



FEATURE

Conquering AI risks

Unpacking and alleviating concerns that threaten AI advancement

Beena Ammanath, David R. Novak, Siri Anderson, and Abha Kulkarni

A REPORT BY THE DELOITTE CENTER FOR INTEGRATED RESEARCH AND THE DELOITTE AI INSTITUTE

AI is already transforming organizations across industries, but emerging risks are generating real unease—and slowing AI adoption. Fortunately, leaders' concerns can be both managed and alleviated.

Introduction

The age of pervasive AI is here.¹ Since 2017, Deloitte's annual *State of AI in the Enterprise* report has measured the rapid advancement of AI technology globally and across industries. In the most recent edition, published in July 2020, a majority of those surveyed reported significant increases in AI investments, with more than three-quarters believing that AI will substantially transform their organization in the next three years. In addition, AI investments are increasingly leading to measurable organizational benefits: improved process efficiency, better decision-making, increased worker productivity, and enhanced products and services.² These possible benefits have likely driven the growth in AI's perceived value to organizations—nearly three-quarters of respondents report that AI is strategically important, an increase of 10 percentage points from the previous survey.

However, a growing unease threatens this rising trendline: Fifty-six percent of surveyed organizations say they plan to slow or are already slowing AI adoptions because of concern about emerging risks—a remarkable level of apprehension considering AI's acknowledged benefits and strategic importance. To better understand the underlying issues at stake, we analyzed respondents' concerns based on three main categories: **confidence** in the AI decision-making process, **ethics** of AI and data use, and **marketplace uncertainties**.³

We aim to explore these concerns in detail and offer insight into what companies can do to competently manage these risks. By doing so, organizations can build the internal assuredness

needed to continue investing and innovating—and increase external stakeholders' trust that AI strategies can be executed as intended.

Fifty-six percent of surveyed organizations intend to slow or are already slowing AI adoptions because of concern about emerging risks.

Understanding the threats to AI progress

Discussion of AI-related risks has been growing in recent years for both business leaders and the general public. Debates about whether AI will kill or create jobs are ongoing; meanwhile, reports of AI bias or failures frequently break into the headlines.⁴ In July 2020, public backlash and widespread criticism shut down AI-powered startup Genderify after only five days. The company claimed that its AI could identify a person's gender by analyzing their name, username, and email address; it pitched this capability to businesses as an enhancement to customer data. The system reportedly included various gender misconceptions and made decisions based on faulty assumptions. For instance, records containing "scientist" returned only a 4.3% probability that the individual was female, despite far higher current representation of women in the sciences.⁵ "Professor" resulted in a 98.4% probability for male, when men hold only 74% of tenured positions in the United States (a percentage that itself reflects inequality).⁶

Examples such as this—and the public backlash that often accompanies their exposure—are making many leaders understandably skittish about how they roll out AI within their organization and for customers, employees, and wider ecosystem partners. To understand more precisely where organizations’ biggest apprehensions originate, we asked survey respondents to rate a list of concerns on a scale from “minimal” to “extreme.” We grouped the listed risks into three categories: *confidence*, *ethics*, and *marketplace uncertainties*.

Confidence refers to whether the company generally believes an AI tool itself is reliable. This includes trusting the insights and decisions made by AI, trust in system security, and being able to

understand, justify, or explain AI’s decision-making process. **Ethics** refers to questioning whether using AI is good and right for society, regardless of AI’s benefit to the business, including issues of data privacy, fairness, bias, and the potential for job loss. **Marketplace uncertainties** are factors that are further outside the company’s direct control, such as the changing regulatory landscape and public or employee opinion. These factors are not necessarily related to the quality of the AI, but they affect how organizations can implement or use it.

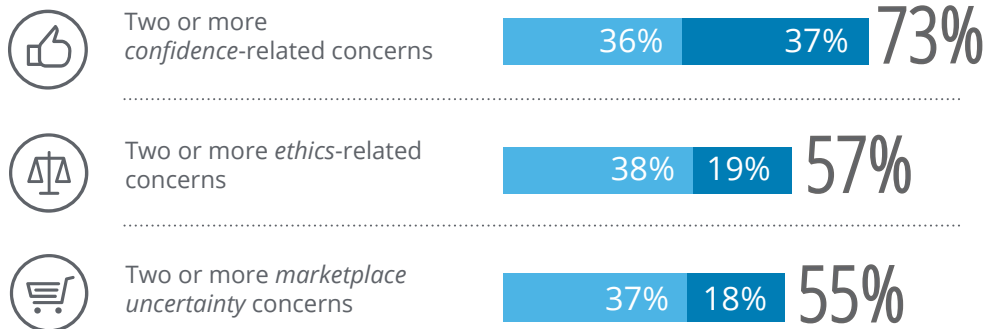
Figure 1 reflects the percentages of respondents who rated two or more concerns in each category as “major” or “extreme.”

FIGURE 1

Major and extreme concerns about AI

■ Major concern ■ Extreme concern

Concerns around AI implementations



Note: This chart reflects the percentages of respondents who rated two or more concerns in each category as “major” or “extreme.”

Source: Deloitte analysis.

CONFIDENCE

As a category, *confidence* in AI systems was the most significant, with 73% of all business leaders surveyed identifying two or more concerns as major or extreme. When looking specifically at the respondents planning to slow their AI adoption, 85% claim two or more major or extreme confidence-related risks.

It's understandable that concerns around confidence could lead to apprehension. They are most tightly related to the quality of the products and services a company provides and thus can contribute substantially to an organization's reputation and public perception; they also can affect the very deliverables a company puts out into the world for consumption. For example, a medical provider could use AI's immense potential to identify cancer types and determine optimal

treatments.⁷ It is possible that AI decision engines could recommend unsafe treatments or misidentify types of cancer. This can result in serious consequences for patients, expose companies to risk, and result in the loss of AI investment dollars. Unsuccessful initiatives can lead to damages more costly than just lost investment—they can reduce the confidence of an organization's employees in further innovation and, worse, reduce the trust of patients or customers.

In the case above, the company reportedly created safeguards to ensure patient safety, reducing the likelihood that it would damage patients' trust. Additionally, the mistakes were used to identify system failures and improve future algorithmic solutions.⁸ If mitigated well from the beginning, this can be turned into a positive element of the path to AI maturity. In fact, there is evidence

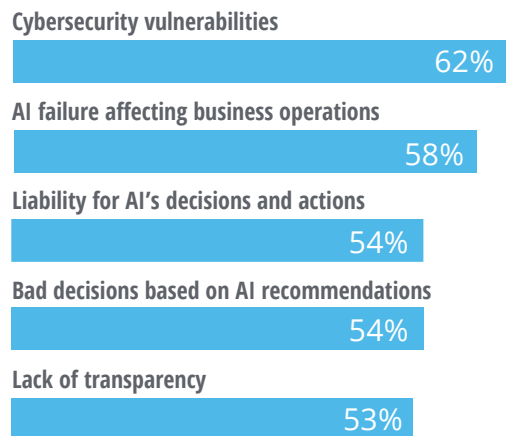
FIGURE 2

AI confidence concerns

■ Extreme or major concern



Overall, how concerned is your organization about the following potential risks associated with your AI initiatives?



Source: Deloitte analysis.

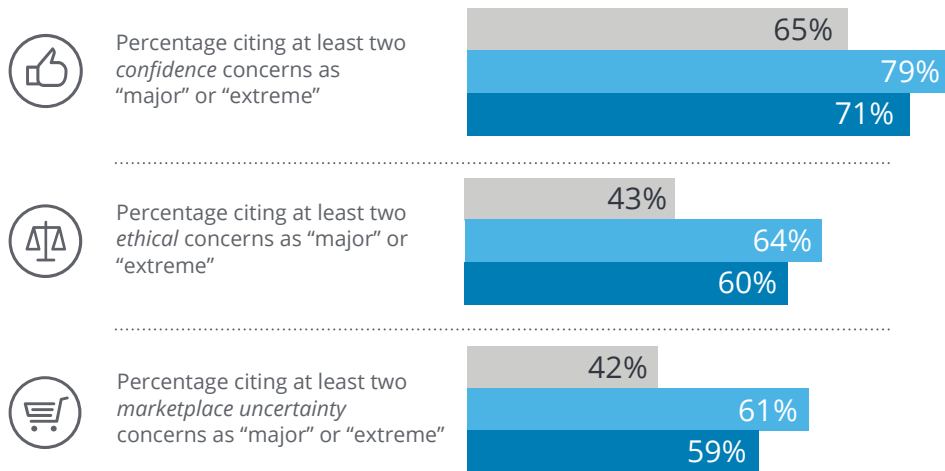
AI MATURITY AND HOW ORGANIZATIONS LEARN TO MANAGE PERCEIVED RISKS

Across the board, as surveyed companies mature in their AI capabilities, the level of concern across *confidence*, *ethics*, and *marketplace uncertainties* follows a relatively bell-shaped trajectory.

FIGURE 3

How maturity changes the level of concern

■ Starters ■ Skilled ■ Seasoned



Source: Deloitte analysis.

Both low- and high-maturity organizations that were surveyed report lower levels of concern, whereas medium-maturity organizations report the highest level of concern across all three categories. This may be because low-maturity organizations don't yet have full vantage of the risks, with many projects still in proof-of-concept or pilot mode. Then, as organizations achieve a medium level of maturity, the challenges might become more apparent, but they may not have developed the capabilities needed to address them yet. Finally, as organizations reach an advanced level, they could have more of the capabilities needed to mitigate those risks, and so their level of concern often decreases once again.

suggesting that as companies mature in their AI capabilities, their level of concern tends to shift as they develop processes, behaviors, or skills to lower their risks and drive more positive outcomes (see sidebar, "AI maturity and how organizations learn to manage perceived risks").

Another significant challenge related to AI confidence is that today's AI systems tend to come with a tradeoff between interpretability and power. Some machine-learning models are so complex

that even highly trained data scientists have difficulty understanding precisely how the algorithms make decisions. As the use of neural networks rises, this issue is becoming even more pronounced, further implicating organizations' ability to justify decisions, mitigate errors, and satisfy regulators.

To help solve this, *explainable AI* is a developing computer science field that seeks to create AI models that are better able to explain themselves.

At present, explainable AI can provide general information about how an AI program makes a decision by disclosing that program’s strengths and weaknesses and the specific criteria the program used to arrive at a decision, and advise on appropriate levels of trust in various types of decisions. Recent recommendations from the U.S. Department of Commerce suggest four principles for assessing AI’s explainability: AI systems need to provide accompanying evidence or reasons for all outputs; systems must provide meaningful and understandable explanations to individual users; system explanations must correctly reflect the system’s process for generating the output; and the system should operate only under the conditions for which it was designed and should not supply decisions to a user without sufficient confidence.⁹ Mastering AI explainability should, in turn, influence an organization’s ability to use AI ethically.

ETHICS

Fifty-seven percent of respondents reported two or more *ethics*-related concerns as major or extreme.

Looking specifically at the respondents planning to slow their AI adoption, 73% cited a major or extreme concern about in at least two ethical areas.

A unique challenge with AI ethics, especially around fairness and bias, is that the data upon which the AI is built can itself be incomplete, biased, or unequal. For this reason, it can be difficult to root out all the unintended biases within a data set, even when a company is well-intentioned.

The humans building AI, analyzing its outputs, and applying its solutions can also fall prey to unintended biases, requiring deliberate actions to mitigate these wherever possible. In a recent health care example, a risk-prediction algorithm using health care spending as a proxy for care ultimately demonstrated racial bias in its results, giving white patients a better chance at benefiting from an extra care program than Black patients.¹⁰ While the algorithm accurately reflected health spending of both groups, it failed to account for economic disadvantages Black people tend to face that often

FIGURE 4

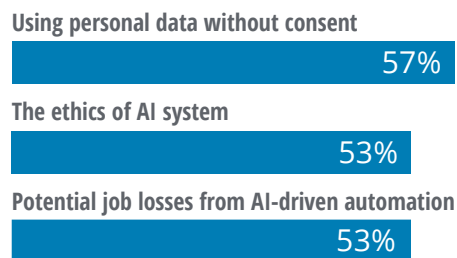
Ethical AI concerns

■ Extreme or major concern



Source: Deloitte analysis.

Overall, how concerned is your organization about the following potential risks associated with your AI initiatives?



affect their health care spending: more expensive emergency room visits, lower insurance coverages, etc.¹¹ This could result in similar spending levels for vastly different levels of need—a reality that the system’s analysis of proxy data missed. In the above example, Black patients were given incorrect risk scores and were excluded from extra care programs at a greater rate.

It is important that program designers thoroughly investigate proxy decisions and variable creation, with safeguards for evaluating whether that use reflects the lived realities of your stakeholders. Further, such examples underscore the need for diverse representation, as well as skills and training among engineers and decision-makers around possible unintended outcomes of algorithm creation.

MARKETPLACE UNCERTAINTIES

Finally, 55% of respondents reported two or more *marketplace uncertainties* as major or extreme concerns. Despite ranking third as a category, this area of concern seems to have an outsized link to

investment behaviors: Of the respondents expecting to slow AI adoption, 71% reported having at least two significant marketplace uncertainties.

The concerns in this category are wide-ranging, from public opinion to the ambiguous regulatory landscape. Fifty-seven percent of survey respondents cite major or extreme concerns about new and changing rules and regulations of AI technologies, such as data privacy, facial recognition, and decision transparency.¹² Fifty-two percent cite major or extreme concerns about customer backlash should they find a flaw or privacy violation in an AI application. Just over half also worry about negative employee perceptions when AI systems are used.

How to navigate the risks inherent in AI adoption

The immense degree of transformation and the vast potential for risk that AI offers can seem daunting—only four in 10 of our survey

FIGURE 5

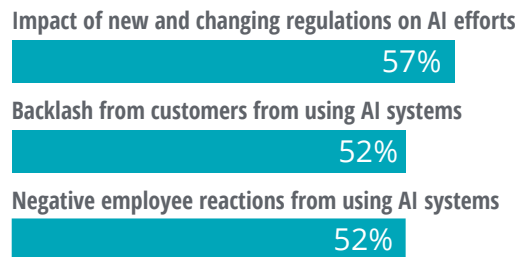
Marketplace uncertainty concerns

■ Extreme or major concern



Source: Deloitte analysis.

Overall, how concerned is your organization about the following potential risks associated with your AI initiatives?



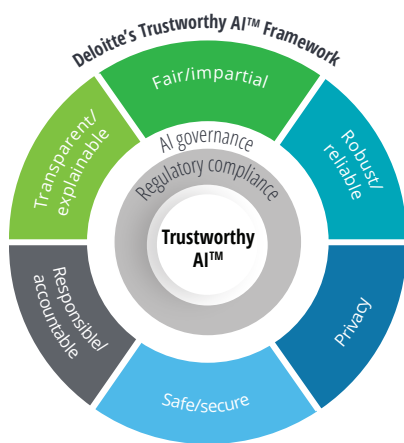
respondents claimed full preparedness.¹³ Given AI technologies’ rapid development, a natural level of uncertainty is sure to linger for some time. However, organizations can look to some current frameworks—and lessons learned from previous technology disruptions—for help in navigating today’s uncertainties.

MANAGING CONFIDENCE AND ETHICS CONCERNS WITH TRUSTWORTHY AI

Of the three categories of concern analyzed previously, two of them—*confidence* and *ethics*—land primarily within the organization’s locus of control. They both ultimately point to the question of whether an organization’s use of AI is trustworthy. To increase confidence in AI, organizations should ensure that their tools and solutions are transparent, reliable, and safe, and that there is a system of accountability in place. Reducing concerns around ethical use will often require evaluating whether tools and solutions are designed for fairness—both in intent and in their effect—and that data use follows clear privacy standards. To help organizations frame AI risks, leaders should develop safeguards that include addressing the six key dimensions highlighted below in figure 6.

FIGURE 6

Trustworthy AI framework



Source: Deloitte's Trustworthy AI services.

Confidence and ethics concerns essentially come down to a question of whether an organization’s use of AI is trustworthy. An organization’s ability to implement this framework largely depends on responsible data management, strong governance standards, and ensuring that a variety of perspectives are in the room to identify and speak out against harmful assumptions.

Recommendation: Focus on data management and governance.

An organization’s ability to skillfully work with data is critical to AI quality and explainability, increasing confidence in AI and leaders’ preparedness to manage its ethical implications. Even today, 40% of surveyed organizations still report “low” or “medium” levels of sophistication across a range of data practices, and nearly a third of executives identified data-related challenges among the top three concerns hampering their company’s AI initiatives.¹⁴ When it comes to governance, only one in five companies surveyed routinely monitor, manage, and improve data quality as part of a formal data governance effort, while only 12% of organizations trust their data to be up to date, and mere 9% believe the data is accurate.¹⁵

Without skillful data management and governance processes in place, an organization could struggle to mitigate risks. By focusing on these two foundational elements, companies can put themselves in a position to successfully implement all six dimensions of trustworthy AI and address the *confidence* and *ethics* concerns that can threaten to slow their adoption.

Recommendation: Insist on diversity of thought.

A company’s ability to implement safeguards also depends on a diversity of perspectives within technology and business stakeholder groups. Data capabilities and governance processes create checkpoints to evaluate compliance on a variety of standards and regulations, but without a diversity of perspectives,

important insights may not surface. Ensuring diversity of thought can help increase the likelihood that potential problems are flagged during the appropriate design and review phases—and not after product launch.

NAVIGATING MARKETPLACE UNCERTAINTIES

Managing *marketplace uncertainties*, those concerns that relate to more external ambiguities, will likely require different approaches. Changing regulations and fluctuating public and employee opinion are always a feature of new and disruptive technology transformations. While companies can't always be in complete control of these, leaders can take steps to navigate them more successfully.

Recommendation: Understand the importance of change management and communication. Public and employee opinion often has little to do with how well AI performs. Thoughtful change management and communication can be critical to successfully navigating this uncertainty. AI will likely transform many organizational roles in the coming years, and helping workers and customers learn new ways of working and engaging will be important to success. An example of an implementation is at Humana, which deployed AI customer assistance agents to answer many of the million-plus calls that often overwhelm customer service agents every month. Humana developed AI agents to handle basic information requests, which accounted for 60% of call volume. This isn't the only way that AI robots equipped with natural language understanding help—they can also assist human agents to gather information behind the scenes, which assisted the Humana call center associates in providing information. By positioning this new technology as

an assistant to human workers, it mitigated the often-perceived threat that some employees may feel toward AI. The company focused on using AI to support (not replace) employees on communication, for example, appropriate expressions of empathy, and on how to collaborate with AI to solve problems. These efforts were focused on trying to maximize the possibilities for success.¹⁶

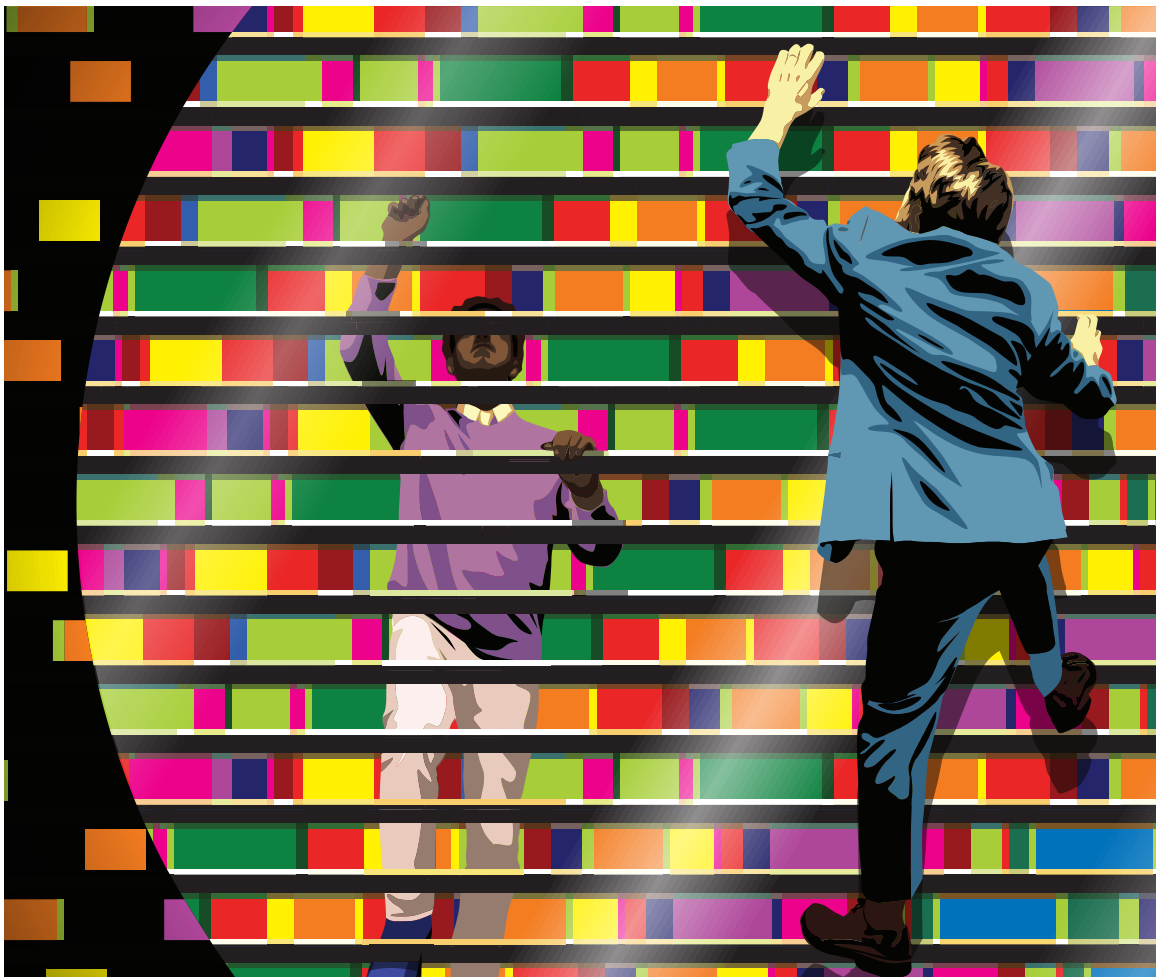
Understanding employee perceptions and planning technology rollouts that mitigate their disruptive nature are often critical for buy-in. Consider working with communication and marketing teams to ensure that both employees and customers adequately understand and see value in the way AI is rolling out. Employing a user-centric approach—instead of forcing new behaviors for cost savings—can help to avoid backlash.

Recommendation: Anticipate the capabilities needed to respond to regulatory shifts. When the European Parliament and European Union created the General Data Protection Regulation (GDPR) in 2016, companies had two years to understand and comply with the regulations before they were implemented in 2018. In the months before the GDPR went into effect, industries that had a history of managing regulations, such as finance, found themselves to be comparatively well prepared.¹⁷ These companies had already built skills and operational capabilities to respond to the required changes related to compliance. As discussions of new AI policy and potential technology regulations increase, now is the time to gain these capabilities so that your organization is ready to both participate in the creation of regulations and respond when they are enacted.

Conclusion

AI holds enormously transformative potential. More than 90% of technology executives surveyed agree that AI will be at the center of the next technological revolution, and that it will take on mundane tasks, allowing people to have greater freedom to pursue more creative work and play.¹⁸

AI's potential value to our world could be too high for business leaders to shrink away due to an overabundance of cautions. With thoughtful development of capabilities and processes, leaders can mitigate the risks and challenges that so many fear. Taking these steps now could set up your organization for significant competitive advantage in the future.



Methodology

This analysis is based on data collected from 2,737 IT and line-of-business executives between October and December 2019. Nine countries were represented, and all participating companies have adopted AI technologies. Respondents were required to meet one of the following criteria: determine AI technology spending and/or approve AI investments; develop AI technology strategies; manage or oversee AI technology implementation; serve as an AI technology subject matter expert; or make or influence decisions around AI technology.

Forty-seven percent were IT executives, with the rest line-of-business executives. Seventy percent were C-level executives: CEOs, presidents, and owners (35%); CIOs and CTOs (32%); and other C-level executives (3%).

To understand more precisely where organizations’ biggest apprehensions originate, we asked survey respondents to rate a list of concerns on a scale from “minimal” to “extreme.” We grouped the listed risks into three categories: *confidence*, *ethics*, and *marketplace uncertainties*.

FIGURE 7

Categories of concern about AI

Categorization	Description of concerns
Confidence	<ul style="list-style-type: none"> ⋮ Lack of transparency—the ability to explain how an AI system arrived at a specific decision ⋮ Liability for decisions and actions made by AI systems ⋮ Making bad decisions based on AI recommendations ⋮ AI failures affecting business operations ⋮ Cybersecurity vulnerabilities in AI-powered systems
Ethics	<ul style="list-style-type: none"> ⋮ Consequences of using personal data in AI systems without consent ⋮ Issues relating to the ethics of AI systems ⋮ Potential job losses from AI-driven automation
Marketplace uncertainties	<ul style="list-style-type: none"> ⋮ New and changing regulations that could affect our AI efforts ⋮ Backlash from customers from using AI systems ⋮ Negative employee reactions from using AI systems

Source: Deloitte analysis.

Endnotes

1. Beena Ammanath, David Jarvis, and Susanne Hupfer, *Thriving in the era of pervasive AI: Deloitte's State of AI in the Enterprise, 3rd Edition*, Deloitte Insights, July 14, 2020.
2. Ibid.
3. Ibid.
4. For a deeper look, see Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (Cambridge: MIT Press, 2018).
5. Catalyst, "Women in science, technology, engineering, and mathematics (STEM): Quick take," August 4, 2020.
6. Bridget Turner Kelly, "Though more women are on college campuses, climbing the professor ladder remains a challenge," Brookings Institution, March 29, 2020; Synced, "AI-powered 'Genderify' platform shut down after bias-based backlash," July 30, 2020.
7. Adam Conner-Simons and Rachel Gordon, "Using AI to predict breast cancer and personalize care," *MIT News*, May 7, 2019.
8. Jo Cavallo, "Confronting the criticisms facing Watson for oncology," *ASCO Post*, September 10, 2019.
9. National Institute of Standards and Technology, "NIST asks A.I. to explain itself," August 18, 2020.
10. Ziad Obermeyer et al., "Dissecting racial bias in an algorithm used to manage the health of populations," *Science* 366, no. 6464 (2019): pp.447–53.
11. Ibid.; Jamila Taylor, "Racism, inequality, and health care for African Americans," Century Foundation, December 19, 2019.
12. Mark MacCarthy, "AI needs more regulation, not less," Brookings Institution, March 9, 2020.
13. Ammanath et al., *Thriving in the era of pervasive AI*.
14. Karthik Ramachandran, *Data management barriers to AI success*, Deloitte Insights, August 7, 2020.
15. MIT SMR Connections and SAS, "How trust delivers value in data, analytics, & AI," January 15, 2019.
16. Kim S. Nash, "Artificial intelligence helps Humana avoid call center meltdowns," *Wall Street Journal*, October 27, 2016.
17. Ponemon Institute, "The race to GDPR: A study of companies in the United States & Europe," April 2018.
18. Edelman, "2019 Edelman AI survey," March 2019.

Acknowledgments

The authors would like to thank **David Jarvis**, **Susanne Hupfer**, **Brenna Sniderman**, and **Natasha Buckley** for their help in developing the ideas of this paper.

About the authors

Beena Ammanath | bammanath@deloitte.com

Beena Ammanath is a managing director with Deloitte Consulting LLP and has extensive global experience in artificial intelligence and digital transformation. She is founder and CEO of Humans for AI Inc. and coauthor of *AI Transforming Business*.

David R. Novak | danovak@deloitte.com

David R. Novak, PhD, is a research manager with the Deloitte Center for Integrated Research. He specializes in developing thought leadership in the areas of humans and technology, artificial intelligence, and sustainability. Prior to joining Deloitte, Novak worked as a professor of communication, as a strategic communication consultant, and in the field of data analytics.

Siri Anderson | sianderson@deloitte.com

Siri Anderson is a researcher and writer at the Deloitte Center for Integrated Research, specializing in developing thought leadership in the areas of future of work and digital reality. Previously, Anderson led the sponsored and contributed content program for Deloitte Digital. Before joining Deloitte, Anderson worked in editorial and strategic communications roles in technology and health care industries.

Abha Kulkarni | abhakikulkarni@deloitte.com

Abha Kulkarni is a senior analyst with the Deloitte Center for Integrated Research, Deloitte Services India Pvt. Ltd. She supports the Future of Work campaign by analyzing workforce data across industries and regions. Her primary focus is on applying quantitative research techniques to enable data-driven research insights.

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

About the Deloitte AI Institute

The Deloitte AI Institute helps organizations transform with AI through cutting-edge research and innovation by bringing together the brightest minds in AI to help advance human-machine collaboration in the Age of With. The Institute was established to advance the conversation and development of AI in order to challenge the status quo. The Deloitte AI Institute collaborates with an ecosystem of industry thought leaders, academic luminaries, startups, research and development groups, entrepreneurs, investors, and innovators. This network, combined with Deloitte's depth of applied AI experience, can help organizations transform with AI. The institute will cover a broad spectrum of AI focus areas, with current research on ethics, innovation, global advancements, the future of work, and AI case studies. To learn more, visit [Deloitte.com](https://www.deloitte.com).

About the Deloitte Center for Integrated Research

The Deloitte Center for Integrated Research focuses on developing fresh perspectives on critical business issues that cut across industries and functions, from the rapid change of emerging technologies to the consistent factor of human behavior. We look at transformative topics in new ways, delivering new thinking in a variety of formats, such as research articles, short videos, in-person workshops, and online courses. To learn more, visit [Deloitte.com](https://www.deloitte.com).

Deloitte.

Insights

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



Follow @DeloitteInsight

Deloitte Insights contributors

Editorial: Matthew Budman, Hannah Bachman, Rupesh Bhat, and Nairita Gangopadhyay

Creative: Victoria Lee and Govindh Raj

Promotion: Maria Martin Cirujano

Cover artwork: Daniel Hertzberg

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

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.