The New Imperative of Responsible AI

For decades, engineers have explored artificial intelligence (AI) to give machines more human-like capabilities. The past 10 years have brought significant progress, from smart assistants like Siri and Alexa understanding our commands and managing simple tasks to driverless cars offering an alternative to taxis in major cities. More recently, generative artificial intelligence (GenAI) has evolved at a breathtaking pace, creating new possibilities for individuals and businesses alike.

Artificial intelligence and its capabilities are fueled by large datasets that can include images, text, audio and numerical data, much of it gathered from the internet, along with the data companies collect about their customers. That's why ethical technology governance to manage AI responsibly should matter to everyone.

Avoiding Unintended Consequences

How does the use and development of AI affect the people on which it's being used? As companies pursue opportunities to make their businesses better and stronger with AI, they should also define initiatives to address the risk that misinformation or bias could create unintended consequences for customers. These consequences are often not a result of malicious intent, but rather a byproduct of how AI systems are designed, trained and deployed. Without initiatives to oversee the approach, companies may be exposing themselves to significant financial and reputational damage that could erode shareholder value.

Here are some of the ways a company's use of AI can have unintended consequences:

- **Imperfect outcomes.** While AI has progressed significantly, saving time and creating greater efficiency and precision with some tasks, it's widely acknowledged that AI can make mistakes. Errors may not be immediately apparent due to the complexity of the underlying algorithms and models can be used to create and disseminate misinformation. Companies should continue to uphold high standards for the integrity, quality and safety of the products and services they are selling. Health insurers, for example, may face lawsuits for allegedly using algorithms to improperly deny care.

- **Biases**. Any direct or indirect discriminatory biases contained in the underlying datasets could lead to outcomes that favor certain groups over others. For example, AI-powered recruitment tools that screen resumes may favor certain characteristics and limit access to the widest and best talent pool for the job. In addition, some companies have paid substantial settlements to resolve allegations of discriminatory practices in areas such as advertising and evaluating medical claims.

- **Privacy and security.** Companies using AI for customer analytics, employee monitoring or predictive modeling may gather and analyze personal data in ways that violate privacy rights. For example, a company may collect data for one stated purpose, such as improving a product, but then use it to train an AI model for an entirely different purpose, which violates the core tenet of many privacy laws. Many AI models are also trained on data that was "scraped" from the public internet. This can include personal blog posts, social media content and even copyrighted material, where sensitive information may be unknowingly incorporated into an AI model. Additionally, companies are managing larger datasets than ever before, which creates an even-greater risk of a data security breach in which sensitive personal information may be exposed to bad actors.

Building trust requires a full command of how an AI system functions and its ability to achieve proper outcomes. AI systems have been evolving to show more levels of decision making. For business outcomes, AI could include the rationale for key decisions to help its customer understand why, for example, a healthcare claim was denied. Businesses could also notify customers when they are interacting with an AI system and provide clear descriptions of the role it plays.

> *Without initiatives to oversee the approach, companies may be exposing themselves to significant financial and reputational damage that could erode shareholder value.*

Hallmarks of Responsible AI

Over the past couple of years, many companies have been considering establishing or managing AI governance frameworks. When we meet with companies to evaluate their AI governance approach, we query them to understand: What are your AI policies and governance structures? Who staffs them and what are key roles and responsibilities? How are they resourced? What frameworks do you use? What have these efforts looked like in practice, and are you assessing and reporting on them regularly? The company engagement often involves an active, multi-year approach that can yield tangible results such as publishing inaugural AI principles, implementing or strengthening AI governance, mitigating algorithmic biases in products and services, and assessing the downstream human rights risks and lifecycle impacts of AI products. We've seen success working with companies to establish basic building blocks to demonstrate that their systems are performing and identifying risks well.

Best practices for companies using AI include establishing formal internal governance structures with board-level oversight of material issues and aligning risk-management programs with established external frameworks such as The National Institute of Standards and Technology (NIST) AI Risk Management Framework. This framework offers a voluntary and structured approach to identifying, assessing and mitigating AI-related risks that emphasizes key characteristics of trustworthy AI. It's a flexible framework that can be applied by companies of all sizes and in any sector.

Companies should have policies, practices and controls that prove these systems work, including:

- Establishing clear governance structures throughout business strategies that rely on AI and disclosing a set of AI principles and commitments.

- Integrating safety, bias and privacy considerations into product life cycles.

- Providing clear insight into the lifecycle processes for product review.

- Creating, strengthening and updating policies and programs for AI applications, to consider individual impacts and unintended consequences and provide methods for resolution or escalation of material issues.

- Disclosing performance metrics, such as how many AI products were reviewed for heightened risks and what percentage were consequently modified or halted.

Transparency and accountability are also important aspects of building trust in the AI outcomes. Companies are making significant investments in AI. Investors should be able to understand and evaluate the company's efforts and progress. Companies should also demonstrate how their risk management processes prepare their businesses to mitigate these risks and capitalize on opportunities.

A Business Imperative

Just as GenAI has become an essential part of business operations, the responsible use of AI has become an essential strategy and component of risk management. And as AI systems become more powerful and embedded in society, investor stewardship to understand company policies and how they align with best practices through company engagements and proxy voting is increasingly critical.

This is not a temporary trend or a passing concern. It represents a fundamental shift in evaluating corporate responsibility and long-term value creation. The companies and investors that lead this transition can help ensure that this transformative technology is developed in a sustainable way that is aligned with long-term value creation.

REAL-WORLD IMPACT: CASE IN POINT

While artificial intelligence presents a spectrum of challenges, from the massive energy consumption of data centers to concerns about worker displacement, our stewardship approach focuses on the ethical considerations and downstream risks AI poses when business outcomes have unintended consequences. Our engagements on this topic often look for assurance, transparency and accountability among companies integrating AI features into their products.

In 2023, we began engaging **Intuit** (INTU) to better understand how their AI governance approach was aligned with best practices. The financial software provider has integrated AI throughout its product offerings in tax and personal finance software as well as its email marketing platform. Following our engagement on the topic, Intuit released public disclosure on AI governance and risk management practices in the first half of 2025.

Investor Perspective: Responsible AI Recommendations

This year, Parnassus led the development of a set of investor recommendations for responsible AI in partnership with other investors. The document is intended to be a helpful reference for companies considering how to approach an ethical technology governance structure.

To learn more, visit www.parnassus.com.