Booz | Allen | Hamilton

ARTIFICIAL INTELLIGENCE

ENGINEERING YOUR AI FUTURE

ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK

Prepared exclusively for the NIST Artificial Intelligence Risk Management Framework Request for Information (86 FR 40810)

CONSULTING | ANALYTICS | DIGITAL SOLUTIONS | ENGINEERING | CYBER

1.0 INTRODUCTION

Booz Allen Hamilton (Booz Allen) is pleased to submit our response to NIST's Artificial Intelligence Risk Management Framework (AI RMF or Framework) request for information (RFI). NIST's leadership in developing standardized frameworks that industry can adopt, and use is critical to safeguard and manage risks associated with AI technologies, products and services.

Artificial intelligence (AI) is an engine for growth and innovation rapidly transforming how businesses, society, and government achieve their respective missions. AI-enhanced capabilities are protecting consumers against cyber fraud, transforming customer experiences, and predicting early onset of diseases. Moreover, AI-enhanced capabilities are making their way onto the battlefield—both physical and virtual—with adversaries that are determined to displace US technological dominance at all costs. This rapid acceleration combined with widely publicized misuses has rightfully seeded concern and attention to enhancing oversight and governance of AI products, services and technologies. Today, however, there is limited policy governing how AI technologies should be developed, implemented and used as well as how risks should be proactively identified and mitigated throughout the lifecycle. Booz Allen welcomes the opportunity to provide input into an industry standard AI RMF.

As the largest provider of artificial intelligence services for the Federal government today[1], Booz Allen provides professional and technical services to design, architect, engineer and integrate AI solutions to accomplish critical missions and to maintain US technological leadership. We support some of our Nation's most high profile and innovative programs—including the Defense Threat Reduction Agency (DTRA) Operations and Integration Directorate and the Joint Artificial Intelligence Center (JAIC)—to transform and advance their enterprise AI initiatives in a deliberate, outcome focused manner to drive mission impact. More broadly, Booz Allen's AI business encompasses:

- An industry leading portfolio of AI/ML projects across civilian, intelligence and defense organizations ranging from early research to large-scale enterprise operations
- Award winning AI research and development teams with leading publications in top academic journals and forums
- A Tech Scouting network and unique partnerships with big-tech AI/ML vendors and non-traditional start-ups (NVIDIA Consulting Partner of the Year 2018-2020; Databricks Federal Partner of Year 2021)

Through our work, Booz Allen recognizes AI is not a single breakthrough technology, but a complex integration of people, processes and technologies that comes with the responsibility to use AI in a way that puts people at the center. To that end, Booz Allen has developed and implemented comprehensive frameworks and approaches—by standardizing lessons learned and best practices developed across our extensive AI portfolio—that we use to guide and govern our own delivery and risk management of AI services and solutions. In response to the key topics and questions outlined in this AI RMF RFI, we describe four relevant pillars of our approach to expand upon: (1) **Responsible AI:** embedding ethical AI principles and toolkits by design into the entire lifecycle; (2) **AI Readiness:** improving AI literacy through training to build trust and adoption; (3) **Management of AI Risk:** governance enforced through a risk management process; and (4) **Operationalizing AI:** de-risking enterprise deployment and technical execution through proven design patterns for data delivery, AI and machine learning (ML) development and operations. These pillars are further described in the following section.

2.0 PILLAR 1: RESPONSIBLE AI

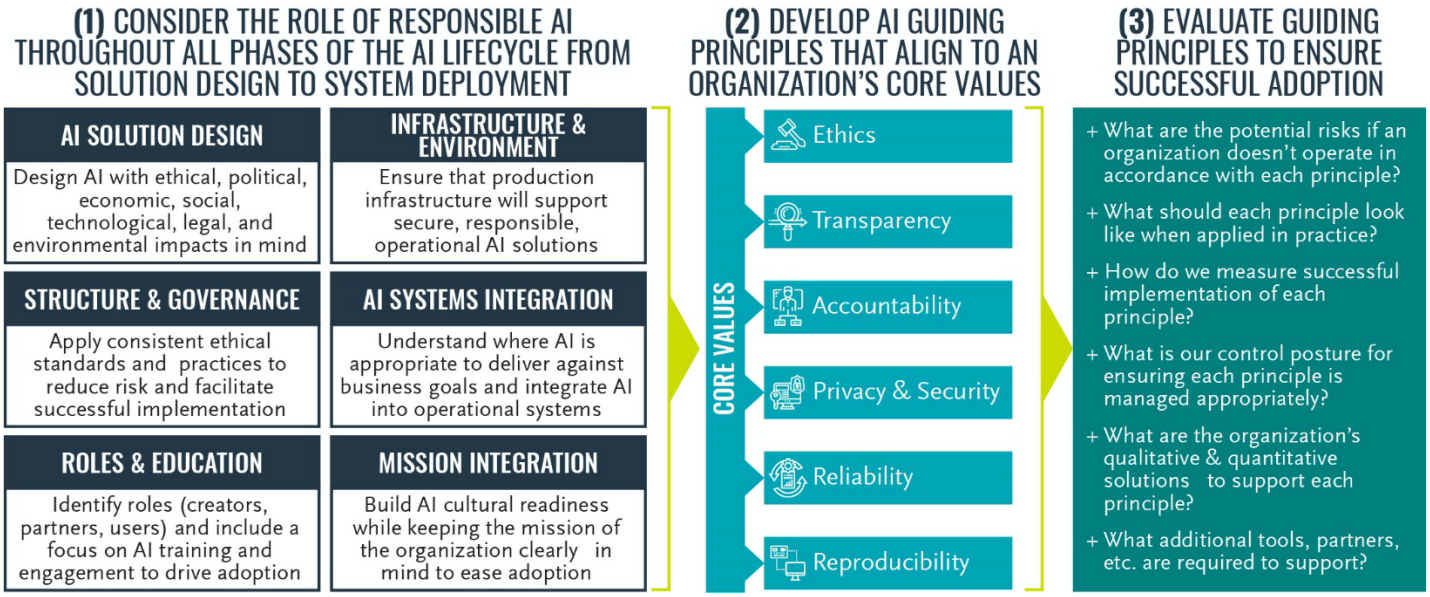
AI ethics is critical due to its consequential impact on individuals, organizations, and societies. Ethical AI is more than trustworthiness, bias, fairness, and other concepts. These are important, but they're merely characteristics of an AI system – neither good nor bad in isolation. As a result, they are ethically agnostic without further context. Ethical AI becomes meaningful only when considering how an AI system will impact other human beings. The question of "*Who*?" is critical. To ensure organizations are properly focused on the *Who*, a thorough impact assessment is necessary to understand the implications and consequences of the technology on stakeholders, users, and others. This assessment should include routinely considering the political, economic, social, technological, legal, and environmental impacts across different stakeholders over time. This analysis should inform how organizations design and build models and select datasets to achieve desired results.

What is considered ethical (or not) needs a source, a compass. It must be based on a set of shared values and principles, and more importantly, on how those values and principles relate to the context of the system for which a fair, interpretable, reliable, robust solution is designed. AI models are meant to be used in the real world – not just in a lab – and the widespread use of AI should achieve mission outcomes while responsibly accounting for human impact and equity. **Ethical AI is about values driving application**, and any organization considering how to manage, design, evaluate and use AI must start by developing guiding principles anchored to mission-driven core values. By aligning AI principles to values, organizations can create a positive impact and reduce unintended harm.

[1] Bloomberg Government Market Analysis

Figure 1 illustrates the steps required to develop and implement ethical AI principles. First, consider the role of responsible AI throughout the entire AI lifecycle, from designing the solution to training the users and integrating AI into production. Successful AI adoption should prioritize ethics early in the process. Second, develop AI guiding principles that align to an organization's shared values, or a society's cultural norms and practices. For example, if "inclusivity" is a core value, an organization may develop a guiding principle that calls for all AI teams to be meaningfully diverse and inclusive, with members from different backgrounds, skills and thinking styles. Finally, evaluate each AI guiding principle to ensure successful implementation of ethical principles in the real world.

FIGURE 1: Approach for the Development & Implementation of Ethical AI Principles



As AI use proliferates, very real ethical challenges become clearer. They range in magnitude and impact, but it's important to remember that real people's lives are often at stake, from medical diagnoses to sentencing recommendations. Prioritizing considerations like fair outcomes, clear lines of accountability when things go wrong, and transparency into how the AI system itself was built and how it operates, will ensure that these tools are used in a manner consistent with an organization's "Responsible AI" values, and help proactively mitigate ethical risks before they arise.

3.0 PILLAR 2: AI READINESS

Successful implementation and adoption of AI depends on a number of factors, including trust and cultural readiness. Establishing trust requires that AI systems align to core values and societal norms; therefore, creating and implementing responsible AI principles is critical for trustworthiness. At the same time, organizations must focus on building a culture of AI adoption that meets users where they are, breaks down technology and organizational barriers to AI, and provides the skills and resources to make the capabilities accessible to all parties. Responsibly developing AI requires a well-integrated cross-functional team. End users, subject matter experts, technical staff, and other functional staff need to work collaboratively throughout the product development lifecycle. Non-AI specific staff, including senior leadership, should have a baseline technical understanding so they can successfully communicate their insights to the technical team. This shared understanding is often achieved through an alignment around the business and mission outcomes that AI capabilities can enable (i.e., exploring the art of the possible) and further validated through iterative proof-of-concepts and prototypes that demonstrate how AI-enhanced capabilities can produce tangible benefits and impact. As an example, the Treasury Department's Bureau of Fiscal Service is developing a prototype capability to test whether artificial intelligence can streamline the annual appropriations process so that agencies can receive money sooner. In particular, BFS is testing if an AI algorithm, using natural language processing techniques, can read PDF documents and extract structured, machine-readable data, which is currently manually processed.[2]

AI staff should consider ethics an ever-present responsibility and have full understanding and appreciation for intended use and desired outcomes. Training stakeholders at all levels on AI fundamentals provides users with an understanding how an AI system works and why it produces the decisions or outputs that it does, which in turn, builds trust and improves adoption. A focus on training so that all employees are AI knowledgeable also provides an organization with access to a larger talent pool, which can be leveraged to build meaningfully diverse and inclusive design and development teams. The example highlighted below demonstrates the impact that an AI training program can have on an organization.

[2] https://federalnewsnetwork.com/artificial-intelligence/2021/02/treasury-pilots-ai-algorithm-to-parse-congressional-spending-bills-faster/

The first step in training AI talent involves conducting an assessment to understand training options and preferences, AI proficiency levels, and relevant workforce demographics and cultural context. A scientifically validated skills assessment helps segment the workforce into groups, each with distinct jobs and training needs, and results of the assessment can be used to show areas of strength and opportunities for improvement through training. Based on the skills assessment, an educational strategy that includes integrated and coordinated curricula designed for different career stages, backgrounds, levels, and technical expertise addresses gaps in foundational skills and introduces specialized AI topics.

AI workforce training empowers a greater number of diverse stakeholders to provide input throughout the full development lifecycle, which ensures that AI projects remain clearly tied to an organization's mission and business needs. Additionally, by incorporating end user perspectives throughout planning and development, AI solutions are more likely to meet user needs and, accordingly, lead to adoption. As cultural readiness grows and AI systems become more widely used to support complex decision making, it will be increasingly important for organizations to prioritize responsible use early in the adoption process. To manage these AI systems, a governance process is critical for meaningfully balancing benefit-to-risk tradeoffs and putting responsible AI principles and values into practice.

AI TRAINING TO IMPROVE AI TRUST AND ADOPTION FOR A FEDERAL HEALTH AGENCY

CHALLENGE: To advance data-driven biomedical discovery, scientists, clinicians and public health specialists must be capable of turning data into insights. Preparing the workforce to understand and apply AI methodologies to advance health care requires data science training programs designed for all career stages, backgrounds, and levels across the organization.

SOLUTION: Booz Allen developed and implemented a tailored workforce development plan for a federal health agency, with the goal of upskilling a diverse workforce while addressing training gaps and cultural barriers. The program was designed to empower staff to improve their skills – regardless of their starting point. This meant ensuring staff were confident in their understanding and ability to apply AI. The team leveraged the Analyze-Design-Develop-Implement-Evaluate (ADDIE) model to ensure content was aligned to documented skills gaps and measurable learning objectives.

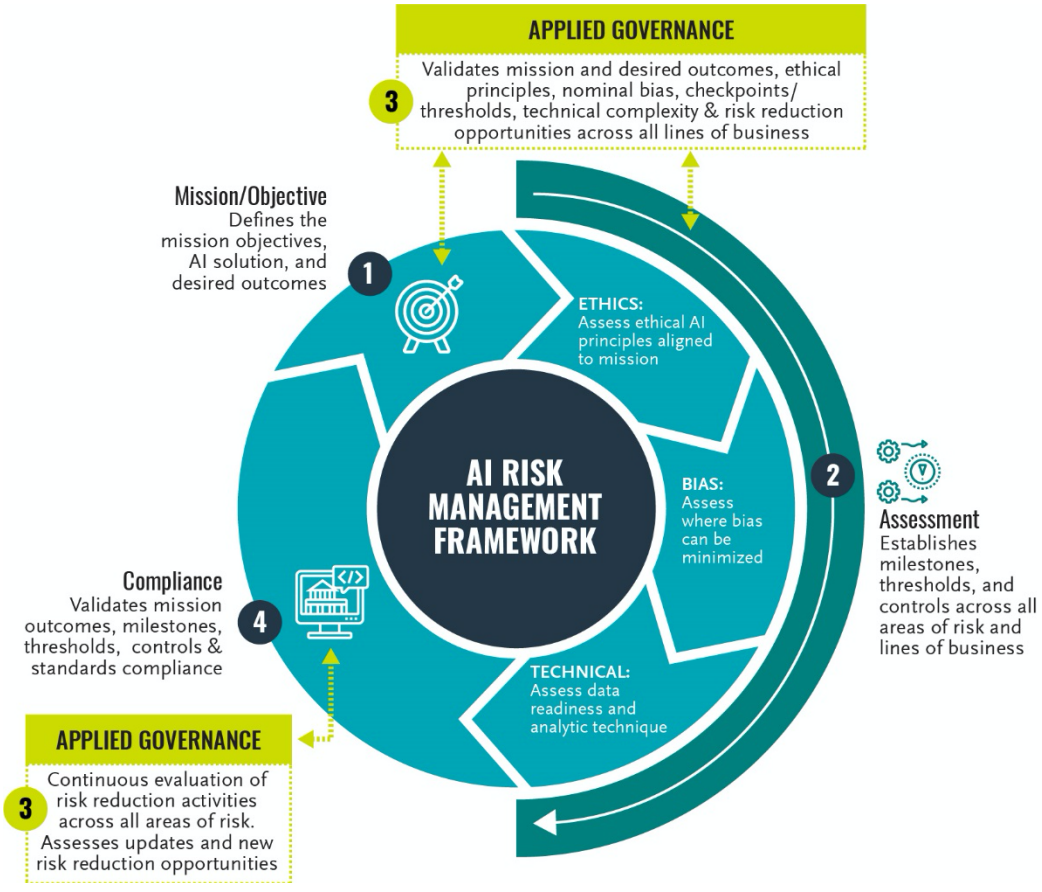
IMPACT: The training program had both immediate and lasting success, engaging 500+ participants. The course catalog featured over 220 options and provided users with an understanding of the training ecosystem to continue their own on-demand learning. The program evaluation survey highlighted how everyone, regardless of their skill level, was able to upskill in a meaningful way that advanced their AI understanding.

4.0 PILLAR 3: MANAGEMENT OF AI RISK

Booz Allen is deeply committed to maintaining an acute awareness of the societal and environmental impacts of our AI systems and applications, and we ensure that our company and our people design AI systems that are grounded in real-world implications. To provide checks and balances and ensure guiding AI principles are being met, we believe that an AI governance process, similar to that described in Figure 2, should be implemented. In addition to ensuring that AI projects are evaluated systematically to mitigate risk, governance builds stakeholder confidence in AI through an organizations' responsible use of AI.

Well-designed controls promote the application of AI through effective management, ensuring that it is meeting performance requirements and is being used in an ethical way.

FIGURE 2: AI Risk Management Framework



1. **MISSION/OBJECTIVE.** Before any application of AI is considered, mission and objective outcomes must be well defined and understood in consultation with a diverse team of technical and domain experts. A well-understood mission will enable oversight controls across logical development, appropriation of data, alignment with ethical principles and organizational values, and assessment of bias to reduce risk in deliverable output. Defined mission objectives allow thresholds to be established that will enable governing bodies to assess drift and deviation from standards established to meet the outcome.

2. **ASSESSMENT.** Whenever machines can perform tasks that would normally require human intelligence, an AI governance process that includes a governing body, which establishes milestones, thresholds and controls, can provide the oversight required to address the unique impacts of AI and reduce risk. While risks are universally relevant across all phases of the AI lifecycle, the appropriate thresholds and controls are tailored across milestones (e.g., R&D model vs. production system) to balance managing risk, innovation and speed. For example, projects in the early R&D phase will benefit from expert review and input around bias mitigation approaches, whereas deployment of an AI solution into production systems may require formal go/no-go approval of model bias controls put in place.

   As described in Pillar 1, ethical considerations and risk management should be integrated throughout all phases of the AI lifecycle. Similarly, bias risks for AI solutions should consider both technical and non-technical factors including use of non-representative data as well as system and model design process that are not inclusive across a diverse group of stakeholders. Finally, in addition to ethical and bias considerations, AI technical development and implementation is highly complex with execution risks that can undermine project success and risk management. These technical risks include access and availability of complete and high-quality data, secure computing infrastructure for model development and deployment, robust software architecture and engineering expertise for scalable system design and implementation, and rigorous vetting and validation processes for models and algorithms.

   Assessing and removing technical execution risks at the onset and throughout delivery will equip project teams with the resources (technology, expertise, etc.) necessary for successful prototyping and production deployment, and mitigate the need for suboptimal workaround solutions that could unintentionally bypass safeguards, policies and guardrails put in place to protect against AI risks.

3. **APPLIED GOVERNANCE.** Without established governing bodies to oversee the application of process, definition of standards, thresholds measurement, and overall compliance, responsible mission outcome cannot be achieved. Establishing governance bodies requires a culture of buy-in to administer and oversee responsible AI solution utility. Lack of oversight or continuous monitoring will result in an AI solution that does not function as desired (rogue AI) and can produce inaccurate or irresponsible outcomes.

4. **COMPLIANCE.** Compliance is achieved when oversight of established milestones, thresholds, and tests can be verified and validated. With a culture of governance, corporate buy-in, and effective oversight, unauthorized changes, lack of versioning and unintended consequences can be mitigated or managed.

An AI Risk Management Framework must evolve as the technology continues to mature and become more standardized, while remaining aligned to an organization's core values and guiding principles. The framework requires a culture of risk ownership through an oversight body that enforces design standards and advises on ethical and technical issues. For successful implementation of the framework, the organization must be ready – educated in the technical utility of AI, forward leaning, transparent and motivated to explore applications of AI to address mission challenges.

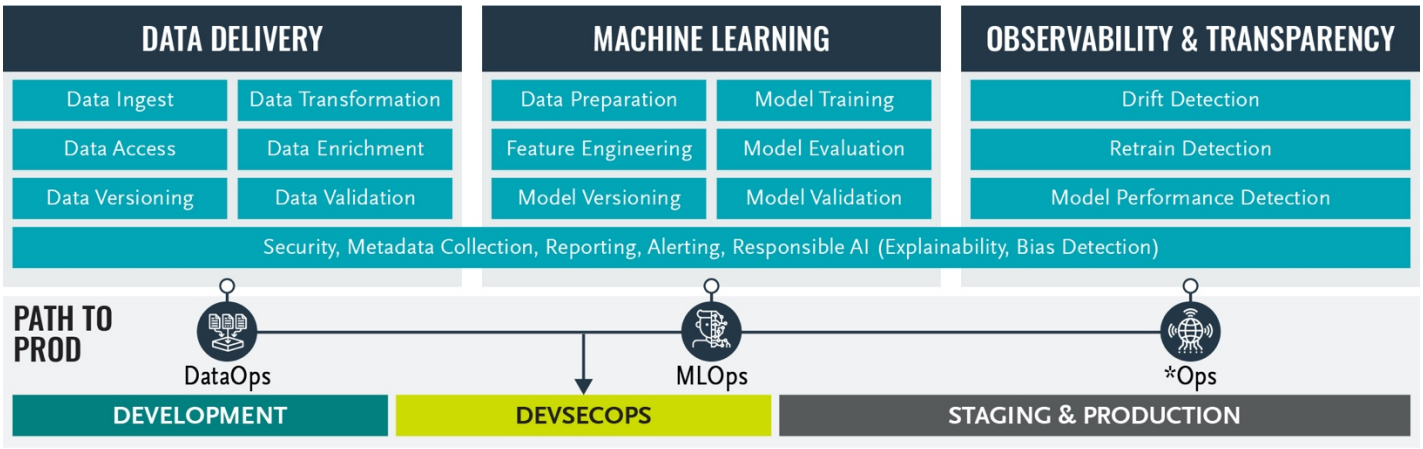
5.0 PILLAR 4: OPERATIONALIZING AI

Effectively operationalizing AI requires organizations to provide resources, guidance, and capabilities that empower teams to accelerate the implementation of enterprise AI solutions. To establish a common understanding of concepts, processes, and terminologies for achieving enterprise AI, Booz Allen developed and published our industry leading AI engineering framework, AIOps. AIOps provides delivery teams with an integrated framework for responsible AI development enabled by an automated and documented modular solution for development and sustainment of AI. Adopting an AIOps framework provides a consistent process to help mitigate risks while enabling organizations to achieve a scalable, sustainable, and coordinated AI enterprise capability in alignment with responsible AI principles and organizational vision and goals. (For additional details on achieving Enterprise AI success through AIOps please reference our O'Reilly publication: https://tinyurl.com/boozallenAIOps.)

To operationalize and deliver responsible, scalable AI solutions, Booz Allen developed an approach that starts with our industry leading AI Reference Architecture (RA), which provides technology agnostic, guiding principles to standardize and accelerate the delivery of AI through common components, capabilities, processes, and terminology (Figure 3). Adopting a standard RA enables an organization to produce real-world solutions from an abstract framework, which drives consistency and standardization, surfacing risk early and supporting mitigation and reduction measures. Our RA—developed through our work with over 120 AI projects across the

federal government—ensures risk reduction in production through technical alignment and reuse of best-of-breed technology, and provides a starting point for any enterprise AI platform.

FIGURE 3: Booz Allen's Reference Architecture



Our AI Reference Architecture provides a proven blueprint to define and assess the critical components within a modern analytics platform, including:

- **DATA DELIVERY:** Covers activities that span ingestion, storage, transformation, enrichment, and delivery of data for analytics at scale. Robust, repeatable data delivery provides a critical foundation for analytics activities, and provides lineage and provenance collection to support data governance.
- **MACHINE LEARNING:** Encompasses the activities that are performed to prepare and transform data into insights and inferences that serve client and other business needs. The RA defines the main capabilities that when combined comprise a holistic ML Workflow.
- **OBSERVABILITY AND TRACEABILITY:** Processes and capabilities for monitoring and reacting to changes in ML workflow performance. This includes tracking performance, detecting when performance shifts beyond acceptable limits, and triggering appropriate actions in response.
- **CROSSCUTTING COMPONENTS:** Supporting analytics capabilities such as Security, Metadata Collection and Explainability that provide core functions supporting the development of analytic services.
- **PATH TO PRODUCTION:** Combines software delivery best practices with Data Operations (DataOps), Machine Learning Operations (MLOps) to establish a robust path to production.

Our framework provides a benchmark to evaluate existing AI architectures and helps identify technical and non-technical risk areas that should be tracked and mitigated for enterprise AI-enhanced capabilities and applications. To evaluate risk, the RA focuses on five key areas, architecture and design, governance, data quality and availability, adoption, security, and how their severity impacts an organization, ensuring risk is managed and trust is developed holistically at the project, mission, and organizational level.

RISK	DESCRIPTION
ARCHITECTURE AND DESIGN	Selecting an architecture that can be setup within an organization's technical and security limit and can be deployed in a timely manner reducing delays. Ensuring the system can support interoperability within the organization and can be reliability operated and maintained through the life of the program.
GOVERNANCE	Ensuring a project has proper oversight and sustainment. Develops rich processes for Configuration Management, Testing and Verification & Validation. Follows proper process to develop and evaluate guiding principles to successfully implement Ethical AI.
DATA QUALITY AND AVAILABILITY	Assessing the program's ability to access data required for its mission, manage its quantity, and evaluate its quality. Evaluating the programs capacity to understand the data and has the resources to do so such as a data dictionary. Ability of the program to correctly evaluate biased data and make the proper corrections.
ADOPTION	The program has defined AI needs that directly support the organizations mission. Proper training and communication plans are in place to support the building of trust and transparency of ethical concerns. User feedback has a key and consistent part in the program's development lifecycle.
SECURITY	Ensure systems are built in a robust and secure manner and incorporating security practices at every stage of development. The organization can properly monitor new complex AU solutions and has techniques in place to combat adversarial AI.

To identify, assess, prioritize and mitigate these risks we employ a standard framework for AI enterprise modernization, which consists of four phases: Assess, Architect, Assemble, and Advance. An enterprise modernization methodology—tailored to AI transformation and operationalization—must bake-in risk identification and mitigation in each phase to ensure that AI capabilities achieve mission impact at scale, speed, quality, and low risk.

PHASE	DESCRIPTION
ASSESS	• Analyze operational requirements, assess the "as is" solution, and derive technical requirements for the target architecture. • Identify the capabilities and components needed to establish a holistic AI solution and establish the architectural foundation for advancement to ML- and AI solutions. • Establish risk governance, key stakeholders, and prioritization process.
ARCHITECT	• Define an extensible AI architecture that supports the mission requirements, accounts for information security requirements needed for ATO. • Identify and tailor foundational components for future AI/ML opportunities. • Surface architecture and design risk, accounting for all 'ilities' and aligned to the solution roadmap.
ASSEMBLE	• Develop and deliver analytic models, tools and solutions. • Establish a robust AI/ML workflow that enables reproducibility and explainability, ensures that ML is responsibly developed, deployed, and applied, increasing confidence, trust, and efficacy. • Iterative identify, prioritize, and mitigate implementation risk.
ADVANCE	• Transition capabilities into operations, performance-based evaluation, and repeatability. • Apply drift detection ensures AI models continue to produce accurate inferences and predictions as datasets and models are introduced or evolve over time. • Automate ethics, bias, and responsible AI pipelines for monitoring and compliance to re-train based on pre-defined triggers. • Continuously monitor, prioritize, and mitigate implementation and operational risk.

6.0 CONCLUSION

Although any new technology could be used improperly if its use is not guided by values, there are unique aspects of AI systems that complicate risk assessment, mitigation and management. While a few organizations have published their ethical AI principles, fewer still have defined explicit controls. For this reason, forward-thinking approaches that support innovation and confidence in AI systems are required. We believe that our AI governance process and Reference Architecture offer something new and valuable for NIST: a practical and transparent way to ensure that AI systems are safe, ethical, and robust for deployment. As the leading provider of AI services and technologies to the Federal government we welcome the opportunity to build an open community of practitioners, and a framework for fair, transparent, and responsible risk management that delivers on the promise of AI.