Trend 03

The Unreal

Making Synthetic, Authentic

FEDERAL TECHNOLOGY VISION 2022

Accenture Federal Services

THE BIG PICTURE

In mid-March, a minute-long video appeared on many major social media sites in which Ukrainian President Volodymyr Zelenskyy, motionless and expressionless, announced his surrender to Russia's invading forces and asked Ukrainian troops to lay down their weapons.

The problem, though, was that Zelenskyy never surrendered nor made such a video. The video was a "deepfake" — a fabricated piece of media created by using artificial intelligence (AI).

This was one of the earliest examples of a weaponized deepfake, intentionally created and circulated on social media to sow confusion and mistrust during war.

While the precise objectives and origin of the fake video remain uncertain, it ultimately had little impact in this instance: the Ukrainian government was anticipating such a tactic and warned the world well before it surfaced. The technical quality of the deepfake was poor and many people could see it for what it was. Once the deepfake appeared, Zelenskyy himself quickly issued a real video saying that Ukraine planned to continue fighting on. Major social media sites quickly took the fake video down.

Just

35%

of consumers are confident or very confident they can recognize or identify deepfake videos or synthetic content.

Nevertheless, the precedent was made and the many potentially malicious implications of deepfakes — especially as their quality improves over time — are becoming clearer to see. But while this example highlights the perils of synthetic data and generative AI, it is important that we not overlook the many opportunities they present us as well.

In fact, it is no exaggeration to say that the continued advancement of AI is highly dependent on the use of synthetic data. That is because AI relies upon extensive, well-defined data to function properly — something frequently in short supply.

Perhaps it is worth pausing here to reflect on just how important AI is becoming to us. Federal agencies use AI today to accelerate regulatory reform; combat fraud, waste, and abuse; identify information security threats; enhance the security and interoperability of information systems; streamline processes for grant applications; model weather patterns; facilitate predictive maintenance; and much more.[1]

85%

of U.S. federal government executives report that their organization is dependent on AI technologies to function effectively.

Or consider, for a moment, the many ways AI has been instrumental in helping us tackle COVID-19. Convolutional neural network (CNN)-based models have helped detect and classify in real time COVID-19 infections in patients employing chest X-ray images. Researchers even developed a fully automatic deep learning system to generate a diagnosis and prognosis of COVID-19 using computed tomography (CT) scans. AI-based methods have been used to track COVID-19 spread over time and place, to conduct disease surveillance by scanning public spaces for people with potential COVID-19 infection, and to enforce various social distancing measures or lockdowns. An AI-enabled mobile app, called Aarogya Setu and launched by the government of India, allows users to check their safety status based on whether they have crossed paths with Covid-19 positive patients.[2] Most importantly, AI helped researchers dramatically accelerate the genomic sequencing of SARS-CoV-2 and its variants and develop potential treatment approaches and even vaccines.[3]

A critical role for synthetic data in our COVID-19 response

But many of these developments could not have been possible without synthetic data. For example, many researchers needed information about how the SARS-CoV-2 virus and its evolving variants, such as delta and omicron, were affecting the human body and public health. Much of this data is collected in patients' electronic medical records. But researchers typically face delays or barriers obtaining national data from medical records because of concerns about preserving the privacy of those patients.

Synthetic data enables researchers to get around such obstacles. Data for a wide array of COVID research, for example, is artificially generated and informed — though not directly derived — by actual patient data. To generate synthetic data, researchers produce an entirely new set of simulated patients that, in aggregate, recreate the exact statistical characteristics of real patient data.

For example, synthetic data might be generated to mimic the aggregated blood pressure, body mass index, and kidney function of a set of real patient data. But the real patients' identities and privacy are protected because the simulated patients have no direct counterparts in the real data.

Understanding these constraints on the use of real patient data, the National Institutes of Health (NIH) in 2021 partnered with the California-based startup Syntegra to generate and validate a nonidentifiable replica of the NIH's extensive database of COVID-19 patient records, called the National COVID Cohort Collaborative (N3C) Data Enclave.[4] Today, N3C consists of more than 5 million COVID-positive individuals. The synthetic data set precisely duplicates the original data set's statistical properties but with no links to the original information so it can be shared and used by researchers around the world trying to develop insights, treatments, and vaccines.[5]

According to two research studies by Washington University School of Medicine in St. Louis, the synthetic data generated from real COVID-19 patients accurately replicated the results of the same analyses conducted on the real patient data. Moreover, not only did the synthetic data accurately reflect the patient characteristics on a broad scale, it also accurately recreated the pandemic's spread and impact over time and in geographic areas that were highly tested for COVID.[6]

"We've shown that we can build sophisticated predictions of what is going to happen in a population with a disease like COVID-19," said co-author and principal investigator Philip Payne, chief data scientist and director of the Institute for Informatics at Washington University. "It is critical that we protect patients' rights to privacy and confidentiality while also responding to the threat posed by COVID-19 in a timely manner. No single institution can address these needs alone. Through the unique capabilities afforded by the use of synthetic data, we are accelerating our efforts to diagnose, treat and, perhaps most importantly, prevent this disease while also demonstrating how we can more effectively respond to future public health emergencies."

Synthetic data is used for a wide variety of use cases, not only when there are privacy concerns around real world data. It can also be used to address shortfalls or gaps in real world data. Take the example of trying to train a machine learning algorithm to calculate a response for a specific scenario that occurs rarely or perhaps not at all. Training a self-driving car to respond accordingly when debris from a truck falls on a highway at night might be difficult if there is little imagery available of such a scenario. Synthetic data allows for data to be created via software to fill such gaps in the available real-world data, boosting the model's overall robustness.

Synthetic data's many use cases

In addition to providing data for edge cases (rare events), synthetic data also can help improve model performance, remove bias in data, reduce the cost of data, and increase the speed of data collection.[7]

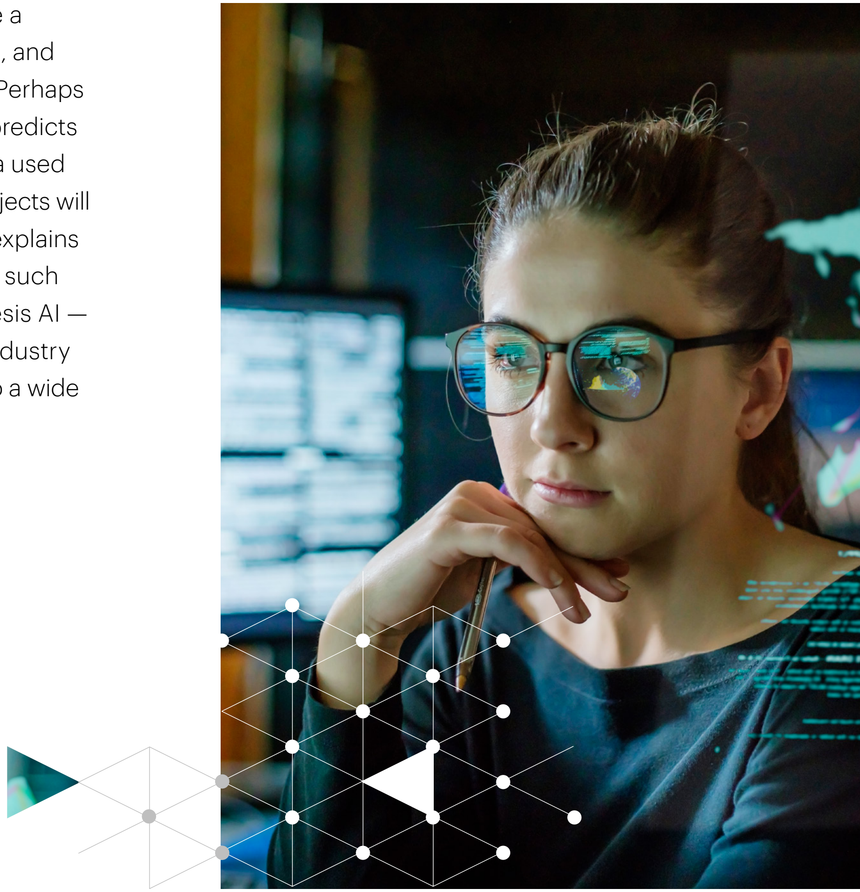
Like healthcare, the financial sector is another industry that is heavily reliant on AI and, in turn, has the potential to benefit from advances in utilizing synthetic data. Insurance companies, banks, credit card companies, and others use AI to lower lending risk in their portfolios, detect and counter fraud, and develop new products and offerings. But data projects like these typically require collaboration and data-sharing, both internally and externally. Privacy fears and compliance headaches can make their own data sets, generated from real-world customer activity, too risky to put to work.

Anonymizing real-world data is not always reliable in assuring individuals' privacy, and encrypting it undermines the data's utility. However, creating synthetic data sets avoids the risk of data leaks and privacy breaches while also overcoming scalability limitations.

> With synthetic data, organizations can rapidly produce high-volume, artificial test data, thereby shortening test cycles and reducing time to production.

Moreover, synthetic data can be generated in many structured and unstructured forms: text, tabular, and even media data such as video, images, and sounds.[8] Google's Waymo, for example, is using AI to generate simulated camera images from sensor data collected by its autonomous-driving vehicles. It is then using those simulated images to train its cars.[9]

In short, synthetic data has become a boon for researchers, data scientists, and organizations that depend upon AI. Perhaps that's why industry analyst Gartner predicts that, by 2024, 60 percent of the data used for AI development and analytics projects will be synthetically generated.[10] It also explains the rapid proliferation of startups — such as MOSTLY AI, Datagen, and Synthesis AI — that are filling out an entirely new industry focused on synthetic data catering to a wide assortment of industry sectors.[11]



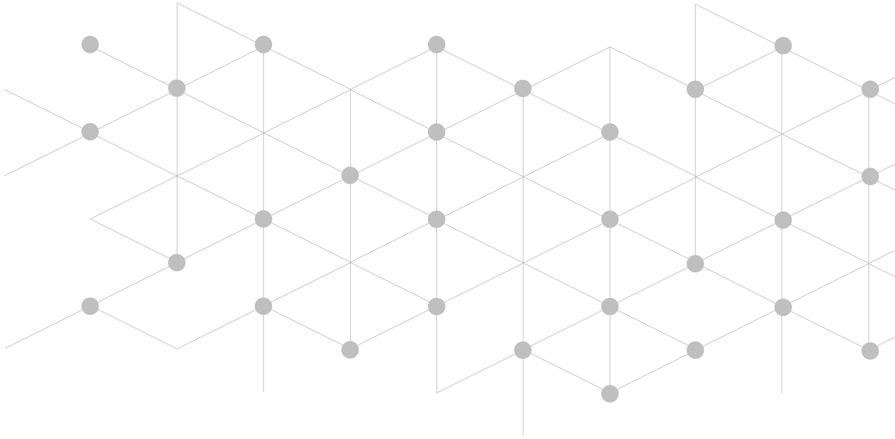
AI as modern-day data factories

But how do we generate synthetic data rapidly, at great scale, and accurately? One way is to use AI itself to generate that data. We call this generative AI. Another important dimension of the unreal world, generative AI can consist of a wide array of data types, including text, visual data, and multimedia. Examples of this are when technologies draw and paint pictures or use information gathered on the internet to create articles. In essence, generative AI enables computers to learn patterns from a large amount of real-world data and generate new content that mimics those underlying patterns.

One common approach to generative AI is through the use of generative adversarial networks (GANS). GANS, fundamentally, are generative modeling architectures that, during training, pit two neural networks — a generator and a discriminator — against each other. The generator network generates new data or content based on the patterns it is learning from real-world source data. As it does so, the discriminator network's task is to differentiate between the source real data and the generated data from the generator. This creates a feedback loop in which the generator constantly learns to produce more realistic data, while the discriminator gets better at differentiating fake data from the real data. GANS can generate a wide variety of content, including images, photographs of human faces, realistic photographs, face aging, and 3D objects.

This powerful architecture has made GANS the primary technology used to create deepfakes, which has justly earned the technology considerable scrutiny. But this same type of generative model can also be easily steered toward highly beneficial use cases. For example, they can be used to improve fairness and remove bias in credit and loan decisions by generating training data that removes biases from protected variables (e.g., gender, race, ...) while the discriminator tries to guess the values of the protected variables based on the generated data. When it becomes impossible for the discriminator to guess these values, the generator has successfully learned to produce debiased data that can be used to train the credit and loan decision model.[12]

This is heady stuff. More and more enterprises are becoming architects of the unreal world. And as they push AI into more collaborative and creative roles, they are blurring the lines between what's real and what isn't. Advances in generative AI, and GANs in particular, are making the creation and use of synthetic data that is incredibly realistic – while being unreal – possible.



A metric for the unreal world: authenticity

As we enter a world with synthetic realness, where AI-generated data — in the form of synthetic data, images, and chatbots, as well as augmented and virtual realities — convincingly reflects the physical world, we are forced to face the questions of what's real, what's not, and perhaps more importantly, when do we care? When we see the news, we want to know if the video of the president is real – but when we watch the latest Doritos commercial, maybe it doesn't really matter. And sometimes, we may actually prefer the unreal, like when we speak to a synthetic nurse about a potential sexually transmitted disease or train an AI model with synthetic data adjusted to counter historical discrimination.

As synthetic realness progresses, conversations about AI that align good and bad with real and fake will shift to focus instead on authenticity.

Instead of asking "Is this real?," we'll begin to evaluate "Is this authentic?" based on four primary tenets:

Provenance – what is the history?

Policy – what are its restrictions?

People – who is responsible?

Purpose – what is it trying to do?

With these principles, synthetic realness can push AI to new heights. By solving for issues of data bias and data privacy, it can bring next-level improvements to AI models in terms of both fairness and innovation.

And synthetic content will enable customers and employees alike to have more seamless experiences with AI, not only saving valuable time and energy but also enabling novel interactions.

That said, using these technologies pushes enterprises — and especially government entities — into controversial terrain. It raises tough questions about how to leverage synthetic data and generative AI in a trustworthy way in the service of government missions – all within the context of bad actors using these same technologies to create deepfakes and disinformation that undermine trust. Like it or not, the unreal world is about to become a part of reality, and the path ahead will be fraught with risk. But authenticity can, and should, be the guide.

THE ANALYSIS

When discussing synthetic data, it is important to think of it as part and parcel to AI. Synthetic data can be used for a wide variety of use cases, both good and bad.

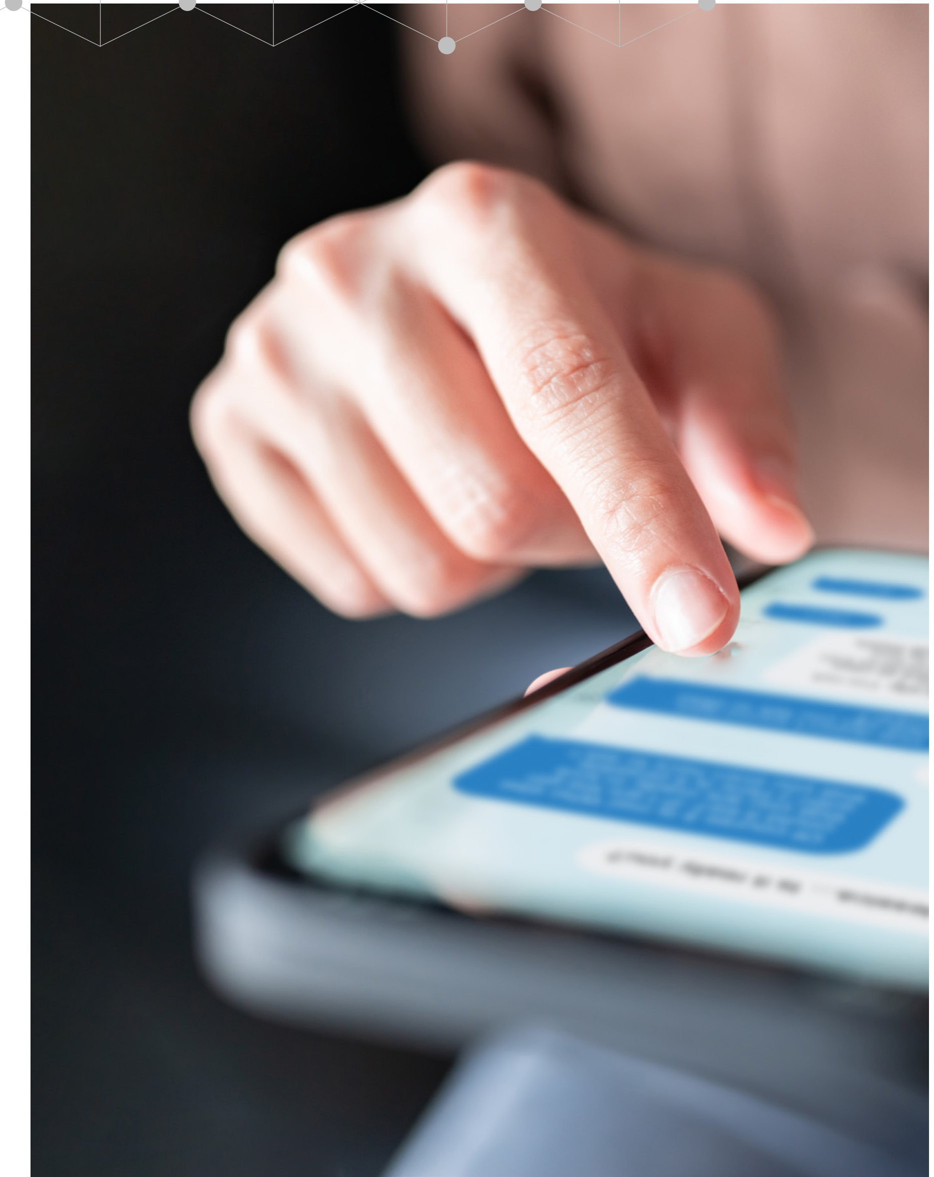
And as AI activity in the federal government continues to quickly ramp up, so too will the use of synthetic data. Here are some areas where we already see federal agencies actively producing and relying on synthetic data to advance their missions.

Customer care

Chatbots are one of the government's biggest use cases for AI, and bot technology is fast evolving. Chatbots must be trained to recognize specific intents in the language they read or hear, such as "I want to find out if my government health insurance will cover this procedure." This intent can be stated in any number of ways.

Synthetically generated text can enable help computers more quickly learn patterns in conversation and produce more effective chatbot responses to the customer inputs they are receiving. This helps speed up the bot development process and lower costs.

Emerging technologies today are pushing the state of the art even further, making bots even more realistic. AI company Hour One, for instance, creates digital characters, based on real people's likeness, that can be shown speaking any text in highly realistic videos (each human receives a micropayment when their character is used, and that use is restricted to safe content).



These characters can play virtual customer service representatives or language teachers, easily converting static text content into video – saving hours of actors' and production crews' time in the studio.

If it sounds like Hour One is creating deepfakes, that's because it is. These are clearly videos of real people saying or doing things they haven't said or done. But in this case, the videos are safe and legal, the actors have given permission for the use of their likeness, and the videos are clearly labeled as computer-generated (Hour One requires that each video disclose that the content is computer-generated and is embedded with an "Altered Visuals" watermark).

The AI-based video bots that Hour One creates seem far from malicious. But does that mean it's okay — is it ethical? This points to the potential challenges and pitfalls that can arise with the government's use of synthetic data.

The proliferation of purposely malicious deepfakes and disinformation has already caused damage, eroding the public's trust in the media they see and read. The public's trust in the technology sector is on a steep decline, reaching all-time lows in 17 out of 27 countries.[13] More to the point, Americans' trust in the government is also at record lows — since 2007, the percentage of citizens saying they can trust the government always or most of the time has not surpassed 30 percent.[14]

> So federal agencies will need to tread carefully and thoughtfully — employing the critical metrics of transparency and authenticity throughout — as they move down this path.

As conversational AI technologies continue advancing and redefining the state of the art in commercial customer experience, we should expect to see federal agencies similarly follow suit as they strive to improve their own citizen experiences around the government services they deliver. As with chatbots today, synthetic data will surely play a key role in developing those technologies in the future.

Open data

One of the earliest federal use cases for synthetic data can be found at the U.S. Census Bureau, which regularly collects vast stores of national data of immense value to many fields of research.

In 2001, the Census Bureau was authorized to integrate person-level micro-data from its longitudinal household survey, called the Survey of Income and Program Participation (SIPP), with IRS tax and earnings data and Social Security Administration retirement and disability benefits data.[15] The resulting data trove offers the most comprehensive information available on how the nation's economic well-being changes over time, and it's a goldmine for academics, researchers, economists, and policy makers.[16] With SIPP data, they can examine, for example, national income distributions, the impacts of government assistance programs, and the complex relationships between government tax policy and economic activity at the local levels.

The problem for Census, however, is that the highly detailed nature of the SIPP data makes it particularly sensitive because the micro-data could be used to identify specific individuals. To make the data safe for public use while also retaining its research value, the Census chose to create synthetic data from the SIPP data sets.[17] The result is the SIPP Synthetic Beta (SSB), a Census Bureau product that was first made public in 2007 and continues to be updated and released periodically.[18]

Federal healthcare

Similarly, the NIH's N3C Data Enclave is an open data initiative aimed primarily at advancing COVID-19 research. Since the database was opened to researchers in September 2020, it has grown to include billions of rows of data representing more than 5 million COVID-19 positive patients, making it the largest open U.S. database of data from patient electronic health records.

Because of its advanced informatics technologies and data linkages to demographic, mortality, and other information, the database helps researchers create clearer pictures of COVID-19 health outcomes among different communities and enables them to find patterns faster than traditional database methodologies allow. Moreover, the N3C Data Enclave has become useful for research well beyond COVID — researchers have used it to improve our understanding of health equity, diabetes, cancer, HIV, rural mortality rates, and chronic obstructive pulmonary disease as well.[19]

Under the open data initiative, scores of federal agencies and subagencies have already made data sets freely available to researchers on Data.gov, the main web portal for open government data.[20] But some data sets cannot be shared because they could reveal personal details of specific individuals. As we see in the Census and NIH examples, synthetic data is an avenue for agencies to bring even more federal data sets to the research community.

As mentioned, synthetic data is increasingly indispensable for medical researchers who must work around data privacy and compliance obstacles. But deep generative models, such as GANS, are also becoming critical tools for creating more robust data sets to train AI-driven diagnostic tools that can be safely applied across many demographic categories.

AI algorithms need to be trained on large, diverse datasets to be generalizable across a variety of populations and to ensure they are not biased in ways that affect their accuracy and reliability. Historical patient data, such as images or scans of certain maladies, often lack the needed diversity and representation to achieve this. For example, a 2020 analysis of data used to train image-based diagnostic AI systems found that approximately 70 percent of the studies that were included used data from three states, and that 34 states were not represented at all.[21]

Algorithms developed without considering geographic diversity, including variables such as disease prevalence and socioeconomic differences, may not perform as well as they should across a varied array of real-world settings.[22] To help address these shortfalls, some developers of AI-based diagnostic tools are looking to GANs. GANs are showing much promise in generating realistic images of skin lesions, pathology slides, colon mucosa, and chest X-rays in a range of imaging modalities, according to recent studies.[23]

For these reasons, federal medical researchers are sure to make greater use of generative images and data in their work. But synthetic data will also carry big implications for federal regulatory agencies in the healthcare arena.

In 2021, the U.S. Food and Drug Administration (FDA) unveiled a new "action plan" for how it will regulate AI-based software as a medical device (AI-SaMD).

The plan recognizes that FDA's traditional processes for regulating software and AI products must change to keep pace with today's innovations – namely, AI, machine learning algorithms and synthetic data embedded in many of today's devices continue to change over time long after they go to market and are in use. With its action plan for AI-SaMD, FDA voiced an intent to review AI/ML software products from pre-market development to their post-market performance. FDA officials also are reportedly considering the use of synthetic data as an approved method for ensuring that data sets used to train AI/ML models in medical devices are diverse and accurately reflect the patient population in which the technology will be used.[24]

Constituent services

The Department of Veterans Affairs intends to leverage synthetic data as part of a new initiative to reduce veteran suicides by 10 percent.

> By parsing synthetic data, department leaders hope they can better clarify veteran challenges and refine predictive risk factors so department programs can be more proactive in addressing the problem.

"It's very clear that we need to use reliable and timely data to identify and address issues impacting our veterans, while also ensuring privacy," said Dr. Carolyn Clancy, assistant under secretary for Health for Discovery, Education and Affiliate Networks at the Veterans Health Administration (VHA). "We believe synthetic data, modeled to precisely mirror real veteran data, while protecting veteran privacy is a great path forward."

THINGS TO LOOK OUT FOR

Can the unreal be trusted?

While synthetic data initiatives can and do help advance federal missions, there are clear risks and challenges to be aware of.

A core concern for any federal agency, of course, will always be trust.

"In the end, adoption of synthetic data becomes an issue of trust in the protected privacy and accurate representation of the original data," wrote the American Council for Technology-Industry Advisory Council (ACT-IAC), a non-profit public-private partnership that promotes the effective and innovative application of technology in government, in a January 2022 whitepaper for the Department of Veterans Affairs.[25]

"Trust can only be engendered through validation efforts, and this may be where the greatest need for policy guidance exists — how can we quantify that the generated synthetic data holds sufficient utility for analytic end uses?"

Just

42%

of consumers believe AI is being used to improve their lives and experiences and only

35%

fully trust how it's being implemented by specific organizations.

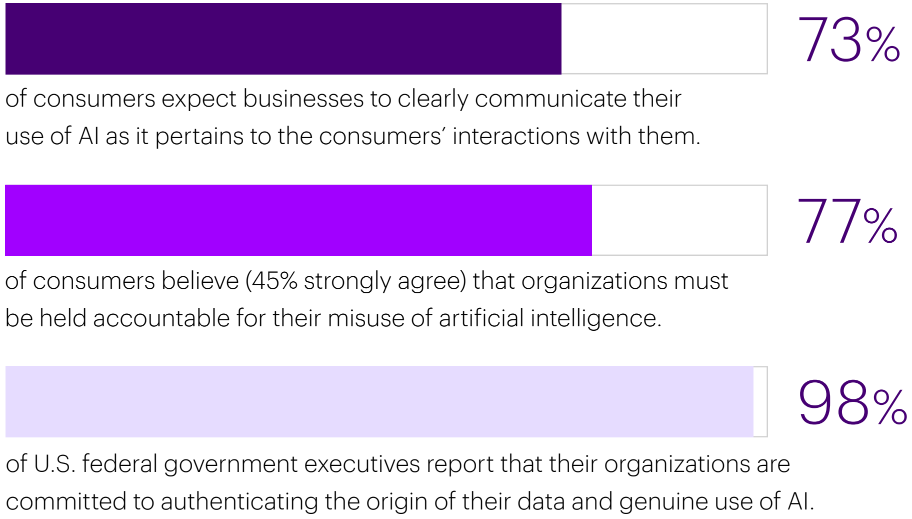
The FDA, for example, in its action plan for regulating AI/ML-based Software as a Medical Device, acknowledges the unique ethical dimensions of regulating this class of products, noting that "AI/ML-based devices have unique considerations that necessitate a proactive patient-centered approach to their development and utilization that takes into account issues including usability, equity, trust, and accountability." In doing so, the FDA said it would make transparency a central tenet in its regulatory approach. "Promoting transparency is a key aspect of a patient-centered approach, and we believe this is especially important for AI/ML-based medical devices, which may learn and change over time, and which may incorporate algorithms exhibiting a degree of opacity," the plan says.[26]

Toward this end, FDA officials met with an advisory committee of patients and caregivers in October 2020 to gain insight into what factors impact their trust in these technologies.[27]

FDA also held a public workshop in October 2021 to elicit additional input from the broader community on how device labeling could support transparency objectives for users.[28] In one published account of that 2021 workshop by Stacy Cline Amin, leader of Morrison Foerster's FDA Regulatory & Compliance practice and a former chief counsel at FDA, participants discussed the idea of labeling AI-based medical devices with something akin to today's food labels — but instead of listing nutritional information, these labels would describe the limitations of a particular device and offer guidance for how to properly interpret and use its data in administering care.[29]

Likewise, in building its SIPP Synthetic Beta product, the Census Bureau was mindful of the challenge of building trust among the public and the research community — again, adopting transparency as a key tactic. The agency published numerous articles explaining in detail its exact methodologies in developing the product as well as how researchers should use SSB data in their research.[30]

Authenticity builds trust



73% of consumers expect businesses to clearly communicate their use of AI as it pertains to the consumers' interactions with them.

77% of consumers believe (45% strongly agree) that organizations must be held accountable for their misuse of artificial intelligence.

98% of U.S. federal government executives report that their organizations are committed to authenticating the origin of their data and genuine use of AI.

Another key to building trust in synthetic data is proactive outreach and dialog with stakeholders. Just as the FDA is engaging patient and caregiver stakeholders as it formulates policy on regulating AI/ML-based medical devices, so too is the NIH consulting with American Indian and Alaskan Native (AI/AN) communities to decide how and whether to make AI/AN COVID-related data available to researchers via the N3C Data Enclave. Currently, AI/AN data in N3C is obscured. Two NIH entities— the National Center for Advancing Translational Sciences (NCATS), which administers the N3C Data Enclave, and the Tribal Health Research Office — are consulting with Tribal stakeholders on whether and how to provide AI/AN data respectfully. "Learning from past examples, our Center wanted to seek prospective support from Tribal Nations in a manner that respects Tribal sovereignty. NCATS decided not to make this data accessible until after it had consulted with Tribal Nations," the Center said on its N3C FAQ website.[31]

Another hurdle is the scarcity of effective policy around the use of synthetic data. Like the FDA, other federal agencies are similarly struggling with how to delineate the appropriate place and role of synthetic data within their purviews. They will need to address questions about how it fits within existing regulatory frameworks; how it can be used to protect patient privacy, address bias in data, and improve medical decision-making; and what steps should be taken to alleviate the trust and accountability concerns that are sure to arise with its use.

This problem is growing urgent for some. While the VA sees much promise in using synthetic data to help reduce veteran suicides, it has yet to put in place the policy guardrails needed to guide the department as it moves forward. The VA even has a cloud-based platform in place to generate synthetic data. But the lack of policy for how and when to use synthetic data complicates the task of standardizing and using synthetic data across the enterprise.[32] In response, the VHA has collaborated with ACT-IAC to help develop policies. The collaboration has resulted in a white paper that proposes how to create and use synthetic data.[33]

The dark side of the unreal

As we see, even when the unreal is put to work in service of productive and well-meaning use cases, there are challenges aplenty. When the use cases turn malevolent, the challenges quickly get more complicated and difficult. This takes us back to the growing problem of discerning and countering malicious deepfakes. AI and ML technologies are reducing the time, cost, and skill sets needed to create deepfakes. And this will only accelerate as computing power and data volumes continue increasing.

It is an open question whether the technologies used to detect deepfakes will be able to keep pace — or whether it even matters. In some cases, the purpose of a deepfake may be to simply fool enough people to create a response. The bigger concern is that the rapid proliferation of deepfakes is undermining our trust in all videos and media, whether genuine or not. When we begin to question what's real or not as a matter of course, truth itself becomes elusive.

There is much to be done to tame this problem. One approach that some are advocating is using blockchain technologies to help ensure provenance and altering of media material.[34] The distributed ledger technology underlying blockchain makes it highly resistant to data modification. In a recent report, the Federal Trade Commission (FTC) notes "...given the limitations of using AI to detect harmful content, it is important to focus on key complementary measures, particularly the use of authentication tools to identify the source of particular content and whether it has been altered. These tools – which could involve blockchain, among other things – can be especially helpful in dealing with the provenance of audio and video materials."[35]

85%

of U.S. federal government executives report that their organizations are planning to mitigate the risk of deepfakes and/or misinformation by preparing proactively and implementing verification mechanisms.

83%

of U.S. federal government executives report that blockchain is going to be critical to their organization's ability to verify the origin of digital content.

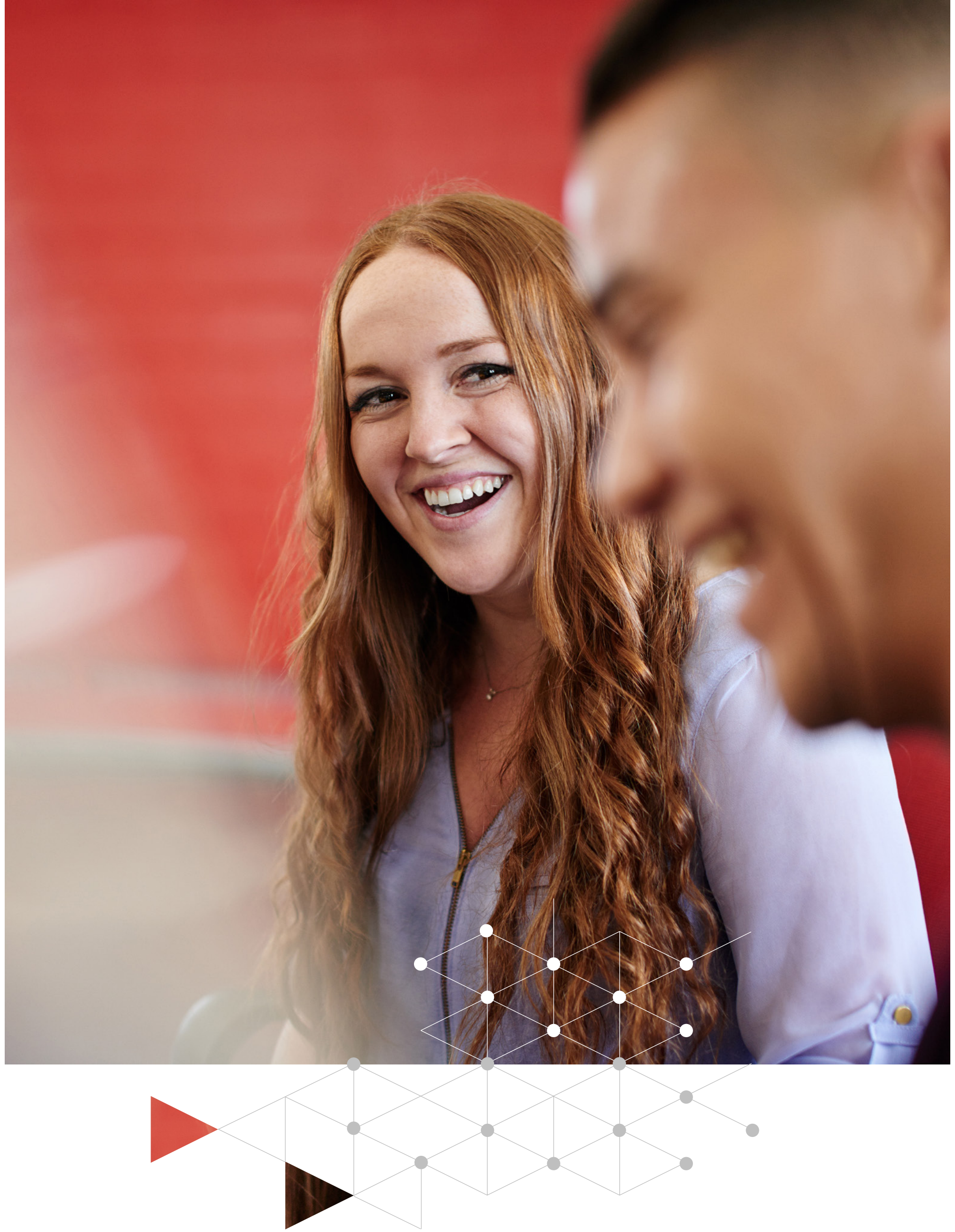
ACTIONS TO TAKE

Elevating authenticity

While synthetic realness has the ability to sow distrust and discord, it also has the power to improve human relationships.

Consider an experiment conducted at Yale University in which participants were put into small groups, each with a humanoid robot, and told to collaborate on a task.[36] The robots were programmed to make occasional mistakes. Some of the robots were designed to admit to those mistakes in a humorous and self-deprecating way – and it was their groups that performed better because the robots improved the participants' communication, allowing them to work together in a more fun and relaxed manner.

In another experiment, people were put into virtual social networks with a few incognito bots and assigned a collaborative task. Again, some of the bots were programmed to make mistakes, and the social networks that they were in grew more flexible in response and consequently outperformed those with bots that did not make errors. This research demonstrates that if designed and deployed in the right ways, AI with human-like qualities like wit and, significantly, imperfection can be used to improve people's performance and strengthen their relationships with each other.



Since we know that being real has no direct bearing on being good, being real should not be the guiding star for business or society. Rather, we propose authenticity as the new compass. Authenticity means being true to oneself and genuine in a way that others can attest to – and more concretely, using generative AI in an authentic way means taking heed of provenance, policy, people, and purpose. By observing these four tenets, businesses can gain confidence not only in their decisions to trust others but also in their use of generative AI such that others can trust them – thus enabling full participation and success in the unreal world.

One way to verify the provenance of digital content and identity – thereby demonstrating authenticity – is through the use of distributed ledger technology (DLT). As an example, Project Origin, led by Microsoft, the BBC, CBC, and The New York Times, is tackling the spread of disinformation using DLT to establish provenance from publishing to presentation.[37]

The Coalition for Content Provenance and Authenticity (C2PA) has built upon this foundation and similar work by the Adobe-led Content Authenticity Initiative (CAI) to propose new standards for authenticating visual media. According to Adobe's Andy Parsons, the goal of the standard is "...so that users can be assured that when media is uploaded with content authenticity, that it is maintained throughout the entire chain of sharing [and] publishing creation, back and forth."[38]

In fact, many analysts are predicting that a large portion of news and video content will be authenticated by blockchain in the coming years. No matter what technologies you use, establishing provenance will be critical as your agency increasingly deals with potential deepfakes and disinformation – and enabling others to establish provenance as they interact with your agency and content will be just as important too.

Next, take stock of your agency's policies with respect to generative AI.

In 2019 for instance, the U.S. state of California passed the BOT Disclosure Law, which states that one must disclose the use of a bot when they are used in communication to sell goods or services or influence a vote in an election.[39] And the EU has drafted legislation to regulate "trustworthy AI," with the purpose of protecting the rights of citizens.[40] The current proposal takes a risk-based approach, banning unacceptable uses of AI and having strict obligations for high-risk use cases. The Business Roundtable has brought together CEOs from some of the largest U.S. corporations to recommend guidance for government regulation of AI.[41] Much of this space is yet to be defined, so where there isn't guidance, agencies will need to define their own policies based on their particular set of missions and business operations, their stakeholders' perspectives, input from Congress, industry best practices, insights from subject matter experts, and their agency's values. And if you are proactive in sharing what works and what doesn't, your agency can be involved in shaping the future of the unreal world – rather than just reacting to it.

From a people perspective, your agency must be prepared organizationally to deal with the challenges that arise with the use of AI.

> Ask yourself, for example, who is responsible for having these tough conversations, and what committees are drafting internal policies?

What departments are using synthetic data or content within the agency, and who will be held accountable if privacy is compromised or customers feel duped? Finally, who will be the point person responsible if your company falls prey to a deepfake or disinformation attack? Having these governance structures in place is imperative to handle the inherent risks baked into the unreal world.

Last but not least, genuine purpose is essential to authenticity. In particular, agencies must define the purpose behind the use of synthetic content, its advantage over non-synthetic content, and the key metrics that can attest to it. For instance, if your agency uses a basic customer service bot simply to cut costs (as opposed to improving availability), there's a good chance it's not living up to its intended purpose of serving customers. However, if the purpose of using synthetic data in a model is to insert counterbias, thereby improving the output of the model, then it could be an authentic use of generative AI. As another example, Soul Machines creates synthetic people that can be used in cases where customers might fear judgment from others and actually prefer to speak to digital people.[42]

Yumi is one such "autonomously animated digital influencer" that answers customers' questions about their skin for P&G's skincare brand SK-II. Yumi's realness can enable more seamless and personal interaction – while simultaneously, its digital essence puts customers at greater ease. Here, the purpose of using a synthetic person is clear on multiple fronts, and as such, is convincingly authentic.

Conclusion

As AI progresses and models improve, enterprises are building the unreal world. But whether we use synthetic data in ways to improve the world or fall victim to malicious actors is yet to be determined. Most likely, we will land somewhere in the expansive in-between, and that's why elevating authenticity within your organization is so important. Authenticity is the compass and the framework that will guide your agency to use AI in a genuine way – across mission sectors, use cases, and time – by considering provenance, policy, people, and purpose. Ultimately, it will unlock new attitudes towards and experiences with AI, unleashing the benefits of the unreal world.

Decision Points

Is your enterprise prepared to take full advantage of unreal world technologies?

- Explore the use of synthetic data. Determine how its advantages could improve existing data strategies, and the algorithms and AI fueled by them, like improving data set quality, reducing privacy risk, and correcting for bias present in historic data sets.
- Identify where unreal content like chatbots or AI-generated images, video, or content could help improve your agency's customer experiences. Find the ways it can create new avenues of connection with your customers, improve the quality of their experiences, and drive new outcomes.
- Pilot the use of unreal technologies to augment the enterprise. Enable employees to leverage them as a partner, enhancing design, simulation, or decision-making capabilities.

How are you protecting your organization and your customers from malicious use of the unreal?

- Identify emerging malicious applications of unreal world technologies before they become a systemic risk. Focus on the veracity and provenance of the information coming into the organization, like potential scams or disinformation, and out of the organization, to ensure unintended falsehoods aren't perpetuated. Consider techniques like the use of distributed ledger technologies to establish provenance.
- Differentiate your use of unreal world technologies from those of threat actors and build trust with your customers by having a clear and communicated purpose. Give people the ability to attest to the genuineness of the agency and its outputs. For example, protect the enterprise from malicious impersonation by incorporating verifiable identity markers throughout your platforms and content.
- Have a plan for how your organization will respond to malicious use of deepfakes or disinformation campaigns against your agency. Explore the most damaging threat scenarios and build the playbook to respond to the events and train, train, train.

How will your enterprise shape the unreal world?

- Authenticity must become an enterprise-wide priority and a C-suite responsibility for generative AI. Know that regulations are formative in this new territory of the unreal world. Have each of your major enterprise functions identify the existing regulations they must adhere to and close the gaps with internal policies that align to agency values. These should be reported up to the accountable C-suite leaders who should maintain a regular agenda item concerning the impact of AI to their programs and business operations, and how to hold it to a higher standard.
- Raise the bar on standards and engage in the standards-making processes. Distrust or harm created by a malicious, careless, or negligent organization or actor in the unreal world could affect how people will embrace and trust the unreal at large. Look for ways to affect the authenticity of the unreal world and hold it to a higher standard.
- Empower your people to not just ask the tough questions but find the tough answers. Exploring the unreal will have implications across security, communications, public affairs, R&D, and beyond. It will be critical for the organization to have a consistent approach to decision making around big topics like security, privacy, safety, transparency, and ethical conduct. A useful starting point is to have specific people or groups be accountable to these answers and ensure that there are effective metrics to monitor the ongoing success and effects of any unreal innovations.

Authors

David Lindenbaum
Machine Learning Director
Accenture Federal Services
in

Jennifer Sample, Ph.D.
Applied Intelligence Growth
& Strategy Lead
Accenture Federal Services
in

Marc Bosch Ruiz, Ph.D.
Managing Director –
Computer Vision Lead
Accenture Federal Services
in

Nilanjan Sengupta
Managing Director – Applied
Intelligence Chief Technology Officer
Accenture Federal Services
in

Shauna Revay, Ph.D.
Machine Learning Center of
Excellence Lead, Strategic Solutions
Accenture Federal Services
in

Viveca Pavon-Harr, Ph.D.
Applied Intelligence
Discovery Lab Director
Accenture Federal Services
in

Behind the Vision

The Accenture Technology Vision takes a systematic look across the enterprise landscape each year to identify evolving technology trends with highest possibilities to disrupt businesses, governments, and societies over the next three years. For 22 years, corporate and government leaders have relied upon this research to prepare their organizations for what's next.

The Accenture Technology Vision is produced by Accenture Labs and Accenture Research. It draws on internal research and analysis, insight from the Technology Vision External Advisory Board, and results of a global survey of 4,660 c-suite executives spanning 23 industries and 24,000 consumers worldwide. Instead of focusing just on the drivers of technological change, the Accenture Technology Vision is distinguished by its examination of the broader themes poised to have the most enduring and transformative impact on how enterprises operate.

The Accenture Federal Technology Vision 2022 applies these insights and findings to the unique challenges and demands facing the U.S. federal government. It features in-depth analysis from more than 50 Accenture Federal Services experts and results of a survey of 200 U.S. federal government executives.

Accenture Federal Technology Vision 2022 Editorial & Marketing Team

John Conley, Riley Panko, Steve Watkins and Katrina "Kat" Szakolczai

Accenture Technology Vision 2022 Editorial & Research Team

Michael Biltz, Ari Bernstein, Julian Dreiman, Maria Fabbroni, Naomi Nishihara, Lara Pesce Ares, and Krista Schnell

Accenture Research (for the Accenture Technology Vision 2022)

Renee Byrnes, Mariusz Bidelski, Gerry Farkova, Harrison Lynch, Sandra Najem, Haralds Robeznieks, Swati Sah, Abira Sathiyanathan, Gabe Schmittlein, and Mélina Viglino

References

1 https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

2 https://www.sciencedirect.com/science/article/pii/B9780323900546000064#bb0380

3 https://www.sciencedirect.com/science/article/pii/S1876034122000144

4 https://www.prnewswire.com/news-releases/syntegra-partnering-with-national-institutes-of-health-nih-and-the-bill-and-melinda-gates-foundation-to-democratize-access-to-the-largest-set-of-covid-19-patient-records-301209504.html

5 https://ncats.nih.gov/n3c/about

6 https://medicine.wustl.edu/news/synthetic-data-mimics-real-patient-data-accurately-models-covid-19-pandemic/

7 https://mostly.ai/all-synthetic-data-use-cases/

https://appen.com/blog/synthetic-data-and-its-role-in-the-world-of-ai/

8 https://www.statice.ai/post/types-synthetic-data-examples-real-life-examples

9 https://venturebeat.com/2020/05/20/waymo-is-using-ai-to-simulate-autonomous-vehicle-camera-data/

10 https://blogs.gartner.com/andrew_white/2021/07/24/by-2024-60-of-the-data-used-for-the-development-of-ai-and-analytics-projects-will-be-synthetically-generated/

11 https://techcrunch.com/2022/05/10/the-market-for-synthetic-data-is-bigger-than-you-think/

12 https://www.eetimes.com/reducing-bias-in-ai-models-for-credit-and-loan-decisions/

13 https://www.cnet.com/tech/tech-industry/peoples-trust-in-tech-is-at-an-all-time-low-edelman-study-says/

14 https://www.pewresearch.org/politics/2022/06/06/public-trust-in-government-1958-2022/

15 https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html

16 https://www.census.gov/programs-surveys/sipp/about.html

17 https://www.nsf.gov/awardsearch/showAward?AWD_ID=1042181

18 https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/SSBdescribe_nontechnicalv7.pdf

19 https://ncats.nih.gov/n3c/about

20 https://data.gov/meta/data-gov-turns-six/index.html

21 https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2021/08/how-fda-regulates-artificial-intelligence-in-medical-products

22 https://jamanetwork.com/journals/jama/article-abstract/2770833

23 https://www.nature.com/articles/s41551-021-00751-8#ref-CR13

24 https://www.jdsupra.com/legalnews/five-key-takeaways-from-fda-s-2271087/

25 https://www.actiac.org/sites/default/files/2022-01/VA%20Synthetic%20Data_0.pdf

26 https://www.fda.gov/media/145022/download

27 Ibid.

28 https://www.jdsupra.com/legalnews/five-key-takeaways-from-fda-s-2271087/

29 https://www.mofo.com/people/stacy-amin.html

30 https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/SSBdescribe_nontechnicalv7.pdf

https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/SSBdescribe_nontechnical.pdf

31 https://ncats.nih.gov/n3c/about/program-faq#privacy-and-security

32 https://governmentciomedia.com/va-leveraging-synthetic-data-improve-suicide-prevention-efforts

33 https://www.actiac.org/sites/default/files/2022-01/VA%20Synthetic%20Data_0.pdf

34 https://www.wired.com/story/the-blockchain-solution-to-our-deepfake-problems/

35 https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf

36 https://www.theatlantic.com/magazine/archive/2019/04/robots-human-relationships/583204/

37 https://www.originproject.info/about

38 https://techhq.com/2022/04/sony-adobe-intel-among-tech-firms-taking-on-deepfakes-with-blockchain-technology/

39 https://www.natlawreview.com/article/california-s-bot-disclosure-law-sb-1001-now-effect

40 https://blog.macfarlanes.com/post/102h1aw/a-gdpr-for-artificial-intelligence

41 https://www.businessroundtable.org/business-roundtable-comments-on-draft-omb-memorandum-to-the-heads-of-executive-departments-and-agencies-on-guidance-for-regulation-of-artificial-intelligence-applications

42 https://www.pymnts.com/commerce-connected/2021/soul-machines-digital-face-connected-economy/

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, Technology and Operations services and Accenture Song — all powered by the world's largest network of Advanced Technology and Intelligent Operations centers. Our 710,000 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 Federal Services

Accenture Federal Services, a wholly owned subsidiary of Accenture LLP, is a U.S. company headquartered in Arlington, Virginia. Accenture's federal business serves every cabinet-level department and 30 of the largest federal organizations. Accenture Federal Services transforms bold ideas into breakthrough outcomes for clients at defense, intelligence, public safety, civilian and military health organizations.
Visit us at **www.accenturefederal.com.**

Business and Consumer Surveys

Accenture Research conducted a global survey of 24,000 consumers to capture insights into their use of, interactions with, and beliefs about technology in their everyday lives. In addition, Accenture conducted a survey of 4,650 C-level executives and directors across 23 industries to understand their perspectives and use of emerging technologies across their organizations. This survey included responses from 200 U.S. federal government executives. The surveys were fielded from December 2021 through January 2022 across 35 countries.