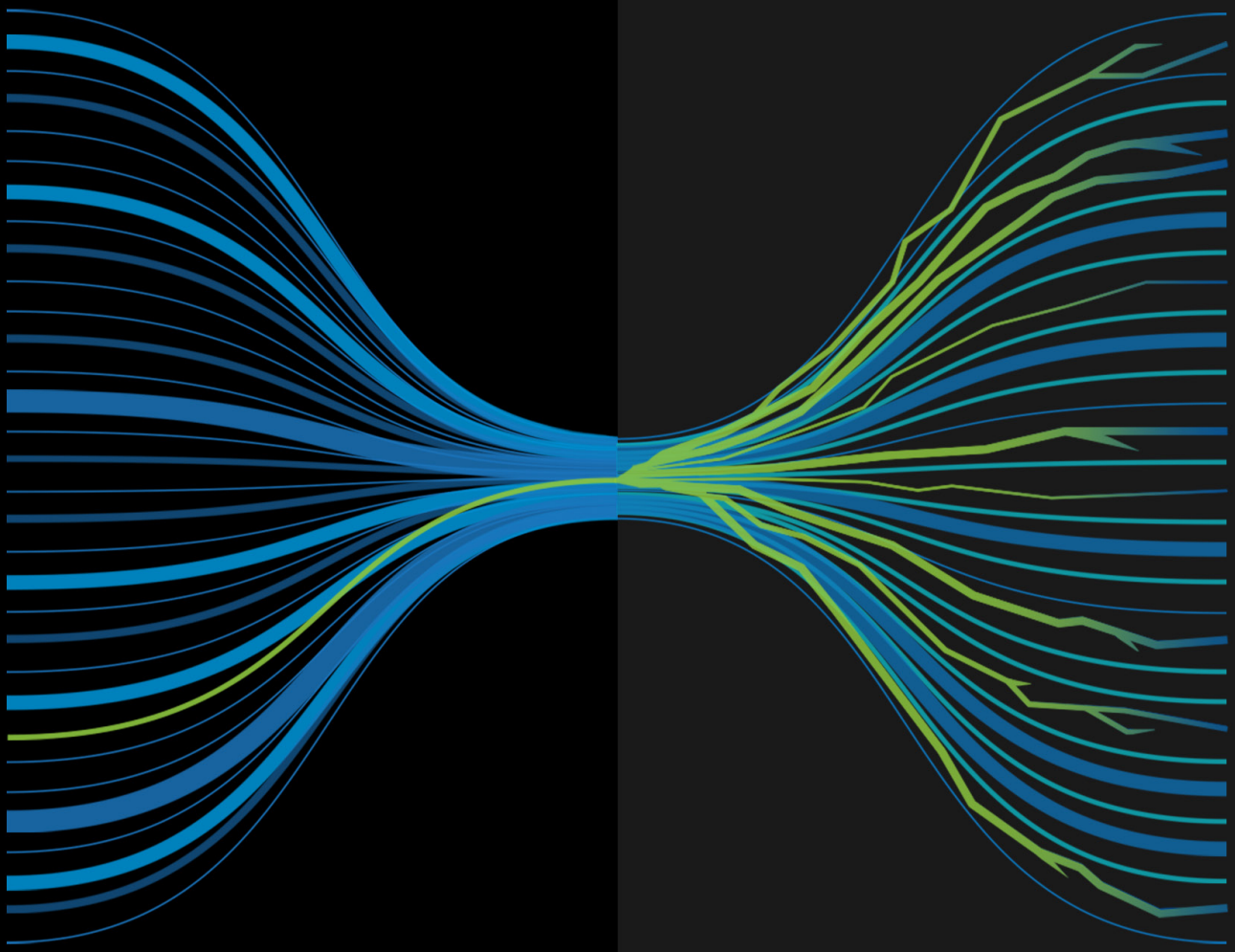
Deloitte. Insights

A report from the
Deloitte Center for Integrated Research



AI model bias can damage trust more than you may know. But it doesn't have to.

About the Deloitte Center for Integrated Research

The Deloitte Center for Integrated Research (CIR) offers rigorously researched and data-driven perspectives on critical issues affecting businesses today. We sit at the center of Deloitte's industry and functional expertise, combining the leading insights from across our firm to help leaders confidently compete in today's ever-changing marketplace. Learn more about our research.

Deloitte Future of Trust

Trust is the basis for connection. It is built moment by moment, decision by decision, action by action. In an organization, trust is an ongoing relationship between an entity and its varying stakeholders. When performed with a high degree of competence and the right intent, an organization's actions earn trust with these groups. Trust distinguishes and elevates your business, connecting you with the common good. Put trust at the forefront of your planning, strategy, and purpose and your customers will put trust in you. At Deloitte, we've made trust tangible—helping our clients measure, manage, and maximize it at every opportunity. Let's talk.

Deloitte AI Institute

The Deloitte AI Institute helps organizations connect all the different dimensions of the robust, highly dynamic and rapidly evolving AI ecosystem. The AI Institute leads conversations on applied AI innovation across industries, with cutting-edge insights, to promote human-machine collaboration in the "Age of With."

Deloitte AI Institute aims to promote the dialogue and development of artificial intelligence, stimulate innovation, and examine challenges to AI implementation and ways to address them. The AI Institute collaborates with an ecosystem composed of academic research groups, start-ups, entrepreneurs, innovators, mature AI product leaders, and AI visionaries, to explore key areas of artificial intelligence including risks, policies, ethics, future of work and talent, and applied AI use cases. Combined with Deloitte's deep knowledge and experience in artificial intelligence applications, the institute helps make sense of this complex ecosystem, and as a result, deliver impactful perspectives to help organizations succeed by making informed AI decisions.

No matter what stage of the AI journey you're in; whether you're a board member or a C-Suite leader driving strategy for your organization, or a hands-on data scientist, bringing an AI strategy to life, the Deloitte AI institute can help you learn more about how enterprises across the world are leveraging AI for a competitive advantage. Visit us at the Deloitte AI Institute for a full body of our work, subscribe to our podcasts and newsletter, and join us at our meetups and live events.
Let's explore the future of AI together.

Contents

Introduction

A LARGE REGIONAL BANK uses a newly developed fraud detection artificial intelligence (AI) algorithm to identify potential cases of bank fraud including anomalous patterns of financial transactions, loan applications, and new account applications. The algorithm is trained on an initial set of data to give an idea of what normal versus fraudulent transactions look like. However, the training data becomes biased by oversampling applicants over 45 years of age for examples of fraudulent behavior. This oversampling continues over a period of months, with the bias growing and remaining undetected. The model becomes more likely to think an older person is committing fraud than reality suggests. Customers are increasingly turned down for loans. Some begin to feel alienated while regulators start to ask questions. Trust is lost, the brand's reputation suffers, and the bank faces significant consequences to its bottom line.

We know model bias is potentially a problem, but do we really know how pervasive it is? Certainly, media outlets write stories that capture the public imagination, such as the AI hiring model that is unfairly biased against women[1] or the AI health insurance risk algorithm that unfairly assigns higher risk scores based on racial identity.[2] But as bad as such examples may be, the AI model bias story hardly ends with what we read in the popular press.

Our research indicates that model bias could be more prevalent than many organizations are aware and that it can do much more damage than we may assume, eroding the trust of employees, customers, and the public. The costs can be high: expensive tech fixes, lower revenue and productivity, lost reputation, and staff shortages, to say nothing of lost investments. In fact, 68% of executives surveyed in Deloitte's recent *State of AI in the Enterprise, 4th Edition* report reported that their functional group invested US$10 million or more in AI projects in the past fiscal year alone.[3] Even internal-facing models can do significant harm and potentially put those millions of dollars of investment at risk.

To solve this problem, we need to go beyond empathy and good intentions. Understanding, anticipating, and, as much as possible, avoiding the occurrence of model bias can be critical to advance the use of AI models across the organization in a way that preserves stakeholder trust. The good news is that there are approaches that organizations can adopt—including technology-based solutions—that can help.

Model bias within your organization may be more prevalent than you know

THE TERM "BIAS" carries many meanings. For the purposes of this study, we may consider Merriam-Webster's definition of bias as "systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others."[4] Generally speaking, AI model bias happens when the training data on which an AI algorithm or model relies is not reflective of the reality in which the AI is meant to operate. In other words, despite the use of the term "model bias," a model is not biased in and of itself; rather, it's the *training data* that renders a model bias. Stuart Battersby, CTO of AI enterprise software company Chatterbox Labs, concurs. "Regardless of context, often, [model bias risk] comes down to the training data," used to inform the model and any training data is vulnerable to bias, according to Battersby.[5] (See sidebar "Organizing the 'wild west' of model bias" for a discussion on the various ways model bias typically presents itself.)

Model bias is particularly troubling in part because it's not always anticipated by organizations or those who are working with the AI models in question. These "weapons of math destruction" as Cathy O'Neil suggests in her book are secret and scalable, which can magnify their danger to an organization and its stakeholders.[6]

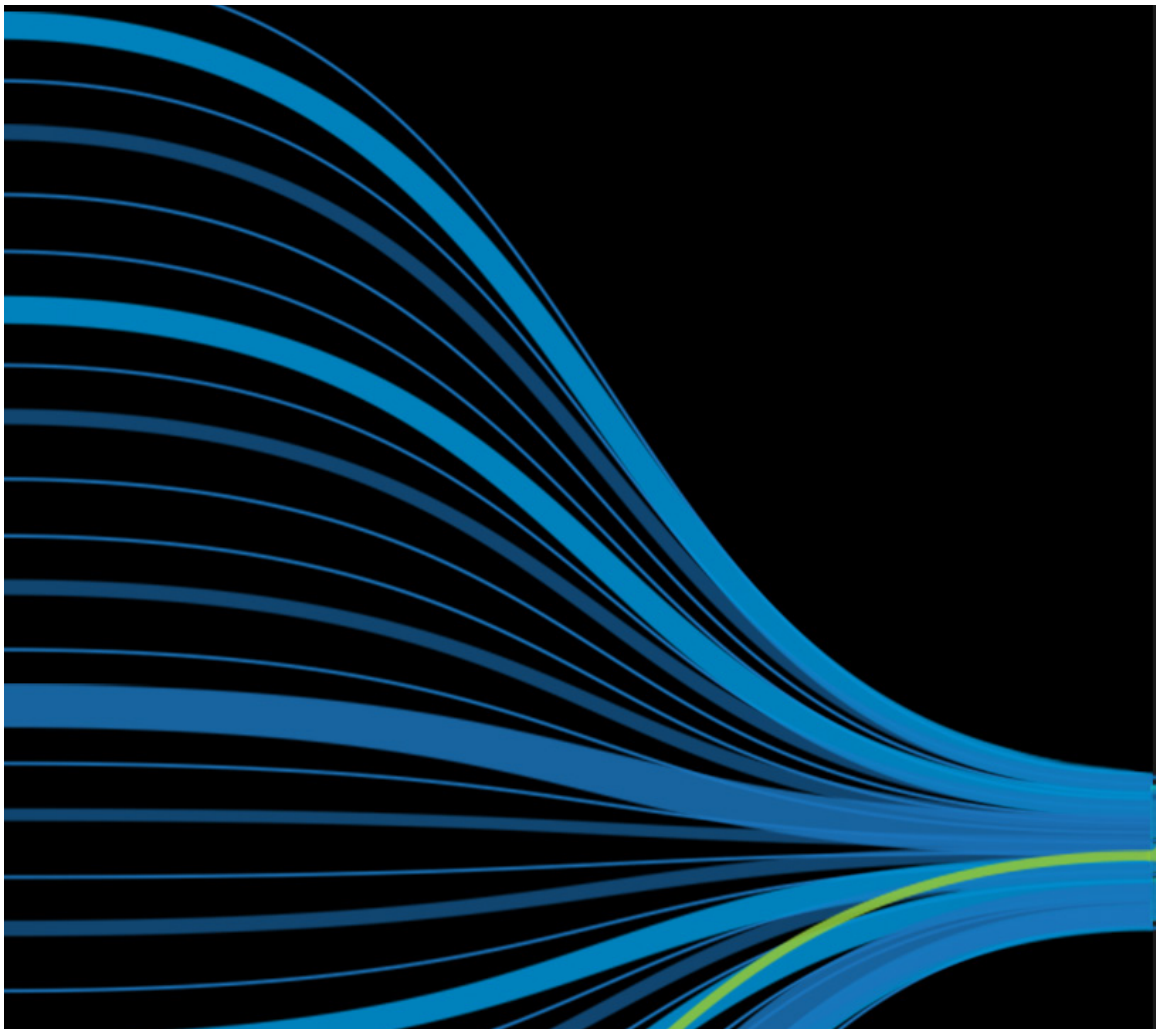
Evidence suggests that some users of AI models may be oblivious to this danger. Consider Deloitte's *State of AI* survey in which some three quarters of overall respondents say they are "confident" or "very confident" that their deployed models will exhibit qualities of fairness and impartiality. A similar share said they are "confident" or "very confident" that their deployed models will exhibit qualities of robustness and reliability.[7] These data points are important because such characteristics as fairness and robustness are the hallmarks of models that operate as they should, without bias.

Model bias is particularly troubling in part because it's not always anticipated by organizations or those who are working with the AI models in question.

Stories of bias found in AI models that speak to societal discrimination and prejudice reside in multiple contexts, including college acceptance decisions,[8] criminal sentencing and parole decisions,[9] and hiring decisions,[10] among many others. Many examples of model bias mentioned publicly relate to bias found in models that serve customer-facing functions. Our research indicates, however, that bias risks are prevalent whether we're

referring to models that affect customers or within the operational or internal part of an organization. Some of these model risks within the "back office" of an organization are often undetected until long after deployment and the accompanying impacts. Indeed, the risk of model bias within an internal operating domain like cybersecurity or compliance may be especially insidious as internal models may not receive the degree of public scrutiny that more outwardly facing deployments may receive, thus delaying their detection. Jayant Narayan, World Economic Forum Artificial Intelligence and Machine Learning Technology Policy lead, says: "Most AI model bias discussion is still on the external facing functions and the use cases of industries that are more customer facing. Companies should reassess bias and risk classification for their internal functions and use cases."[11]

Put another way, AI model bias is domain agnostic. In all of its forms, it can occur anywhere an AI model is deployed, regardless of context. Where context does matter, as we'll discuss, is in the impact of model bias on trust.



ORGANIZING THE "WILD WEST" OF MODEL BIAS

Several classes or archetypes of model bias emerged during our research. We identify two main groups of biases based on the type of action that impacts the model: "Passive" bias—where bias is not the result of a planned act—and "active" bias—where the bias occurs because of human action, either with or without intent and, even when intentional, often without *negative* intent. Both types of bias can manifest in different ways, and both should be considered when developing strategies to mitigate model bias risk. In characterizing bias in the classification that follows, we use our own terms as well as terms that are commonly observed in social science and technology literature.[12]

Passive bias

Examples of passive bias may include:

- *Selection bias*: Overinclusiveness or underinclusiveness of a group; insufficient data; poor labeling. An example of selection bias may be found in an AI model trained on data in which a particular group is identified with a certain characteristic at a higher rate than objective reality justifies.
- *Circumstantial bias*: Training data staleness; changing circumstances. An example of circumstantial bias may include a predictive AI model trained on data that was accurate originally but is no longer accurate because of changing realities or "facts on the ground."
- *Legacy or associational bias*: AI models trained on terms or factors associated with legacies of bias based on race, gender, and other grounds, even though unintentionally. One example is found in a hiring algorithm trained on data that, while not overtly gender-biased, refers to terms that carry a legacy of male association.

Active bias

Examples of active bias may include:

- *Adversarial bias*: Data poisoning; post-deployment adversarial bias. A hostile actor, for example, gains access to a model's training data and introduces a bias for nefarious objectives.
- *Judgment bias*: Model is trained properly, but bias is introduced by a model user during implementation by way of misapplication of AI decision output. For example, a model may produce objectively correct results, but the end user misapplies those results in a systemic fashion. In that sense, judgment bias differs from other model biases in that it is not the direct result of flawed training data.[13]

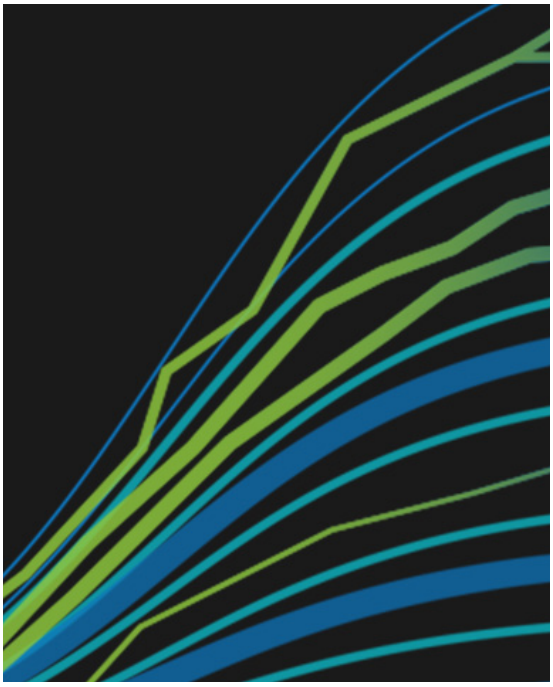
The above grouping is far from exhaustive or definitive; other bias characterizations exist. Such speaks to the evolving and still nascent understanding of what model bias is and how it occurs.

The trust connection: Model bias may be exponentially more damaging than you know

THE IMPACT OF AI model bias can cascade across an organization by impacting its decision-making and trust with stakeholders. Decision-making and trust are two separate but interrelated concepts. Trust is the foundation of a meaningful relationship between an organization and its stakeholders at both the individual and organizational levels. Trust is built through actions that demonstrate a high degree of competence and intent, that result in exhibited capability, reliability, transparency, and humanity. Competence is foundational to trust and refers to the ability to execute, to follow through on your brand promise. Intent refers to the reason behind your actions, including fairness, transparency, and impact. One without the other doesn't build or rebuild trust. Both are needed.

When a poor decision is made based on faulty analysis from biased data, an organization risks losing trust with stakeholders who may be relying on a model's advice. This could manifest, for example, in board members who lose trust in an executive team that recommends an unprofitable project or employees who question the hiring of a less qualified candidate.

Once a decision error occurs and trust breaks down with a given stakeholder, that stakeholder's behavior can change. For an employee, this could mean less engagement at work, for a customer, lower brand loyalty or, for a supply chain partner, less willingness to recommend the business to others. These behavioral changes can have a meaningful impact on organizational performance, possibly limiting sales, productivity, and profitability. Ultimately, the lack of trust can prevent a company from fulfilling its goals and purpose with stakeholders.



Consider the bank to which we referred at this paper's outset. In that example, AI model bias impacts decision-making in leading a bank to make unfair assumptions about older credit applicants and, as a result, avoid selling products to the older, underserved market. The reverse could also be true with bias leading a bank to grant loan applications to younger applicants who are actually engaging in fraud. And once this bias is known—even if the bank made efforts to correct it—bank professionals may lose confidence in the output of the algorithm. Indeed, they may lose confidence in AI models more generally. As a result, they may avoid important business decisions such as pursuing actual cases of fraud.

Multiple stakeholders are impacted by the model bias in this example. This bias, if it leads a bank to underserve the older banking customer, may alienate a constituency. This would put their trust and patronage at stake. It may also jeopardize the trust and business of other customers who become aware of and are offended by this bias, even if not directly affected. Because this bias may run afoul of various regulatory and statutory requirements as found in the Equal Credit Opportunity Act, it may damage the trust of regulatory authorities in ways that could result in civil penalties that affect the bottom line.[14] Ultimately, the consequences of this model bias could harm the bank's reputation and bottom-line performance.

This is just one of many examples of the consequence to decision-making and trust when AI models are unfairly biased (figure 1). The impact of AI model bias is typically not limited to one stakeholder group. On the contrary, the faulty decisions that result most often impact multiple stakeholder groups and can negatively influence their willingness to trust an organization. This context within which that bias takes place—the set of decisions, stakeholders, and behavioral changes that result—can define the stakes and cost to the organization.

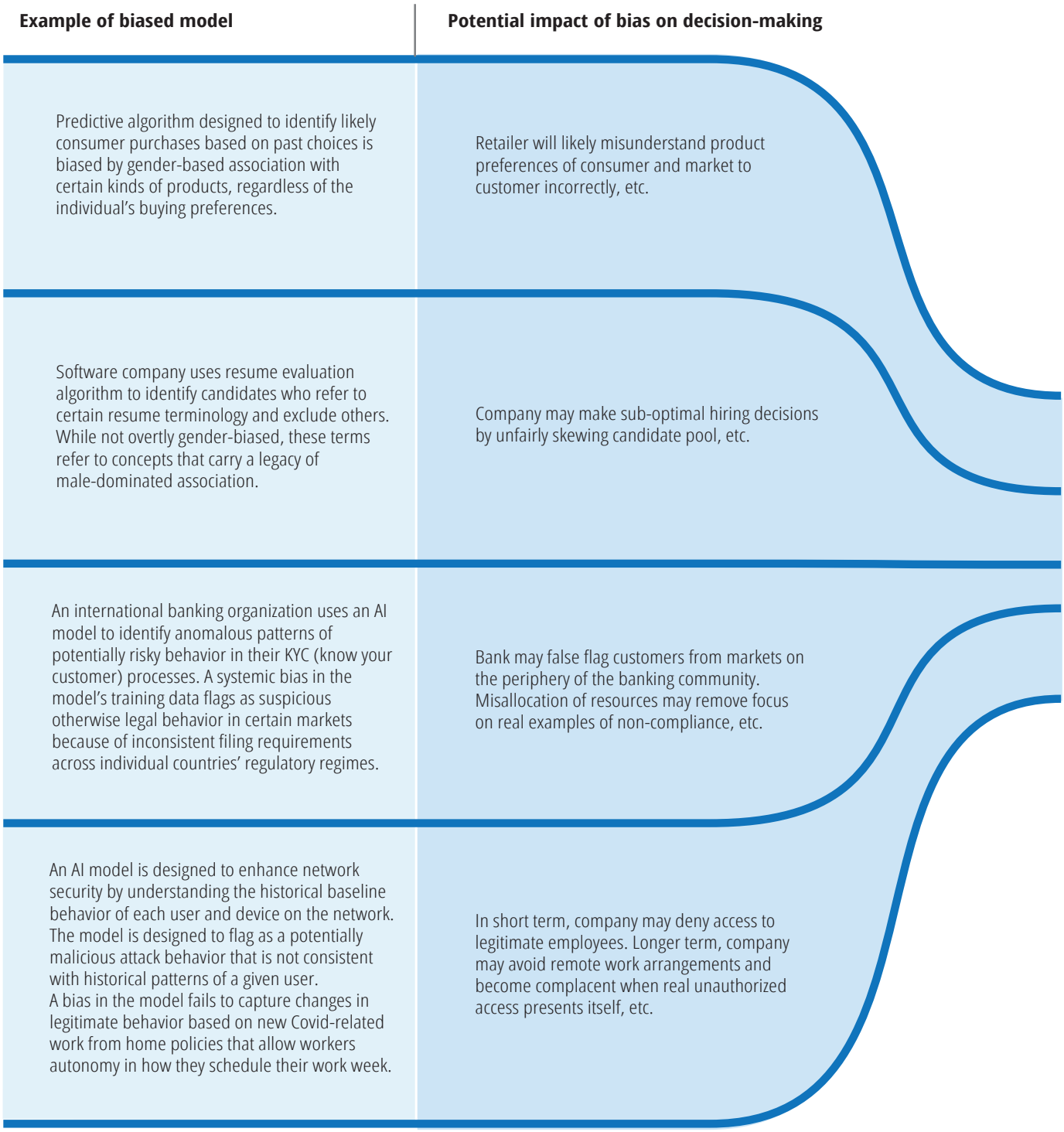
The impact of AI model bias is typically not limited to one stakeholder group. On the contrary, the faulty decisions that result most often impact multiple stakeholder groups and can negatively influence their willingness to trust an organization.

To illustrate the individual character of model bias, we depict a few different case scenarios in which the nature of model bias could manifest and how decision-making and trust might be affected as a result. (figure 1)

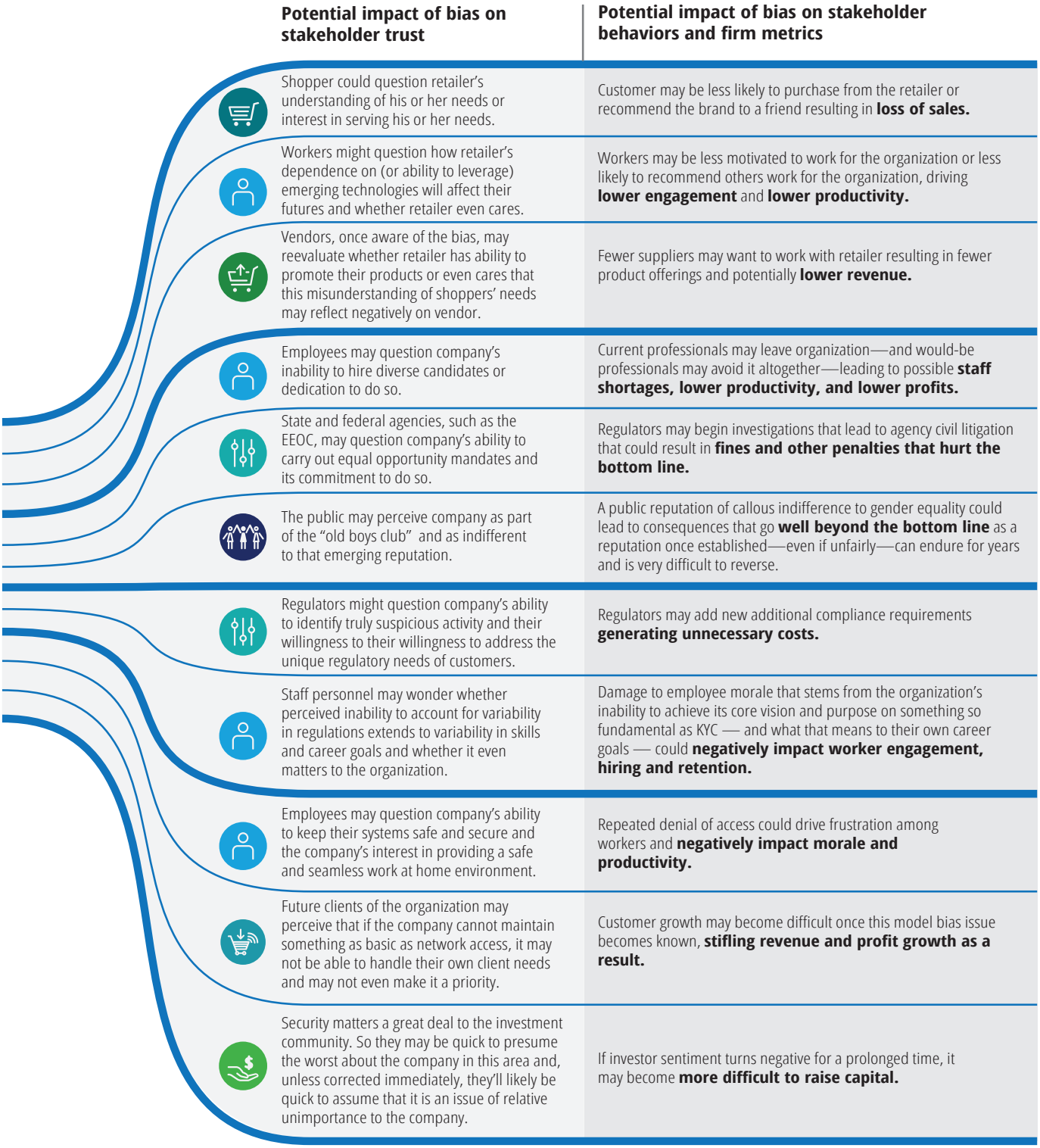
FIGURE 1

Model bias scenarios and their potential impact on decision-making and trust

Customers Employees Suppliers Regulators Community Potential customers Investors



Source: Deloitte Analysis.[15]



Potential impact of bias on stakeholder trust	Potential impact of bias on stakeholder behaviors and firm metrics
Shopper could question retailer's understanding of his or her needs or interest in serving his or her needs.	Customer may be less likely to purchase from the retailer or recommend the brand to a friend resulting in **loss of sales.**
Workers might question how retailer's dependence on (or ability to leverage) emerging technologies will affect their futures and whether retailer even cares.	Workers may be less motivated to work for the organization or less likely to recommend others work for the organization, driving **lower engagement** and **lower productivity.**
Vendors, once aware of the bias, may reevaluate whether retailer has ability to promote their products or even cares that this misunderstanding of shoppers' needs may reflect negatively on vendor.	Fewer suppliers may want to work with retailer resulting in fewer product offerings and potentially **lower revenue.**
Employees may question company's inability to hire diverse candidates or dedication to do so.	Current professionals may leave organization—and would-be professionals may avoid it altogether—leading to possible **staff shortages, lower productivity, and lower profits.**
State and federal agencies, such as the EEOC, may question company's ability to carry out equal opportunity mandates and its commitment to do so.	Regulators may begin investigations that lead to agency civil litigation that could result in **fines and other penalties that hurt the bottom line.**
The public may perceive company as part of the "old boys club" and as indifferent to that emerging reputation.	A public reputation of callous indifference to gender equality could lead to consequences that go **well beyond the bottom line** as a reputation once established—even if unfairly—can endure for years and is very difficult to reverse.
Regulators might question company's ability to identify truly suspicious activity and their willingness to their willingness to address the unique regulatory needs of customers.	Regulators may add new additional compliance requirements **generating unnecessary costs.**
Staff personnel may wonder whether perceived inability to account for variability in regulations extends to variability in skills and career goals and whether it even matters to the organization.	Damage to employee morale that stems from the organization's inability to achieve its core vision and purpose on something so fundamental as KYC — and what that means to their own career goals — could **negatively impact worker engagement, hiring and retention.**
Employees may question company's ability to keep their systems safe and secure and the company's interest in providing a safe and seamless work at home environment.	Repeated denial of access could drive frustration among workers and **negatively impact morale and productivity.**
Future clients of the organization may perceive that if the company cannot maintain something as basic as network access, it may not be able to handle their own client needs and may not even make it a priority.	Customer growth may become difficult once this model bias issue becomes known, **stifling revenue and profit growth as a result.**
Security matters a great deal to the investment community. So they may be quick to presume the worst about the company in this area and, unless corrected immediately, they'll likely be quick to assume that it is an issue of relative unimportance to the company.	If investor sentiment turns negative for a prolonged time, it may become **more difficult to raise capital.**

Model bias should be addressed in a proactive and holistic way

ONCE AN INCIDENT of model bias is found, the organization should "get under the hood" to assess the nature of the bias (including its causes), the ways it's already affected decision-making and, ultimately, stakeholder trust, and how to prevent its reoccurrence. As Chatterbox Lab's Battersby says, "You want to really get to the root cause as to why you have that bias and what that means within your organization in order to prevent it from occurring again."[16] With that said, reacting to a bias already in place is far less preferable than anticipating and preventing the bias from originating at all—or at least before deployment. Ted Kwartler, vice president of Trusted AI at DataRobot, puts it this way: "Finding bias in models is fine, as long as it's before production. By the time you're in production, you're in trouble."[17]

The following set of guideposts can help organizations anticipate AI model bias across contexts. Such guideposts can help an organization to deploy AI models in ways that are fair and transparent.

1. *Educate all within the organization about the potential for AI model bias risk*. Even among those most directly involved in the development and deployment of AI models, biases are not always front of mind. For others throughout the organization, model bias is often an abstraction that only becomes an issue after the bias and its accompanying impacts become obvious. Leaders and workers within the organization—throughout the C-suite and beyond—should understand the strategic imperative that model bias represents because everyone throughout the organization can be affected by it. Such education should target end users of the model across departments such as marketing and HR, so they can be alert to the potential for bias to exist and cautious that they don't unintentionally introduce a bias through faulty implementation.

2. *Establish a common language to discuss model risk and methods to mitigate it*. Trustworthy AI, also known as ethical or responsible AI, shares common themes in the development and use of AI applications. These themes include fairness, transparency, reliability, accountability, safety and security, and privacy. Such themes provide a common language and lens for evaluating and mitigating AI risks, including model bias. Organizations can consider these themes when designing, developing, deploying, and operating AI systems. Each of these themes articulates an aspect of what, together, makes for trustworthy AI. Each supports the organization's ability to deploy AI models competently and with the right intent.[18]

3. *Ensure that humans who are most impacted by the model are "in the loop" when developing the model.* Our research reveals that humans tend to believe in the accuracy of AI model decisions without any real understanding of how the model works or was developed.[19] This is an especially precarious practice when model bias enters the picture. Each part of the AI model life cycle should routinely reflect a partnership between the technology and all stakeholders. "Bias can be managed if there's a human in the loop," says Chatterbox Labs CEO Danny Coleman. But humans in the loop are not just those who develop and deploy the models. It's also about the end consumers of the model's decision outputs. They should be as much a part of how a model is developed (understanding what it can and cannot do) as anyone to mitigate the potential damage to trust that a problem can bring. Coleman calls this "managing stakeholder expectations."[20] And this involvement of stakeholders should start at the model conception stage. Preeti Shivpuri, Deloitte Canada Trustworthy AI leader, puts it this way: "Engaging consultations with different stakeholders and gathering diverse perspectives to challenge the status quo can be critical in addressing inherent biases within data and making AI systems inclusive from the start."[21]

4. *Include process and technology as well.* "Bias is a challenge. It's always going to be there. But I think the best way to solve for it is with a people, process, and technology approach." So says Chatterbox Lab's Coleman.[22] Humans play an integral role in the AI development life cycle and bias mitigation. But humans are only a part of a larger, integrated schematic that makes trustworthy AI possible.

Ted Kwartler, vice president of Trusted AI at DataRobot, puts it this way: "Finding bias in models is fine, as long as it's before production. By the time you're in production, you're in trouble."

In other words, any solution to the challenge of AI model bias should be *holistically* based on an integration of people, process, and technology. No one aspect of this three-legged stool is necessarily more important than another. Human judgement is important, as we mentioned. Process provides a sense of order and discipline to AI model governance. It includes monitoring and correcting for model bias that, together, help form the sequential steps of operationalizing machine learning models, sometimes referred to as "MLOps."[23] Technology, for its part, is the third leg of the three-legged stool. Without it, the model (and any model bias) would not exist. But technology is also part of the solution. Software platforms are now being developed that can help organizations uncover bias and other vulnerabilities, and help ensure that a model operates fairly.[24]

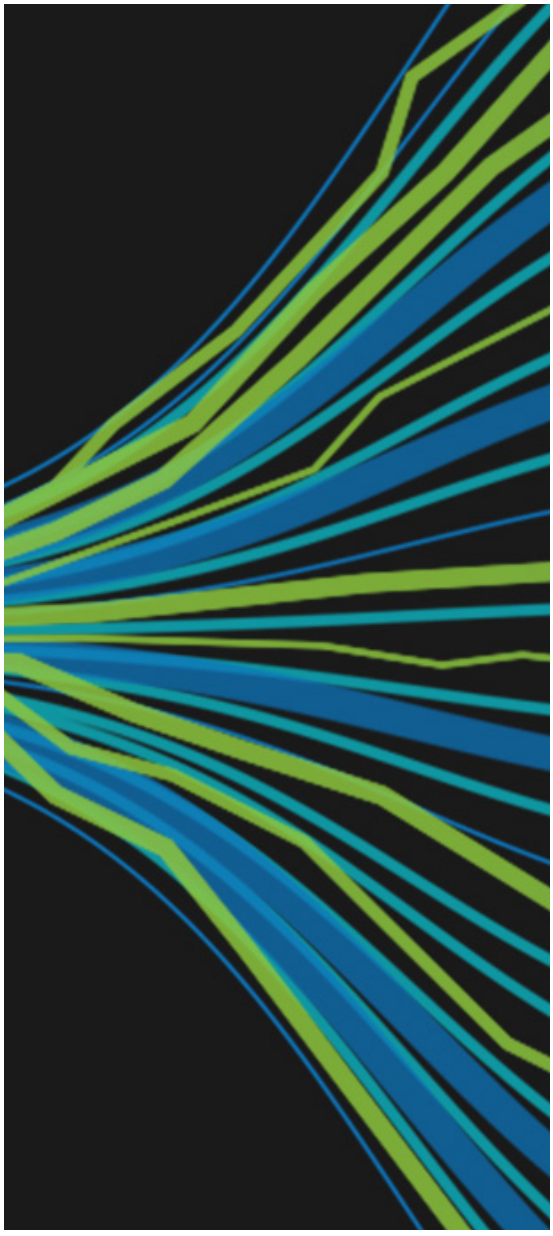
Moving forward with intention

BUILDING TRUST WITH stakeholders is a multifaceted, complex challenge. We are all connected. When trust breaks down with one stakeholder, others become aware and may change their behaviors as well.

AI and trust share an inseparable relationship. Trust cannot flourish in an environment that relies on flawed AI and even the most unbiased AI model can provide decision outcomes that matter very little if they serve an untrusting environment. The primary reason that organizations should think about AI model bias is that—more than many issues—bias has the potential to undermine this relationship.

Organizations should meet the challenge of AI model bias with the sense of urgency that such a consequential issue deserves. To some, model bias may seem like an emerging, far-flung abstraction. But it is real. And the damage it can cause to stakeholder trust is real, whether organizations focus on it or not.

But there is a path forward. Organizations have at their disposal the tools and resources to help address the challenge of AI model bias before it manifests—through a holistic approach that includes education, common language, and unrelenting awareness. The organization that chooses a proactive approach now will likely have a leg up on the organization that is required to take a reactive approach later.



Endnotes

1. Corinne Purtill, "Algorithms learn our workplace biases. Can they help us unlearn them?," *New York Times*, March 12, 2020.
2. Starre Vartan, "Racial bias found in a major health care risk algorithm," *Scientific American*, October 24, 2019.
3. Beena Ammanath et al., *Becoming an AI-fueled organization: Deloitte's State of AI in the Enterprise, 4th edition*, Deloitte Insights, October 21, 2021.
4. Merriam-Webster, "Bias," accessed November 2, 2021.
5. Based on personal interview.
6. Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, New York: Crown, 2016.
7. Ammanath et al., *Becoming an AI-fueled organization*.
8. Lilah Burke, "The death and life of an admissions algorithm," Inside Higher Ed, December 14, 2020.
9. Karen Hao, "AI is sending people to jail—and getting it wrong," *MIT Technology Review*, January 21, 2019.
10. Miranda Bogen, "All the ways hiring algorithms can introduce bias," *Harvard Business Review*, May 6, 2019.
11. Based on personal interview.
12. For additional reading on algorithmic biases, see, for example, Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, Brookings, May 22, 2019; Wenlong Sun, Olfa Nasraoui, and Patrick Shafto, "Evolution and impact of bias in human and machine learning algorithm interaction," *PLoS One 15 no.8, 2020;* David Danks and Alex John London, "Algorithmic bias in autonomous systems," presented at 26th International Joint Conference on Artificial Intelligence 2017, accessed November 2, 2021.
13. For more on the interplay between AI systems and human judgment, check, for example, Avi Goldfarb and Jon Lindsay, *Artificial intelligence in war: Human judgment as an organizational strength and a strategic liability*, Brookings, November 2020; Ajay Agrawal, Jashua Gans, and Avi Goldfarb, "Prediction machines: The simple economics of artificial intelligence," Harvard Business Review Press, 2018.
14. Federal Trade Commission, "Equal Credit Opportunity Act," accessed November 2, 2021.
15. AI model bias scenarios are illustrative and no reference to actual case examples is intended.
16. Based on personal interview.
17. Ibid.
18. Deloitte, "Trustworthy AI: Bridging the ethics gap surrounding AI," accessed November 2, 2021; Kate Schmidt and Matt Furlow, *Investing in trustworthy AI*, Deloitte, accessed November 2, 2021.
19. See, for example, Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos, "To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making," *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1, 2021, pp. 1-21.
20. Based on personal interview.

21. Based on personal interview; Deloitte, "Building trust in AI: How to overcome risk and operationalize AI governance," accessed November 2, 2021.
22. Based on personal interview.
23. Beena Ammanath et al., *MLOps: Industrialized AI*, Deloitte Insights, October 23, 2020.
24. Deloitte has partnered with Chatterbox Labs to develop one such solution. Deloitte Consulting LLP, "Deloitte AI Institute teams with Chatterbox Labs to ensure ethical application of AI," *PR Newswire*, March 15, 2021; Jett Oristaglio, "Introducing DataRobot bias and fairness testing," DataRobot, December 15, 2020.

Acknowledgments

The authors wish to thank **Danny Coleman** (Chatterbox Labs), **Stuart Battersby** (Chatterbox Labs), **Ted Kwartler** (DataRobot), and **Jayant Narayan** (World Economic Forum) as well as Deloitte professionals **Preeti Shivpuri** (Deloitte Canada), **Alexey Surkov**, **David Schatsky**, **Tasha Austin, Susanne Hupfer**, **David Jarvis**, **Rod Sides**, and **Brenna Sniderman** for the time they spent in sharing their invaluable insights. The authors would also like to thank **Saurabh Rijhwani** for his marketing support and **Negina Rood** for her vital research support. The authors extend a special thanks to **Kate Schmidt** for her ongoing insights and support throughout the development of this paper.

About the authors

Don Fancher | dfancher@deloitte.com

Don Fancher is a Deloitte Risk & Financial Advisory principal with Deloitte Financial Advisory Services LLP. He serves as the global leader of Deloitte Forensic as well as the coleader of Deloitte's Legal Business Services practice. Fancher has more than 30 years of experience assisting clients and leading practices in forensic, dispute consulting, and legal transformation. He currently leads more than 4,500 Deloitte professionals around the world serving clients in areas such as financial crime, disputes and investigations, business insurance, discovery, data governance, legal transformation, and contract life cycle management.

Beena Ammanath | bammanath@deloitte.com

Beena Ammanath is executive director of the Global Deloitte AI Institute and leads Trustworthy AI & Ethical Tech at Deloitte. She is the author of the upcoming book releasing in spring 2022—*Trustworthy AI*—which helps businesses navigate trust and ethics in AI. Ammanath is an award-winning senior executive with extensive global experience in AI and digital transformation, spanning across e-commerce, finance, marketing, telecom, retail, software products, services, and industrial domains. Ammanath is also the founder of nonprofit, Humans For AI, an organization dedicated to increasing diversity in AI.

Jonathan Holdowsky | jholdowsky@deloitte.com

Jonathan Holdowsky is a senior manager with Deloitte Services LP and part of Deloitte's Center for Integrated Research, managing a wide array of thought leadership initiatives on the issues of strategic importance to clients within the consumer and manufacturing sectors. His current research explores the promise of emerging technologies such as additive and advanced manufacturing, Internet of Things, Industry 4.0, cloud, AI, and blockchain and digital assets. Holdowsky is also deeply engaged in Deloitte's *Future of Trust* initiative that examines the role that trust plays in every domain of human activity.

Natasha Buckley |nbuckley@deloitte.com

Natasha Buckley is a senior manager with Deloitte Services LP and part of Deloitte's Center for Integrated Research. She leads research projects exploring trust, digital transformation, and the future of work. Her current research focuses on trust measurement and driving trust across the enterprise with different stakeholder groups.

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

Beena Ammanath
Executive director of Deloitte AI Institute | Deloitte Consulting LLP
+1 925 474 7139 | bammanath@deloitte.com

Beena Ammanath is executive director of the Global Deloitte AI Institute and leads Trustworthy AI & Ethical Tech at Deloitte.

Don Fancher
Principal | Deloitte Risk & Financial Advisory | Deloitte Financial Advisory Services LLP
+1 404 220 1204 | dfancher@deloitte.com

Don Fancher is a Deloitte Risk & Financial Advisory principal with Deloitte Financial Advisory Services LLP. He serves as the global leader of Deloitte Forensic as well as the coleader of Deloitte's Legal Business Services practice.

The Deloitte Center for Integrated Research

Jonathan Holdowsky
Senior manager | The Deloitte Center for Integrated Research | Deloitte Services LP
+1 617 437 3198 | jholdowsky@deloitte.com

Jonathan Holdowsky is a senior manager with Deloitte Services LP and part of Deloitte's Center for Integrated Research, leading thought leadership initiatives that explore the promise of emerging and disruptive technologies.

Natasha Buckley
Senior manager | The Deloitte Center for Integrated Research | Deloitte Services LP
+1 617 437 2585 | nbuckley@deloitte.com

Natasha Buckley is a senior manager with Deloitte Services LP and part of Deloitte's Center for Integrated Research. She studies how companies across industries and geographies are progressing in their digital journey.

Deloitte.
Insights

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

Follow @DeloitteInsight

Deloitte Insights contributors

Editorial: Andy Bayiates, Emma Downey, Dilip Poddar, and Ribhu Ranjan

Creative: Kevin Weier, Sonar Meena, and Sanaa Saifi

Audience development: Roshni Thawani

Cover artwork: Kevin Weier

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