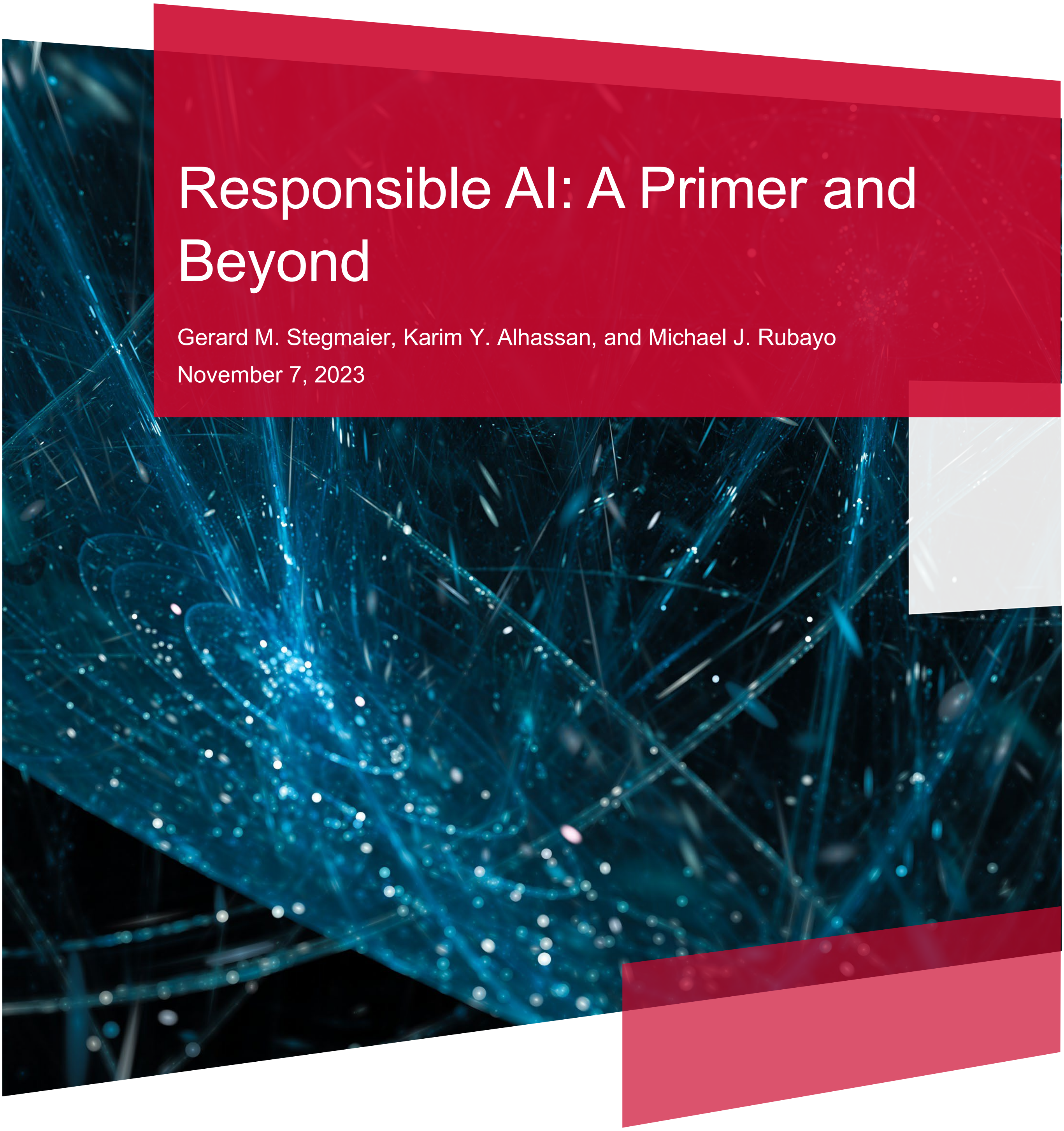
Responsible AI: A Primer and Beyond

Gerard M. Stegmaier, Karim Y. Alhassan, and Michael J. Rubayo

November 7, 2023



ReedSmith
Driving progress
through partnership

Responsible AI: A Primer and Beyond

Introduction

Artificial intelligence and machine learning has been powering applications, workflows, and business processes for decades, quietly transforming how we live and work. However, until recently, the use of artificial intelligence has remained in the background, reserved for software developers and technologists. The advent of ChatGPT and similar generative artificial intelligence tools ("GenAI") has literally and figuratively changed the conversation putting the power of this technology at the fingertips of consumers. This radical change has in turn driven furious introspection and discussion at the highest levels of enterprise. McKinsey & Company estimates that artificial intelligence could add trillions to global GDP.[1] Due to the potential for such vast productivity gains, it is predicted that 80% of enterprises will have used GenAI in some form by 2026.[2]

It is estimated that artificial intelligence could add trillions to global GDP. The potential for productivity gains is so vast that 80% of businesses will use some form of GenAI by 2026.

Of course, with such hype comes vast scrutiny. Globally, governments and regulators have ratcheted up their focus on the use of AI, from the Federal Trade Commission reminding businesses that it is "focusing intensely on how companies may choose to use AI technology,"[3] the Securities and Exchange Commission warning of the inevitability of financial crisis,[4] and entire countries banning the use of particular GenAI applications due to privacy concerns.[5] In addition to regulatory scrutiny, the novelty of GenAI use cases and the uncertainty around the application of current legal regimes and principles has invited litigation and regulatory risk, from copyright claims based on image outputs to common-law privacy claims predicated on training datasets.

Moreover, due to the inherent technological limitations associated with GenAI – and the "blackbox" way in which these services operate – the potential for operational and reputational harm abounds, with incorrect – so called hallucinations – biased, harmful, or otherwise damaging outputs generated by the use of such applications. Beyond the quality of outputs, the way in which these outputs are generated, especially the data on which they are trained and how that training occurs, poses additional challenges.[6]

Given these risks, balancing the potential of artificial intelligence against its potential risks and costs begs important questions about corporate governance generally and the emerging discipline of artificial intelligence governance in particular. Here, we are not writing about the type and nature of regulations that should exist or what governments and regulators should or should not do. For business, the practical questions today are not whether and when to adopt this technology but how to ensure it is used effectively *and responsibly*. This means figuring out how to define, establish, and implement governance frameworks around the use of AI, ensuring that the policies and procedures which are adopted serve their purpose.

What is Artificial Intelligence Governance and Responsible Artificial Intelligence?

Broadly speaking, artificial intelligence governance means the ability to direct, manage, and monitor the use of artificial intelligence within an enterprise. A hallmark of effective artificial intelligence governance is that it is holistic and reaches across business lines and through organizational silos. To be clear, effective governance of artificial intelligence is not merely a tech or IT issue. And, like other effective specific governance initiatives, it benefits from being more broadly aligned to an enterprise's broader corporate governance requirements and oversight. Lastly, effective governance often begins at the top – originating in the boardroom and informing and influencing overall business strategy and risk management. Responsible artificial intelligence often means more than mere artificial intelligence governance because the name itself suggests an ethical component which may span duties or obligations beyond compliance or looking out solely for the interests of shareholders. In this respect, responsible artificial intelligence may look beyond the strict confines of corporate law and Delaware law in particular.[7]

Effective governance of artificial intelligence is not merely a tech or IT issue.

Elements of an Effective AI Governance Framework

At its core, an effective AI governance framework should be no different than other governance mechanisms already rooted within enterprises. Although the risks associated with AI may present distinct challenges – as does the use of any new technology – the general principles underpinning a successful framework around the use of AI remain the same.

AI governance frameworks frequently use a process involving:

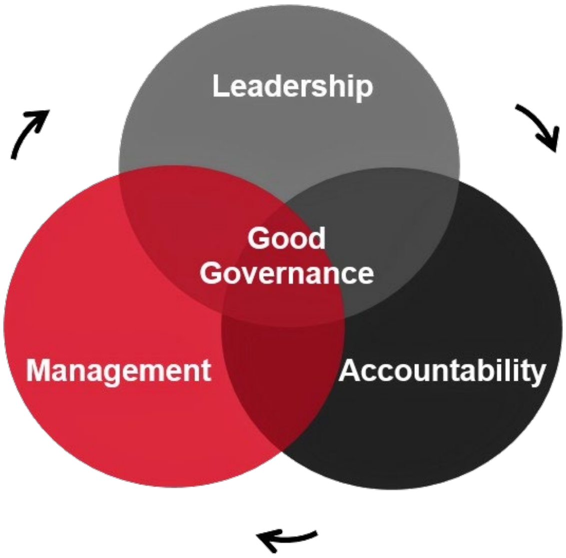
- determining and memorializing AI governance scope;
- inventorying and monitoring AI deployment, usage and potentially effects and consequences;
- developing and implementing specific AI policies and procedures and updating and reconciling other policies and procedures as appropriate; and
- deploying, testing and updating controls.[8]

And, most importantly, each of these steps represents part of an iterative process occurring over time. This cycle ideally takes into account broader enterprise risk considerations and reflects them in artificial intelligence governance as a subset of those broader risks. In a nutshell, things will happen, mistakes will be made but appropriate “north stars,” followed deliberately, using continuous learning and improvement, can result in effective governance, including AI governance. This, in turn, can contribute dramatically to the success of an enterprise. In fact, one recent report found that certain high-performing enterprises focused on responsible AI development and use were able to create more than 50% revenue growth while outpacing their peers in regards to ESG and customer experience metrics.[9]

- **AI Governance Scope.** The first step towards effective artificial intelligence governance is defining the contours of in-scope applications and activities. Put another way, to regulate AI, you must know what AI is. For example, the Biden Administration recently defined artificial intelligence as “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.”[10] Although artificial intelligence might generally be seen by the public as any technology intended to replicate human intelligence, the definition of AI – as used in the governance context – benefits from being both industry and business specific. For example, for an enterprise primarily engaged in marketing activities, the governance framework may only treat GenAI applications in scope while a financial services business may choose instead to focus on applications with decision-making capabilities. In short, the defined scope of any framework is more likely to be effective if it seeks to mitigate business-specific risks in a tailored, deliberate manner.
- **Inventory and Monitoring.** Having defined what artificial intelligence is in the context of the business (scope), the next practical step is to then determine with sufficient granularity how, where and when it is being used. Not only is baseline assessment frequently critical to initiating effective governance but recognizing that many assessments are stale before they are even completed requires a thoughtful approach. Resources are limited and assessments which are too detailed (or not detailed enough) often will set the tone for the entire initiative. The inadequacy or pain learned from data mapping exercises or, for those old enough to remember, initial Sarbanes-Oxley controls reviews, provides important opportunities to leverage internal learning from previous projects including recognizing that good governance projects are often fundamentally interdisciplinary. As the pace and ubiquity of AI-powered applications hastens, the potential for shadow sprawl is increased, with the risk of unmonitored use escaping the boundaries established by the framework.
- **Policies and Procedures.** As with any governance framework, policies and procedures are the day-to-day oxygen of an effective AI governance framework, setting the rules of engagement and informing and incentivizing employee behavior with the aim of creating consistency. The maxim of perfect practice leads to perfect performance comes to mind. Whether embedded in an existing governance, risk management and compliance framework or developed independently, artificial intelligence-related governance policies and procedures should expressly outline permissible and prohibited activities, outline processes, and articulate expectations. Often, where possible, existing policies and procedures may create a tremendous source of leverage. For example, where an enterprise maintains a third-party governance framework for onboarding external vendors and applications, the integration and prioritization of AI-enabled applications may be considered. Doing so may leverage existing enterprise functions while ensuring AI-specific risks are properly accounted for and mitigated as a subset of larger IT, security, data protection and related governance.
- **Controls.** To ensure that the identified program components are being properly considered and implemented, properly functioning internal controls are essential. A first step in implementing sufficient internal controls involves the formation of a centralized function, whether a working group or other body charged with implementing the framework. Such groups in many organizations will also often receive direct oversight and monitoring by the board of directors often through or aligned to existing corporate governance oversight. In this respect guidance from the National Association of Corporate Directors on cybersecurity governance informs an instructive approach.[11]

Clear leadership buy-in and direction from directly responsible individuals with the support of the board will facilitate greater consistency in decision making, cohesion, and focus, while ensuring that governance remains flexible and responsive to external changes. Moreover, the centralized body or working group should possess sufficient authority to establish and update policies and procedures, conduct internal audits, and categorize and define risks. Importantly, these groups also provide an opportunity to ensure inclusion of all relevant stakeholders while addressing broader business strengths, to help ensure effectiveness. Additionally, the resulting internal controls often will benefit increasingly from automation. The rapid growth of software and automation to facilitate governance, risk, and compliance management both generalized and specific to artificial intelligence can augment and help ensure the proper existence of and functioning of the controls.[12]

The rapid growth of software and automation to facilitate governance, risk, and compliance management both generalized and specific to artificial intelligence can augment and help ensure the proper existence of and functioning of the controls.



Key Questions for Businesses Deploying AI and Seeking to Implement Responsible Artificial Intelligence.

The volume of artificial intelligence-related announcements by governments, regulators, and business has been astounding.[13] Keeping track, much like keeping track of artificial intelligence usage in the enterprise, is difficult, if not impossible and could be wasteful. At the same time, legal requirements, which are a core component of effective governance, will need to be monitored and addressed. With this in mind, we present what we think are some practical starting points for organizations which may not have formed a specific artificial intelligence governance initiative or an effort to ensure responsible use of AI in their enterprises. Significantly, many of the activities and areas to focus upon will be able to leverage security, privacy and prior data governance and other governance initiatives using the process identified above or a similar approach.

On, October 30, 2023, the G7, as part of the Hiroshima Process, released certain principles for the promotion of trustworthy and innovative AI development and use.[14] This statement built on the AI principles put forth by the Organization for Economic Co-operation and Development ("OECD"), an intergovernmental organization comprising of 38 member countries. Many of their recommendations are not necessarily clearly directed at organizations, governments, or regulators but rather set forth what we would call affectionately "commandments" for responsible artificial intelligence. We have sought to distill here, or quote directly, those we think are most useful to keep top of mind and when leading or seeking executive and board sponsorship of an artificial intelligence governance effort.

- **Assess Risks and Build and Deploy Responsibly.** "Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle."
- **Manage Privacy, Data, IP Inputs.** Implement appropriate data input measures and protections for personal data and intellectual property and similar issues both in development, deployment and usage.

- **Monitor Usage and Consequences**. Listen for “patterns of misuse after deployment, including placement on the market.”
- **Focus on Quality/Efficacy and Use Warning Labels and Guardrails**. “Publicly report advanced AI systems’ capabilities, limitations and domains of appropriate and inappropriate use to support ensuring sufficient transparency, thereby contributing to increased accountability.”
- **Apply Reasonable Risk Management Principles**. Develop, implement and disclose, as appropriate, AI governance and risk management principles and policies and procedures grounded in a risk-based approach – including privacy, security, quality and mitigation measures.
- **Practice Secure Software Development Life Cycles**. Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.
- **Practice Practical Provenance**. For us this means train and build artificial intelligence responsibly with respect to issues like privacy and intellectual property rights and also consider ways to ensure users can easily identify what’s AI-generated versus human or human-hybrid. The OECD says directly “[d]eploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content” but this is only one aspect of provenance. One way to visualize this commandment is to recognize that in any good recipe there are ingredients, preparation techniques and then there’s the presentation.

It should be obvious that these seven commandments inter-relate and in some respects may even be marginally duplicative. At the same time, the OECD’s decision to enumerate them separately may be an important signal about how government and regulators will ultimately view failure to adhere to them or may be prone to make them prescriptive requirements.

In addition to these commandments which may be directly relevant for AI-practicing organizations there are others which may tie in more broadly to the specific mission, vision and values of an organization. While our explication of these may be a bit cheeky we think presenting them this way emphasizes their practicality and orientation towards the common good or in increasing parlance “Responsible AI.”

- **Play Well with Others**. Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems, including with industry, governments, civil society, and academia. By playing well with others we can also collaborate better and learn from them.
- **Figure Out What’s Real. Pay Attention. Mitigate or Avoid Bad Stuff.** The OECD advises that responsible artificial intelligence should prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures. Note: Ensuring these risks are real and their mitigation is proportional to relevant benefits and burdens and who should carry them is and will remain a fundamental challenge for all policymaking.
- **First Things First. Eat the Frog**. Prioritize encouraging or investing in “the development of advanced AI systems to address the world’s greatest challenges, notably but not limited to the climate crisis, global health, and education.”
- **All Better Together**. “Advance the development of and, where appropriate, adoption of international technical standards.”

Conclusion

In short, AI – and GenAI in particular – promises to bring enterprise-wide advantages and efficiencies in various domains and at vast scale. However, without an effective governance mechanism in place, regulatory, operational, and reputational harms may mount, neutralizing the benefits while slowing and preventing innovative use cases. As such, to prepare for the AI age, enterprises should consider carefully adopting a principle-based, business-specific framework to harness the vast benefits while reducing and managing identified risks.

Gerard M. Stegmaier

Partner, Reed Smith
Tech & Data
+1 202 414 9228
gstegmaier@reedsmith.com
X: @1sand0slawyer
LinkedIn:
linkedin.com/in/gerardstegmaier

Karim Y. Alhassan

Associate, Reed Smith
Tech & Data
kalhassan@reedsmith.com
LinkedIn:
linkedin.com/in/karim-alhassan-b48732139/

Michael J. Rubayo

Associate, Reed Smith
Tech & Data
mrubayo@reedsmith.com
LinkedIn:
linkedin.com/in/michael-rubayo/

[1] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights

[2] https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026

[3] https://www.ftc.gov/business-guidance/blog/2023/05/luring-test-ai-engineering-consumer-trust

[4] https://www.ft.com/content/8227636f-e819-443a-aeba-c8237f0ec1ac

[5] https://www.bbc.com/news/technology-65139406

[6] https://www.wired.com/story/ai-chatbots-can-guess-your-personal-information/#:~:text=The%20AI%20models%20behind%20chatbots,scammers%20or%20to%20target%20ads.&text=The%20way%20you%20talk%20can,re%20talking%20to%20a%20chatbot.

[7] *See* Gerard M. Stegmaier and Courtney E. Fisher, *Caveat Director: Things May Not Always Be What They Seem in the Court Room, Server Room and Board Room—Analyzing Cybersecurity Challenges in Corporate Governance, a 2023 Update* (discussing privacy, security and fiduciary duties under Delaware law and highlighting how technology-related risk could increasingly be deemed "mission critical" regulatory risk).

[8] *See* e.g., National Institute of Standards and Technology, AI Risk Management Framework (released January 2023), available at https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

[9] *The Art of AI Maturity, Advancing from Practice to Performance*, available at https://www.accenture.com/gb-en/insights/artificial-intelligence/ai-maturity-and-transformation.

[10] *See* Biden Administration Executive Order on "safe, secure, and trustworthy artificial intelligence", October 30th, 2023, available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[11] *See* Stegmaier and Fischer, *supra* note 7, at 2-726.

[12] https://www.insightpartners.com/ideas/responsible-ai-governance/

[13] *See* e.g., International Association of Privacy Professionals, *AI Governance: What is Being Governed?,* October 25, 2023, available at https://iapp.org/news/a/ai-governance-what-is-being-governed/?mkt_tok=MTM4LUVaTS0wNDIAAAGPC-_MN3SAi8IdbDGAQC-e5hjlEhbMZLrNbEJv-e7RbV7ftAFeuAE16vKzCk3nHPA7xCUIaJlSIOed1CgGGzsXkD613fiqRCLLWsyb4aFHyh-L; Biden Administration Executive Order on "safe, secure, and trustworthy artificial intelligence", October 30th, 2023, available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/; G7 Leaders' Statement on the Hiroshima AI Process, October 30th, 2023, available at https://www.mofa.go.jp/files/100573466.pdf.

[14] G7 Leaders' Statement on the Hiroshima AI Process (October 30, 2023), available at https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process