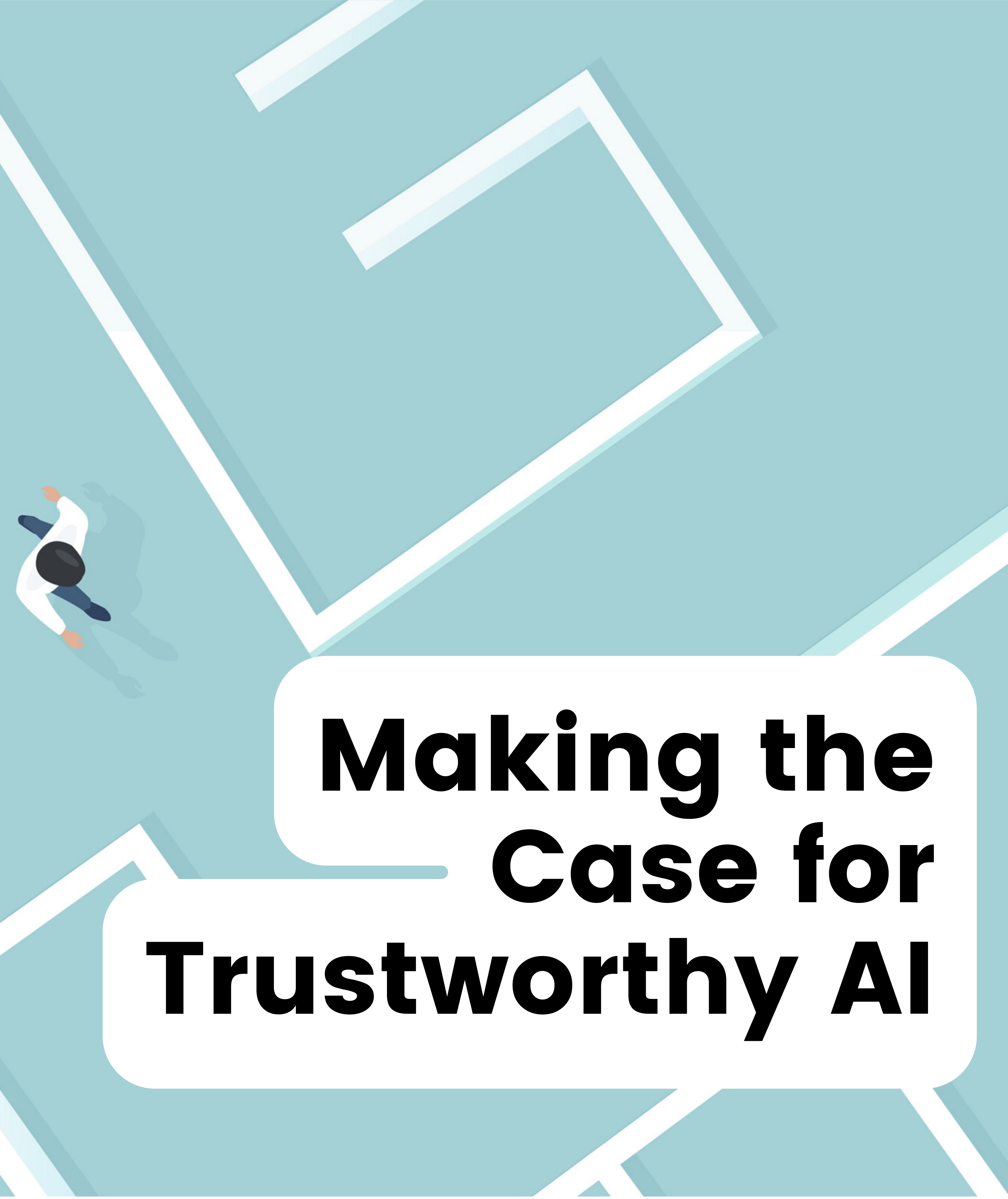


Making the Case for Trustworthy AI

ASPEN DIGITAL
aspen institute

Deloitte.

TABLE OF CONTENTS

03 Making the Case

04 Motivations

07 Challenges

10 Looking Ahead

14 Conclusion

MAKING THE CASE

Artificial Intelligence (AI) is being adopted and implemented across a broad range of industries and applications, from telecom and transport to groceries and governments. These AI systems can be prominent elements of a product or service, or they may augment how organizations operate behind the scenes. Regardless of how apparent the integration of AI may be, these applications of advanced automation systems raise important questions about liability, responsibility, and ethical impacts on the public. This inquiry is referred to as trustworthy AI.

Between December 2021 and March 2022, Aspen Digital & the Deloitte AI Institute convened three roundtable discussions to address the business case for trustworthy AI. This series of dialogues brought together global business leaders, including CIOs across Fortune 100 companies, in conversation with civil society and government leaders to share insights about how their organizations are approaching the challenges and opportunities that trustworthy AI presents. While some organizations may be early in their implementation journey and others already have established practices, many recognize that there is still work to be done in operationalizing trustworthy AI practices across their institution.

This brief addresses the primary motivations for organizations to help ensure their AI efforts meet standards of trustworthiness, the challenges teams face, and the future outlook expressed by the global business, civil society, and government leaders involved.

MOTIVATIONS

The participants in the AI roundtables expressed a variety of motivations for ensuring their organizations' AI efforts are trustworthy, including values-alignment, business opportunity, and regulator risk. But many noted reputational risk as the key motivator.

AI conjures a uniquely evocative image in the public consciousness. Longstanding concerns about losing control of these complex systems have become increasingly prominent. While the technologies that constitute AI in a business context remain a far cry from the fictional artificial intelligences of HAL or the Terminator, the speed and scale afforded by intelligent automation is nonetheless consequential. Several notable examples of AI-related harms have captured public attention in recent years highlighting issues of bias, flawed performance, and worker displacement.

Public distrust about development and management of AI is an obvious concern for consumer-facing companies, where demand can be negatively impacted. But all companies—whether direct-to-consumer or business-to-business—can face reputational risks associated with AI harms. As AI becomes more ubiquitous across industries, competition for talent becomes more fierce. Negative press coverage can factor into whether data scientists, machine learning engineers, AI product managers, and other key employees choose one employer over another. According to the 2021 Edelman Trust Barometer, 61% of employees “choose, leave, avoid, or consider employers based on their values and beliefs.” Some organizations recognize that trustworthy AI investments are a differentiator in their recruitment strategy, allowing them to persuade members of this highly sought-after talent pool to join (or stay on) their staff.

> ***According to the 2021 Edelman Trust Barometer, 61% of employees “choose, leave, avoid, or consider employers based on their values and beliefs.”***

Negative press can also be a motivator for policymakers and their staff, who are compelled to prioritize public wellbeing and address the concerns of constituents who may not understand these technologies. Grace Simrall, Chief of Civic Innovation and Technology at Louisville Metro Government, spoke to the urgency of the issue. “We were on a top ten list that we did not want to be on,” she said, referencing findings from a 2019 report that predicted Louisville would be one of the ten metropolitan areas in the United States with the highest share of jobs at risk of automation. This instigated a multi-year initiative to educate and upskill residents and workers on AI. Public interest in topics such as “responsible AI,” “trustworthy AI,” and “AI ethics” has been growing and is driving action at the state and local levels, although federal policy hasn’t moved as quickly.

While some organizations may wait for regulation to instigate their trustworthy AI efforts, many hope to get out ahead of it. They may see trustworthy AI efforts as merely an extension of their existing business practices, especially in traditionally regulated industries like financial services. These industries may serve as leaders or models for other types of organizations. One participant referenced Model Risk Management as a potential blueprint for developing trustworthy AI processes in other fields. Trustworthy AI could not only be a differentiating aspect of products and services, but lead to an entirely new industry in its own right. There is a burgeoning field of AI auditing startups that promise to make it easier for businesses to identify potential AI risks and more firms are offering trustworthy AI services to their clients.

In addition to the business opportunities and the reputational and regulatory pressures, many organizational leaders also report feeling personally motivated to take action on AI to align with their own values and those of their organization. Alissa Cooper, Vice President and Chief Technology Officer for Technology Policy and a Fellow at Cisco Systems, noted that the emotional salience in the narratives of some AI harms may motivate decision-makers to take a stronger sense of responsibility, especially as compared to other more abstract technology issues. “People can connect with it more easily than they would with, for example, a data breach,” she said. Another participant also recognized that trustworthy AI efforts often go hand-in-hand with the mission of the organization. They emphasized strategies that “ally the moral arguments [for trustworthy AI] with larger business goals.”

WHAT'S MOTIVATING ORGANIZATIONS

- **Alignment with Core Organizational Values**
- **Reputation Management**
- **Anticipating Regulation**
- **Staff Satisfaction & Retention**

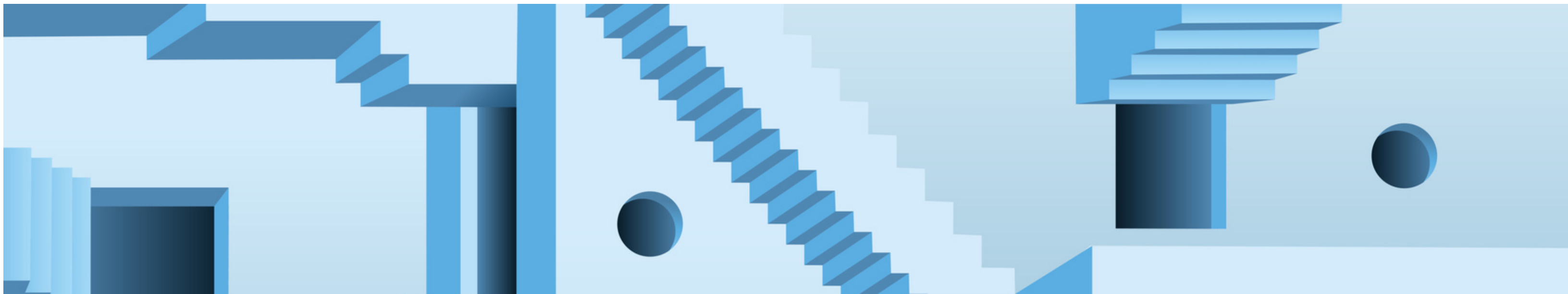
CHALLENGES

Although commitments to trustworthy AI principles have become more popular among corporate and government entities, substantive implementation in business operations often requires considerable investment. The roundtable discussions highlighted some of the hurdles of operationalizing trustworthy AI throughout an organization, particularly outside the tech industry.

One challenge is that many people working within organizations that are using AI consider trustworthy AI responsibilities beyond their purview and that trustworthy AI efforts are solely the domain of technologists. At the same time, the technologists developing and deploying these systems may lack sufficient knowledge of the impacts of these systems to meaningfully assess their consequences. This is further complicated when AI systems do not directly involve decisions or data about people where impacts might be more relatable or obvious to practitioners. It can be unclear to employees whether and how trustworthy AI principles should be applied resulting in an accountability gap.

Some stakeholders may not even realize they are using AI technologies to achieve their business goals.

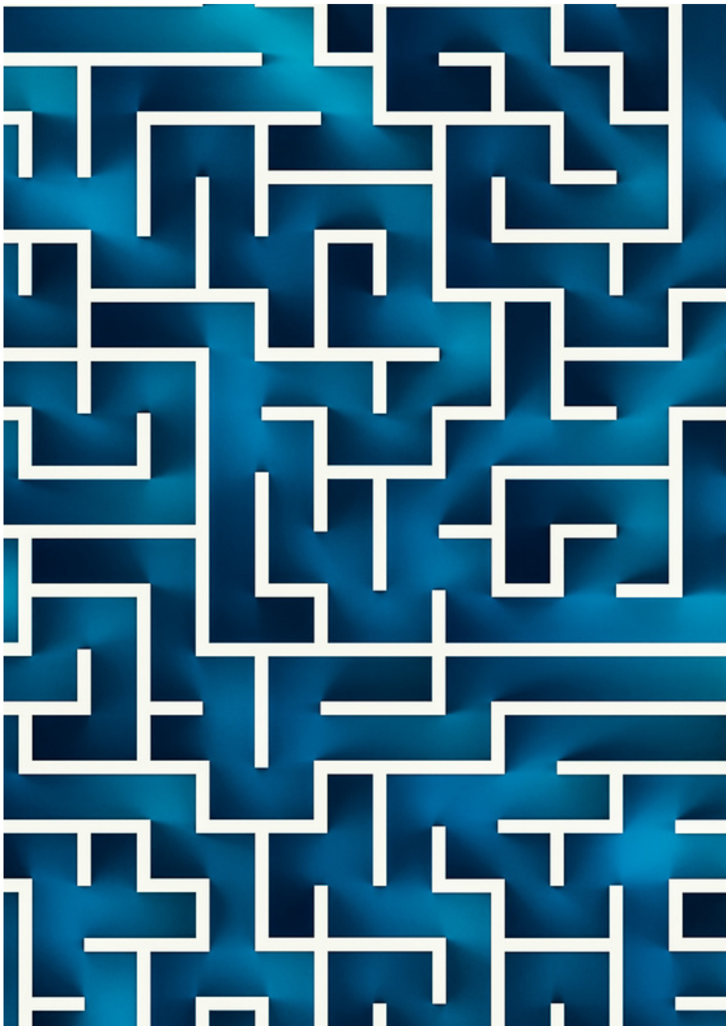
This confusion is compounded by ambiguity around terminology. The term “AI” has been used to describe and promote a wide range of products and services including static models built with machine learning, continual learning models, as well as some more traditional rules-based systems. At the same time, some key stakeholders–including organizational leadership–may not realize that they are using AI technologies to achieve their business goals. Even for organizations that create these systems, defining terms and specifying rules about development can be a challenge. One participant described the situation in their own organization: “We had very strict rules about how to use data. What we found when we did an internal audit is that there were over 200 interpretations.”



Without clear guidance on how to evaluate AI systems for risk of potential harm, resources may be misallocated and more critical systems could be overlooked. Recognizing that they lack the internal capacity, many organizations turn to external sources for AI technologies (and associated trustworthy AI practices) but may lack expertise to adequately evaluate providers. One participant remarked "You're only as good as your supplier"; another recognized that "We may not have the contractual rights to peek into [technology providers'] software, nor do we have the expertise to (that's why we're hiring them)!"

Currently, there are no established industry or regulatory standards against which to evaluate AI systems. A variety of benchmarks exist for measuring system performance in specific areas such as fairness, accuracy to the training data, and computational cost, but they can sometimes promote contradictory design choices. Selecting which of these benchmarks are appropriate for a particular system requires both technical knowledge and ethical judgements. For the organizations that decide to build in house, hiring the specialized talent needed to make these sorts of decisions can be expensive. Fortune magazine reported that the median annual salary for a data scientist in the United States in 2020 was over $164,000. Even for organizations with appropriate skills on staff, providing employees with the training, time, and authority to carry out trustworthy AI commitments can come at a significant cost.

Another more difficult to quantify cost of trustworthy AI is the perception that it will negatively impact innovation. Many practitioners may see trustworthy AI considerations as a hindrance to their productivity; the narrative associated with these activities is often one of burdensome compliance and unnecessary paperwork. Yet technologists who have experience with clear trustworthy AI guidelines often report that they can help to eliminate ambiguity which hastens innovation. Companies and governments aiming to implement trustworthy AI practices across their organization may need to overcome false perceptions in order to build support and fully integrate these processes into their organizations' operations.



"Trustworthy AI efforts may not be cordoned off within the realm of any individual department. It can no more be the sole responsibility of lawyers and compliance specialists than it can be constrained to engineers and technologists."

LOOKING AHEAD

Although operationalizing trustworthy AI continues to be a challenge, many organizations are undaunted. Resources, playbooks, and tools have proliferated in recent years. In this context, the participants in the AI roundtables identified multiple strategies for expanding adoption of trustworthy AI practices—ranging from internal processes to greater information sharing across the industry. None of these strategies are mutually exclusive, and all may be incorporated as part of an organization's overall trustworthy AI approach.

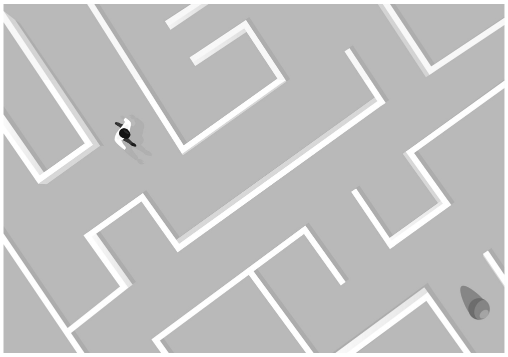
One theme that resonated with many of the roundtable participants was that of "whole-org responsibility." It is important that trustworthy AI efforts not be cordoned off within the realm of any individual department. Trustworthy AI can no more be the sole responsibility of lawyers and compliance specialists than it can be constrained to engineers and technologists. Cultivating this "whole-org responsibility" means aligning teams across the organization and unifying them around agreed-upon goals. Company values may guide these alignment efforts. By building upon the existing structures and buy-in around an organization's mission and purpose, it may be possible to instill a sense of accountability among all workers.

Leadership buy-in is a key to achieving this values-driven alignment across an organization, but can be difficult to achieve. High-level executives may struggle to recognize how trustworthy AI efforts are relevant to their business because they lack understanding of how and where these automated systems are being used. Other times, they may fail to connect trustworthy AI practices with their existing business priorities.

LENSES FOR VIEWING TRUSTWORTHY AI

Participants in the roundtable series identified three lenses through which senior leaders can become champions of trustworthy AI efforts:

Diversity, Equity, and Inclusion
Many organizations are already undertaking efforts to increase diversity, equity, and inclusion in their workplaces and their products. Framing trustworthy AI through this lens can help leaders to orient around the bias and justice aspects of trustworthy AI.



Productivity
When existing tools aren't working or are leading to declines in user satisfaction (whether those users be staff or customers), trustworthy AI efforts may be positioned as a component of productivity. This approach may be resonant for companies seeking to improve internal processes as well as those incorporating AI into products and services.

Privacy & Legal Risk
Although federal policy in the US was recognized as slow-moving, other jurisdictions are increasingly enacting trustworthy AI requirements. The European Union's AI Act was an early mover in this space. In addition to regulatory compliance, privacy and legal risks may motivate business leaders when interacting with partners who may have more stringent guidelines in place.

Leadership buy-in is not just important for aligning organizational efforts. It can also motivate workers to feel safe pursuing trustworthy AI activities within their respective roles. When practitioners see trustworthy AI efforts as supported from the top, workers have more permission to engage without being perceived as an unwelcome critic. Achieving leadership-level value-alignment can allow trustworthy AI work to be recognized as helping the organization to achieve its goals rather than acting as a hindrance.

Even with values-alignment, however, knowledge and skill are still key to trustworthy AI success. Vikram Somaya, Chief Data Analytics Officer for PepsiCo noted, “We understand that just the ‘top-down’ approach is not enough. This has to happen at the place where the rubber is hitting the road.” Educating and empowering people with knowledge about how AI is used and the trustworthy AI issues relevant to their work can be essential for making values-oriented alignment actionable. This knowledge is not yet widespread. Although there are an increasing number of professionals entering the workforce with certifications and education in these topics, for many organizations, developing trustworthy AI expertise means training existing employees for whom these may be new concepts.

Participants in the roundtable series shared a variety of strategies including executive training from outside consultants, incorporating trustworthy AI modules in other company-wide training experiences such as annual cybersecurity trainings, and even peer-mentorship models. One participant described an effort to for scale expertise across a large multinational company. In this organization, they developed a community of mid-career professionals to serve as internal digital ambassadors. These employees were given additional training and education as part of a “train the trainer” effort. These ambassadors were then empowered to share their knowledge inside their organizations to help their colleagues better understand the technical and ethical considerations of these technologies.

”

Another participant described a different "train the trainer" model in which each product or development manager partners with an inclusion specialist to make decisions about how to implement the organization's values and principles in their technology. The manager then reports back to their product or development team the reasoning behind the decisions made. This deliberative process was noted as a key element to successfully mitigating AI risks. Even in organizations that do not build technologies internally and instead source AI tools and expertise from outside, these types of employee deliberation and education efforts can be beneficial. As one participant noted, "The more folks know what the risks are, the more they can make better purchasing decisions."

These models may continue to see wider adoption as more organizations recognize trustworthy AI considerations as essential to their overall strategies. Gartner predicted in 2021 that within two years all workers across AI development and training will be required to demonstrate some level of expertise in trustworthy AI. While formalized standards do not yet exist, there are many approaches and experiments underway. As one participant described it, "We are building the plane as it flies." Roundtable participants emphasized the importance of pilot programs and sandbox environments, as well as the psychological safety of organizational cultures that not only tolerate but embrace failure. "You need to be safe to fail and to learn from failure."

The "whole-org" approach to trustworthy AI can also bring another benefit: greater diversity. Compared to the private sector overall, tech workers have tended to skew more white, Asian, and male. By educating and involving a wider-range of employees in trustworthy AI efforts, more diverse perspectives may shape the implementation of these technologies. Participants noted that diverse teams may also promote more productive discussion and deliberation. One participant noted that when a wider range of perspectives are represented there is more deliberation about decisions in general which helps prevent mistakes that are costly to identify and address later in the AI development and deployment process.

Finally, while internal efforts are essential to operationalizing trustworthy AI within an organization, participants in the roundtables also recognized cross-organizational coordination as an important strategy. One participant noted, “there isn't a single [governance] framework for this, and so you don’t know what people's expectations are going to be… building a common set of expectations nationally and internationally makes a lot of sense.” This is a widely held view. According to the 2021 Investing in Trustworthy AI report by the Deloitte AI Institute and U.S. Chamber of Commerce, 79 percent of AI professionals surveyed indicated that standards for performance and reliability of AI algorithms were high priority. Roundtable participants expressed optimism about multi stakeholder organizations and building relationships with trusted consultants. Pooling resources and expertise was noted as an especially attractive strategy for small-to-medium sized organizations.

CONCLUSION

Ultimately, despite challenges in operationalizing trustworthy AI, more and more organizations and leaders are inspired to act. Remaining challenges are largely issues of knowledge gaps or ill-defined accountability. With greater values-alignment, educational opportunities, and cross-functional collaboration, leading organizations can continue to make progress. As Grace Simrall, Chief of Civic Innovation and Technology at Louisville Metro Government put it, “if we are open to learning about it and try not to repeat the mistakes of the past, this is what's going to move us forward.”

Thank you to the Deloitte AI Institute for their leadership and support of this initiative.

Individuals are encouraged to cite this report and its contents. In doing so, please include the following attribution:

B Cavello, ”Making the Case for Trustworthy AI.” The Aspen Institute, Sept. 2022. CC BY-NC.
https://creativecommons.org/licenses/ by-nc/4.0/.

ASPEN DIGITAL
aspen institute

Deloitte.