Effective AI Management Unlocks Innovation

A purpose-built AI Strategy and AI Governance, mitigates risks while amplifying benefits.

As artificial intelligence (AI) and machine learning (ML) technologies move from experimental to mainstream, federal agencies will face escalating pressure to leverage the benefits of modernization while managing its inherent risks. Many agencies lack internal risk assessment practices suited for the still-evolving AI/ML discipline, which the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence[1] addresses through the creation of a permanent chief artificial intelligence officer (CAIO) at every agency. Therefore, agencies across the federal ecosystem need a plan for strong AI management—one that simultaneously contains risks and develops a governance approach that maximizes potential benefits.

Guidehouse is helping agencies address these overlapping concerns and develop a sustainable and forward-looking plan for AI management. Our strategy, tightly aligned with the National Institute of Standards and Technology's Artificial Intelligence Risk Management Framework (NIST AI RMF) and President Biden's executive orders on AI, helps organizations develop safeguards to govern, map, measure, and manage AI in ways that will address the risks of AI systems in practice.

Developing an AI Strategy

The single most important strategic step in an AI risk management and governance strategy is simply to commit to a need for a cohesive strategy. Without an internal authority, a governance framework, and trusted expert advisors, AI governance will inevitably be unfocused. Untethered, policies may prioritize preventing headline-grabbing risks (even if those risks are not germane to the agency's AI needs) and prompt agencies to clamp down on potentially beneficial innovations in the name of safety. Inconsistent policies make it challenging to proceed with AI projects in a compliant, approved manner.

[1] "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," October 30, 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Guidehouse

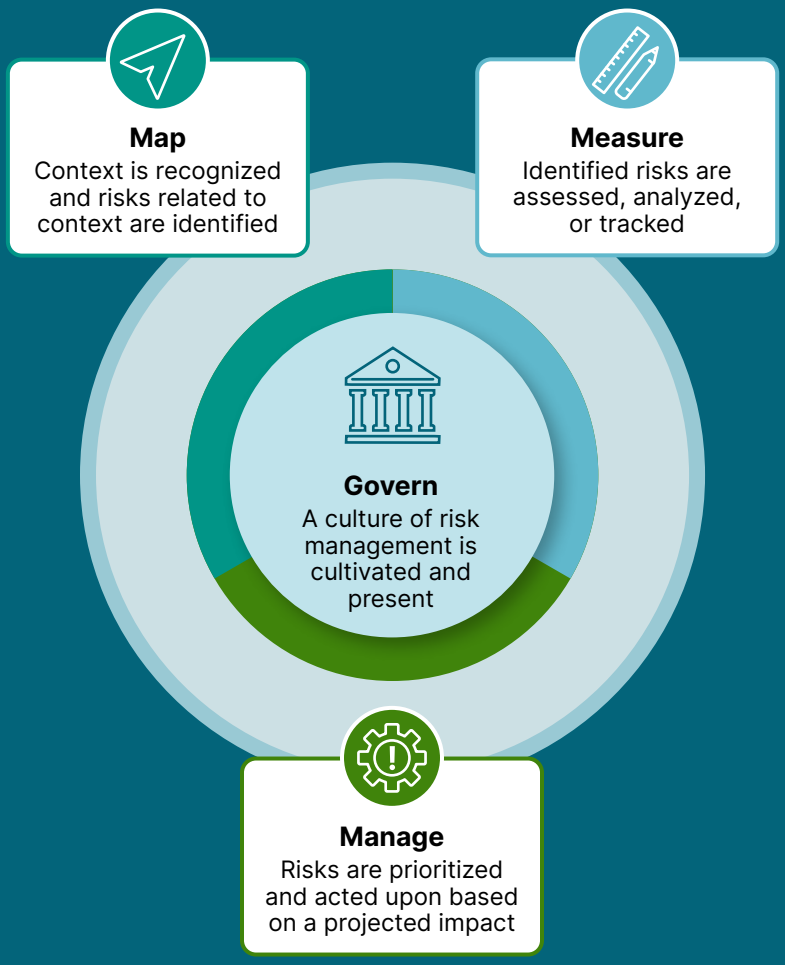
Agencies may lack the internal expertise to differentiate between unacceptable and worthwhile risks. Independent practitioners, including experts at Guidehouse, can provide context and field-tested lessons as AI governance is established and reviewed. Outside experts are also essential to the process of verifying and validating AI models, including validating the models' basic assumptions, data ingestion, and results.

Internal resources and expertise are also critical to AI governance. Before the publication of the October 31, 2023, Executive Order, the office of the chief data officer (CDO) in some agencies served as the clearinghouse for AI management, risk assessment, and guidance within that agency. Now, the new chief artificial intelligence officer will directly lead this strategy, including driving AI innovation, managing AI risks, and implementing priorities in past and future AI-related executive orders. The CAIO should also continue any existing work to de-silo AI regulations and controls already present in other parts of the agency.

The new executive order specifies a number of duties for the CAIO, including:

- **PROMOTION OF RESPONSIBLE INNOVATION** This includes guiding investment and R&D efforts as well as addressing human rights and worker training.
- **PRIVACY** The CAIO is responsible for keeping the agency's use of AI from infringing on civil liberties and ensuring that data procurement and retention follows applicable laws and policies.
- **ACCOUNTABILITY** This includes watermarking or labeling AI outputs when appropriate, controlling against deceptive or exploitative output from generative AI, and meeting ongoing reporting requirements.

The NIST AI RMF[2] is designed to be comprehensive enough for federal entities while remaining flexible enough to be applied by organizations of many sizes and missions. The NIST AI RMF organizes AI management operations around three equally weighted functions, plus one cross-disciplinary core.



- **MAP** Establish the context for the use of AI, including its intended beneficial uses and potential negative impacts.
- **MEASURE** Establish metrics to measure risks identified in the mapping stage, rate the trustworthiness of the AI-generated results, and track the collection and application of feedback across AI's field of influence.
- **MANAGE** Decide how to proceed based on the stated goals and purpose of the AI weighed against the information gathered in the map and measure stages. Continually assess the risks and benefits introduced by third-party models and data sources.
- **GOVERN** In the NIST AI RMF model, the govern step overlaps with all three other functions. This step includes putting formal statements of legal and regulatory requirements and risk management and accountability practices in place. It also establishes expectations for a safety-first approach to AI usage.

The NIST AI RMF also provides helpful templates for assessing types of potential and actual AI risk based on the harm impact of that risk (harm to people, harm to an organization, harm to an ecosystem, etc.) This mapping of potential risks to potential harms is key to prioritizing response and monitoring and to ensuring that resources devoted to risk reduction pay off in the form of averted harm.

[2] "Artificial Intelligence Risk Management Framework (AI RMF 1.0 ," January 2023, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (nist.gov).

How Guidehouse Can Help

Guidehouse brings experts in agency operations and emerging AI science to bear on AI governance challenges. In each stage, we bring interpretations, perspective, and field-tested results to address the growing need for agency-specific AI management.As trusted advisors to several federal departments and agencies, Guidehouse can assess current AI management practices and recommend changes that suit an agency's specific needs. This helps us take existing frameworks, including NIST AI RMF, and quickly adapt them to the unique day-to-day and long-term mission circumstances that each agency faces.

As AI practitioners, Guidehouse has developed extensive structure and rules enforcement across internal experiments in generative AI and AI automation. And as experienced modelers, Guidehouse experts can help refine the hypothesis being tested by an AI model, or explore other available datasets to uncover information that could add value or be a richer source of predictive power than the data already under consideration. In addition, we can assess the potential endurance of a proposed model's predictions, and, if necessary, spur the development of better models that can iterate over time and be of greater use for a longer period.

As data experts, we can assess the safety and robustness of a model's training set with more refinement than an untrained practitioner. We can also monitor the drift and skew in the datasets training an AI model over time and flag potential issues in performance and bias that those drifts might introduce.

Regulators and other supervisory bodies are often keenly interested in the explainability of AI models. Depending on the chosen algorithm and model design, explainability can be difficult to achieve. Guidehouse can help ensure that any model approaching production is suitably explainable to meet internal and external stakeholder requirements. This helps people speak credibly to the potential benefits and harms of a model's outputs and ensures that the agency itself clearly understands the risks associated with both inputs (data sources) and outputs.

Finally, not all problems are best solved with AI. Guidehouse recognizes both the strengths and weaknesses of artificial intelligence. We help agencies develop the optimal solution to data-driven problems, incorporating AI when appropriate.

Advantages of a Comprehensive AI Management Strategy

Taking control of AI governance and putting the right people on task is about more than avoiding inefficiencies, decision paralysis, and siloed operations. The process helps unlock value that might otherwise be hidden or delayed, including:

EFFECTIVE MANAGEMENT AND GOVERNANCE PROCESSES Clarity and transparency in the workforce is essential to an effective AI strategy. As new policies, procedures, and technologies are implemented, it's critical to support staff to build their capabilities and understand appropriate and approved AI usage.

Taking control of AI governance ... helps unlock value that might otherwise be hidden or delayed.

IMPLEMENTING BEST PRACTICES Guidehouse experts can help separate proven best practices from popular misconceptions. For example, the most powerful AI is based on neural networks and deep learning and those models are extremely hard to explain but at the same time may produce the best results. Expert review helps improve AI model transparency and provide clear guidance on when the model should and should not be used. A sensitive model with obscure inner workings can still have sound predictive value and provide benefit. Discarding all such tools out of hand could lead to missed opportunities or wasted resources spent reinventing something that already exists in a more mature form.

STREAMLINED HANDLING OF COMPLEX, HIGH-SECURITY TASKS Some models work with extremely sensitive inputs or produce outputs with high-stakes implications for an individual, community, or nation. As a result, the process of reviewing (adjudicating) these models is complex and often time-consuming. Working with experienced AI professionals can marshal the model through the right review process in an expedited and effective manner.

IMPROVED UNDERSTANDING OF FALSE POSITIVES AND FALSE NEGATIVES No model can be 100% accurate. There will always be some risk of both false positives and false negatives. With expert guidance, however, agencies can understand the tradeoffs of models tuned to accept more of one than the other and document the impact incorrect predictions could have on the agency, on affected individuals, and on the broader ecosystem. Guidehouse experts can engage to establish appropriate thresholds and tolerance for these errors. Doing so before an AI model is built can improve the performance of an AI solution as well as stakeholder satisfaction with its results.

Agencies face significant challenges in articulating and maintaining top-notch AI management practices. Guidehouse, in conjunction with agency CAIOs, CDOs, and evolving guidance from NIST and other industry partners, can help navigate these difficulties in a timely manner.

To learn more, review the most up-to-date AI risk management framework documentation from NIST, then contact a Guidehouse AI and data management professional to discuss mapping these practices to your organization.

Contacts

April Fordyce, Director
Defense & Security Advanced Analytics
afordyce@guidehousefederal.com

Nong Nai, Director
Cybersecurity
nnai@guidehousefederal.com

About Guidehouse

Guidehouse is a leading global provider of consulting services to the public sector and commercial markets, with broad capabilities in management, technology, and risk consulting. By combining our public and private sector expertise, we help clients address their most complex challenges and navigate significant regulatory pressures focusing on transformational change, business resiliency, and technology-driven innovation. Across a range of advisory, consulting, outsourcing, and digital services, we create scalable, innovative solutions that help our clients outwit complexity and position them for future growth and success. The company has more than 16,000 professionals in over 55 locations globally. Guidehouse is led by seasoned professionals with proven and diverse expertise in traditional and emerging technologies, markets, and agenda-setting issues driving national and global economies. For more information, please visit guidehouse.com.

guidehouse.com/services/data-analytics-intelligence

@GHTechSolutions linkedin.com/showcase/guidehouse-technology-solutions/