June 2023

Building Robust RAI Programs as Third-Party AI Tools Proliferate

by Elizabeth M. Renieris, David Kiron, and Steven Mills

AUTHORS

Elizabeth M. Renieris is guest editor for the *MIT Sloan Management Review* Responsible AI Big Idea program, a senior research associate at Oxford's Institute for Ethics in AI, a senior fellow at the Centre for International Governance Innovation, and author of *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse* (MIT Press, 2023).

David Kiron is an editorial director at *MIT Sloan Management Review* and coauthor of the book *Workforce Ecosystems: Reaching Strategic Goals With People, Partners, and Technology* (MIT Press, 2023).

Steven Mills is a managing director and partner at BCG, where he serves as the chief AI ethics officer.

CONTENTS

The risks and failures of AI systems are more palpable and numerous than ever, but organizations are at risk of falling behind.

Introduction

In just a few short months since its release, OpenAI's ChatGPT tool has catapulted the capabilities, as well as the ethical challenges and failures, of artificial intelligence into the spotlight. Countless examples have emerged of the chatbot fabricating stories, including falsely accusing a law professor of sexual harassment and implicating an Australian mayor in a fake bribery scandal, leading to the first lawsuit against an AI chatbot for defamation.[1] In April, Samsung made headlines when three of its employees accidentally leaked confidential company information, including internal meeting notes and source code, by inputting it into ChatGPT.[2] That news prompted many companies, such as JPMorgan and Verizon, to block access to AI chatbots from corporate systems.[3] In fact, nearly half of the companies polled in a recent Bloomberg survey reported that they are actively working on policies for employee chatbot use, suggesting that a significant share of businesses were caught off guard and were unprepared for these developments.[4]

Indeed, the fast pace of AI advancements is making it harder to use AI responsibly and is putting pressure on responsible AI (RAI) programs to keep up. For example, companies' growing dependence on a burgeoning supply of third-party AI tools, along with the rapid adoption of generative AI — algorithms (such as ChatGPT, Dall-E 2, and Midjourney) that use training data to generate realistic or seemingly factual text, images, or audio — is exposing them to new commercial, legal, and reputational risks that are difficult to track.[5] In some cases, managers may lack any awareness about the use of such tools by employees or others in the organization — a phenomenon known as *shadow AI*.[6] As Stanford Law CodeX fellow Riyanka Roy Choudhury puts it, "RAI frameworks were not written to deal with the sudden, unimaginable number of risks that generative AI tools are introducing."

Organizational RAI programs are struggling to keep pace with technical advancements in AI.
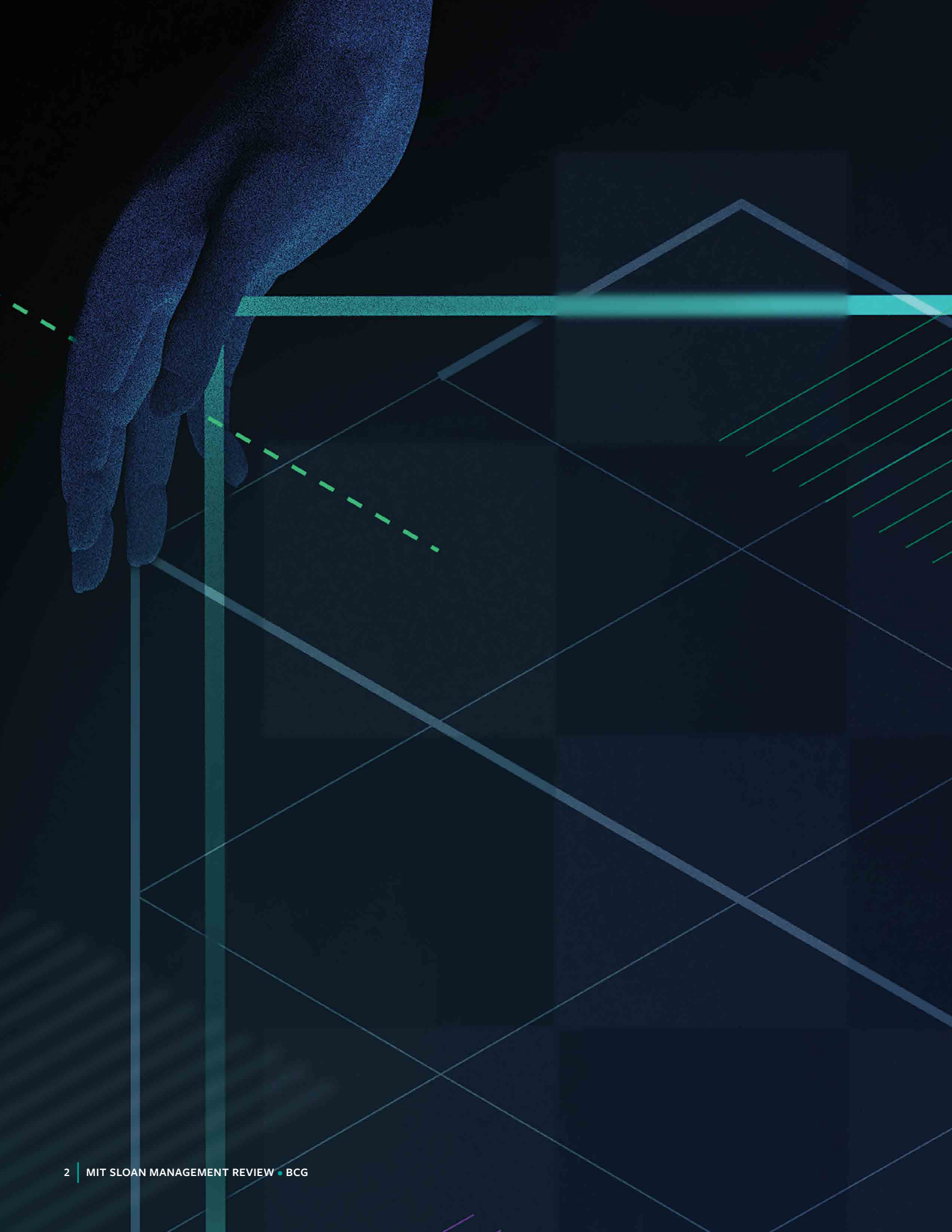
This trend is especially problematic for organizations with RAI programs that are primarily focused on AI tools and systems that they design and develop internally. The fact is, the vast majority of organizations we surveyed use third-party AI tools, and a majority rely on them exclusively, having no internally developed AI of their own. Proactively evaluating external solutions becomes necessary to anticipate and preempt, not just retrospectively address, AI failures that stem from third-party technologies, which according to our research account for more than half of all such failures. As Linda Leopold, H&M Group's head of responsible AI and data, observes, "Responsible AI programs should cover both internally built and third-party AI tools. The same ethical principles must apply, no matter where the AI system comes from. Ultimately, if something were to go wrong, it wouldn't matter to the person being negatively affected if the tool was built or bought."

The fundamental issue is that organizational RAI programs are struggling to keep pace with technical advancements in AI. These advancements are growing the ecosystem of available third-party AI solutions and making it easier to use AI throughout the organization, but they are also expanding the scope and complexity of risks that RAI programs must address. As companies evolve their approach, partly in response to an increasingly fierce race to deploy AI, some are reorganizing their responsible AI teams. Others appear to be scaling back internal resources devoted to responsible AI as part of a broader trend in industry layoffs. These reductions in RAI investments are happening, arguably, when they are most needed. RAI is even a White House-level concern, with the Biden administration recently unveiling a set of initiatives designed to "promote responsible American innovation in artificial intelligence and protect people's rights and safety."[7]

This *MIT Sloan Management Review* and Boston Consulting Group (BCG) report is based on our second consecutive year of conducting a global survey, interviewing C-suite executives, and gathering insights from an international panel of AI experts, including academics and practitioners, to help us understand how RAI is being implemented in organizations worldwide. Last year, we published a report titled "To Be a Responsible AI Leader,

A GROWING GAP BETWEEN RAI LEADERS AND NON-LEADERS

As organizations worldwide continue to mature their RAI efforts, the gap between Leaders and Non-Leaders appears to be growing. (For more details about the construction of this maturity index, see "About the Research," page 4.) This year, 13% more of the organizations we surveyed can be characterized as RAI Leaders compared with last year.[i] This may be attributable to more organizations widening the scale of their RAI programs — that is, the extent to which RAI policies, processes, and approaches are implemented and adopted across the organization as opposed to on an ad hoc or semi-ad hoc basis. For example, 7% of Non-Leader organizations we surveyed this year had enterprisewide approaches to RAI, compared with none in 2022. Moreover, the average maturity scores of both Leaders and Non-Leaders improved in 2023 (up 8% and 2%, respectively) over 2022. In other words, while organizations across the board grew more mature compared with last year, those in the Leader group matured 6% faster, further widening the gap between Leaders and Non-Leaders.

Focus on Being Responsible," in which we concluded that successful RAI efforts actually may have more to do with being a responsible organization than they do with AI as a technology.[8] This year, we focused more narrowly on the extent to which organizations are addressing risks stemming from the use of internally and externally developed AI tools, such as generative AI.

Our research suggests that RAI programs must do a better job of tracking and mitigating the risks of third-party AI use and continue to iterate and adapt to technical advancements in AI. It also suggests that now is the time to double down on, and invest in, a robust RAI program. In this regard, CEOs can play a significant role in building an organizational culture that supports RAI efforts among managers, employees, vendors, and other partners, including demonstrating a willingness to look for and identify potential AI failures. CEOs can be critical to enacting meaningful and lasting investments that support a robust and effective RAI program in the face of a rapidly evolving threat landscape.

Third-Party AI Risks on the Rise

Our research reveals that organizations worldwide are highly reliant on third-party AI — AI tools or algorithms designed and developed by another entity that an organization buys, licenses, or otherwise accesses for its own internal purposes or as part of an offering to its customers. The vast majority (78%) of organizations surveyed this year report accessing, buying, licensing, or otherwise using third-party AI tools, including commercial APIs, pretrained models, and data. In fact, more than half (53%) of organizations surveyed rely exclusively on third-party AI tools and have no internally designed or developed AI technologies of their own. As Nitzan Mekel-Bobrov, eBay's chief AI officer, observes, "Third-party AI tools, including open-source models, vendor platforms, and commercial APIs, have become an essential part of virtually every organization's AI strategy in one form or another."

As with AI more generally, third-party AI tools can expose organizations to all manner of risks, including reputational damage and the loss of customer trust, financial losses, regulatory penalties and compliance challenges, and litigation. In other words, outsourcing AI from third parties doesn't inoculate organizations from these hazards. On the contrary, more than half (55%) of all AI-related failures stem from third-party AI tools, leaving organizations that use them vulnerable to unmitigated risks. UNICEF's digital policy specialist, Steven Vosloo, says the extent to which an organization's RAI program addresses third-party AI risks in practice depends on the rigor of its program, and he notes that "determining how to fully assess the risks (real or potential) in third-party AI tools can be challenging."

In fact, despite widespread reliance on third-party AI tools, organizational RAI programs may be failing to account for the substantial risks they pose. A fifth (20%) of organizations that use third-party AI tools fail to evaluate the risks at all. (SEE FIGURE 1, PAGE 5.)

Philip Dawson, head of AI policy at Armilla AI, cautions that "enterprises have not fully adapted their third-party risk management programs to the AI context or challenges of safely deploying complex systems like generative AI products. Many do not subject AI vendors or their products to the kinds of assessments undertaken for cybersecurity, leaving

FIGURE 1

**Despite Widespread
Use, Risks From Third-
Party AI Tools Are Not
Often Assessed**

While more than three-
quarters of the organizations
we surveyed use third-party
AI tools, more than half of
AI-related failures stem from
the use of these tools.

78%
use third-party
AI tools

53%
use third-party
AI tools *exclusively*

20%
fail to evaluate the risks
stemming from
third-party AI tools

55%
of AI-related failures stem from
third-party AI tools

We define third-party AI tools as AI tools that an organization accesses, buys, or licenses.

them blind to the risks of deploying third-party AI solutions." Acknowledging the shadow AI problem, Simon Chesterman, senior director of AI governance at research institute AI Singapore, says that one of the biggest challenges is "that we don't know what we don't know." Third-party AI makes this all the more challenging when teams across an organization are able to engage vendors without oversight.

While there is no silver bullet for mitigating third-party AI risks, or any type of AI risk for that matter, employing a wide variety of different approaches and methods to evaluate third-party tools appears to be more effective than using fewer approaches. For example, organizations that employ seven different methods are more than twice as likely to uncover AI failures compared with those that use only three (51% versus 24%). These approaches may include the specific evaluation of a vendor's RAI practices, contractual

language mandating adherence to RAI principles, vendor pre-certification and audits (where available), internal product-level reviews (where a third-party tool is integrated into a product or service), and adherence to relevant regulatory requirements or industry standards. Oarabile Mudongo, a policy specialist at the African Observatory on Responsible AI, observes that "to effectively address the risks associated with third-party AI tools, RAI programs should include a comprehensive set of policies and procedures, such as guidelines for ethical AI development, risk assessment frameworks, and monitoring and auditing protocols."

Regulations Raise the Stakes

Addressing the risks of AI use, including the mounting risks stemming from third-party AI tools, is paramount,

given that the regulatory landscape is evolving almost as rapidly as the technology itself. Many new AI-specific regulations are taking effect, and new ones are being drafted in their wake. It's not just happening on a national level; states and localities are beginning to act too. For example, laws in New York, Illinois, and Maryland, as well as draft legislation introduced in California and a handful of other states, address the use of AI tools in the context of hiring and employment.[9] In Europe, the much-anticipated AI Act and corresponding AI Liability Directive will impose stringent requirements on AI systems deemed to be "high risk," as well as general-purpose systems like AI chatbots, and make vendors liable for any damage to consumers.[10] Data protection authorities worldwide are also assessing the compatibility of AI tools with existing laws.[11]

In fact, a number of existing laws already apply to the use of AI even though these regulations might not specifically address AI use. Examples include consumer protection laws, nondiscrimination laws, and data protection and privacy laws, among others. About half (51%) of the organizations we surveyed report being subject to non-AI-specific regulations that nevertheless apply to their use of AI, including a high proportion of organizations in the financial services, insurance, health care, and public sectors. Organizations subject to such regulations account for 13% more RAI Leaders than organizations not subject to them. They also report detecting fewer AI failures than their counterparts that are not subject to the same regulatory pressures (32% versus 38%). This finding, which applies across RAI maturity levels, may be due to better overall risk management practices. For example, "With clients in regulated industries such as financial services, we see strong links between model risk management practices predicated on some sort of external regulation and what we suggest people do from an RAI standpoint," observes Triveni Gandhi, responsible AI lead for AI company Dataiku.

Our research suggests that RAI maturity may actually contribute more to perceived preparedness for new AI-specific regulations than an existing culture of compliance with other regulations. In this vein, it is noteworthy that a majority of RAI Leaders (72%) feel *prepared* for emerging AI-specific regulations, while a majority (60%) of organizations subject to non-AI-specific regulations that apply to their use of AI feel *unprepared* for them. One reason may be that when AI failures do occur, they are more likely to take a bigger toll on organizations that are operating under more scrutiny and in higher-risk environments or industries. For example, highly regulated organizations face significantly more impact from AI failures than organizations in unregulated industries in terms of reputational damage (27% versus 20%), financial loss (24% versus 16%), regulatory compliance challenges (20% versus 13%), and litigation risk (14% versus 4%). (SEE FIGURE 2, PAGE 7.)

These findings highlight the urgency for all organizations, regardless of their existing regulatory environment, to prepare for new and emerging AI-specific regulations. Our research suggests that the best way to do so is by quickly scaling and maturing RAI programs to anticipate and address AI risks, including the risks of third-party AI tools.

Now Is the Time to Double Down on RAI

Our research clearly suggests that with AI-related risks proliferating and global regulatory scrutiny intensifying, organizations cannot afford to scale back or even maintain the status quo when it comes to their RAI efforts. As the U.S. Federal Trade Commission recently cautioned, "Given [the] concerns about the use of new AI tools, it's perhaps not the best time for firms building or deploying them to remove or fire personnel devoted to ethics and

Having CEO engagement in an RAI program can help an organization identify and address existing risks and contribute to a greater sense of preparedness for the future.

Use of Third-Party AI Tools

77%

of organizations in industries subject to non-AI-specific regulations use third-party AI tools

82%

of organizations in industries not subject to non-AI-specific regulations use third-party AI tools

Failures From Third-Party AI Tools

24%

of organizations in industries subject to non-AI-specific regulations have experienced failures

36%

of organizations in industries not subject to non-AI-specific regulations have experienced failures

Types of Failure From All AI Tools

■ Industries Subject to Non-AI-Specific Regulations
■ Industries Not Subject to Non-AI-Specific Regulations

Reputational Damage

27%
20%

Financial Loss

24%
16%

Regulatory Compliance Challenges

20%
13%

Litigation Risk

14%
4%

We define third-party AI tools as AI tools that an organizations accesses, buys, or licenses.

FIGURE 2

Organizations in Industries Subject to Regulations Experience Fewer Failures From Third-Party AI Tools

Organizations report failures from all types of AI tools. While a high proportion (77%) of organizations in industries subject to non-AI-specific regulations use third-party AI tools, they experience 12% fewer failures from those tools than their counterparts in industries not subject to these regulations do.

---

responsibility for AI and engineering."[12] Instead, now is the time for organizations to double down on, and affirm their commitments to, RAI. CEOs can play a key role in ensuring an organization's commitment to RAI, as well as sustaining the necessary investments.

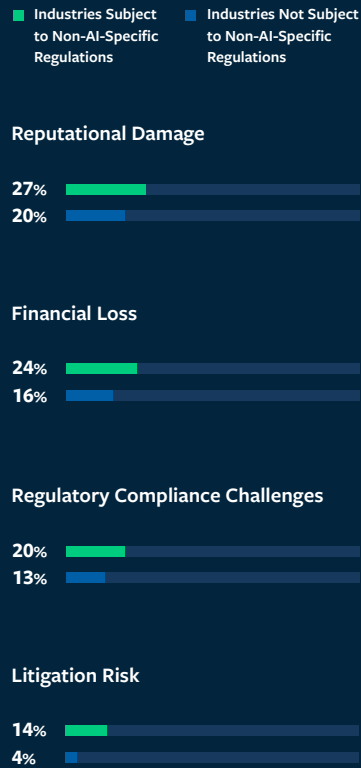Having CEO engagement in an RAI program can help an organization identify and address existing risks and contribute to a greater sense of preparedness for the future. Organizations whose CEO takes a hands-on role in RAI efforts (such as by engaging in RAI-related hiring decisions or product-level discussions or setting performance targets tied to RAI) report **58%** more business benefits than organizations with a less hands-on CEO, regardless of Leader status. **(SEE FIGURE 3, PAGE 8.)** What's more, organizations with a CEO who is directly involved in RAI are more likely to invest in RAI than organizations with a hands-off CEO (**39%** versus **22%**). In sum, meaningful CEO engagement in RAI matters for all organizations; the deeper the engagement, the greater its benefits appear to be.

	CEO is directly involved	CEO is not directly involved
Better products or services	71%	52%
Accelerated innovation	66%	45%
Long-term profitability	57%	31%
Brand differentiation	48%	28%
Improved recruiting and retention	39%	19%
Likely to invest in RAI	39%	22%

Direct CEO Involvement Leads to More Benefits

Organizations that directly involve the CEO in their RAI initiatives realize 58% more business benefits than those that don't involve their CEOs.

At a time of urgency for organizations to reaffirm their commitments to RAI, few companies have CEOs that are directly engaged with efforts to make their RAI programs more robust and effective. Only about a quarter (28%) of organizations we surveyed report having a CEO who takes a hands-on role in RAI efforts, which can be critical to helping a company negotiate trade-offs (perhaps only perceived trade-offs) between getting RAI right and its related financial costs. This state of affairs is perhaps why the current U.S. presidential administration recently summoned the CEOs of leading AI companies to the White House to discuss responsible AI, reminding them that "the private sector has an ethical, moral, and legal responsibility to ensure the safety and security of their products."[13]

Recommendations

1. MOVE QUICKLY TO MATURE RAI PROGRAMS.
As the gap widens between Leaders and Non-Leaders, organizations must move quickly to mature their RAI programs to keep up with the changing landscape. This includes broadening the scale and scope of their RAI program and ensuring that it applies organizationwide rather than on an ad hoc or partially ad hoc basis. Not maturing in this context means falling behind, which may result in reduced competitiveness as rivals go on to realize more business benefits.

2. PROPERLY EVALUATE THIRD-PARTY TOOLS.
The use of third-party AI tools is widespread across organizations and will only continue to grow with the adoption of generative AI and other advancements, thus introducing significant legal, commercial, and other risks. While there is no silver bullet to mitigating these risks, organizations should adequately and continually evaluate the use of third-party AI using a variety of tools and methods as part of an effective RAI program; in fact, the more methods an organization uses, the more effective it is likely to be.

3. TAKE ACTION TO PREPARE FOR EMERGING REGULATIONS.
Organizations in regulated industries appear to have better practices around risk management, which may, in part, contribute to better RAI outcomes and greater business benefits.[14] And yet, even these organizations feel unprepared for the AI-specific regulations that are on the horizon. As these new AI-specific regulations come online, all organizations can stand to benefit from the kind of structured approach to risk management that an effective RAI program offers, including addressing the use or integration of third-party AI tools.

4. ENGAGE CEOS IN RAI EFFORTS TO MAXIMIZE SUCCESS.
Our research shows that CEO engagement can boost the benefits of RAI programs and, in turn, help to mitigate the risks of AI use, including the use of third-party AI tools. But the type of engagement matters. Organizations where CEOs play an active role in RAI programs through hiring, target setting, or product-level discussions experience significantly more benefits. As risks mount and regulations proliferate, it will become more important for CEOs to directly engage in RAI efforts to boost the benefits and minimize the risks.

5. DOUBLE DOWN AND INVEST IN RAI.
Now is not the time to cut back on resources or teams devoted to ethical or responsible AI, or even to sustain RAI efforts at the same level as last year. Over the past year, AI's adoption has soared and so, too, have the risks associated with the technology. As a result, now is the time to invest in RAI and ramp up efforts to scale RAI programs — because as these risks mount, so, too, does the urgency of RAI.

Conclusion

The AI landscape has changed dramatically over the past year, and what were mere research and development projects have suddenly become commercial deployments. The technology, and generative AI in particular, has gained rapid adoption through both consumer-facing and enterprise-grade tools through a wide variety of use cases and applications. Alongside this rapid adoption, the risks and failures of AI systems are more palpable and more numerous. Many companies were caught off guard by the spread of shadow AI use across the enterprise. At the same time, regulators are beginning to apply existing laws to these commercial deployments and new AI-specific regulations are coming online, intensifying regulatory scrutiny.

In this climate, not investing in RAI is tantamount to falling behind and exposing your organization to material risk. And while it may feel as though the technology is outpacing your RAI program's capabilities, the solution is to increase your commitment to RAI, not walk away from it. If anything, these recent developments demonstrate the urgency of RAI, which is now even a White House-level priority. They also reaffirm the key insight from our report last year — namely, that being an RAI Leader has more to do with being a responsible organization than with AI technology itself.

REFERENCES

1 P. Dixit, "U.S. Law Professor Claims ChatGPT Falsely Accused Him of Sexual Assault, Says 'Cited Article Was Never Written,'" Business Today, April 8, 2023, www.businesstoday.in; and T. Gerken, "ChatGPT: Mayor Starts Legal Bid Over False Bribery Claim," BBC, April 6, 2023, www.bbc.com.

2 M. DeGeurin, "Oops: Samsung Employees Leaked Confidential Data to ChatGPT," Gizmodo, April 6, 2023, https://gizmodo.com.

3 A. Lukpat, "JPMorgan Restricts Employees From Using ChatGPT," The Wall Street Journal, Feb. 22, 2023, www.wsj.com.

4 J. Constantz, "Nearly Half of Firms Are Drafting Policies on ChatGPT Use," Bloomberg, March 20, 2023, www.bloomberg.com.

5 "Generative AI," BCG, accessed May 24, 2023, www.bcg.com.

6 J.K. Bickford and T. Roselund, "How to Put Generative AI to Work — Responsibly," BCG, Feb. 28, 2023, www.bcg.com.

7 "Fact Sheet: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation That Protects Americans' Rights and Safety," The White House, May 4, 2023, www.whitehouse.gov.

8 E.M. Renieris, D. Kiron, and S. Mills, "To Be a Responsible AI Leader, Focus on Being Responsible," MIT Sloan Management Review and BCG, Sept. 19, 2022, https://sloanreview.mit.edu.

9 L. Mearian, "Legislation to Rein In AI's Use in Hiring Grows," Computerworld, April 1, 2023, www.computerworld.com.

10 "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts," European Commission, April 21, 2021, https://eur-lex.europa.eu; and "Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)," PDF file (Brussels: European Commission, Sept. 28, 2022), https://commission.europa.eu.

11 "DPAs, Global Stakeholders Mull AI Regulation," IAPP, April 21, 2023, https://iapp.org.

12 M. Atleson, "The Luring Test: AI and the Engineering of Consumer Trust," Federal Trade Commission, May 1, 2023, www.ftc.gov.

13 "Statement From Vice President Harris After Meeting With CEOs on Advancing Responsible Artificial Intelligence Innovation," The White House, May 4, 2023, www.whitehouse.gov.

14 S. Mills, S. Singer, A. Gupta, et al., "Responsible AI Is About More Than Avoiding Risk," BCG, Sept. 20, 2022, www.bcg.com.

i The samples used to construct the maturity index differed from 2022 to 2023 due to different individuals surveyed each year.

MIT SLOAN MANAGEMENT REVIEW

At *MIT Sloan Management Review (MIT SMR)*, we explore how leadership and management are transforming in a disruptive world. We help thoughtful leaders capture the exciting opportunities — and face down the challenges — created as technological, societal, and environmental forces reshape how organizations operate, compete, and create value.

MIT SLOAN MANAGEMENT REVIEW BIG IDEAS

MIT SMR's Big Ideas Initiatives develop innovative, original research on the issues transforming our fast-changing business environment. We conduct global surveys and in-depth interviews with front-line leaders working at a range of companies, from Silicon Valley startups to multinational organizations, to deepen our understanding of changing paradigms and their influence on how people work and lead.

BOSTON CONSULTING GROUP

Boston Consulting Group (BCG) partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we work closely with clients to embrace a transformational approach aimed at benefiting all stakeholders — empowering organizations to grow, build sustainable competitive advantage, and drive positive societal impact. Our diverse, global teams bring deep industry and functional expertise and a range of perspectives that question the status quo and spark change. BCG delivers solutions through leading-edge management consulting, technology and design, and corporate and digital ventures. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, fueled by the goal of helping our clients thrive and enabling them to make the world a better place.

BCG X

BCG X is the tech build and design unit of BCG.

Turbocharging BCG's deep industry and functional expertise, BCG X brings together advanced tech knowledge and ambitious entrepreneurship to help organizations enable innovation at scale.

With nearly 3,000 technologists, scientists, programmers, engineers, and human-centered designers located across 80-plus cities, BCG X builds and designs platforms and software to address the world's most important challenges and opportunities.

Teaming across our practices, and in close collaboration with our clients, our end-to-end global team unlocks new possibilities. Together we're creating the bold and disruptive products, services, and businesses of tomorrow.
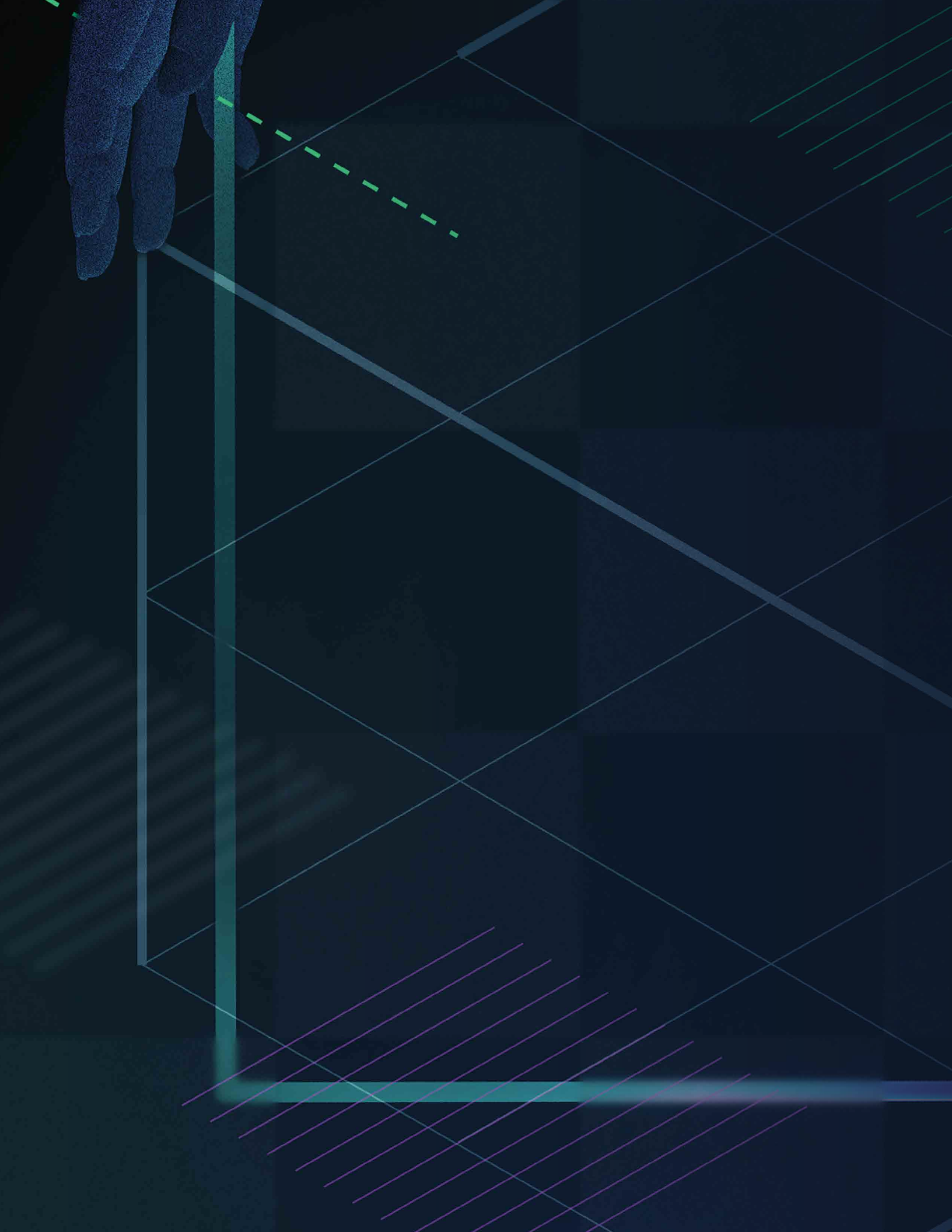
---

sloanreview.mit.edu/big-ideas

MITSloan
Management Review

> SLOANREVIEW.MIT.EDU/BIG-IDEAS

MITSloan
Management Review