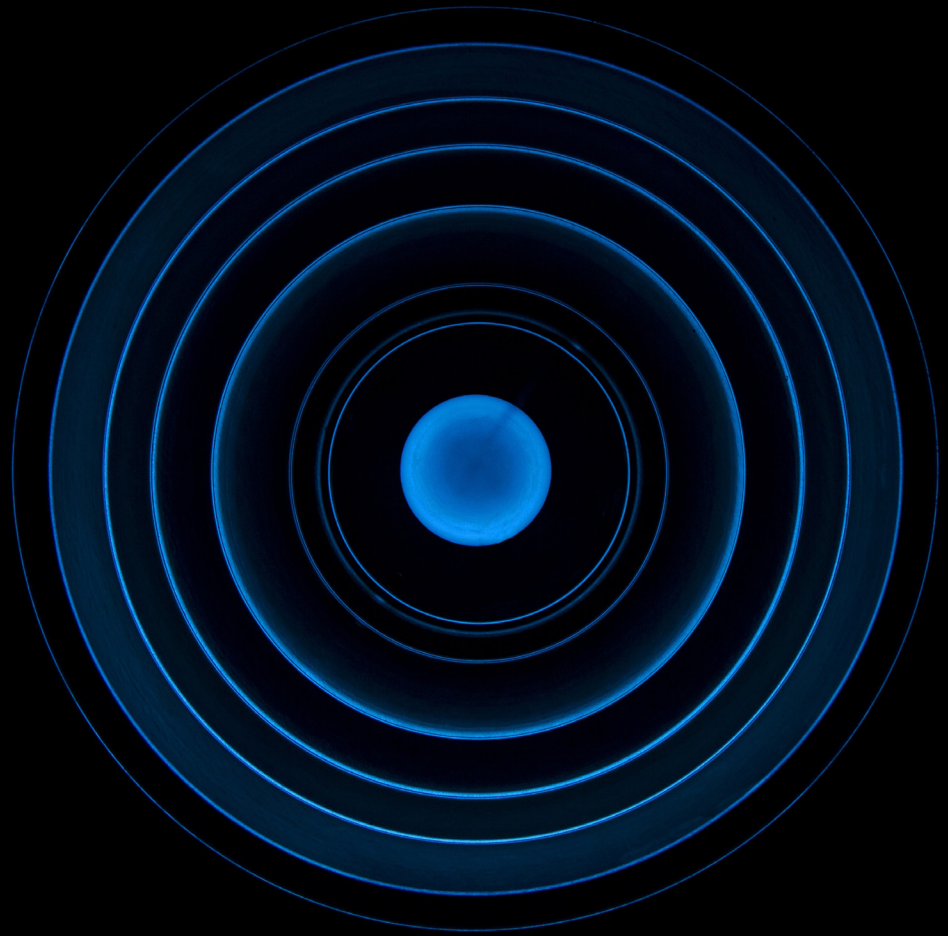
Deloitte.

Trustworthy AI in practice

A report by the Deloitte AI Institute

About the 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 a 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 meet ups and live events. Let's explore the future of AI together.

www.deloitte.com/us/AIInstitute

Trustworthy AI in practice

There is no shortage of pontificating and handwringing over the ethics of artificial intelligence (AI), and views range from a future of abundance to dystopia. Often, the matter is reduced to concerns over bias. While that is a valid issue, it is just one of several dimensions of trust that should have purposeful treatment and effective governance.

The use cases for AI are proliferating across every industry. We see novel AI solutions emerging in areas such as hiring and workforce management, medical diagnosis and treatment recommendations, supply chain efficiency and constraint management in manufacturing, and internal controls for financial reporting. As solutions are deployed, there is a vital need to ensure that these powerful tools can be trusted.

Trustworthy AI does not emerge coincidentally. It takes purposeful attention and effective governance. Indeed, the path from conceiving an AI use case to deploying the model at scale is paved with critical decisions based on careful assessment of impact, value, and risk. Creating and using Trustworthy AI takes more than a discrete tool or a periodic review. It should have a larger governance structure that permeates the entire organization. Taking an end-to-end view, what's needed is an alignment of **people, processes, and technologies** that together promote effective AI governance, and ultimately, AI solutions we can trust.

When an organization orients its AI initiatives toward an intentional focus on ethics and trust, the reward is often a greater capacity to promote equity, foster transparency, manage safety and security, and in a structured way, address the ethical dimensions of AI that are relevant and important for each use case and deployment. To reach this future state, there are a number of considerations and priority issues in mobilizing people for AI governance, enhancing processes and controls, and using technology to bolster trust.

Mobilizing people for AI governance

Across the AI life cycle, there are many critical stakeholders, each of whom brings a diverse perspective and priorities. Whether it is an executive, a plant floor operator, or an IT professional, each stakeholder has a role to play in promoting Trustworthy AI. Some important areas for attention include:

Roles, responsibilities, and accountabilities

Define who is a stakeholder for AI outcomes and how they are meant to participate in the AI life cycle. This includes not just data scientists and AI architects but also implementation leads, development leads, data governance managers, peer reviewers, security experts, and users throughout the lines of business. Also consider the stakeholders impacted, including employees, partners, vendors, and customers. Importantly, the roles, responsibilities, and accountabilities should be documented, establishing clear expectations as they relate to AI ethics and trust.

Education and communication

The workforce should have structured opportunities like knowledge transfer sessions to learn about and understand AI governance. Whether it is through meetings, corporate AI charters and vision statements, online training, or other approaches, a curriculum can lead employees through what constitutes ethics and trust in AI, where the employee fits into the life cycle, and how they can contribute to governance and change management. One approach toward this end is to augment existing ethics training that employees often complete to include AI ethics education. On an organization-wide basis, companies should continually communicate with employees, partners, and customers concerning AI activities. This simultaneously keeps stakeholders informed while also establishing a dialogue. Trustworthy AI is a mindset that should be embedded at all levels in the organization's culture.

Role-specific upskilling

Across every business unit, there is a set of knowledge, leading practices, and skills that are vital to governing AI. Compliance and regulatory leads need familiarity with emerging regulations and laws in regions where a business operates. Human relations professionals should contend with how Trustworthy AI impacts the workforce and third parties. Business domain experts should collaborate with computer and data scientists into translating business objectives into practical AI applications. Executives seek the bigger picture of how trusted AI programs contribute to enterprise strategy. These and other stakeholders need new skills that are specific to their practice area. The result is that upskilled employees working with AI can be better equipped to provide insights and guidance on how the solutions impact business processes and decision-making.

Enhancing processes and controls for trust

Operationalizing Trustworthy AI typically requires creative thinking among business leaders, critical analysis throughout every stage of the AI life cycle, and reliable assurance that the tools and the system around them are meeting the relevant dimensions of trust. Every business is different and faces challenges and priorities unique to their business strategy and objectives, which results in different opportunities to successfully leverage AI capabilities. As such, there is no one-size-fits-all framework for effective AI processes. Instead, conceiving and implementing processes takes key activities for devising the right processes to govern the enterprise's AI programs.

Define the vision

A catalyzing tactic is bringing the organization's leaders together to develop a holistic, equitable approach to creating and using Trustworthy AI. This is not simply a sporadic "business as usual" C-suite meeting. Rather, in a conducive setting with focused goals, leadership can identify the AI's purpose and how and whether it is delivering its intended outcomes. This is an opportunity to think through priorities, encourage multiple viewpoints and ideas, and integrate ethics and governance issues with the organization's business strategy and objectives.

One way to perform this activity is turning to Deloitte's Greenhouse® Experience, which is a structured approach to foster breakthrough ideas in a consciously designed environment using design thinking, analytics and insights, and tested collaboration tools, techniques, and frameworks. Importantly, once a vision is established, it should be documented and shared with leadership, employees, and partners, and the vision should be updated as needed as the organization matures and scales its AI programs.

Identify the risks

Risk analysis is familiar territory for business leaders, and the same principles apply for AI. Risk management strategies and ongoing model risk assessments can help the enterprise prepare for and guard against external sources that could negatively affect the model and the business strategy. A comprehensive and ongoing risk strategy and execution program begins with risk assessments. The assessments inform a risk management plan that serves as the basis for designing risk mitigation controls and responses, which enables human monitoring and decision-making. These assets and activities can collectively support AI model and cloud infrastructure robustness and reliability, address system vulnerabilities, and treat other important dimensions of trust (e.g., privacy, safety, security.).

Identify the gaps

To know how to make process changes, the business should understand where gaps exist in AI risk controls. Building on the risk analysis, organizations can implement or adapt processes and controls to support AI governance. Importantly, AI governance encompasses more than coding of the model. It also includes the broader infrastructure necessary for successful implementation and oversight of AI models. Key components and questions in AI governance may include:

- Data governance and processing – How does the organization obtain and maintain data used in AI models?
- Security – How does the organization secure data used by AI models (as well as the models themselves) from adversarial attack?
- Organizational governance – How does the organization define testing objectives and monitor and evaluate results from controls and processes over AI models?
- Model development and evaluation – How does the organization validate AI model performance?

These components will likely leverage existing infrastructure, but it is important to incorporate Trustworthy AI concepts within AI governance as organizations adapt to a business environment where AI has a pervasive impact on core business processes. With this approach, organizations can identify gaps and develop effective solutions while still driving program progress, speed, and efficiency.

Validate performance

Business leaders need confidence that AI models perform as expected and are in line with business strategy and regulatory compliance. This requires increased transparency, which can be achieved through a rigorous validation process. Validation includes model testing, assessing whether documentation adequately describes the theory and design of the AI algorithm, and ongoing monitoring. Model bias and degradation can be identified, assessed, and mitigated through the validation process.

Early involvement of validation activities helps ensure continuous participation and feedback in business discussions and contributes to a holistic review and monitoring process. The goal is to minimize the potential for error and loss, implement controls, anticipate risks, define achievable use cases, and apply independent assurance by testing the tools and system. Enterprises may employ these capabilities in-house, although the business might also turn to a professional services firm to share leading practices across a multitude of use cases and help accelerate and advance the organization's efforts.

While much of the work of transforming the workforce and processes to foster Trustworthy AI is rooted in human planning and decision-making, there is clearly a role for technology.

Bolstering trust with technology

Continuous, effective oversight should include monitoring AI tools with equally innovative technology solutions. This allows the organization to evaluate whether an AI tool is performing as intended and in line with the relevant dimensions of trust. Executives need this capability for real-world AI assessments, such that they can measure and evidence performance, value, and trustworthiness–which then provides outputs and measures that can be governed.

There is a pronounced challenge when working with "black box" AI, whose inner workings may defy transparency and explainability. Building a model that is intrinsically transparent may be feasible for some use cases, but it may also lead to less accurate outcomes given a trade-off between accuracy and transparency. One approach to navigating this balance is selecting a technology solution that pairs with AI tools to interrogate the model's performance and deliver an assessment and validation. In addition to model evaluation technologies, enterprises may also look to solutions in other technology areas, including AI data management (e.g., synthetic data generation), privacy, cybersecurity, regulatory compliance, risk, and post-deployment monitoring.

For all enterprises, the application of validating technology can improve AI assessments. Importantly, the product of technology-enabled assessments is not just improvements to the model itself. By deeply understanding how a model performs, an organization can be more effective with other components of AI governance.

Board considerations

Like any risk topic, the ethics of AI and the impact on equity are areas of particular interest to those charged with governance. Some implications and outcomes, such as reputational risk, require board consideration and attention. When delving into the more granular details of governance, however, points of concern may be better weighed and addressed by a risk committee or an audit committee. Of particular interest to audit committees in the coming years will likely be consideration of AI relative to internal controls over financial reporting. This is especially the case as the technology moves from automation of financial reporting tasks (e.g., data population into forms) to more advanced areas, such as decision-making.

Whether built in-house or by a third party, it is important for the audit committee to understand whether AI is being used in the Internal Control over Financial Reporting (ICFR) process, and if so, the level of delegated authorization that the tools are providing to the overall ICFR process. For example, if AI is highlighting key terms in a contract, but the accountant continues to review the entire contract, this may not require significant oversight. However, if the AI is making authorization decisions for transactions (e.g., investment, payment, etc.), this may require a more in-depth understanding of the decision rights and management's monitoring controls. The reality is that regardless of whether AI is used in ICFR specifically, the governance aspects are similar relative to trust, brand, and equity-all of which are on the agenda for public company boards.

Increasingly, audit committees should be informed of the implications of new technologies via educational sessions and continual updating. From there, the audit committee is better equipped to ask the right questions, some of which include:

- What are the objectives of the AI model, and how does it align or enable business strategy and objectives?
- Who should be a part of the governance group, and how should it be structured?
- What is the risk assessment process regarding AI, and what are the identified risks associated with relevant AI use cases?
- How does the organization monitor risks, processes, and controls on an ongoing basis?
- Are there controls in place to ensure the AI use cases do not exhibit biases that could be illegal, discriminatory, or unethical?
- How much authority or judgment is delegated to the relevant AI models?

Looking ahead to a trustworthy future

These approaches for orienting people, processes, and technologies are demonstrated tactics for operationalizing Trustworthy AI. Yet, knowing a solution is often simpler than implementing it, and when confronting an evolving technology with enormous potential for business benefit, enterprises can benefit from guidance and support. A collaborator with deep experience in Trustworthy AI development and application can offer an independent perspective and critical knowledge.

Deloitte uses a three-pronged approach to enabling AI governance:

- Evaluating roles and responsibilities, implementing change management, and conducting training.
- Offering services around AI strategy and governance, risk management, and AI assessments.
- Using tools to test and monitor your AI models.

We know business leaders are seeking 20/20 vision on potential high-risk exposure areas, advice and recommendations on AI controls, and guidance throughout the AI life cycle in the areas of feasibility, discovery, modeling, acceptance, and integration. Deloitte's Trustworthy AI™ framework provides a cross-functional approach to identify many of the key decisions and methods needed for appropriate procedures.

Just as valuable is proven excellence in AI auditing. Deloitte was recently recognized by the International Accounting Bulletin for a Digital Innovation of the Year for its Omnia Trustworthy AI module, which provides guidelines and guardrails for organizations to design, develop, deploy, and operate ethical AI solutions. The module is a unique combination of our Trustworthy AI™ framework, digital evaluation tools, and independent insights from Deloitte specialists.

Across the spectrum of professional services-from strategy consulting to advisory support to audit and assurance-Deloitte can offer the approach and underlying offerings that help you take the right steps, make sound decisions, and implement the processes and programs that help ensure you are able to not just find value in AI but trust it as well.

Authors

Beena Ammanath
Executive Director
Global Deloitte AI Institute
Email: bammanath@deloitte.com

Brian Cassidy
US Audit & Assurance
Trustworthy AI leader
Partner | Deloitte & Touche LLP
Email: bcassidy@deloitte.com

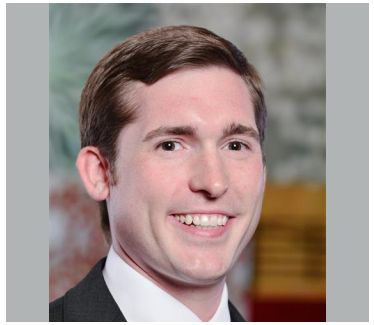
Ryan Hittner
Risk & Financial Advisory
Managing Director
Deloitte & Touche LLP
Email: rhittner@deloitte.com

Alexey Surkov
US Risk & Financial Advisory
Trustworthy AI leader
Partner | Deloitte & Touche LLP
Email: asurkov@deloitte.com

Contributors

John Fogarty
Audit & Assurance
Senior Manager
Deloitte & Touche LLP
Email: johfogarty@deloitte.com

Sanjana Jain
Risk & Financial
Advisory Manager
Deloitte & Touche LLP
Email: sanjajain@deloitte.com



Zach Bowman
Audit & Assurance
Senior Manager
Deloitte & Touche LLP
Email: jambowman@deloitte.com

Deloitte.

About Deloitte