Deloitte.
Insights

FEATURE

Trustworthy open data for trustworthy AI

Opportunities and risks of using open data for AI

Tasha Austin, Kara Busath, Allie Diehl, Pankaj Kishnani, and Joe Mariani

DELOITTE AI INSTITUTE FOR GOVERNMENT AND THE DELOITTE CENTER FOR GOVERNMENT INSIGHTS

Many government and nongovernment agencies are releasing massive amounts of open data that can be used to train AI models and unlock huge value for society. Yet, organizations need to be cognizant of the risks and ensure that open data offers a safe path to future AI.

EARLY IN HER career, Fei-Fei Li, now professor of computer science at Stanford University, recognized that an algorithm would not be able to make better decisions unless the underlying data reflects real-world data. Her solution was to map the entire image library of the world. The result of the 2.5 years of effort was ImageNet, a collection of 14 million images.[1]

Published in June 2009 at a computer vision conference in Florida, ImageNet's open dataset quickly became the basis of an annual challenge to see which algorithm would have the lowest error rate in identifying images.[2] In the inaugural competition, held in 2010, every team had an error rate of at least 25%. However, by combining the techniques of deep learning with the massive set of training data available with ImageNet, researchers sent error rates tumbling. By 2017, the last year of the competition, the error rate was less than 3%.[3] ImageNet provided a big boost to AI—the dataset is credited with the resurgence of deep learning.[4] The same marriage of deep learning with massive datasets has been central to advances like self-driving cars, facial recognition, cyber defense, and predicting traffic congestion.[5]

To accelerate the development of AI, many government agencies, nonprofits, think tanks, and even for-profit companies release massive amounts of open data that can be used to train AI models; and the push for agencies to release open data has only increased since the enactment of the Foundations of Evidence-Based Policy Making and Open Data Act in 2018.[6] Opening up data for AI use can unlock huge value for society—from finding cures for lethal diseases, to combatting climate change, to effectively responding to crisis, the potential is immense.

Yet, for all its benefits, open data also carries risk. Open data can certainly accelerate AI development, but using massive public datasets to train models can unintentionally undermine privacy or perpetuate encoded biases. Even the pioneering ImageNet data faced some of these risks as creators removed people-related categories and blurred individuals' faces to try to protect their privacy.[7] For open datasets released by the public sector, government leaders should be cognizant of the risks and take steps to ensure that open data offers a safe path to future AI.

Opportunities and risks of using open data for AI

GOVERNMENTS COLLECT VAST amounts of data on everything from health care to housing, economic development to national security. Government agencies also produce and release data such as census figures, financial market information, weather data, transportation routes, and more.[8]

These large public datasets can help train predictive models that can create value for public and private sectors and, most importantly, constituents. For instance, government data on health care can help doctors, hospitals, and pharmaceutical companies improve existing treatment options and even create novel cures. A machine learning model based on real-world, open data played an instrumental role in the clinical trial process of a COVID-19 vaccine by recommending where trial participants should be drawn from based on where virus hotspots were likely to emerge during the trial.[9] Timely data can also predict faster transportation routes in real time, measure the impact of public transit, and reduce traffic.[10]

Open data can not only be used to create AI, but also to accelerate the development of new AI models. For example, the ImageNet dataset has been a key tool in accelerating AI model development for computer vision and deep learning researchers around the world.[11] Open datasets can help accelerate AI development in two ways. First, they can reduce data monopolies—where one company or agency controls all sources of data on an issue—which stymie AI innovation by limiting access to needed data. Second, they can save the time and expense involved in collecting, aggregating, and storing data, allowing researchers, entrepreneurs, and government agencies to spend more time on solving problems.

But open datasets also carry with them the imprint of how they were created. These datasets contain critical information reflecting a valuable historical record of transactions. But if those historical records are incomplete or reflect historical biases, they might train future AI models to recreate those biases. When using AI to make critical decisions, three main categories of risks come into play:

These large public datasets can help train predictive models that can create value for public and private sectors and, most importantly, constituents.

Risk of inbuilt bias

While AI can do many incredible things, the more we use it, the greater the chance that bias may creep into decisions based on it. A key source of such biases is the underlying training data that fuels algorithms. Technologist Maciej Ceglowski argues that AI models trained on historical data can unintentionally perpetuate historical systemic unfairness.[12]

Three types of dataset biases are common: interaction bias, latent bias, and selection bias. *Interaction bias* arises when an algorithm is trained on a dataset which provides limited interaction with varying demographics. For example, facial recognition systems that are trained primarily on the faces of white men are significantly more likely to misidentify the faces of women or minorities.[13] In *latent bias*, algorithms trained on historical data may stereotype. For example, using historic college admissions data of student recruitment may unintentionally lead to the perpetuation of historical disparities in college attendance by gender or race.[14] *Selection bias* occurs when a certain group is overrepresented in a dataset and another underrepresented. In the health sector, for example, a growing body of research indicates how lack of patient data on people of specific ethnicities has led to cancer detection models with differing degrees of accuracy depending on skin color.[15]

Risk to privacy

The great benefit of having massive amounts of data publicly available for AI development, however, is counterbalanced by the risk that this data may contain personal information that could intrude on individuals' privacy. AI's ability to track patterns also makes it highly effective at reidentifying personal data in anonymized datasets, causing significant privacy concerns. For example, within an hour a researcher was able to identity the home addresses of New York taxi drivers from an anonymized dataset of trips in the city.[16] Similarly, a health department's open data on medical billing could be linked with other open data such as year of birth, number of children, and birth dates to reidentify people from anonymized data.[17]

Risk to security

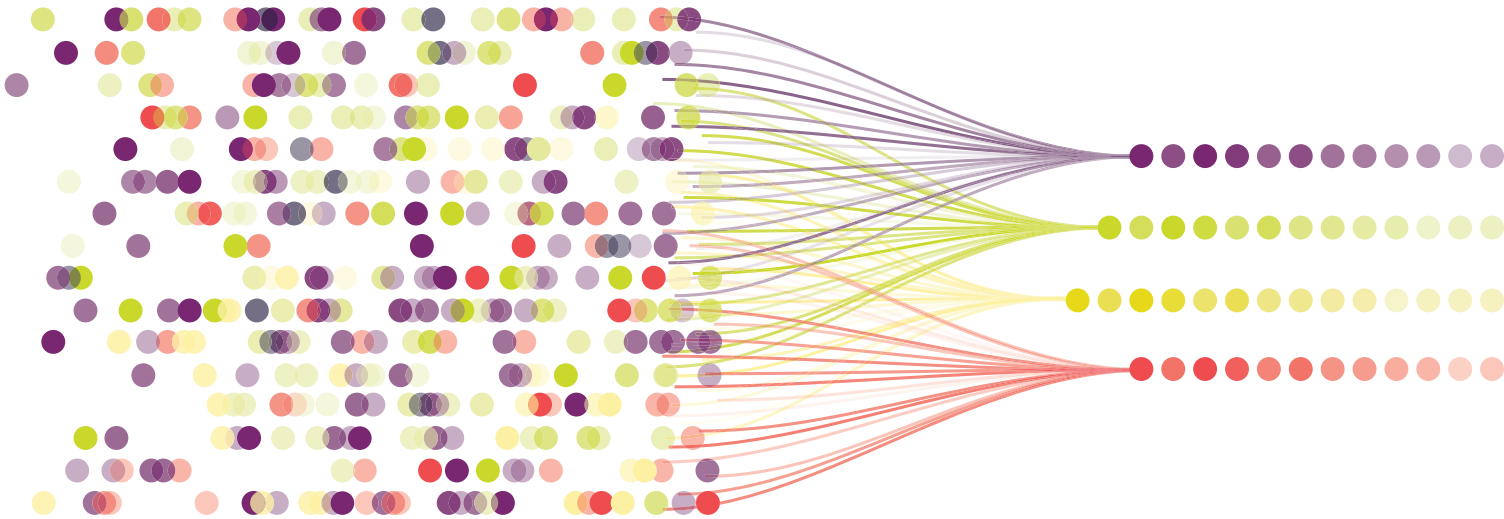
Making training data publicly available can not only pose a threat to individual privacy but can also open up avenues for compromising the security of AI models built from the data by providing an additional vector for hackers to attack. In cases where open datasets are created by the public or open to public changes, attacks can use *data poisoning*, where false values are introduced into an otherwise secure open dataset. In other cases, the mere availability of the training data can be used by attackers. If bad actors have knowledge of how an AI model has been trained, they can subtly change inputs to manipulate the model's outputs. One study examined the risk to medical imaging software from adversarial attacks that subtly modify images. The changes were undetectable to the human eye but could lead to deep learning systems misclassifying images up to 100% of the time.[18] Such attacks can have grave consequences, as many organizations, including government agencies, release open datasets for medical images to improve diagnosis and treatment.[19]

Data governance at each stage of the AI life cycle

TO OVERCOME BIAS, privacy, and security risks and use open data in a trustworthy manner, agencies should play an active role to protect the data from both intentional tampering and unintentional inaccuracies. With a few key controls at every stage of the AI life cycle, government leaders can harness the benefits of accelerated AI and open data while preserving their integrity and accuracy.

Data gathering and preparation: Adopting data standards

Bias, privacy, and security risks can crop up at any point in the AI/ML life cycle; therefore, data scientists and developers should test for them throughout the development life cycle. It is possible to identify potential sources of risk within a dataset early on, especially with open datasets. Chief data officers can institutionalize the use of tools such as data cards to help data scientists document key information about the datasets. These cards can include information on the composition of data, the motivation behind putting the dataset together, and intended use cases. Data tagging allows developers to better understand data lineage, how it has been transformed over time, and its original context, allowing them to make more appropriate use of it in training models. Apart from data cards, chief data officers should emphasize on assessing the accuracy of data labels in open datasets. A study by MIT found an average of 3.4% errors across 10 popular open datasets sets, including ImageNet. The volume of errors ranged from 2,900 to over 5 million in the analyzed datasets.[20]



While controls such as data cards and assessment of data labeling errors can help govern data use within an organization, open data standards can help do so across an entire ecosystem. These are reusable agreements that make it easier for people and organizations to publish, access, share, and use better quality data.[21] Standards help data scientists and stewards thoroughly understand their datasets and thus make informed decisions as to whether they are ready to be used for training an AI model. Organizations, such as the Open Data Institute, have published guides designed to help organizations create shared vocabularies, taxonomies, and ontologies that can help fuel data exchange. In the health sector, open data standards have had a huge impact on supporting the response to the COVID-19 pandemic. As the central coordinating body for clinical terminology standards, the National Library of Medicine (NLM) has helped medical professionals collect patient data in a standardized way that ensures a base of comparison with other electronic health records (EHRs), allowing the health community to better track, diagnose, and treat the disease.[22]

While controls such as data cards and assessment of data labelling errors can help govern data use within an organization, open data standards can help do so across an entire ecosystem.

Model development: Employ explainable models with transparency

Many AI algorithms are commonly referred to as black boxes, as it can be difficult even for the creators of a model to know why it reached a certain conclusion. Organizations should focus on creating transparent algorithms or offer explanations for their outcomes.

While it may not be possible to completely explain the mechanism of the algorithm for many types of deep learning, generating different kinds of explanations about how the model worked can help people in different roles work with the model more effectively.[23] For example, one set of explanations can be for those impacted by an AI model's outputs. Such explanations are used to build trust and acceptance by explaining why a loan application was approved or rejected, for example. For an AI model developer, on the other hand, a more detailed explanation may be needed to help with debugging or improving an AI model.[24] The explanation for system developers or technical staff (such as data scientists) should help them identify when their models may be making spurious correlations, leading to poor in-production performance. The explainable model can also identify whether the problem originates from the model or from issues with the underlying data, such as under-representation of certain groups. This level of transparency can also be a critical safeguard to the security of the model in that it can help reveal when an outcome may have been the result of adversarial attempts at manipulating the model.

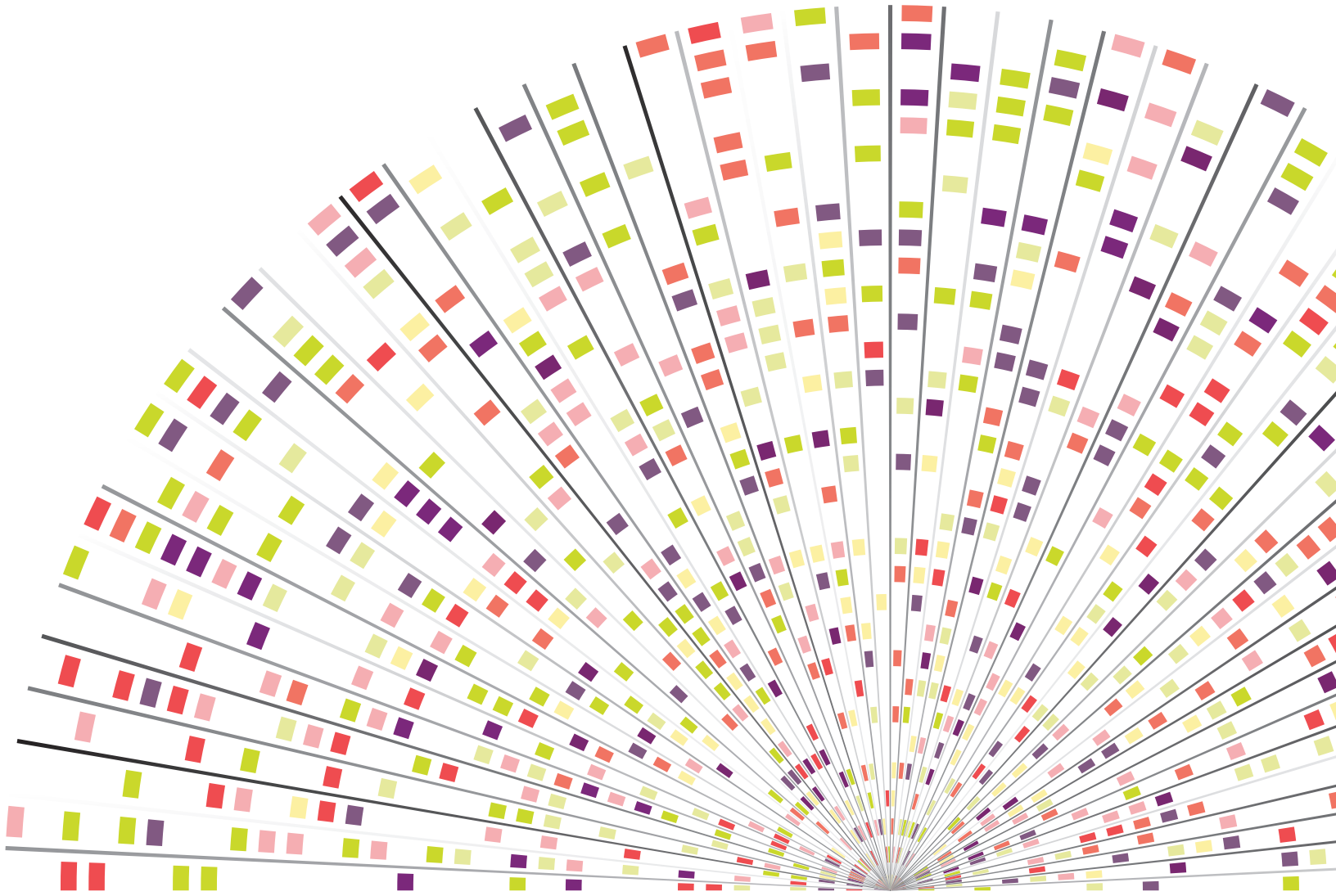
Such rules and other metrics can help data scientists determine if their model has a disparate impact on a race or sex. If such a metric flags a potential bias, strong understanding of the data used to train the model can help correct it. In the case of a lack of data representing a race or sex, the model developers could seek additional open data sources or collect data to supplement their training dataset.

As agencies look to develop more explainable models, they may have to balance trade-offs between accuracy and explainability. Simple algorithms based on linear regression, rule-based classifiers, or decision trees would be easier to explain, but complex algorithms could be more accurate because of their ability to model complex relationships between predictors.[25] Whether to prioritize accuracy or explainability would partly depend on the use case of algorithms. If an algorithm is used to approve or disapprove loans, grants, or patents, then the ability to explain the decision would give applicants a chance to improve input variables such as on-time payments. On the other hand, in cancer detection, patients are likely to value accuracy over whether the algorithm is easily explainable or not.

Model deployment: Apply corrections to mitigate imperfect data

Ensuring trustworthy AI is not confined to identifying the right data to train AI models. Risks exist throughout the life cycle, and while some of them can be identified and mitigated before training, others are discovered throughout the iterative process of model training, testing, and evaluation.

For example, developers can compare an AI model's outputs against set metrics only after it has been created. Metrics can help AI model developers determine if their model has an adverse impact on a protected class such as age, race, or sex. The US Equal Employment Opportunity



Commission (EEOC) developed one such rule—the four-fifths rule—to screen for adverse impacts in human resources decisions.[26] This rule states that adverse impact can be determined as a "selection rate for any race, sex, or ethnic group which is less than four-fifths (80%) of the rate for the group with the highest rate."[27] For instance, if a company hires 40% of male applicants for a specific role but the selection rate for female applicants is 20% for the same role, then the selection process can be judged as biased because the impact ratio is 0.5 (20% divided by 40%) which is less than 0.8 or 80%.[28]

But all is not lost if such biases are detected either in the model or the underlying data. Just as glasses can correct poor vision, data correction can address bias in models. For example, a cross-functional team of Deloitte professionals tested a public dataset of mortgage and loan applications for data and model bias. The analysis identified potential sources of historical representation bias within the original dataset and confirmed this hypothesis by finding indications of disparate impact in loan origination rates for applicants that identified as having two or more minority races, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, or Black or African American. To mitigate this bias, the team applied preprocessing bias mitigation techniques such as variable repair (i.e., modification of variable distributions in the training dataset) and were able to reduce model outcome bias at minimal cost to overall model accuracy.

Getting started

OPEN DATA CREATES myriad opportunities to accelerate AI development. As agencies release more open datasets, AI models will likely use them to drastically improve government operations and services. Agencies can create trustworthy AI by using data governance, deploying explainable AI models, and applying corrections to minimize the risk of bias even as they accelerate AI's deployment.

As agencies release more open datasets, AI models will likely use them to drastically improve government operations and services.

To get started, chief data officers should take the following steps that can improve the reliability of their data and AI programs:

- Build relationships with academia, industry, and other government agencies to ensure their organization has access to the latest tools and procedures for data governance and explainable AI.

- Promote data standards and tools that can help data scientists evaluate which datasets are appropriate for AI. For example, standards such as data cards can provide information on the context of a dataset's creation, allowing researchers to decide if it is a good fit for the model they would like to build, while tools that can tokenize data can help ensure both privacy and accuracy when dealing with sensitive datasets.

- Adopt MLOps and other process controls to help institutionalize data governance at every stage of the AI life cycle. MLOps are the set of automated pipelines, processes, and tools that streamline steps of AI model construction. In our survey of more than 500 government executives, respondents indicated that documenting and enforcing MLOps make organizations better prepared to navigate privacy and ethical risks arising from AI.[29]

- Agencies can conduct an extensive impact assessment of their open datasets to mitigate any privacy risks. The assessments can help organizations decide whether to release datasets to the public and, if released, what privacy measures should be taken.[30]

With these and other steps, government leaders can make use of open data to accelerate AI, more confident that it will bring the transformational benefits of AI to government and constituents while mitigating their exposure to new risks.

Endnotes

1. GE Healthcare, "What is ImageNet and why 2012 was so important," August 21, 2019.
2. IEEE Computer Society, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, accessed October 29, 2021; Dave Gershgorn, "The data that transformed AI research—and possibly the world," Quartz, July 26, 2017.
3. Jessi Hempel, "Fei-Fei Li's quest to make AI better for humanity," *Wired*, November 13, 2018.
4. Gershgorn, "The data that transformed AI research—and possibly the world"; GE Healthcare, "What is ImageNet and why 2012 was so important"; Khari Johnson, "ImageNet creators find blurring faces for privacy has a 'minimal impact on accuracy," *VentureBeat,* March 16, 2021.
5. Hempel, "Fei-Fei Li's quest to make AI better for humanity"; NYU Dispatch, "5 examples of using AI/deep learning for the government and public sector," 2021.
6. Chief Information Officer Council, "Foundations for Evidence-Based Policymaking Act of 2018," accessed October 29, 2021.
7. Kaiyu Yang et al., "A study of face obfuscation in ImageNet," arXiv, March 14, 2021.
8. Sonal Shah and William D. Eggers, *Introduction: The government CDO—Turning public data to the public good*, Deloitte Insights, October 12, 2018.
9. Terri Park, "Behind Covid-19 vaccine development," *MIT News*, May 18, 2021.
10. Sonal Shah and William D. Eggers, *Introduction: The government CDO—Turning public data to the public good*.
11. Will Knight, "Researchers blur faces that launched a thousand algorithms," *Wired*, March 15, 2021.
12. Olivier Thereaux, "Using artificial intelligence and open data for innovation and accountability," Open Data Institute, December 20, 2017.
13. Steve Lohr, "Facial recognition is accurate, if you're a white guy," *New York Times*, February 9, 2018.
14. Sara Stivers, "AI and bias in university admissions," *ISM Perspectives* magazine, Fall 2018.
15. Christophe Olivier Schneble, Bernice Simone Elger, and David Martin Shaw, "Google's Project Nightingale highlights the necessity of data science ethics review," *EMBO Molecular Medicine* 12, no. 3 (2020).
16. Alex Hern, "New York taxi details can be extracted from anonymised data, researchers say," *Guardian*, June 27, 2014.
17. Chris Duckett, "Re-identification possible with Australian de-identified Medicare and PBS open data," ZDNet, December 18, 2017.
18. Charles Q. Choi, "Medical imaging AI software is vulnerable to covert attacks," *IEEE Spectrum*, June 4, 2018.
19. Jessica Kent, "NIH makes largest set of medical imaging data available to public," HealthITAnalytics, July 23, 2018.
20. Kyle Wiggers, "MIT study finds 'systematic' labeling errors in popular AI benchmark datasets," *VentureBeat,* March 28, 2021.
21. Open Data Institute, "Open standards for data," accessed October 29, 2021.
22. Dianne Babski, "Health data standards: A common language to support research and health care," National Library of Medicine, January 27, 2021.

23. P. Jonathon Phillips et al., *Four principles of explainable artificial intelligence*, National Institute of Standards and Technology, US, Department of Commerce, August 17, 2020.

24. Ibid.

25. Natasha M., "Accuracy and bias in machine learning models," Big Data Made Simple, January 24, 2019.

26. Nathan Mondragon, "What is adverse impact? And why measuring it matters," HireVue, March 26, 2018.

27. Prevue, "Adverse impact analysis/four-fifths rule," May 6, 2009.

28. Sara Kassir, "Bias in conventional hiring tools: Understanding the status quo," pymetrics, June 9, 2021.

29. Edward Vanburen, William Eggers, Joe Mariani, Pankaj Kamleshkumar Kishanani, "Scaling AI in government: How to reach the heights of enterprisewise adoption of AI," Deloitte Insights, forthcoming.

30. Christopher Wilson, *Managing data ethics: A process-based approach for CDOs*, Deloitte Insights, February 7, 2019.

About the authors

Tasha Austin | laustin@deloitte.com

Tasha Austin is a principal in Deloitte's Risk and Financial Advisory business and the director of Deloitte's Artificial Intelligence Institute for Government. She focuses on amplifying Deloitte's capabilities and services in key areas such as trustworthy/ethical AI, provides insight-driven solutions to her clients, and is responsible for elevating Deloitte's thought leadership and digital presence in AI to the federal market. Tasha also leads Deloitte's strategic engagement initiatives with HBCUs and bridging the digital divide in under-resourced communities by working with nonprofit organizations to help advance equity and promote social justice.

Kara Busath | kbusath@deloitte.com

Kara Busath is a senior manager in Deloitte's Government & Public Services Risk & Financial Advisory practice. A statistician specializing in financial data management–related areas, she has significant experience developing complex data analytics to validate financial data, statistical sampling, quantitative methodologies including statistical modeling and regression, experimental design and analysis, financial compliance testing, and data-quality integrity testing. Over the past 13 years, Kara has served many agencies in the Department of Defense in their Audit Readiness efforts.

Allie Diehl | aldiehl@deloitte.com

Allie Diehl is a senior consultant in Deloitte's Analytics & Cognitive practice focused on growing AI strategy and adoption across the federal government. She has lectured on AI at Georgetown University and helped design an advanced degree class. She graduated Phi Beta Kappa from Johns Hopkins University with a double major in economics and international studies as a Hodson Scholar. She has previously worked as an equities research analyst and an actuarial analyst.

Pankaj Kishnani | pkamleshkumarkish@deloitte.com

Pankaj Kishnani is a researcher with the Deloitte Center for Government Insights. He specializes in emerging trends in technology and their impact on the public sector.

Joe Mariani | jmariani@deloitte.com

Joe Mariani leads research into emerging technologies for Deloitte's Center for Government Insights. His research focuses on the intersection of culture and innovation in both commercial businesses and government organizations. His work has appeared in the National Academy of Sciences, World Economic Forum, *US News & World Report, Wall Street Journal, Cyber Defense Review, Marine Corps Gazette*, and more. His previous experience includes work as a consultant to defense and intelligence organizations, high school science teacher, and Marine Corps intelligence officer.

Contact us

Our insights can help you take advantage of change. If you're looking for fresh ideas to address your challenges, we should talk.

Industry leadership

Edward Van Buren
Director, Deloitte AI Institute for Government | Deloitte Consulting LLP
+1 571 882 5170 | emvanburen@deloitte.com

Edward Van Buren is the Strategic Growth leader—Artificial Intelligence (AI) for Deloitte Consulting LLP's Government & Public Services (GPS) industry and the executive director of the Deloitte AI Institute for Government. He works with technology companies and other strategic partners to develop solutions, harnessing the power of AI/ML for federal, state, local, and higher education clients.

Tasha Austin
Principal | Deloitte & Touche LLP
+1 571 882 5479 | laustin@deloitte.com

Tasha Austin is a principal in Deloitte & Touche LLP's Government and Public Services practice and has nearly 20 years of professional services experience involving commercial and federal financial statement audits, fraud, dispute analysis and investigations, artificial intelligence, and advanced data analytics.

Center for Government Insights

William Eggers
Executive director | Deloitte Center for Government Insights | Deloitte Services LP
+1 571 882 6585 | weggers@deloitte.com

William Eggers is the executive director of Deloitte's Center for Government Insights, where he is responsible for the firm's public sector thought leadership. His most recent book is *Delivering on Digital: The Innovators and Technologies That Are Transforming Government.*

About Deloitte AI Institute for Government

The Deloitte AI Institute for Government is a hub of innovative perspectives, groundbreaking research, and immersive experiences focused on artificial intelligence (AI) and its related technologies for the government audience. Through publications, events, and workshops, our goal is to help government use AI ethically to deliver better services, improve operations, and facilitate economic growth.

We aren't solely conducting research—we're solving problems, keeping explainable and ethical AI at the forefront and the human experience at the core of our mission. We live in the Age of With—humans with machines, data with actions, decisions with confidence. The impact of AI on government and its workforce has only just begun. To learn more visit Deloitte.com.

About the Deloitte Center for Government Insights

The Deloitte Center for Government Insights shares inspiring stories of government innovation, looking at what's behind the adoption of new technologies and management practices. We produce cutting-edge research that guides public officials without burying them in jargon and minutiae, crystalizing essential insights in an easy-to-absorb format. Through research, forums, and immersive workshops, our goal is to provide public officials, policy professionals, and members of the media with fresh insights that advance an understanding of what is possible in government transformation.

Deloitte.
Insights

Sign up for Deloitte Insights updates at www.deloitte.com/insights.

Follow @DeloitteInsight

Deloitte Insights contributors
Editorial: Ramani Moses, Emma Downey, and Nairita Gangopadhyay
Creative: Sonya Vasilieff, Rahul Bodiga
Audience development: Alexandra Kawecki
Cover artwork: Sonya Vasilieff

About Deloitte Insights

Deloitte Insights publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte Insights is an imprint of Deloitte Development LLC.

About this publication

About Deloitte