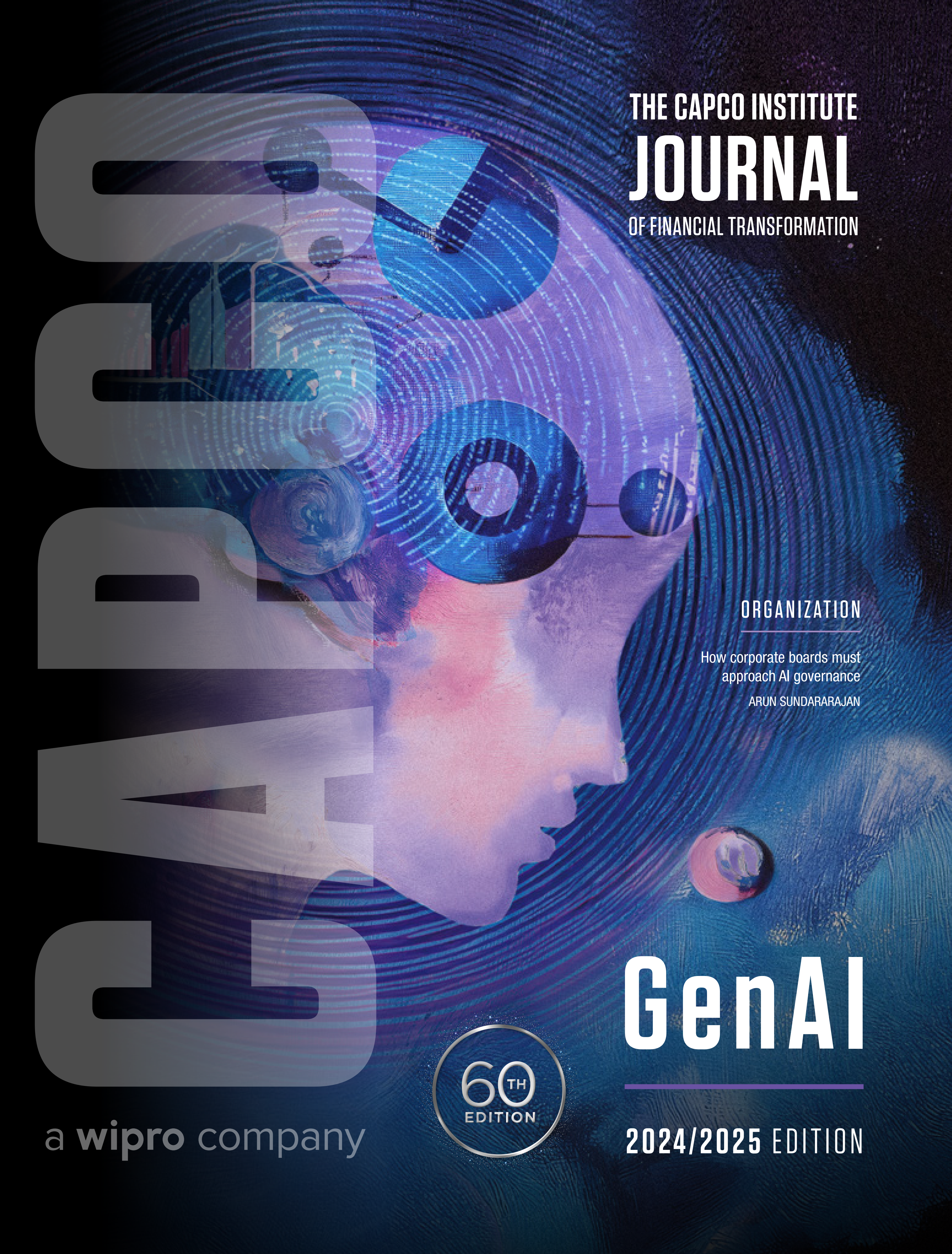


THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

ORGANIZATION

How corporate boards must approach AI governance

ARUN SUNDARARAJAN

GenAI

60TH EDITION

CAPCO

a wipro company

2024/2025 EDITION

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

TECHNOLOGY

ORGANIZATION

REGULATION

CAPCO CEO WELCOME

DEAR READER,

Welcome to our very special 60th edition of the Capco Journal of Financial Transformation.

The release of this milestone edition, focused on GenAI, reinforces Capco's enduring role in leading conversations at the cutting edge of innovation, and driving the trends shaping the financial services sector.

There is no doubt that GenAI is revolutionizing industries and rapidly accelerating innovation, with the potential to fundamentally reshape how we identify and capitalize on opportunities for transformation.

At Capco, we are embracing an AI infused future today, leveraging the power of GenAI to increase efficiency, innovation and speed to market while ensuring that this technology is used in a pragmatic, secure, and responsible way.

In this edition of the Capco Journal, we are excited to share the expert insights of distinguished contributors across academia and the financial services industry, in addition to drawing on the practical experiences from Capco's industry, consulting, and technology SMEs.

The authors in this edition offer fresh perspectives on the mindful use of GenAI and the implications of advanced GenAI on financial markets, in addition to providing practical and safe frameworks for boards and firms on how to approach GenAI governance.

The latest advancements in this rapidly evolving space demonstrate that the potential of GenAI goes beyond automating and augmenting tasks, to truly helping organizations redefine their business models, processes and workforce strategies. To unlock these benefits of GenAI, I believe that firms need a culture that encourages responsible experimentation and continuous learning across their organization, while assessing the impact of the potential benefits against a strategic approach and GenAI framework.

I am proud that Capco today remains committed to our culture of entrepreneurialism and innovation, harnessed in the foundation of our domain expertise across our global teams. I am proud that we remain committed to our mission to actively push boundaries, championing the ideas that are shaping the future of our industry, and making a genuine difference for our clients and customers – all while ensuring to lead with a strategy that puts sustained growth, integrity and security at the forefront of what we do.

I hope you'll find the articles in this edition both thought-provoking and valuable as you create your organization's GenAI strategy and future direction. As we navigate this journey together, now is the time to be bold, think big, and explore the possibilities.

My greatest thanks and appreciation to our contributors, readers, clients, and teams.

Anne-Marie Rowland

Annie Rowland, **Capco CEO**

HOW CORPORATE BOARDS MUST APPROACH AI GOVERNANCE

ARUN SUNDARARAJAN | Harold Price Professor of Entrepreneurship and Director of the Fubon Center for Technology, Business, and Innovation, Stern School of Business, New York University

ABSTRACT

As the landscape of artificial intelligence (AI) evolves rapidly, AI oversight by corporate boards is essential for managing AI exposure and complying with new AI laws. Competitive pressure to stay ahead in the AI race is inducing CEOs to embrace innovation aggressively, making board oversight especially critical. This paper presents a framework for corporate boards that identifies some key AI governance dimensions and provides guidelines for assessing their organizational risk and regulatory likelihood. The dual lenses of risk and regulation can simultaneously aid a board in prioritizing governance aspects to pay attention to and in choosing a robust oversight strategy. Mapping the risk-regulation matrix shapes appropriate recommended oversight strategies, ranging from proactive self-regulation and compliance monitoring to more passive wait-and-watch strategies. The paper further provides a structured way to navigate the evolving regulatory and governance landscape while unshackling boards from the subjectivity and imprecision of terms like "responsible" or "ethical" AI, leading to oversight that aligns with a company's unique risk profile and industry-specific regulatory context, while recognizing that AI governance touches a range of topics, from technology, intellectual property and sustainability to audit, measurement, and risk assessment.

1. INTRODUCTION: THE EVOLVING LANDSCAPE OF AI GOVERNANCE

The landscape of AI governance has become decidedly more multifaceted over the last two years. Before 2022, two issues – data privacy and algorithmic bias – were a primary focus of both internal corporate governance and government legislation efforts. Most saliently, the E.U.'s General Data Protection Regulation (GDPR) redefined consumer data protections globally, while also introducing key E.U.-specific requirements on algorithmic profiling, the transparency of algorithms, and the detection of potential biases in automated decision systems. GDPR inspired parallel legislation in countries ranging from the U.K. (the 2018 Data Protection Act) to Brazil (the 2020 LGPD[1]). In parallel, China's 2021 Personal Information Protection Laws required that the use of personal information in automated decision making does not lead to discriminatory treatment. While the ambitious Algorithmic Accountability Act in the U.S. is unlikely to become federal law, a growing number of state and local laws (including, for example, New York City's Local Law 144 – 2021) are mandating actions to mitigate algorithmic bias. Meanwhile, long-standing anti-discrimination laws in many countries have translated into requirements that machine learning systems not use "protected attributes" as features of training data, and there have been self-regulatory efforts by organizations worldwide to minimize the replication of these attributes from combinations of other training data features.

Generative AI (GenAI) has made this governance landscape substantially more complex. The inherent unpredictability of GenAI creates an array of issues of robustness: occasional "hallucinations" in AI output are now a reality that must be managed rather than an error that can be corrected, and generated content must align with organizational brand. Blurring lines between the quality of human- and AI-generated

[1] Lei Geral de Proteção de Dados Pessoais

content raises the question of whether an organization must be transparent about attributing machine-generated content. More broadly, the notion that one can aspire to make one's AI "transparent" is an increasingly utopian ideal in an era of large language models (LLMs) with trillions of parameters. There are new governance issues around appropriate training data for LLMs, from copyright infringement to the leakage of corporate intellectual property. The enormous energy needs of AI infrastructure challenge sustainability goals, while workforce displacement issues seem poised to take center-stage as the capabilities of AI become more human-like. Meanwhile, the challenges of fairness and privacy remain: the ascendance of GenAI has raised novel and subtle possibilities for unintended bias, while discussions around data privacy have become more nuanced, with separate attention needed to consumer data protection, training data governance, and preserving the intellectual autonomy of human workers.

Many excellent and current AI governance guidelines exist for governments and policymakers.[2] However, for a corporate board, navigating oversight in this multifaceted and evolving governance environment is a significant challenge. Some boards struggle to assess whether AI governance is a strategic role or a control role, and whether a dedicated new AI committee is necessary or if AI-related oversight can be subsumed by standing risk or audit committees. Broad subjective phrases like "responsible AI" and "ethical AI" induce lengthy discussions about the scope of what constitutes responsible or ethical behavior while compounding uncertainty about the connection of responsible AI to broader corporate social responsibility.

As this article will explain, breaking down AI governance into its specific dimensions can significantly enhance clarity, and assessing each of these dimensions through the dual lenses of risk and regulation can simultaneously aid a board in prioritizing them and in choosing a robust oversight strategy.

2. SOME KEY DIMENSIONS OF AI GOVERNANCE

The set of specific issues that might fall under the broad umbrella of AI governance is evolving. I discuss some of today's most salient dimensions in what follows. These are arranged in no particular order, and as I will explain later, there is no ranking of importance inherent in the order in which they are presented. Put differently, there is no absolute prioritization – relative importance is specific to an organization, and further, can emerge only from a process of assessing risk, reinterpreting existing laws in the AI context, and anticipating industry-specific regulation.

2.1 AI alignment

The use of AI implies a ceding, to varying extents, of autonomy in what the humans in an organization do. This makes it important to ensure that this autonomy does not lead to a divergence between organizational values, goals, or culture and the choices made by AI systems. A useful dichotomy is between "content alignment" and "decision alignment".

- **Content alignment:** involves ensuring that the generated "content" of an AI system is aligned with an organization's objectives or principles. For companies like Google or OpenAI that create general-purpose GenAI, this involves ensuring that AI output does not inadvertently create unacceptable content ranging from hate speech to prohibited topics. For most other companies that adapt these GenAI systems into business applications, content alignment will focus more on ensuring that the output of these applications, whether from a conversational AI system interacting with clients or a system being used to generate marketing content, is aligned with the brand and image of the organization.
- **Decision alignment:** involves ensuring that "decisions" that are delegated to an AI system are aligned with organizational goals. Such alignment has for many years been the focus of companies creating self-driving automobiles and have brought philosophical discussions like those of the "trolley problem" into mainstream business debates.[3] For most other companies, issues of decision alignment may be more frequent in lower stakes situations – for example, about the nature of decisions a customer service chatbot makes about product refunds or rebates when conversing with a customer, or decisions about recommendation/advertising targeting.

2.2 Intellectual property (IP) governance

To understand the most important IP governance issues related to AI, one must first recognize that the growing scale of AI systems leads to a build-versus-buy managerial assessment that is elevated to being a governance issue due to the proliferation of open-source models like the Llama LLM series released to the public by Meta (formerly Facebook) and a range

[2] https://tinyurl.com/mrtke9tu
[3] https://tinyurl.com/y965aen2

of models developed by academics and others available on repositories like Hugging Face. Choosing open source is cost-effective, allows greater IP control over customized systems, and places transparency choices more squarely in the hands of the organization. However, it can create quality control and security issues,[4] and can require in-house AI talent beyond the reach of many, impeding future progress for an organization not on the scientific cutting-edge of AI. In contrast, relying on a vendor like OpenAI or Google can be extremely expensive as an organization's AI use scales, can lead to opaqueness being a default rather than a choice, and, in some cases, may lead to lock-in that can constrain innovation and increase future cost uncertainty.

A deeper IP issue arises when one unpacks how shared GenAI technologies play a growing role in building AI applications for specific organizational uses. We are accustomed to AI systems being trained on structured sets of proprietary outcomes. However, large language models (LLMs) and other GenAI systems for images and video are trained on massive datasets that often encompass the entirety of humanity's available digitized content. For example, it is believed that OpenAI's GPT models are trained on all publicly available digital written content. Now consider the typical way in which most organizations will adapt a general-purpose system like LLMs for their specific purposes (for example, to create a customer service chatbot that understands the company's products, or an AI system for employees knowledgeable about the company's human resources policies and practices). One approach involves customizing an LLM developed by a company like OpenAI or Google using corporate specific knowledge (a process called "fine-tuning"). Although the AI systems that emerge from this process are proprietary to the company, corporate IP has, in a sense, been absorbed into the model's parameters. A different approach involves "augmenting" what is sent to a (non-proprietary) LLM with fragments of internal documents or past relevant conversations "retrieved" from an internal knowledge management system (a design often implemented using what is called "retrieval augmented generation" or RAG). Again, unless the company develops and hosts its own LLM, company knowhow is being sent (albeit in small chunks) to an external system. Whichever strategy a company chooses, the IP challenge is clear – this kind of tacit knowledge transfer requires careful oversight and thought.

2.3 Training data governance

The governance issues around training data that lead to the creation or use of a company's AI systems do not stop with the IP challenges discussed above. Oversight of the possible liabilities that a company may face on account of the training data used in its AI systems is also essential. Again, this is a multifaceted issue.

- An organization must determine the extent to which it is aware of all the data that may have been used to train the systems used by its AI applications. If using shared GenAI infrastructure like OpenAI's GPT or Google's Gemini, it must also consider whether to be prepared for regulatory demands that associated training data be made "transparent", either to a regulator or to the public.
- It is also almost certain that the training datasets of all LLMs and image generating systems have included "copyrighted" information used without the explicit permission of the copyright holders.[5] Although courts in the U.S. may eventually deem this use of copyrighted content "fair use", this is neither certain nor internationally applicable. Some countries like Singapore already have explicitly legislated the use of copyrighted information for AI model training, others like Australia have far more restrictive definitions of fair use than that of the U.S. The uncertainty and variance in how different countries will resolve the question of fair use makes this a key governance issue, since the direct liability associated with regulatory shocks could be significant. Even if an organization is not training its own LLMs, there may be substantial indirect costs if these shocks lead to unexpected changes in the availability or performance of the LLMs that one's AI systems depend on.
- Over time, organizations will increasingly use the output of their employees as training data for new or improved AI systems. For example, employees may be permitted to create "digital twins" that raise productivity by writing in their style or voice, draft contracts, or serve as chatbot substitutes when the employee is unavailable. Although this idea of a digital replica may seem like science fiction, it is increasingly feasible with today's AI technologies. An organization that is capturing and encoding the human capital of its workforce in AI systems must think through and implement a framework that regulates use, longevity, and value sharing from such systems.

[4] https://tinyurl.com/4drcjxjb
[5] https://tinyurl.com/yu7kadd9

2.4 Model explainability

Boards must often contend with the extent to which they insist that the AI systems their organization uses generate output whose logic can be explained. Over the last 20 years, artificial multi-layered neural networks (often called "deep learning" systems)[6] have become the favored model for building machine learning systems. Their superior performance comes with a hidden cost, because "explaining" the logic of their underlying statistical models is impossible. For example, an organization using a deep learning system for loan approval may be unable to explain why the system turned down a specific loan application. In contrast, a simpler underlying model based on logistic regression[7] that places weights on different features could allow an organization to explain that it was the income level or the credit score that led to the decision, but such an explainable system will almost certainly make less profitable decisions. This landscape is further complicated by GenAI systems, not just because their workings are not amenable to explainability, but because it is highly likely than any organization that is not a tech giant is instead reliant on systems built by companies such as OpenAI, Google, Anthropic or Meta, and is thus limited in its quest for explainability by the choices made by its AI vendors.

2.5 Model transparency

Independent of explainability, an organization may face internal or external pressure to make the workings of its AI systems "transparent". For example, in its early days, Uber faced pressure to make the details of its surge pricing algorithm visible to users and regulators. An insurance company using AI to price its products and set premiums may consider whether to explain the logic of this process to all its consumers. Similarly, an investment firm using AI to make trading decisions may face transparency pressure from regulators towards creating a system-wide view to assess contagion risks. Beyond the explainability-performance trade-off associated with neural networks, transparency can have competitive impacts as the performance of the AI systems becomes an increasingly important source of advantage. And again, the transparency options of an organization will be limited by the transparency choices made by its GenAI vendors like OpenAI, Google, Anthropic, or Meta.

2.6 AI robustness

AI has always been less predictable than its deterministically programmed counterparts. This is a natural consequence of the paradigm – a machine learning system that makes predictions based on a probabilistic statistical model will always have some associated unpredictability. There is no absolute way around this trade-off because a completely predictable machine learning system has little value – the unexpectedness of predictions and their departures from what human analysts may come up with is what makes them useful.

As the underlying statistical models have become larger and more complex, the associated unpredictability has grown. It is widely recognized that LLMs tend to "hallucinate", confidently providing information that is imagined and incorrect. Since LLMs generate new and original content through a process of successive next-word prediction,[8] these hallucinations will never be eliminated and must instead be managed. The governance of AI robustness thus involves balancing the trade-off between creativity and human-likeness on the one hand, and accuracy on the other, especially for AI systems that are customer-facing. Appropriate systems for conflict resolution and due process must be determined if, for example, a customer is provided with incorrect information about a refund or an interest rate by a customer service chatbot, or an employee makes vacation plans based on an outdated policy provided by an internal LLM-based human resources system.

A related dimension of robustness will relate to managing more subtle "traits" of underlying LLMs, especially in an environment where new versions are released with increasing frequency. These new versions typically report improved performance based on a variety of standardized benchmarks. Applications built on top of LLMs must then decide whether to take advantage of these improvements or stay with a tried-and-tested older model, a decision often taken without clarity about more subtle behavior changes that the transition may induce. Recent research[9] has shown, for example, that the Fall 2024 version of OpenAI's GPT4 (named o1), while outperforming its predecessor on most standardized metrics, demonstrates a dramatic drop in the human-like trusting behaviors that it displays. As LLMs form the basis for a growing number of high-stakes commercial systems, their increasing opacity and complexity can lead to hidden fault lines, adding another layer of complexity to the governance of AI robustness.

[6] https://tinyurl.com/mrxamrj7
[7] https://tinyurl.com/3k5vke35
[8] https://tinyurl.com/mvyejty6
[9] Li, Sedoc, and Sundararajan, unpublished.

2.7 Machine attribution

One of the most common uses of GenAI is to generate new written and visual content, from marketing and advertising material to customer communications. Video generating AI will soon be ubiquitous. Large-language models also excel at summarizing written content. Granted, tactical decisions about the right mix of human- and machine-generated materials may receive executive focus organically, but there is an associated governance choice of attribution – whether to reveal the AI versus human provenance of public-facing content, and if so, in what situations. It may seem natural to label an AI-generated video as having been AI-generated, but what about an AI-generated summary of user reviews, a marketing document that was generated with the aid of AI but with some human participation, or an advertising image that was human-created but hyper-personalized using AI? Insufficient machine attribution could lead to customer backlash, while excessive attribution could create the impression of inauthenticity.

Additionally, as AI agents take over larger fractions of synchronous and conversational customer interaction, a related machine attribution issue that requires clear governance is whether to inform a customer when they are interacting with an AI agent rather than a human. Today, most AI-driven customer interaction systems, from automated voice systems to website chatbots are easily recognizable as being non-human. As the human-machine lines blur in the coming years, many of these choices will be driven by government regulation, since this is an issue high on the legislative agenda, but boards must nevertheless proactively ensure that their organization makes choices on this front that are aligned with their brand and values.

2.8 Algorithmic bias and inclusivity

AI systems tend to reflect, or even amplify, the biases present in the data they are trained on. In simple terms, absent active intervention, biases that exist in society – whether related to gender, race, or socioeconomic status – can be inadvertently encoded into AI systems. For example, an AI-driven recruitment tool might favor candidates of a certain background because it was trained on historical hiring data that reflected existing inequalities. This issue has grown in prominence as AI has taken on greater decision making roles in areas like hiring, lending, and law enforcement.

Bias in AI systems is not a new issue. As machine learning proliferated in real-world settings, the potential to reproduce discriminatory outcomes has been widely recognized over the past decade. A variety of cases have received extensive media coverage, from predictive policing tools unfairly targeting certain communities and bail decision systems possibly displaying bias in denial to healthcare algorithms exhibiting racial biases in treatment recommendations.

With the emergence of GenAI, however, these challenges have taken on new dimensions. As discussed, LLMs create new content after being trained on large, diverse datasets. Their training data includes vast amounts of internet data, unmoderated content with a higher likelihood of biased views and dialog. Thus, the parameters in a GenAI model might reflect cultural stereotypes and gender biases that are subtle but eventually have widespread influences. It is difficult to isolate and address these biases by altering training datasets due to their enormity and opaqueness.

While a board might simply be tempted to ask that their GenAI be created in a way that aligns its "views" with the organization's values, LLMs operate in a way that makes it difficult to directly change their "beliefs". Unlike a human being, an LLM does not consciously hold beliefs; instead, it generates responses based on statistical associations derived from training data. As a result, when a generative model starts displaying biased behavior, there is no direct way to correct its underlying worldview. Instead, developers are forced to add increasingly complex sets of guardrails – specialized programs and machine learning systems that check output – to try to mitigate harmful or biased outputs. These guardrails involve varied techniques and policies that attempt to filter or guide the responses generated by the model. While these methods can be effective to some extent, they are not foolproof, and surrounding an AI system with an increasingly complex web of guardrails increases its fragility.

2.9 AI use and sustainability

AI consumes a growing fraction of the electricity of countries in which its hardware infrastructure is based. By some estimates, the power needs of AI in the U.S. will eventually exceed those of New York City, and AI data centers are projected to constitute close to 40% of the total increase in U.S. power demand by 2030.[10] For AI producers like OpenAI, Microsoft, Google, and Meta, this already creates a significant

[10] https://tinyurl.com/5dp39r5z

sustainability challenge. For example, since ChatGPT was released, Microsoft has scaled back and fallen short of its sustainability goals,[11] while aggressively seeking out alternative sources of sustainable power, including recently striking a deal to use the entire 837MW output of the fabled and recently recommissioned nuclear power plant at Three Mile Island in Pennsylvania.[12] Every organization must assume that their AI usage will grow dramatically in the coming years, and that each new generation of AI will be increasingly power-hungry. Examining the sustainability footprint of one's AI providers while balancing the quest for innovation with the organization's sustainability goals requires careful thought and oversight.

2.10 AI workforce displacement and transition planning

It is widely anticipated that changes in the mix of activities between machines and humans will cause a significant percentage of the workforce in the U.S., Western Europe, and Japan to transition to a new occupation over the coming two decades. Some estimates suggest that by 2030, one in 16 workers will need a new occupation due to AI workforce displacement.[13] Corporations must decide how proactive to be in supporting their employees to adapt, grow, and invest in their skills.

"Reskilling" is something that is seen as a cost driver today but may be central to a brand's identity in the future. A useful parallel comes from how corporate approaches to sustainability or responsible labor practices have evolved. A couple of decades ago, both were seen as part of corporate social responsibility, choices that drove up costs rather than profits. Today, people make consumption choices based on a brand's sustainability positioning and may shun companies with unfair labor practices. A decade from now, the same may be true about responsible workforce transition policies.

Educational funding from governments has traditionally focused on early-career development. One might argue that corporations are uniquely positioned to create opportunities for mid-career reskilling that align directly with their evolving needs. However, this requires more than just offering skill-based training programs. Just as universities prepare students for their first careers with a broad range of experiences beyond the classroom, corporations should build reskilling programs that go beyond mere technical training. These programs should include mentoring, career coaching, networking opportunities, and branded credentials. By providing these additional elements, corporations can help employees build confidence, develop professional networks, and explore new career paths.

A deeper governance issue relates to human intellectual autonomy.[14] Today's AI technologies hold the potential to decentralize access to a wide range of skills and productive capabilities, empowering millions to follow entrepreneurial pursuits while fostering the rise of a new generation of AI-driven professionals – from educators and healthcare providers to investment advisors and data scientists. As discussed briefly in Section 2.2, as AI systems within an organization progressively encapsulate the human capital of a workforce, if individuals cannot assert a level of ownership over their personal generative processes, talents, or expertise, we may face a future where intelligence and skills become overly commoditized and centralized. This could render humans unable to reap the economic rewards of their own human capital investments, reducing the benefits of AI to a select few rather than the broader population.

While this list of governance issues is lengthy, it is by no means exhaustive. For example, a board must consider how AI changes its existing governance approaches to cybersecurity and data privacy. And over time, new AI capabilities are bound to bring new governance challenges. Addressing them requires a nuanced assessment of organizational risk and a delicate balance between self-regulation and compliance. I unpack these points in greater detail in the following section.

3. AI GOVERNANCE OVERSIGHT: RISK AND REGULATION

Oversight of all these dimensions of AI governance can be a challenge for any board. To prioritize, each AI governance dimension should be evaluated through two critical lenses.

The first lens is the level of risk that the AI governance issues associated with a dimension might pose to one's specific organization. For example, there may be little or no actual risk posed to an organization that does not operate in the technology space if they choose not to make transparent the fact that they are using publicly available training datasets. In

[11] https://tinyurl.com/ym6zpm66
[12] https://tinyurl.com/3fu2674z
[13] https://tinyurl.com/mr2h7pvd
[14] https://tinyurl.com/mtufyu5v

contrast, leaking of key proprietary intellectual property due to flawed choices associated with letting a vendor fine-tune a version of their LLM to create a customer service chatbot could be quite damaging. Clearly, for dimensions that pose a higher organizational risk, careful thought must be given to risk mitigation strategies and a higher level of oversight is warranted.

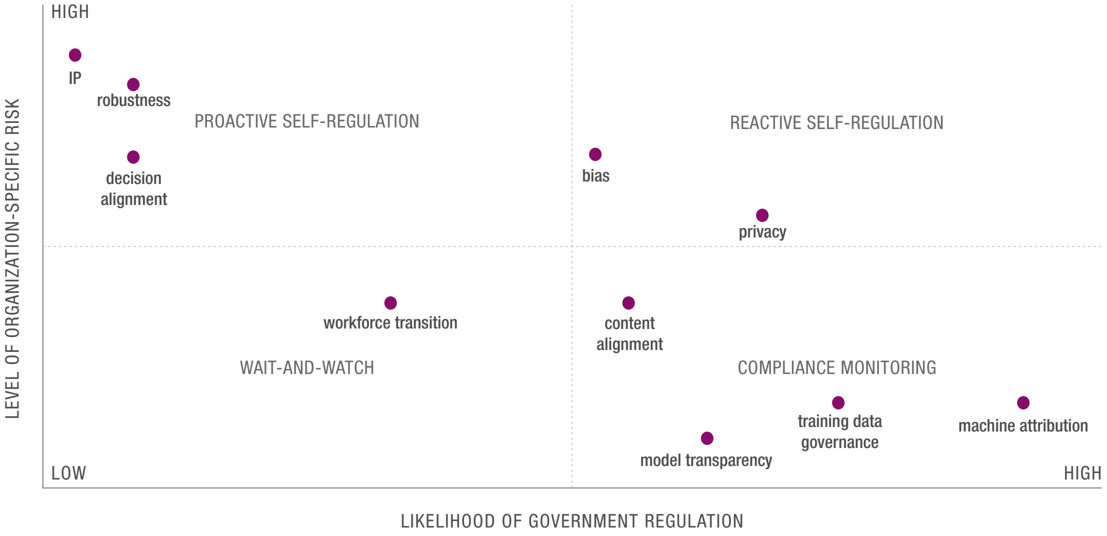
The second lens is the likelihood that the dimension will be subject to government regulation in the near future. For example, it is highly likely that there will be government regulation relating to machine attribution from several agencies and jurisdictions. In contrast, it is unlikely that governments will find it necessary to create new legislation relating to the boundaries around a company's IP ownership when their data is used to train an AI system, tending instead to rely initially on existing IP laws and the bilateral contracting regime.

Placing each governance dimension according to its relative risk and regulatory likelihood clarifies the landscape of AI governance for a corporate board. An illustrative example of such a mapping is provided in Figure 1.

Importantly, there is no universal placement of these governance dimensions in the risk-regulation space – this is necessarily an **organization-specific** assessment. For example, an AI vendor like OpenAI faces significant risks associated with training data governance, while a company in the oil industry may face lower risks on this dimension. Similarly, it is unlikely that governments will demand model transparency from the customer service chatbots of a consumer packaged goods company, but more likely they will consider this for AI systems that interact with financial markets and whose actions may affect the risk of financial contagion.

For a technology giant like Google or Meta, content alignment represents a high-risk dimension because the company's AI-generated content is widely disseminated and has the potential to have significant repercussions if untrue or misaligned with a country's value system. In contrast, a financial institution like a bank may view decision alignment as a higher-risk dimension because the decisions made by AI systems in the context of lending, risk assessment, or customer service can have direct and profound impacts on customers, regulatory compliance, and financial stability. Similarly, model explainability might be a relatively low-risk dimension for a manufacturing organization that uses AI primarily for internal process optimization. However, for an insurance company using AI to set premiums or approve claims, model explainability could be crucial, given the need to explain decisions to both customers and regulators. Similarly, AI robustness may be a top priority for companies developing mission-critical AI systems, such as those in aerospace or autonomous driving, where failure could have catastrophic consequences, while those in industries like retail, where AI use is largely for product recommendations

Figure 1: The risk-regulation matrix for AI governance



and targeting advertisements, this dimension might be important but not existential, allowing for a more measured approach to governance.

Depending on where each dimension lands, there are four broad oversight strategies that a board can consider.

3.1 Wait-and-watch

If a governance dimension is assessed as having both **low** organizational risk and a **low** likelihood of regulation, the recommended approach is to "wait and watch". In this scenario, boards should do some planning and monitoring but prioritize the dimension lower on their governance agenda. For instance, consider the dimension of AI and sustainability for an organization whose AI use is not especially resource-intensive. Choices relating to the source of electricity used in this company pose low risk, and it is unlikely that there will be new pertinent regulation targeted specifically at the sustainability of the power used specifically for AI. The "wait and watch" approach allows a board to focus its governance attention elsewhere while staying informed about potential future shifts.

Of course, adopting a "wait and watch" strategy does not mean neglecting the governance dimension entirely. Monitoring the pulse of technological advancements that might affect the dimension is important. For example, five years ago, AI robustness was not on the radar of most companies or governments, but the recent rapid advances in GenAI have moved it on to the front burner.

3.2 Compliance monitoring

When a dimension presents **low** organizational **risk** but carries a **high** likelihood of **regulation**, boards should adopt a "compliance monitoring" approach. The goal here is to anticipate regulatory requirements and ensure the organization is ready to comply once those requirements are formalized. Boards might also consider whether compliance is likely to involve sufficiently high costs to warrant participating in the shaping of eventual regulation.

Machine attribution serves as a good example of a dimension in this category for many organizations, wherein absent regulation, the risks associated with attributing content as AI-generated, rather than human-created, are relatively low, especially if the content is non-sensitive or non-public-facing. However, driven by concerns about transparency and misinformation, governments are gradually requiring the attribution of machine-generated content, perhaps viewing it as low-hanging fruit and a relatively non-controversial way to dip their toes into AI regulation. As AI agents assume larger fractions of conversational customer interactions and are imbued with greater economic autonomy, machine attribution will remain high on the regulatory priority list. Thus, monitoring regulatory developments closely and establishing internal processes that can be scaled up for compliance is prudent. This might include tracking proposed regulations in key markets and maintaining flexibility in labeling content as AI-generated. The emphasis here is on efficient allocation of resources – preparing to comply without overcommitting to a dimension that presents limited internal risk.

3.3 Proactive self-regulation

For governance dimensions with **high** organizational **risk** but a **low** likelihood of **regulatory** intervention, boards must insist that their company be proactive about crafting an internal self-regulatory regime. Waiting for regulations that may never arrive or viewing these governance dimensions as lower priority because of their absence on the government regulatory radar would be a mistake. Instead, organizations must take the lead in assessing risks and defining a governance framework.

Decision alignment and intellectual property governance are prime examples of dimensions that fall into this quadrant for many companies. In sectors like finance or healthcare, decisions made by AI systems can have profound impacts on customers. The organizational risk associated with misaligned decisions is significant. Proactive self-regulation in this context involves active red-teaming to ensure that decision making by any new AI system is aligned with the organization's values and strategic goals. Creating internal standards for decision making transparency, establishing protocols for human oversight, and implementing safeguards to ensure that AI decisions can be adjusted when necessary are additional tactical steps that could help.

In certain settings, a board might consider asking the organization to take the lead in setting self-regulatory standards for their industry or sector, and creating a coalition that shares the same self-regulatory approach. For example, a group of companies may have greater leverage than any individual one if, perhaps through an industry consortium, they define and dictate shared standards around the boundaries of corporate IP when models are fine-tuned or sensitive information is sent to an external LLM in a RAG implementation.

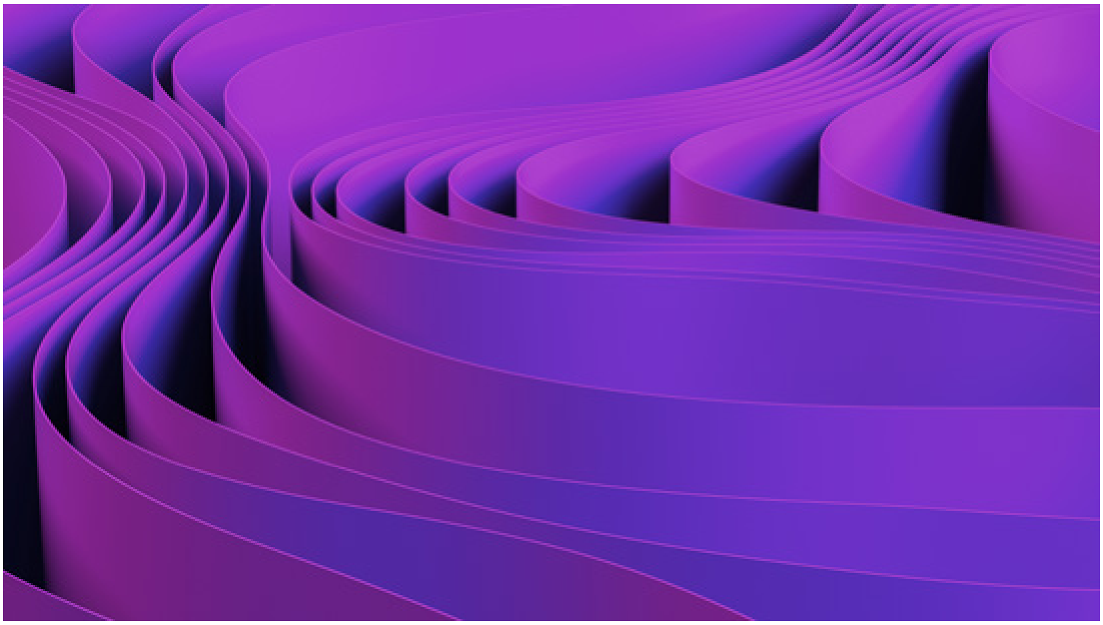
3.4 Reactive self-regulation

Finally, dimensions that exhibit both **high** organizational **risk** and a **high** likelihood of **regulation** must, of course, be given clear oversight priority, but a question that may arise is how to balance developing governance internally with anticipated external compliance. One approach would be to catalyze the active developing and implementation of internal governance policies, but to take a flexible rather than rigid approach while committing significant resources to shaping the expected government regulation towards aligning it with internal approaches. For a company like Google or Meta, who produce AI-generated content that reaches billions of users, ensuring that content is brand-aligned and does not inadvertently promote harmful or inappropriate material is both a high-risk issue and one already facing regulatory headwinds. In this case, internal steps like investing in content moderation technologies and establishing clear policies on acceptable content should be taken in parallel with active engagement with regulatory bodies to shape emerging standards. For governance dimensions in this quadrant, ensuring that internal self-regulatory approaches can be modified to meet new legal requirements as they emerge is crucial. Actively seeking dialogue with policymakers and contributing to, or leading, industry standards can also help align future regulations with existing internal practices, reducing the compliance burden associated with regulatory changes if they occur.

4. CONCLUSION: NAVIGATING AI OVERSIGHT

The framework provided in this paper lowers the complexity and obtuseness of AI governance by breaking it down into specific dimensions, an important first step towards prioritizing oversight. The dual lenses of risk and regulation can simultaneously help a board rank which aspects to pay attention to and choose a robust oversight strategy – from wait-and-watch and compliance monitoring for dimensions identified as having lower organizational risk to either reactive or proactive self-regulation for higher-risk dimensions, depending on the likelihood and imminence of government intervention. A board that merely monitors or discusses the latest AI legislation like the E.U.'s AI Act at a high level is providing insufficient oversight and control. Further, the relative prioritization of these different facets of AI governance must be specific to the company and industry. The importance of a tailored approach becomes apparent when considering that each organization has unique needs, hazards, and regulatory exposures, making it essential for boards to evaluate their specific context carefully.

Boards must aim to have at least one member sufficiently well-versed in the digital realm who can monitor the landscape and surface possible issues independent of the executive team. In parallel, the conversation about creating an AI governance committee should happen sooner rather than later. Many



organizations may be tempted to subsume AI oversight into an existing or standing committee like the audit committee, the risk committee, or the technology committee. However, as this article makes clear, AI governance overlaps with numerous specialized areas, from technology, intellectual property and sustainability to audit, measurement, and risk assessment. Having a dedicated committee lowers the risk of pursuing governance that is too deep and narrow, creates more robust oversight, and may be especially helpful in organizations with substantial AI investments or those operating in highly regulated industries. Such a committee can ensure that eventual AI governance has appropriately informed focus and control. A board-level committee can also facilitate a deeper understanding of emerging AI issues while ensuring that governance is balanced appropriately and judiciously with the executive team's desire to pursue more rapid or aggressive AI innovation.

Finally, boards would be well served by investing considerable thought during the phase in which they map their AI governance dimensions into the risk-regulation matrix, actively seeking appropriate executive, expert, and legal input to aid risk assessment and understand the likely legislative landscape. Elevating the importance of this step lends credibility to the idea that AI is a board priority and allows what follows to be undertaken with greater confidence. The ensuing oversight will then be targeted more prudently and boards can guide their management teams towards focusing executive attention where it matters most.

This document was produced for information purposes only and is for the exclusive use of the recipient.

This publication has been prepared for general guidance purposes, and is indicative and subject to change. It does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (whether express or implied) is given as to the accuracy or completeness of the information contained in this publication and The Capital Markets Company BVBA and its affiliated companies globally (collectively "Capco") does not, to the extent permissible by law, assume any liability or duty of care for any consequences of the acts or omissions of those relying on information contained in this publication, or for any decision taken based upon it.

ABOUT CAPCO

Capco, a Wipro company, is a global management and technology consultancy specializing in driving transformation in the energy and financial services industries. Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on LinkedIn, Instagram, Facebook, and YouTube.

WORLDWIDE OFFICES

APAC
Bengaluru – Electronic City
Bengaluru – Sarjapur Road
Bangkok
Chennai
Gurugram
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

MIDDLE EAST
Dubai

EUROPE
Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
Glasgow
London
Milan
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto

SOUTH AMERICA
São Paulo

THIS UNIQUE IMAGE WAS GENERATED USING MID-JOURNEY, STABLE DIFFUSION AND ADOBE FIREFLY

WWW.CAPCO.COM

CAPCO
a wipro company