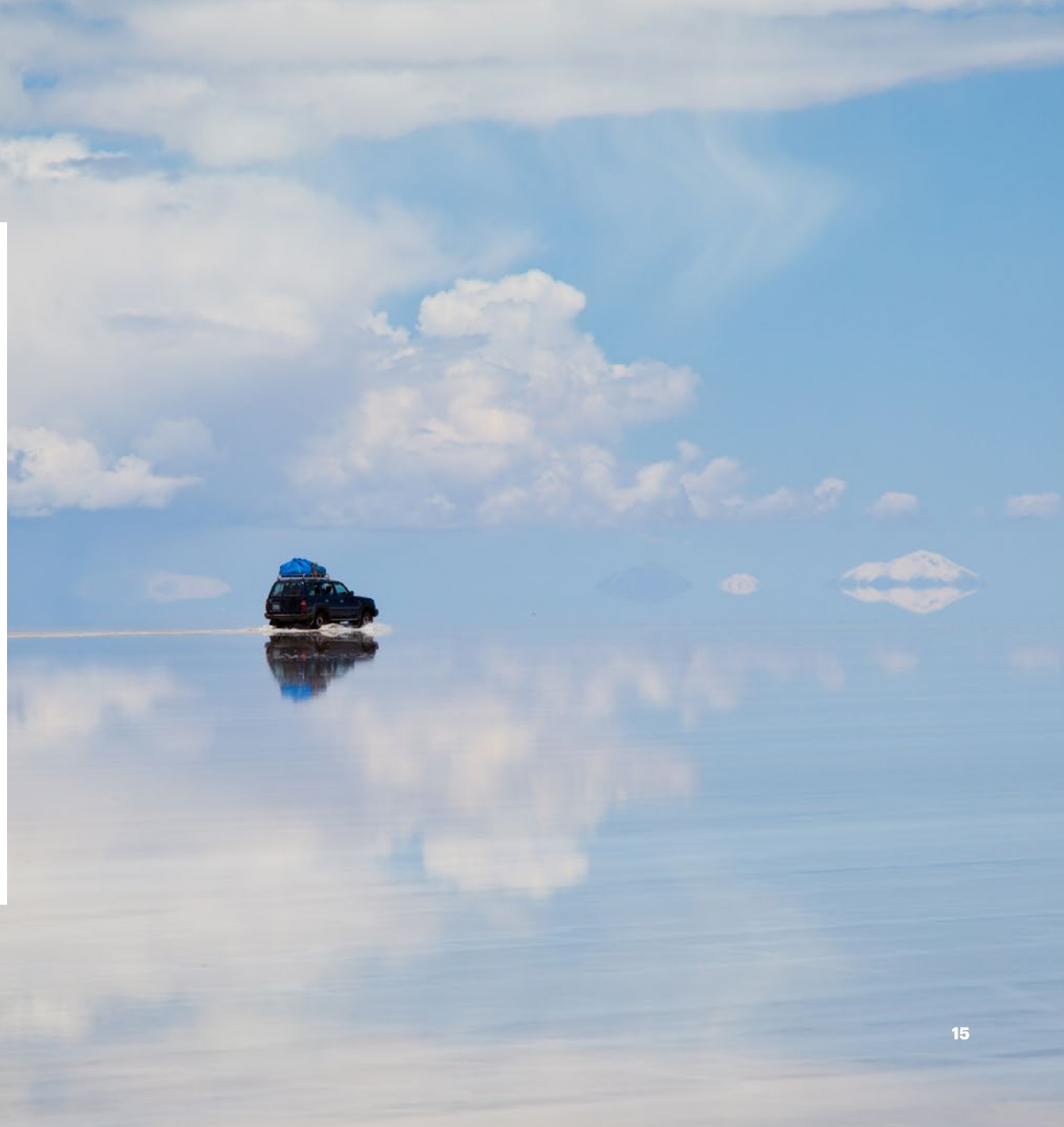
“Don’t let the perfect be the enemy of the good,” says Danks. “The very first step that’s taken is not going to solve the problem with deepfakes in one fell swoop. Instead, we need to think, ‘How can we make some progress? How do we get the worst of the worst out of the system, and then how do we get a little bit better?’ Recognizing that it’s a dynamic system, it’s going to take time.

“If we do it in a knowledge-sharing environment, I think that we can start to make real progress,” says Danks, Without it just being the problem where everybody just throws up their hands and says ‘It’s somebody else’s job to fix this.’



Looking forward

Every technological innovation has potential for abuse. Given how heavily our societies and business models rely on information, deepfakes perhaps have more potential for misuse than most. It's shortsighted, though, to focus only on the potential for negative impact—and businesses doing so are leaving value on the table. We must address deepfakes, to be sure. And we need to think through the ethical implications whenever we apply GANs and other underlying methods, just as we should with any technology. But businesses have much to gain from creative, thoughtful use of deepfake technologies. In everything from product development to healthcare to the entertainment industry, synthetic data offers real value. How will you capture it?



Contacts



Marc Carrel-Billiard

Senior Managing Director,
Accenture Technology
Innovation and Accenture Labs

marc.carrel-billiard@accenture.com



Edy Liongosari

Growth and Strategy Lead,
Accenture Technology
Innovation

edy.s.liongosari@accenture.com

Contributors

Manish Ahuja

Malek Ben Salem

Ronak Bhatia

Medb Corcoran

Luca Costabello

Laura Degioanni

Neville Dubash

Andy Fano

Jer Hayes

Laetitia Kameni

Alex Kass

Mike Kuniavsky

Neil Liberman

Rory McGrath

Lisa O'Connor

Sanjay Podder

Nisha Ramachandra

Shubhashis Sengupta

Michelle Sipics

Dylan Snow

Steven Tiell

Richard Vidal

Xu Zheng

About Accenture

Accenture is a global professional services company with leading capabilities in digital, cloud and security. Combining unmatched experience and specialized skills across more than 40 industries, we offer Strategy and Consulting, Interactive, Technology and Operations services — all powered by the world’s largest network of Advanced Technology and Intelligent Operations centers. Our 569,000 people deliver on the promise of technology and human ingenuity every day, serving clients in more than 120 countries. We embrace the power of change to create value and shared success for our clients, people, shareholders, partners and communities. Visit us at www.accenture.com

About Accenture Labs

Accenture Labs incubates and prototypes new concepts through applied R&D projects that are expected to have a significant impact on business and society. Our dedicated team of technologists and researchers work with leaders across the company and external partners to imagine and invent the future. Accenture Labs is located in seven key research hubs around the world: San Francisco, CA; Washington, D.C.; Dublin, Ireland; Sophia Antipolis, France; Herzliya, Israel; Bangalore, India; Shenzhen, China and Nano Labs across the globe. The Labs collaborates extensively with Accenture’s network of nearly 400 innovation centers, studios and centers of excellence to deliver cutting edge research, insights and solutions to clients where they operate and live. For more information, please visit www.accenture.com/labs

This content is provided for general information purposes and is not intended to be used in place of consultation with our professional advisors.

Copyright © 2021 Accenture. All rights reserved. Accenture and its logo are registered trademarks of Accenture.

This document refers to marks owned by third parties. All such third-party marks are the property of their respective owners. No sponsorship, endorsement or approval of this content by the owners of such marks is intended, expressed or implied.

References

1. <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>
2. <https://github.com/sentdex/GANTheftAuto/>
3. <https://www.theguardian.com/news/shortcuts/2019/aug/13/danger-deepfakes-viral-video-bill-hader-tom-cruise>
4. <https://www.c4isrnet.com/information-warfare/2019/08/29/how-the-pentagon-is-tackling-deepfakes-as-a-national-security-problem/>
5. <https://www.globenewswire.com/news-release/2020/12/03/2139202/0/en/Digital-Domain-Adds-Decades-to-David-Beckham-for-New-Malaria-Must-Die-Campaign.html>
6. <https://www.youtube.com/watch?v=GIM8xPotAkg>
7. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>
8. <https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum>
9. <https://ai.facebook.com/blog/reverse-engineering-generative-model-from-a-single-deepfake-image/>