Insights from July Webinar

Navigating Risk in the Age of AI.

Executive Summary

As AI systems integrate into critical operations at breakneck speed, the conversation around governance has shifted from optional to essential. With AI comes a new array of challenges: ethical dilemmas, operational uncertainties, and rising regulatory pressure.

Companies now stand at a pivotal inflection point. The task ahead isn't simply to scale AI—but to do so responsibly, even as global rules are still in flux. Effective AI governance must evolve into a dynamic framework that enables innovation while protecting against unintended consequences.

Strong governance frameworks make AI systems more explainable to regulators, more reliable to partners, and more valuable to end users. They foster trust—a prerequisite for widespread adoption—and reduce legal exposure while improving model performance. Most importantly, they create the conditions for sustainable competitive advantage.

The Unique Systemic Risks of AI

Governing AI is uniquely complex because AI itself behaves differently from traditional software. Unlike legacy systems that follow explicitly defined rules, AI—especially large language models and other forms of machine learning—generate outputs probabilistically, based on patterns in vast datasets. This makes them powerful and flexible, but also unpredictable and opaque.

That unpredictability introduces systemic risks—challenges that are difficult to detect, hard to model, and capable of cascading across entire organizations. The more deeply AI is embedded in essential functions, the more these risks demand serious strategic attention.

The Pillars of Trustworthy AI

The Five Pillars of Trustworthy AI have emerged as a global consensus across academia, industry, and government bodies. These principles can become the yardstick for every downstream policy and review. They are best viewed as a shared, evolving standard rather than a proprietary framework. They reflect a broad interdisciplinary agreement about the essential requirements for building AI systems that people—and regulators—can trust. In other words, they offer the foundational principles that any AI governance program must address:

1. **Explainability:** Results must be clear enough that domain experts can verify outcomes. “If the outcome can’t be explained, it can’t be trusted,” he emphasized.
2. **Fairness:** Models must be evaluated to ensure they don’t perpetuate discrimination or bias.
3. **Transparency:** Organizations need visibility into training data and the logic of their models.
4. **Robustness:** Systems must be wrapped in a secure control environment capable of withstanding tampering and attacks.
5. **Privacy:** Sensitive information must be protected from leaking or misuse

Here's a look at some of the biggest risks:

Opacity and Lack of Explainability
Many modern AI systems operate as black boxes—delivering statistically accurate results without offering clear insight into how decisions are made. This opacity is particularly problematic in sectors like healthcare, criminal justice, and finance, where regulators and other stakeholders demand transparency and accountability.

Privacy Violations and Data Leakage
AI models, especially large generative models, are prone to memorizing parts of their training data, potentially exposing sensitive personal information such as medical records, financial details, or names. This poses serious regulatory concerns, particularly under privacy laws like the *General Data Protection Regulation (GDPR)* enacted by the European Union in 2018 that governs how personal data must be handled and the *California Consumer Privacy Act (CCPA)* a U.S. state-level privacy law enacted in California in 2020 that grants consumers rights regarding the personal information that businesses collect about them.

Bias and Discrimination
AI systems frequently reflect and perpetuate biases present in historical or web-scraped training data. These biases can manifest in ways that disproportionately disadvantage specific groups, such as women, minorities, or people from lower socioeconomic backgrounds—particularly in domains like hiring, lending, or law enforcement.

Model Drift and Fragility

AI systems can degrade over time as the external environment shifts—due to changes in user behavior, market conditions, or adversarial tactics. These drifts may go unnoticed until the system fails in high-stakes scenarios such as supply chain optimization, autonomous control, or fraud detection.

Over-Reliance on Centralized Foundation Models

An emerging systemic risk in AI stems from the widespread reliance on a narrow set of foundational models, particularly large language models (LLMs. This creates a single point of failure: a vulnerability or flaw in one commonly used model can quickly cascade across an enterprise.

Security Vulnerabilities

As AI models become more capable and accessible, they are increasingly targeted by malicious actors. Common attack vectors include adversarial examples, prompt injection, data poisoning, and model extraction—each of which can undermine model integrity or leak proprietary information.

Loss of Human Oversight

As AI systems increasingly automate decisions, there's a risk that humans may either over-trust the technology or fail to intervene when it matters. This can lead to ethical breaches, operational errors, or failure to detect edge cases.

These unique risks demand equally novel approaches to governance. Traditional IT controls—while necessary—are insufficient on their own. Effective AI governance requires an integrated approach that spans technical safeguards, operational practices, and strategic oversight.

Creating AI Governance Frameworks

So how do companies get started? It begins with defining their AI risk threshold—a clear understanding of what's acceptable and what's not. This process requires an understanding of an organization's strategic goals, regulatory environment, and ethical principles—and then setting clear thresholds for AI behavior and performance.

Clarifying the objectives of AI deployment is foundational. What does the organization hope to achieve with AI? And what trade-offs are considered reasonable in pursuit of those outcomes? These answers help determine which risks are worth taking and which require mitigation.

This risk tolerance should align with the company's broader approach to enterprise risk and its risk policies. It's not enough for AI governance to operate in isolation; it must be integrated into existing decision-making, oversight, and compliance frameworks. Once risk appetite is clarified, the organization can begin mapping potential hazards—such as data misuse, systemic bias, regulatory breaches, or performance degradation—and assess how likely and impactful those risks may be.

To support this work, risk assessment tools like scenario modeling, impact scoring, and threat simulations can help companies quantify the tradeoffs and prioritize action accordingly. One critical dimension of this assessment is determining which systems are high-risk, and therefore require the highest levels of control. Financial Service organizations are increasingly adopting a tiered approach to AI risk management, recognizing that not all systems pose the same level of threat.

High Scrutiny: Where human well-being, rights, or systemic stability may be affected.

- Does this AI impact individual rights, safety, or livelihoods?
- Could biased outputs result in unfair or discriminatory outcomes?
- Is this used in critical sectors like healthcare, finance, criminal justice, or hiring?
- Would failure of the AI cause harm to individuals or the public?
- Is the AI making autonomous decisions without human oversight?

Examples: Loan approval, medical diagnosis, facial recognition for law enforcement

Moderate Scrutiny: Where outcomes matter, but aren't life-altering.

- Does this AI need to explain its output to users or regulators?
- Is it used in decision support (vs. decision-making)?
- Could errors impact trust or operational performance but not safety or rights?
- Is the AI used to influence (not decide) things like marketing, forecasting, or customer support?

Examples: Chatbots, demand forecasting, employee engagement scoring

Low Scrutiny: Where the impact on people is minimal or indirect.

- Is the AI primarily used to improve back-office or operational efficiency?
- Would errors have limited or no impact on people?
- Is this use case internal-facing or limited to automation of routine tasks?

Examples: Invoice processing, document classification, inventory management

Quick Decision Tree:

Is it consumer-facing or public-facing?

- **Yes:** Higher scrutiny likely
- **No:** Possibly low scrutiny

Can it affect someone's health, finances, job, or legal status?

- **Yes:** High scrutiny
- **No:** Proceed to next

Does it need to be explainable or auditable?

- **Yes:** Moderate scrutiny
- **No:** May be low scrutiny

Would errors cause minor inefficiencies or major harm?

- **Minor:** Low
- **Major:** High

Translating Governance into Practice

Once risk tiers are defined, companies can begin translating abstract principles into actionable policies. This often starts with internal frameworks that define transparency requirements, documentation standards, and accountability structures. For instance, organizations may create tiered access policies for different model types, mandate risk classification protocols, or form cross-functional ethics review boards to oversee high-impact use cases.

Several global frameworks can support this transition:

- The **NIST AI Risk Management Framework** offers a lifecycle-based approach for identifying, managing, and mitigating AI risks before harm occurs.
- **ISO/IEC 42001**, the first international standard for AI governance, provides a comprehensive guide for aligning AI operations with legal and ethical norms.
- **ISACA's AI Auditing Guidelines** equip IT and compliance teams to evaluate AI system integrity and ethical alignment.
- The **EU AI Act**, passed in 2024, is the world's first comprehensive AI regulation. It sets strict requirements for organizations operating within or selling into the EU—marking a decisive shift from self-regulation to enforceable accountability.

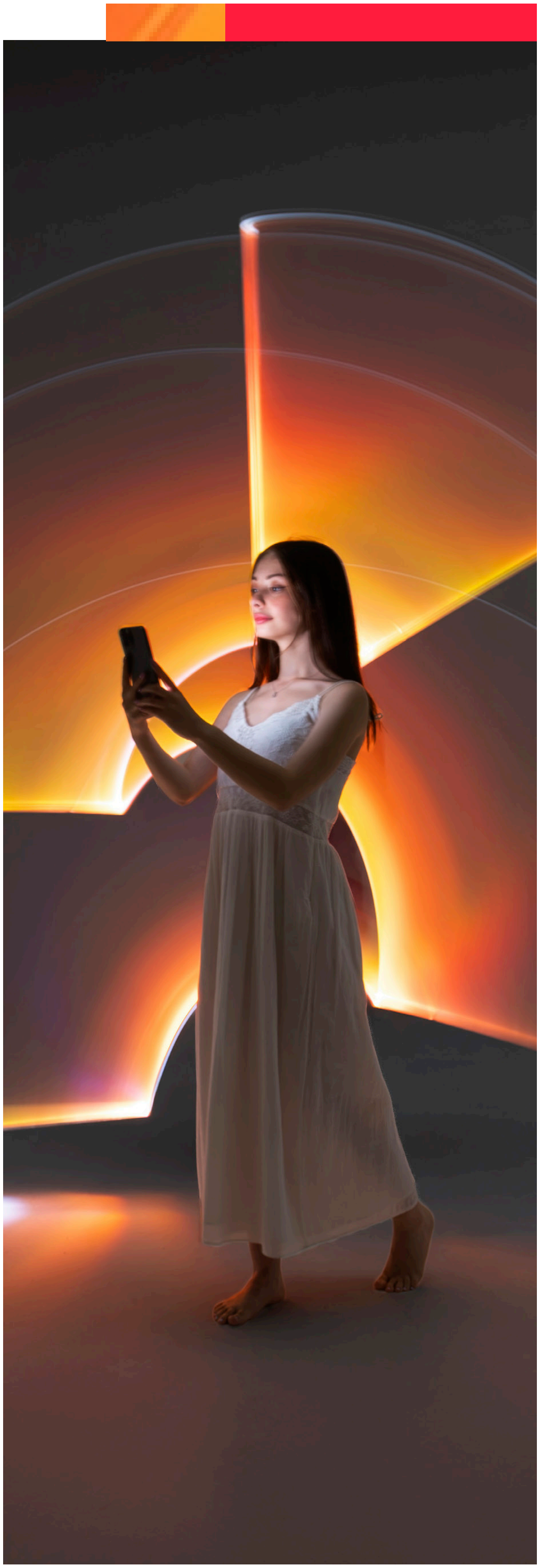
Beyond frameworks, companies should embed responsible AI practices into their operating culture. That includes building diverse governance teams that bring together expertise from legal, technical, compliance, and business functions. It means proactively addressing ethical considerations—fairness, transparency, privacy by design—throughout the AI lifecycle. Continuous performance audits, employee training programs, and open internal dialogue are all essential to ensure responsible use. And, in complex or high-risk domains, organizations should not hesitate to seek external guidance from AI ethics professionals, regulatory advisors, and third-party auditors.

Looking Ahead: Governance as Strategic Advantage

Ultimately, AI governance is not about slowing progress. It's about ensuring that innovation unfolds in ways that are responsible, resilient, and aligned with human values. When approached strategically, AI governance becomes the catalyst for confidence, not constraint. It ensures AI systems are accountable to regulators, dependable for partners, and aligned with end-user expectations. Governance lays the groundwork for transparency and trust, which are foundations of broad acceptance, while minimizing risk and reinforcing performance.

At its core, it enables responsible experimentation and positions AI as a long-term, strategic advantage.

As AI becomes an indispensable part of business and society, the organizations that thrive will be those that treat governance not as a burden, but as a blueprint for long-term success.



> As AI becomes an indispensable part of business and society, the organizations that thrive will be those that treat governance not as a burden, but as a blueprint for long-term success.

View Webinar Recording →

About RGP
RGP is a global professional services firm with nearly three decades of experience helping the world's top organizations—from Fortune 50 to fast-moving startups—solve today's complex business problems. A trusted partner to CFOs and finance leaders, we deliver the talent, consulting, and outsourced services solutions you need to grow faster, work smarter, and keep up with change—all through a flexible model and global network of experts.