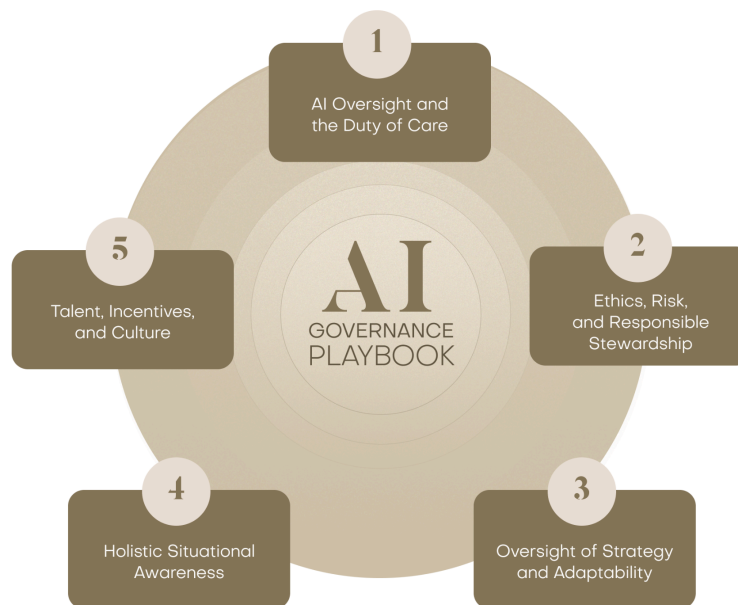


The Athena Alliance AI Governance Playbook for Boards and Management



INTRODUCTION

The Challenge:

We are collectively embarking on an unprecedented global experiment, an era of rapid technological advancement, where data algorithms can shape destinies. The rise of artificial intelligence (AI), especially Generative AI (“GenAI”), in the public consciousness is infused with excitement for what might become possible, and at the same time, fraught with anxiety and uncertainty for the outcomes that may be unleashed.

This much is clear: the outcome will depend heavily on the integrity of our governance structures, the quality of our processes and the intentional focus of the people who lead them.



AI and other exponential technologies are poised to permeate nearly every facet of society, and their application must be informed by a holistic spectrum of voices that reflect the needs and experiences of those stakeholders. The Athena Alliance AI Playbook serves as a compass for executives and board members, guiding them through ethical quandaries and strategic decisions to harness AI's potential responsibly. It is informed by a collection of business perspectives that contributes additional and needed context, texture and perspective to the conversation around AI ethics, use and oversight. We hope it is one of many points of view that are elevated in service to the larger mission: widespread adoption of AI that takes into account a broad range of experiences, needs and perspectives.

The publication of this Playbook coincides with several notable conversations about the responsible use of technology. The European Union's (EU) approval of the pioneering AI Act marks a regulatory milestone in AI governance and a call to action for boards and leaders alike. In the US, the White House issued an Executive Order on the safe, secure and trustworthy use of AI and followed up with guidance on key actions. Upheaval at OpenAI following CEO Sam Altman's dismissal and rehiring over the course of a head-spinning five days, triggered a widespread debate in many board rooms centered on philosophical differences around the adoption speed and use of AI. United States Congressional hearings on a bill aimed to curb the negative effects of technology on young people included striking images of parents holding photos of children who died from actions linked to their use of social media.

It was a sobering reminder of our obligation as leaders to consider risks and potential consequences on the front end of adopting new technology, when we have the most opportunity to get it right. The writing is on the wall: governance models built to address past challenges are no longer sufficient to address the complexity we face today and into the future. The responsibility for this lies not just with the board, but also with the executives and leadership teams who enact strategy and manage day-to-day risk.



The Authors:

The Athena Alliance is a collective of executive and board-level women who represent a diversity of backgrounds, expertise, life experience, and perspectives. We are a community focused on impact. We believe that business, society and humanity are at a time of deep transformation. Many of the old tools, structures, and norms we have relied on for decades are no longer fit for purpose. What connects us is the belief in the imperative for more human-centered, inclusive, and accountable leadership as well as a deep sense of community and the desire for authentic and intentional conversations about our most pressing challenges.

We know that when diverse voices have a seat at the table, different questions are posed, and a wider view of risks and opportunities is surfaced. This in turn leads to more robust discussion and more effective solutions to an organization's most pressing challenges. We have been struck by the lack of diversity in early discussions on exponential technologies, particularly the absence of the voices of women and people of color.¹ As AI's societal impacts unfold, the influence will be transformational, profound, and far-reaching. Addressing these shifts requires a diversity of voices to inform what tools and strategies we deploy, how we assess risks and opportunities and how we manage downstream impacts. In current conversations playing out about unintended consequences of other technologies like social media, we see not only an opportunity but an imperative to have a conversation about responsible governance now, while our ability to influence the responsible implementation of AI will have the most impact.

This Playbook builds upon the great work that has been done to date to help leaders understand and prepare for the impacts of AI including that of The Alan Turing Institute, The Stanford Institute for Human Centered Artificial Intelligence, The Centre for Humane Technology, MIT Computer Science & AI Lab and NIST, to name just a few². This Playbook contributes additional context and perspective to this discussion, emphasizing the need for responsible AI adoption that considers a wide range of experiences and viewpoints and balances business growth with long-term sustainability in its many forms.

¹ [Fei-Fei Li and the binders full of women in AI | The AI Beat | VentureBeat](#)

² See additional resources under Pillar 1 for links to these



The Approach:

In this Playbook, we lay out key pillars of governance and oversight that are critical to creating a culture of responsible AI use. The framework offers a balanced approach that navigates the thin line between safety and innovation, short-term profits, and long-term value creation, with the true driver of business at its core – people.

In this Playbook, the term AI is used broadly to include foundational AI platforms, machine learning, GenAI and large language models. AI's power and attendant ethical considerations – from bias and human rights considerations to privacy, disinformation, and copyright issues – are heady. Understanding this is critical to fostering a culture of responsibility and transparency as well as navigating the delicate balance between progress and innovation and the impacts on both quality of life and the long-term sustainability of people and planet. The recommendations here are rooted in a larger duty of care that ensures that each step an organization takes, it approaches with wisdom, intentionality and a vision rooted in long-term value creation and the greater good.

Because this Playbook is written by women who represent a variety of functions and leadership roles, it is useful for anyone, and everyone interested in understanding and overseeing the implementation of new technologies. The framework laid out in the Playbook is meant to provide a first step and designed to be agile and adapt as our understanding of AI impacts unfolds. It lays a critical foundation that we will then build upon over the coming months and years. We will supplement with case studies and more in- depth guidance for functional leaders on considerations for tactical application of AI and other exponential technologies.

During this critically important phase of AI adoption, the Playbook aims to ensure diverse viewpoints are heard, and contributors of varied backgrounds who are making significant contributions to new technology have an equal seat at public, private and social decision-making tables.

EXECUTIVE SUMMARY

Artificial Intelligence (AI) and other exponential technologies are poised to transform nearly every aspect of life as we know it. These times demand a thoughtful, proactive, and vigilant understanding of AI's transformative capabilities and potential repercussions, and call for a paradigm shift in boardroom strategies. The *Athena Alliance AI Governance Playbook (the "Playbook")* – a referenceable, holistic, practical and dynamic guide for board members and executives – is designed to help leaders at the top of organizations navigate the risks and rewards of exponential technology now and in the coming years by laying out a human and sustainability-focused lens on governing this new and complex business landscape. While the Playbook focuses more specifically on AI, it can be applied to the governance of multiple forms of emerging technology. The goal of the Playbook is to provide a human-centered, multidisciplinary view of the key governance issues of such technology and to give leaders a framework for how to apply effective oversight, even as our collective understanding of the uses, implications, opportunities and risks continues to evolve.



Authored by senior women executives and board members, and shaped by a broad range of stakeholders, it lends a critical voice that has heretofore been insufficiently included in shaping the conversation about AI governance. The guidance in the Playbook reflects insights and expertise from dozens of women representing a diversity of industries, sectors, and roles with experience in everything from data, technology, product, talent, culture, risk and ethics, to legal, regulatory, governance, strategy and ESG. We are board members, C-suite leaders, executives, technologists, founders, strategists, and advisors who are all keenly focused on how the coming AI revolution will shape nearly every aspect of our businesses and our lives. The aim of the Playbook is to support board members and company leaders in ensuring their organizations deploy safe, relevant, compliant, and responsible AI policies and products.

This Playbook merges ethical, regulatory, and strategic considerations with an emphasis on long-term value, responsible stewardship and comprehensive situational awareness of converging forces. Key takeaways include the necessity of adaptable AI governance frameworks, robust talent strategies and proactive tech risk intelligence. It is imperative that all business leaders, regardless of functional area, strive to align AI advancements with risk not only risk oversight and strategic growth, but ethical principles and societal well-being. This Playbook is an indispensable resource for leaders seeking a sustainable, proactive approach to AI governance.

The AI Playbook is organized as follows:

1. Introduction
2. Executive Summary
3. The Five Pillars of Organizational AI Governance
 - PILLAR 1 - AI Oversight and the Duty of Care
 - PILLAR 2 - Ethics, Risk and Responsible Stewardship
 - PILLAR 3 - Oversight of Strategy and Adaptability
 - PILLAR 4 - Holistic Situational Awareness
 - PILLAR 5 - Talent, Incentives and Culture
4. Each pillar contains the following sections:
 - (a) Athena's Point of View (POV)
 - (b) Considerations and Recommendations
 - (c) Key Actions
 - (d) Questions Board Members Should Ask Management
5. Conclusion and Acknowledgments
6. Appendix and Additional Resources

THE FIVE PILLARS OF ORGANIZATIONAL AI GOVERNANCE

Outlined below are the Five Pillars of Organizational AI Governance, applicable across various entities – including businesses, societal organizations, and government bodies. These pillars primarily focus on the board's role but are equally vital for management to understand and embed throughout their organizations.

- (1) AI Oversight and the Duty of Care
- (2) Ethics, Risk and Responsible Stewardship
- (3) Oversight of Strategy and Adaptability
- (4) Holistic Situational Awareness
- (5) Talent, Incentives and Culture



1. Pillar One - AI Oversight and the Duty of Care

a. *Athena's POV*

The work of the board has long been rooted in the concept of duty of care. At the most basic level, duty of care means a moral, ethical or legal obligation to ensure the safety or well-being of others. In the context of a board of directors, it refers to the level of competence and business judgment expected of a board member and requires directors to act in good faith and make informed decisions in service to the organization and its stakeholders.

The board's fundamental duty of care is, at its essence, a duty to be informed.

As with cyber and other risks rooted in technology, the duty of care extends well beyond the tech stack to include regulatory risk, people and culture, brand health, M&A, the geopolitical landscape, stakeholder considerations and strategic growth. The board should ensure it is fit for purpose to provide effective oversight of AI through this multi-faceted lens. It reinforces the power of a commitment to trust-building, transparency, regulatory adherence, and future-readiness as it relates to the company's technology footprint.

To this end, the board should work with management to create an AI governance framework. Together they should consider critical aspects including accountability and risk assessments, cybersecurity, data management, fairness and ethics, human capital oversight, privacy, transparency and securing trust of customers and other key stakeholders. These considerations



are critical for every organization to understand, but we acknowledge the capacity to respond may differ based on organization size and budget. The recommendations and questions laid out here provide a guidepost for all boards, who can then determine how best to prioritize and resource them based on individual circumstances.

While the duty of care is expressly called out as a legal responsibility of the board, it is equally applicable to the senior leadership of the organization. It reinforces the importance of a continuous and ongoing conversation between the board and management about AI strategies, focusing on ethical and legal implications, third-party collaborations, and long-term value creation. This dialogue should be informed by continuous learning and staying up to date at an oversight level of AI advancements and their broader business and societal impacts. Company leadership should have a firm grasp on the evolving landscape of technology risks, such as privacy and cyber considerations, AI-related biases, disinformation, and data provenance and quality. They should engage the board as active partners in building a forward-thinking perspective.

b. Considerations and Recommendations

With the rapid development and accessibility of AI, this foundational pillar of “care” is vital not merely as a principle but also as a potentially existential imperative for companies and the stakeholders and communities they impact. Many of the ways companies may explore AI use are pre-product-market-fit, and where there is no set playbook or standards, and ecosystem impact are not yet clear. For example, when using AI to increase productivity, initially this might mean saving time to do the same task, whereas later could impact headcount planning. Even further down the road, it could impact entire job categories and even industry sectors. Emerging technology regulation often lags, and compliance today doesn’t guarantee compliance tomorrow. Every company will need to build their own POV on the bar they want to meet, and if that’s above and beyond current regulatory regimes. A long-term focus allows space to test, builds tolerance for ambiguity, and gives a clearer picture of risks and tradeoffs.

The duty of care is frequently met by asking thoughtful and relevant questions that get to the heart of core issues related to strategy and risk. This is especially true in the context of AI. Care in this context is expressed by balancing innovation with integrity, foresight with ethics and growth with sustainability – whether in the products companies develop, the technologies employees use or the overall culture. Care also means being ruthlessly curious about the latest technological developments, analyzing their potential implications and staying aware of changes in laws and regulations.

Board members should consider the wide-ranging effects of AI on all stakeholders. This includes employees whose work will likely be transformed by AI; customers, who trust in the safety and fairness of AI-driven products and services; and suppliers, who might be required to adapt to new AI-integrated systems. It also encompasses the wider community that could be affected by the company's use of AI, from local economies to global markets. In overseeing AI strategies, board members must balance the promise of these technologies with the expectations and rights of these stakeholders.

Directors must strive for a reasonable level of understanding of the technological resources available such that they can ensure ethical constructs are in place and guide implementation to meet emerging and sustained business and strategic needs. Agility is the name of the game and constant adaptation is critical.

c. *Key Actions*

We offer the following recommendations for directors:

- *Adopt a Learner's Mindset by Engaging in Foundational and Continuous AI Education:*
This could include courses on key AI topics, including applications, implications, and ethical considerations. Follow blogs and articles by thought leaders who are knowledgeable about AI and how it impacts your industry, adjacent industries, and the world more broadly. Regular updates through courses, live demonstrations of AI tools and practical examples of internal AI use cases will help in understanding this complex and evolving field and how it applies to your organization. Engage with experts who have in-depth knowledge about specific areas of AI, staying abreast of emerging capabilities, trends, and urgent news. This education should not be a one-time endeavor but a continuous process, enabling directors to act on a fully informed basis, understanding both the opportunities and the challenges posed by AI.
- *Engage in Strategic Discussions on AI Implementation:*
Initiate comprehensive discussions with the executive team about the current and future state of AI within the organization and among competitors. Understand the strategies for AI implementation, including appropriate human-machine teaming to leverage the respective strengths of each. Identify how the board can support achieving these strategic goals. Actively engage the management team as a thought partner in considering both the opportunities and the risks to the organization.
- *Foster Collective Intelligence Through "Divide and Conquer:"*
Map AI to the work of the key committees and assign specific AI-related areas of focus to each. These committees, both standing and ad hoc, should reflect AI's dynamic nature in their scope, and their charters should reflect this expanded purview. For example, the compensation committee might evaluate AI's impact on employee compensation structures, while the audit committee could examine its implications for financial and technological risk management. The nominating and governance committee should enhance governance structures and board succession planning to meet the moment. Committees focused on CEO succession planning and ESG concerns should also incorporate AI considerations into discussions of their priority items. Regular reviews of AI-related risks should be a staple item on the board's agenda, ensuring the board's approach evolves alongside AI technology. This collaborative approach ensures holistic oversight of AI governance and allows the board to leverage diverse perspectives in decision-making. If the board does not already have a standing technology or risk committee, it should consider establishing a dedicated subcommittee for AI oversight.

- *Align AI with Ethical and Strategic Objectives:*
Ensure that the adoption and implementation of AI technologies aligns with the company's values of corporate integrity, stakeholder interests and long-term value creation. Consider how these map to trust-building, transparency, company culture, brand safety, regulatory compliance and creating a sustainable, future-ready organization.
- *Board Evaluation and Succession Planning:*
Align the board's skillsets matrix with the technology strategy inclusive of AI. Update board evaluations to include assessment of board education and engagement around emerging trends, including AI.

d. Questions Board Members Should Ask Management

1. How does the board stay abreast of the latest AI technological advancements and their potential impact on our industry? How are we gathering, evaluating, and monitoring best practices in a rapidly evolving and dynamic environment?
2. Do we have adequate and diverse representation of technical and legal expertise, both within the board and accessible externally, to oversee complex AI initiatives effectively and compliantly?
3. Is our board skillsets matrix aligned with advances in technology? What capabilities should be added to the board and when? How often does this need to be reevaluated/updated?
4. Are we as a board focusing an appropriate level of time and attention on matters related to AI and related technologies? If not, how will we shift priorities?
5. Do we understand how AI oversight of both strategy and risk maps to each of the key committees? Have the charters been updated to reflect any change to the committee purpose, objectives, authority, and processes?
6. Are we sufficiently educated on the company's AI systems and governance to provide effective oversight, and if not, what steps do we need to take?
7. What kind of board education program makes most sense to ensure a baseline of fluency in issues of AI and related tech as they relate to our industry and business model?
8. Does our annual strategy offsite make space to consider balancing the trade-offs between innovation and rigorous risk management measures?



2. Pillar Two – Ethics, Risk and Responsible Stewardship

a. *Athena's POV*

The rise of GenAI and related technologies create a whole host of enhanced risks and ethical considerations that must be understood and managed with intention. This requires a thoughtful, integrated, and multi-disciplinary approach to AI governance.

The board should apply (for itself) and demand (from management) high ethical standards, responsible stewardship and compliance in the development, application and implementation of AI and related technologies.

This principle serves as a guiding force for responsible innovation and stewardship, ensuring that technological advancements align with the company's ethical commitments and regulatory obligations via policies and procedures. There are critical considerations about bias that, if not approached with intention, can supercharge blind spots that at best can hamstring strategy and at worst can do real harm. Design, data collection and deployment must be conducted thoughtfully and include a diversity of demographics and perspectives.

Boards should ensure that AI considerations are properly integrated into a company's governance, strategy, and product development via an appropriate system of enterprise risk management (ERM). The risks of introducing AI into products and processes - intellectual property (IP) leakage, loss of privacy or mishandling of sensitive data, brand or reputation damage and a dynamic regulatory environment - aren't new. But the power, potential and ubiquitous integration of AI across our work and personal lives raises the stakes. Boards and leadership should assume it increases the likelihood of such risks and work to create a culture of responsible AI use across the organization.

Finally, it is essential that the organization works with trustworthy partners within their ecosystem. The board's role in shaping the AI vendor selection policy is critical. This comes down to establishing and overseeing guidelines that ensure vendors align with the company's ethical standards, data management practices and privacy norms. This strategic oversight is essential to ensure that AI integration is not only technologically sound but also ethically grounded and legally compliant.

b. Considerations and Recommendations

The board should ask for frameworks that capture a full spectrum of AI applications within the company, ensuring that each application aligns with both ethical norms and business objectives. This includes how AI is integrated into various domains, such as compliance, HR, product development, customer service, finance, legal and marketing. Oversight should also extend to third parties to ensure their practices align with corporate policy, data management, model monitoring and code governance.

The board must ensure that data provenance, model design, training and implementation, and AI applications are governed by principles that prioritize fairness, transparency, and respect for privacy. They should ensure the company has a keen understanding and line of sight to data provenance. For example, US federal courts have held that artwork generated by an AI is not eligible for copyright protection³. The same analysis would likely apply to code generated by AI. It is possible then that proprietary code that would typically be subject to copyright protection would not be afforded this protection if the code is generated by a public-facing AI. At the same time, the company must ensure it safeguards its own intellectual property, especially given the substantial investment required for AI technologies.

In addition, the board should consider how the company will handle disclosures about AI-related issues whether mandatory or voluntary. This includes proxy statements, earning calls, investor updates, annual reports, company websites and social media. As regulator and other stakeholder scrutiny of AI continues to grow, the board should inquire as to whether the company's statements about use, risks, opportunities, etc. are consistent across all of these various channels. Crisis scenario planning should include tech risk scenarios and clearly articulate the interconnectedness of risk mitigation, crisis management, business continuity and data protection plans. Key considerations should include unintended consequences, false positives, implicit bias in training data, assumptions, diversity (or lack thereof) in AI developers, lack of regulation, inadequate regulatory framework, data provenance and copyright challenges, lack of common standards and oversight.

The core of responsible organizational AI stewardship, as well as compliance with all applicable regulatory and legal requirements, begins with a holistic, lifecycle approach to the integration, existence, deployment and usage of AI (and its inputs – especially, data) within its systems, products and policies. Boards can pressure test for adequate oversight of AI by asking how it is being deployed in the organization, what data is being used and how it is being applied. The board should understand at a high level the potential risks at each phase of any process that employs AI from design, data collection and employment to product life cycle management and end of life. The appendix provides a suggested approach to breaking down this lifecycle into its component parts and phases that can be a guide for asking questions that get to the heart of AI stewardship.

³ [THALER v. PERLMUTTER et al, No. 1:2022cv01564 - Document 24 \(D.D.C. 2023\) :: Justia](#)

c. *Key Actions*

We offer the following recommendations for directors:

- *Assess and Prioritize Risks Based on the AI System Life Cycle:*
Apply an adaptive risk management framework that prioritizes high-impact risks within the specific context. Do not treat all AI risks the same.
- *Approach Evolving Regulations Proactively:*
Engage legal or regulatory expertise to actively stay abreast of and ensure compliance with all relevant laws and regulations for all of the countries and jurisdictions in which the company does business and where employees and workers reside. This approach should include a proactive eye on how global regulations about AI might evolve, especially when considering expansion into new markets or the development of new product offerings. This approach is vital to maintain legal compliance and ensure the ethical and responsible deployment of AI technologies in a globally diverse environment.
- *Consider Partners and Other Third Parties:*
Create a framework for evaluating M&A, partnerships, and integration opportunities. This should extend across the supply chain and network of partners, ensuring that AI ethics and compliance are deeply integrated at every level. Apply special scrutiny to third-party AI tools used by the company and ensure they adhere to the highest legal, privacy and ethical standards. Consider regular audits and risk assessments of these tools to maintain compliance and align with the company's code of ethics. This includes formulating policies against the use of AI in creating deceptive content or perpetuating biases, those that ensure inclusivity and diversity in AI training datasets and align AI applications with the company's core values and mission.
- *Monitor AI Continuously:*
Ensure a protocol for continuous monitoring of AI usage, including how to address and learn from failures and unintended consequences.
- *Bring the Best of Both Humans and AI:*
Make explicit when human oversight is required and will be used. For example, the board should ask management to establish policies for human oversight of the data used for training AI models, including when and how corrective actions will be taken as needed.
- *Foster Stakeholder Trust:*
Ensure proactive communication and education on data use policies and procedures, including will or will not be used for AI training and adoption.

d. Questions Board Members Should Ask Management

1. Are we aiming to simply meet industry-specific ethical standards, or do we aspire to exceed them and be an industry leader? What role does stakeholder feedback play in refining our compliance strategies?
2. How does management ensure that the team entrusted with AI model creation, training, and evaluation has adhered to appropriate and relevant legal and ethical standards and industry best practices? What metrics and processes are in place to ensure compliance with our AI governance policies?
3. Have we appropriately coupled ethics, AI and broader technology issues with ERM? For example, what measures are in place to ensure AI model outcomes do not inadvertently violate existing laws, including those related to labor, anti-discrimination, data privacy, etc.?
4. What is our process for monitoring the development and implementation of AI tools? For example, how do we identify when AI outputs shift from the expected to unintended or outlier cases (e.g. "model drift")? Once a product or service has been deployed, how does management ensure the continuous monitoring of ethical, legal compliance and quality standards? Do we have tools and policies in place to detect malicious or fraudulent use of AI?
5. In cases where we rely on third-party AI tools, how do we ensure compliance with legal, privacy and ethical standards? How frequently is the use of these tools audited?
6. Can management articulate their approach to appointing and properly resourcing a high- functioning interdisciplinary operational team to conduct continuous ERM that includes constant evaluation of technological risk, including emerging and frontier risk?
7. What personnel and functions are represented in the initial AI design and development, build or buy decision? Is anyone from legal, ethics, compliance, privacy, and quality present together with engineers, data scientists, enterprise architects and software developers? Does the team have sufficient diversity to help mitigate bias and blind spots? How does design and development take into consideration the perspective of the end user?
8. Do we have robust data policies ensuring defensible intellectual and privacy rights, licensing and permissions for data used in AI models and applications? Do we understand the provenance of the data used to build and feed AI models? How do we delineate between human-created IP, which is protected by copyright, and the AI-created portions that may not be protected under current copyright laws?
9. For publicly listed companies, do risk factor disclosures in the 10K or other required disclosures accurately reflect AI risk? Are our voluntary disclosures – whether on the website, in the proxy statement in CEO/leadership statements in earnings calls, etc. – consistent across channels and crafted with an eye toward regulatory and stakeholder scrutiny?



10. Do we anticipate increased cyber vulnerabilities because of AI adoption? Are our existing cyber security frameworks and protocols sufficiently agile to take into consideration the implications of and enhanced risks posed by AI and other emerging technology? For example, is the vetting of third-party software suppliers sufficiently rigorous to ensure proper cyber risk management?
11. What is our strategy for crisis management in the context of AI? Have we anticipated the public relations aspects of our AI initiatives, including unintended outcomes such as data misuse or data leaks? Have we considered red teaming exercises to identify potential unintended consequences?
12. What measures are in place to identify and manage potential conflicts of interest in our AI initiatives?
13. Do we have robust procedures for addressing and resolving conflicts of interest and vendor associations, especially in the context of AI?

3. Pillar Three - Oversight of Strategy and Adaptability

a. *Athena's POV*

The board should monitor the effectiveness of the company's AI practices as they relate to enabling strategy and make changes as needed that reflect and anticipate the dynamic environment. An adaptable approach is vital to stay aligned with both technological advancements and the competitive landscape, to ensure the company's AI initiatives remain effective and responsible. For example, can we differentiate between what AI applications are table stakes or operational versus those that strategically fuel growth? Have we considered what level of accuracy is acceptable for a product release or use of third-party tools?



3 Oversight of Strategy and Adaptability

At the heart of this pillar is the balance between innovation and responsibility.

Have we considered what level of accuracy is acceptable for a product release or use of third-party tools? This impacts competitiveness as well as brand and credibility, so it should be discussed and weighed carefully. Additionally, the board should encourage and facilitate cross-functional collaboration in AI strategy formulation. Data silos are antithetical to adequately managing complex, interconnected risks and to harnessing potential opportunities. By integrating AI initiatives across business functions, the company ensures a unified and effective approach to AI, aligning technological innovation with the company's overarching goals and values. For example, a board may consider organizing questions to management using a matrix of internal vs. external AI usage overlaid with the various functions of the organization.

b. *Considerations and Recommendations*

Navigating the challenges of data access in vendor systems, especially in the face of disruptions such as database intrusions or ransom demands, is another critical area. The board should guide management in developing strategies to manage these risks, including setting appropriate contractual obligations for data and IP protection with vendors. This strategy is essential for ensuring the integrity and security of data used in AI systems, because "AI is not a product nor a 'one-off' service, but a system delivered dynamically through multiple hands."⁴

⁴ [Expert Opinion: Regulating AI in Europe](#)



Identify areas of experimentation with AI in a controlled environment (with specific scope), keeping in mind risk mitigation and ensuring safe exploration. If areas in the business are intending to use AI to disrupt the market, they should consider scope, environmental factors, and the decision-making framework, including the level of human oversight vs. autonomy.

The board should foster an environment where cross-functional stakeholders and diverse community representatives are involved in AI strategy and risk assessment. Evaluate AI applications from various perspectives, aligning with global standards of fairness and inclusivity. This fosters a more effective and holistic strategic lens that can identify both gaps and opportunities.

By proactively adapting governance practices to the evolving AI landscape, the board ensures the company is prepared for future advancements, striking a balance between innovation, ethical stewardship and social responsibility.

c. Key Actions

We offer the following recommendations for directors:

- *Align AI Strategy to Broader Business Strategy:*
Ensure the AI strategy follows from the business strategy, including people, operations and social impact strategies. Consider how responsible innovations through use of AI might boost brand value and competitive standing.
- *Envision Possible Futures to Enable Strategy:*
Consider a range of potential future states to help inform and shape an AI strategy tied to the business strategy. What might happen if competitors move faster? How might factors like geographic expansion or new product development impact growth strategies? If AI transforms many roles significantly and productivity increases exponentially for some tasks, for example as Goldman Sachs predicts,⁵ what possibilities might that create for long-term value creation?
- *Look to the Past to Help Inform the Future:*
Considering how the company has previously adapted to transformative moments can provide a helpful lens to inform strategy. How did your business respond to the dawn of the internet age? Or the rise of cloud computing? Where did you adapt appropriately? Looking back, what would you have done differently? Faster? Slower? What opportunities did you capture or miss?

⁵<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>

- *Identify Areas of Experimentation with AI in a Controlled Environment:*
If your organization has no prior experience with AI, it can be helpful to start small with a few low-risk use cases, learn from them and then plan accordingly for the larger, more complex and likely more value-add opportunities. An example of a low-risk use case with safe exploration might be having marketing teams test different GenAI solutions, with human review and refinement to outputs and measure productivity and message effectiveness.
- *Foster Stakeholder Engagement and Global AI Alignment:*
Establish platforms for dialogue with diverse stakeholders and align the company's AI strategies with international governance standards. Multinational and global organizations should consider how best to balance opportunity and complexity if, for example, different countries require different AI solutions to align to the business strategy, regulatory environment, governance standards and culturally acceptable applications of AI.
- *Encourage Cross-Functional Collaboration in AI Strategy Development:*
Promote interdisciplinary collaboration to ensure holistic integration of AI across the company's strategies and operations. This can help ensure the strategy crafted is more robust and considers opportunities and constraints across business units and products, IT systems, regulation and compliance, data availability, talent bench strength, customer needs and norms, and other key areas applicable to your organization.
- *Refine and Adjust as the Future Unfolds:*
The pace of AI and other exponential technologies indicate the need to provide clear and continual oversight of strategy. It is important to remember Amara's law⁶, that we are likely to underestimate the impact of AI on a five or 10-year horizon yet also overestimate what will happen in the next 12 months. Having a defined approach to monitoring assumptions and outcomes of the business strategy can help ensure appropriate oversight and adaptability to changing circumstances.

⁶ <https://www.pcmag.com/encyclopedia/term/amaras-law>

d. Questions Board Members Should Ask Management

1. How do we measure the effectiveness and efficiency of our AI applications against our business goals?
2. What is our long-term strategic vision for AI, and how are we preparing for future AI advancements? Is this being clearly communicated to employees? What about other key stakeholders?
3. Where do we want to be on the continuum of AI adoption? Do we want to be an industry leader or simply ensure we stay current and in compliance? How will we measure our success against our aspirations?
4. How are we leveraging AI to drive innovation and enhance our competitive positioning in the market?
5. What are the intended uses of AI at the company? Are these prioritized and funded in alignment with the company's objectives?
6. What are competitors and closely related organizations, or peer groups, doing with AI and does our strategy reflect our competitive aspirations and market standing?
7. What are our short, medium, and long-term objectives with regards to AI and related technologies? Where will that put us in the market? What KPIs will we measure to know if we are not moving fast enough?
8. How are we ensuring that our AI governance practices are adapted over time to be in line with global best practices and industry standards?
9. Have we considered how geopolitical factors might impact supply chains and therefore our strategy? For example, have we considered availability of chips and other key components or computing power availability, talent (whether employees or offshore and nearshore resources) and relationships with vendors and partners?

4. Pillar Four – Holistic Situational Awareness

a. Athena's POV



Board members must recognize that the technologies they oversee, particularly AI, have short and long-term social and environmental footprints and implications. They should apply a keen, inclusive, and strategic lens to the review and assessment of enterprise-wide risk management to include environmental, social, governance and technology issues, risks and opportunities. Exponential technologies present their own challenges around environmental and social issues. Bias can influence training models and thus supercharge inequities; Data and computing power create environmental considerations around resource consumption.

Where discrepancies arise between policies governing technology and the company's values, directors should actively seek to realign them. This mindset is foundational to a long-term value creation approach to AI governance. In embracing these responsibilities, the board does not just govern AI within the company; it also shapes the wider societal impact of these technologies.

b. Considerations and Recommendations

The board should ensure the company's ERM system includes a comprehensive and inclusive approach to identifying and understanding all relevant risks and opportunities impacting key stakeholders and longer-term value creation. This means:

- a. The company should understand its core, material technology issues and opportunities deeply, both tactically and strategically, so it can properly assess its universe of technology risks and opportunities.
- b. The company should understand the intersection of its core tech issues, risks and opportunities with other aspects of sustainability and social issues. For example, what does a more data-intensive business model mean for its carbon footprint, and is it acceptable from an environmental standpoint? Does this scrutiny extend to other parts of the supply chain? In the privacy realm, how much say do the subjects of data

collection have over how their data is shared and treated? What is the company doing to ensure that the data it is using or generating for its products or services is analyzed for provenance, quality, balance and lack of bias and discrimination?

- c. Management should identify and properly resource an appropriate cross-functional/operational team to conduct continuous evaluation of risks related to environmental and social considerations as well as emerging and frontier tech risk most relevant to the business.

c. *Key Actions*

We offer the following recommendations for directors:

- *Integrate ESG, AI and ERM:*
The board must ensure that management is fully considering and integrating these issues into their ERM system together with their other risks and opportunities related to sustainability and social issues.
- *Pressure Test Risk and Ethics:*
The board must question management on how it approaches the provenance, equity, integrity, fairness and safety of the AI it develops and/or deploys with a focus on overseeing risk, ethics and impact
- *Consider Environmental Impacts:*
The board, in partnership with management, must ensure that AI applications are developed and utilized in ways that are sustainable and contribute positively to the company's environmental goals. This means considering the energy consumption of data centers, the lifecycle of AI technologies and the potential for AI to contribute to more efficient operations and reduced waste. Effective oversight is not just about mitigating harm but leveraging AI to actively foster environmental stewardship, thereby aligning technological progress with the urgent need for ecological preservation and sustainability.
- *Keep People Impacts Central to Strategy:*
Some of the critical social issues in ESG regarding AI that boards and management should consider include (a) workplace and talent impacts of the deployment of AI for assistance in hiring, promotion, firing and diversity, equity, and inclusion (DEI) decisions; (b) deployment of AI in identifying, understanding and addressing human rights issues in the company's supply chain; (c) ensuring strategies are in place to ensure AI systems are accessible and inclusive to a diverse user base; (d) ensuring the product life cycle from design to implementation is created with an eye toward mitigating bias; (e) weighing and taking precautions to prioritize privacy concerns of stakeholders.



d. Questions Board Members Should Ask Management

1. What steps are we taking to foster a culture of ethical, responsible and safe AI use within our organization? How does our current and future AI strategy align with corporate social responsibility commitments?
2. What are the systems and data sources most material to our business and how do they relate to our risks, opportunities, and stakeholder impacts?
3. Does company management understand the intersection of its core tech issues, risks and opportunities with key stakeholder concerns to include environmental and social considerations? How is our AI and related technology strategy aligned with our strategies concerning social issues and sustainability?
4. How are we assessing and managing the impact of our AI applications across different stakeholders? Are we taking special care to consider impacts to historically marginalized groups?
5. How do we ensure the appropriate level of transparency and accountability regarding AI algorithms, especially for sensitive decisions like hiring or law enforcement?
6. How does the tech risk profile intersect with climate impact?
7. Have we considered the risk of technology introducing bias into a product, process, or system?
8. How do we vet AI vendors for ethical data sourcing and responsible AI solutions?
9. What processes and established frameworks are in place to uphold corporate responsibility and integrate lessons learned into the development of future products and services

5. Pillar Five - Talent, Incentives and Culture

a. Athena's POV



Much of the focus on AI has been on technology. But the application of AI, and thus its impacts, are ultimately driven by the people, not machines. Therefore, it is critical that the company not simply align talent strategy with its broader AI and technology systems strategy, but ideally create them in tandem. Then there is the question of culture. If people are a critical factor enabling AI within any organization, culture is what focuses them on the right metrics, behaviors, and outcomes.

Culture is a key driver of both short-term and long-term value creation. People are at the heart of AI application and adoption. To harness AI capabilities to drive growth and realize their mission, vision and values, companies must focus deeply and intentionally on both.

The board should have a clear understanding of how talent strategy holistically ensures the company is fit for purpose in rapidly evolving times. Implications of AI and exponential technology should factor heavily when evaluating the CEO and senior leadership and when considering succession planning. Key to success is a regular review of the organization's talent strategy in the context of advances in AI. Across the employee lifecycle from recruiting, screening, and hiring to learning and development, performance management and DEI, the company should consider how AI and talent intersect. It is imperative that organizations take a multi-stakeholder approach to managing tech risk, reflecting both builders and users. Put simply: responsible use, application, data validation and deployment of AI is the responsibility of every employee, and company culture should reflect this.

Compensation should reflect this new reality by appropriately balancing the potential risks and opportunities of AI and related tech. We encourage transparency on how compensation, accountability and ethics align as they relate to AI and associated technologies. Shareholders and key stakeholders should understand the commitments a company has made to the responsible use of AI and related tech and understand how compensation and accountability align to serve these ends.



Additional cross-checks will help to ensure alignment with the overall privacy and security strategy of the business as well as the organization's values and social impact strategy. These should be reflected in how employees and key leadership are evaluated and compensated.

b. Considerations and Recommendations

Ensure that management has conducted a holistic review of AI's impact on talent throughout the organization, aligned to the chosen AI strategy. Boards should ask to see detailed workforce planning strategies designed to ensure the company is prepared for the long-term implications of AI on the workforce, while assessing short-term talent needs.

For example, in the short term, employees may fear job loss leading to reduced participation in initiatives; Skepticism or even cultural norms such as valuing deep expertise and high-touch interactions may deepen resistance to change. In the long term, organizations will need to rethink how to train and prepare workers differently for increasing responsibility if historical approaches such as leaning on low-level repetitive tasks are transformed or eliminated by AI.

Consider how AI and talent intersect across the employee lifecycle from recruiting, screening, and hiring to learning and development, succession planning and performance management. Ensure the organization is competitive in terms of benefits, compensation, talent development and overall diversity of the team.

Workforce plans should take into consideration the evolving regulatory landscape for the countries and regions in which the company operates and where employees are based. The talent strategy should address how/when upskilling and reskilling may be needed, implications for short- and long-term workforce planning, include workforce training on risk management when using AI and consider new skills that will be needed in the future as AI use cases mature.

Fully enabling the talent within the organization is only part of the equation. Remember the oft cited and evergreen Peter Drucker quote "culture eats strategy for breakfast." The rapid evolution and potential competitive advantages of AI mean that talent within the organization must adapt quickly. All of this requires a thoughtful approach to company culture, one that centers agility, adaptability, feedback, curiosity, and ethics. There are several key factors the board and leadership should consider when assessing and addressing company culture in the context of AI:

Unsanctioned Uses

One of the most urgent areas of AI risk to be addressed is the unsanctioned use of AI tools by employees, executives, and board directors. Multiple studies⁷ have shown that a significant portion of employees use AI tools at work for work purposes, yet only a small fraction of companies provide any guidelines or guardrails on responsible and ethical use, unintentionally putting their organizations at risk.

⁷ <https://www.fishbowlapp.com/insights/70-percent-of-workers-using-chatgpt-at-work-are-not-telling-their-boss/>
<https://www.bcg.com/publications/2023/what-people-are-saying-about-ai-at-work>
<https://www.salesforce.com/news/stories/ai-at-work-research/>

Given the prevalence of employees using AI and Generative AI for work purposes, boards must move quickly to ensure their companies have policies in place for safe, transparent, and responsible use of these tools. Once policies are in place, ongoing education and awareness raising is key, ensuring employees understand the “why” behind a given policy. Employee misuse often comes down to lack of awareness – for example, an employee who wants to deliver the best outcome by using an AI-powered tool inadvertently or unwittingly putting sensitive company information into the public domain.

It is important to acknowledge that it is not practical to expect that an organization can block all unsanctioned use of AI tools, so companies must determine what guardrails to put in place, balancing productivity gains with organizational risk. Incident response teams should have specialized training, or outside subject matter experts they can turn to, to handle allegations of AI-related misuse.

Unintended Consequences

Despite the most rigorous planning processes, AI introduces the potential for unintended consequences including misuse of data, blind spots and miscalculations. Executives should be held accountable in cases where inappropriate risk mitigation practices are in place and unethical approaches are used, as defined in the organization's own policies or code of conduct.

c. *Key Actions*

We offer the following recommendations for directors:

- *Evaluate Talent Strategy Impacts:*
Once your AI strategy has been aligned to your business strategy, review your talent strategy. What additions or modifications should be made to the overall talent strategy, if any? What should be reprioritized? What should be adjusted? How will you start preparing your talent today?
- *Conduct a Talent Gap Assessment:*
Identify critical gaps in talent and skills that must be filled in the short term. Determine how you will fill those gaps, considering a multi-pronged approach that can include reskilling and upskilling, workforce augmentation, external speakers, and expert advisors. In the short term, consider embedding AI champions in the business or empowering those who are passionate about AI to take on projects that help fill immediate needs as a stretch assignment.



- *Examine Training From Multiple Angles:*
Consider whether AI education programs are inclusive and relevant to everyone, not just for people in traditional "tech" roles. For example, can the larger team broadly articulate how AI maps to organizational strategy? Have we integrated AI-related issues into larger training and compliance programs?
- *Solicit Employee Feedback:*
Consider adding questions about AI to traditional HR surveys that measure staff satisfaction and concerns. This tool will allow the board to measure how AI issues (both the vision and the challenges) get communicated throughout the organization.
- *Prepare For the Unexpected:*
Review company's code of conduct to expressly address AI ethics. Develop a process for how unanticipated AI outcomes will be handled, considering different types of potential consequences and accountability measures that will be applied in each case.

d. Questions Board Members Should Ask Management

Talent:

1. Who within the C-suite has primary ownership of the AI risks and opportunities related to talent? Are they capable and educated on this subject? If not, what is the plan to strengthen the capability of the executive team to handle this area? How will they stay abreast of evolving updates?
2. How is our DEI strategy aligned to our strategy for AI and related technologies?
3. How does the company's AI strategy intersect with critical talent functions, including onboarding, ongoing training, change management and fostering a culture of agility and ethics?
4. Who within the board has primary ownership for governance and oversight of how AI is used in connection with talent? What is the frequency of updates?
5. If some roles will be eliminated or their scope significantly diminished due to the introduction of AI, what will be the impact on adjacent roles?
6. If entry-level or other roles are expected to transform in significant ways due to AI, how will talent be trained and prepared for promotion opportunities?
7. Which executives are managing AI? Does the company have the right combination of technology, regulatory and risk management experts to responsibly develop and govern AI guidelines and procedures?
8. Have we aligned our C-suite talent strategy with the overall strategic plan and the impact of emerging trends to that plan? What existing or anticipated skill set gaps do we have on the leadership team, and how will they be addressed?
9. Has management structured and incentivized their cross-functional staff to work in an interconnected and proactive way on tech risk mitigation, crisis management, business continuity and data protection through scenario planning, lessons learned etc.?

Culture and Compensation:

10. How might AI shift individual relationships to work and the team? What broader workplace culture trends and shifts should we anticipate and plan for?
11. Are our security and ethics reporting functions, including whistleblower protocols, prepared and capable of handling AI-related incidents?
12. How is our AI strategy aligned to our broader business strategy and risk tolerance, and how is this reflected in our compensation policies?



13. How are we planning for and mitigating potential unintended consequences of our AI strategy? Does our compensation policy reflect this?
14. How are we providing transparency regarding accountability to pursue AI and related technologies in ethical ways, and how is this reflected in our compensation philosophy?
15. Do our compensation philosophy and policies reflect the intersection of AI and related technologies with our sustainability goals and commitments around social issues?

CONCLUSION

This Playbook represents a compelling call to action for every executive and board member. It emphasizes their indispensable role in steering organizations through this evolving AI era with foresight, integrity and a deep sense of responsibility. The five pillars of AI governance provide a robust framework for ethical decision-making, talent management and risk intelligence. As AI continues to reshape the business and social landscape, the Playbook underscores the critical role of each decision-maker's contribution toward achieving the best outcomes for their organizations and society at large.

ACKNOWLEDGMENTS

We would like to acknowledge and thank all the authors of this piece: Meghan Anzelc, Andrea Bonime-Blanc, Erin Essenmacher, Joyce Li, Heidi Lorenzen, Carrie Anderson, Cyd Brandvein, Stephanie Goutos, Shaara Roman, Avril Ussery Sisk and SuiLin Yap.

Special thanks to Meghan Anzelc, Andrea Bonime-Blanc, Erin Essenmacher, Joyce Li and Heidi Lorenzen for co-chairing our working group and leading this effort.

Additional thanks to those who have read and provided their wisdom, insights and feedback: Brenda Bernal, Robyn Bew, Coco Brown, Dd Budiharto, Rochelle Campbell, Catrina Cuevas, Karen Davis, Anjali Dooley, Traci Granston, Holli Harris, Diane Honda, Patti Humble, Christi Jakubik, Uma Jayaram, Sarika Khanna, Lisa McGill, Susan Paddock, Natalie Pierce, Katie Roberts, Alex Shulman-Peleg, Matthew Sekol, Michelle Tinsley and Anita Wang.

RECOMMENDED RESOURCES

Pillar 1 - AI Oversight and the Duty of Care

1. [Center for AI and Digital Policy \(CAIDP\)](#)
2. [Artificial intelligence | The Alan Turing Institute](#)
3. [Stanford University Human-Centered Artificial Intelligence \(HAI\)](#)
4. [Center for Humane Technology](#)
5. [MIT Computer Science and Artificial Intelligence Laboratory](#)
6. [AI Risk Management Framework | NIST](#)
7. [Fei Fei Li. The Worlds I See: Curiosity, Exploration and Discovery at the Dawn of AI. MacMillan 2023.](#)
8. [Joy Buloamwini. Unmasking AI: My Mission to Protect What is Human in a World of Machines. Random House 2023.](#)
9. [The women in AI making a difference. TechCrunch 2024.](#)

Pillar 2 - Ethics, Risk and Responsible Stewardship

1. [Artificial Intelligence and Algorithmic Fairness Initiative. US Equal Employment Opportunities Commission \(EEOC\).](#)
2. [Blueprint for an AI Bill of Rights. The White House Office of Science and Technology \(OSTP\).](#)
3. [Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring. DOJ Civil Rights Division.](#)
4. [The Ethics and Compliance Initiative | Ethics.org.](#)
5. [The Markkoula Center for Applied Ethics, UC Santa Barbara](#)
6. [The AI Alliance](#)
7. [Partnership on AI](#)
8. [Policy Perspectives on Artificial Intelligence | The Business Roundtable](#)
9. [Digital Trust Resources | World Economic Forum](#)

Pillar 3 - Oversight of Strategy and Adaptability

1. [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. The White House.](#)
2. [EU AI Act: first regulation on artificial intelligence](#)
3. [AI regulation around the world, from China to Brazil. The Washington Post 2023.](#)
4. [AI and the Role of the Board of Directors. Harvard Law School Forum on Corporate Governance 2023.](#)
5. [Artificial Intelligence Risk Management Framework. National Institute of Standards and Technology \(NIST\).](#)
6. [How Boards Can Accelerate their Learning Curve in AI and Cybersecurity. Spencer Stuart 2023.](#)
7. [MIT Computational Law](#)

Pillar 4 - Holistic Situational Awareness

1. [Cutting-Edge Technology: Continuous and Holistic Governance of Cloud and AI. NACD 2024.](#)
2. [Governing Exponential Technology: The Board's Role](#)
3. [The ESG Advocate \(Substack Newsletter\)](#)
4. [Sustainability, Finance and Technology: How Markets Rethink the Integration of ESG. Routledge 2022.](#)

Pillar 5 - Talent, Incentives and Culture

1. [How generative AI can boost highly skilled workers' productivity. MIT Sloan 2023.](#)
2. [Artificial intelligence: Digital humanities, business, & society | Nature](#)
3. [What do employers expect staff to know about AI? BBC 2024.](#)
4. [AI at Work: Building a Future Ready Workforce. Legal Business World 2023.](#)

APPENDIX

Supplement to Pillar 2 - Ethics, Risk and Responsible Stewardship:

There are potential risks to consider at each phase of the AI lifecycle, from inception to end of life management. A key consideration across the entire lifecycle is connected data silos cannot adequately address interconnected risks. Getting the company's data estate in order is a key competitive enabler in the market. A more detailed breakdown of key considerations across each phase follows.

Phase 1: Product Strategy and Design

Nearly all product strategy is – or will be – affected by AI, so the integration of ethics and compliance at the outset of design is essential. This includes “buy vs build” decisions and ensuring the right balance between return on investment (ROI), risk management and alignment to corporate strategy. Ensuring this integration requires a collaborative culture between several key functions including ethics, DEI, legal, privacy and quality control experts as well as engineers, data scientists and software developers. These teams must assess whether to develop AI solutions in house or procure them externally, weighing the ROI against ethical and compliance risks. Their combined expertise ensures that ethical considerations and compliance standards are deeply embedded in design strategies. Companies that do not have all of this expertise in-house may need to leverage outside expertise to complement and augment internal resources. This comprehensive approach is crucial in the decision-making process, ensuring that every development stage not only upholds the highest ethical and compliance standards, but also aligns with the company's strategic and financial objectives.

Phase 2: Data Collection & Preparation

The data collection and preparation stage should involve specialists responsible for overseeing data quality and data governance. These experts bring vital perspectives on privacy, ethics and compliance, which are essential for maintaining the integrity and ethical use of data. Their role is to ensure that data handling practices align with ethical standards and regulatory compliance, setting a strong foundation for responsible AI development.

Phase 3: Prompt Engineering, Model Training & Evaluation [Pre-Release]

Model training and prompt engineering should be done thoughtfully to ensure fairness, mitigate bias and make good faith efforts at anticipating potential unintended consequences. The aim is to develop AI models that are not only effective but also align with ethical guidelines and regulatory expectations. Team leads should pressure test whether models have been trained to equally weigh key demographic considerations and consider what other perspectives may be needed to mitigate bias in the system.

Phase 4: Model Deployment & Monitoring [Post-Release]

Once the model (product or service) has been deployed, embed continuous monitoring and feedback loops in the lifecycle review process for key areas including ethics, legal and regulatory compliance, product safety and production quality. This process yields critical data, lessons learned and can be a powerful risk mitigation tool. It also fosters a culture of continuous improvement, including the ability to identify and report problems, concerns, potential ethical,

legal, regulatory and other compliance issues. This phase should also include scenario planning, risk identification and mitigation and crisis management/incident planning.

Phase 5: Product Lifecycle Management & Transition Plan

Effective product lifecycle management hinges on continuous adherence to evolving legal and ethical standards. This also involves a strategic approach to managing user data and AI outputs, guided by a strong ethical framework. The focus is on ensuring that the evolution of the AI product is both responsible and sustainable, keeping in mind the long-term implications on the organization, users and society.

Phase 6: End of Life/End of Service

The end of life or service phase, particularly in cases of product obsolescence or defects, requires a strategic approach to data retention and model repurposing. This phase should include a comprehensive review of data retention practices from an ethical standpoint and explore responsible strategies for the decommissioning or transformation of the models. It should also include a review of impacted workflows and downstream tools to ensure a smooth transition when AI tools are removed. This is a crucial stage for learning from past experiences and applying these insights to future development of products and services.

Supplement to Pillar 5 - Talent, Incentives and Culture:

The topic of potential unintended consequences of AI is complex and nuanced. Below is provided as a more detailed breakdown of some of the unintended consequences, including those that have arisen historically.

There is the potential for unintended consequences, despite the most rigorous planning process. Historically, these have included the following:

- Misuse of AI outputs by users, either internal or external (e.g. customers), whether intentional or unintentional. This leads to outcomes that were unintended and potentially harmful depending on the context and application. Examples include Microsoft's chatbot Tay,¹⁰ Air Canada's loss in small claims court and¹¹ Chevrolet's lack of guardrails in their chatbot.¹² Higher-risk consequences include vulnerabilities in custom GPTs that can expose sensitive data¹³ and the use of GenAI for more sophisticated phishing attacks, including through traditional approaches like email¹⁴ and new approaches such as cutting-edge deepfakes using audio and video.¹⁵

¹⁰ <https://www.pcmag.com/news/microsoft-puts-tay-chatbot-in-time-out-after-racist-tweets>

¹¹ <https://www.forbes.com/sites/marisagarcia/2024/02/19/what-air-canada-lost-in-remarkable-lying-ai-chatbot-case/?sh=6d9c8a1a696f>

¹² <https://www.businessinsider.com/car-dealership-chevrolet-chatbot-chatgpt-pranks-chevy-2023-12?op=1>

¹³ <https://www.wired.com/story/openai-custom-chatbots-gpts-prompt-injection-attacks/>

¹⁴ <https://securityintelligence.com/x-force/ai-vs-human-deceit-unravelling-new-age-phishing-tactics/>

¹⁵ <https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>

- The past doesn't predict the future. Since AI tools are developed on existing data and generally presume future data will be similar, unexpected outcomes can occur when this relationship no longer holds. This can happen when there is a significant change to a business or to its customer base or behaviors. Many industries experienced this type of shock at the start of the Covid-19 pandemic, when many patterns of behavior changed quickly and dramatically.¹⁶ AI solutions may not have worked as expected under these new conditions, and companies without appropriate monitoring in place may have been caught off-guard.
- Negative impact on employee morale, organizational culture and overall competitiveness. In 2018, Google non-renewed a Pentagon contract after employee uproar over Project Maven,¹⁷ an initiative to use AI with drone imagery to identify military targets. In 2023, the CEO of a small Indian company made headlines¹⁸ when 90% of staff were replaced with AI. Competitive standing is a concern not just for companies but for countries as well,¹⁹ and those slow to adapt are at risk.
- Unexpected volume of usage creates unanticipated challenges. Scale and speed effects not considered in the design and implementation of AI tools can cause problems when those effects are larger than anticipated. Robinhood's challenges with GameStop stock trading in 2021 is one such example.²⁰
- AI tools are brittle in ways that are not well understood, causing cascading effects. Knight Capital Group's \$440M+ error from a software glitch²¹ and subsequent acquisition by Getco²² in 2012 may have happened over a decade ago, yet it is a prescient warning to companies of the necessity of monitoring and validation of outcomes at all stages of the process.

¹⁶<https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/>

¹⁷<https://www.nbcnews.com/news/military/google-halt-controversial-project-aiding-pentagon-drones-n879471>

¹⁸<https://www.businessinsider.com/ai-ecommerce-ceo-layoff-support-staff-copy-paste-jobs-unsafe-2023-10?op=1>

¹⁹ <https://www.axios.com/2023/07/27/us-china-squander-ai-lead>

²⁰ <https://www.cnbc.com/2021/02/17/robinhood-faces-lawsuits-after-gamestop-trading-halt.html>

²¹<https://archive.nytimes.com/dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/>

- Guardrails don't adequately consider edge cases or unusual circumstances. While GenAI tools for generating news have been around for over a decade,²³ companies must carefully consider the potential for stakeholder backlash. Vanderbilt University's public relations gaffe²⁴ is a clear example of an unintended consequence, and the firing of Sports Illustrated's CEO²⁵ after accusations of publishing AI-generated articles is an illustration of the fallout that can arise from a misstep. As companies integrate AI capabilities into more aspects of their business and more of their vendors and suppliers do the same, organizations must carefully consider the potential for costly errors and improve guidelines and guardrails when they do occur.

²² <https://www.cnbc.com/id/100325960>

²³ <https://www.npr.org/templates/story/story.php?storyId=122424166>

²⁴ <https://www.cnn.com/2023/02/22/tech/vanderbilt-chatgpt-shooting-email/index.html>

²⁵ <https://www.bbc.com/news/world-us-canada-67619015>