disc
Data Integration Support Center
WestEd

2025
June

SPOTLIGHT

Building a Secure AI Environment

IMPLEMENTATION PLAN

AUTHORS:
Baron Rodriguez
Laia Tiderman
Sara Kock

As public agencies evaluate generative artificial intelligence (GenAI) integration, the fundamental challenge extends beyond technology selection to comprehensive risk management and operational security. Successful AI implementation requires balancing innovation potential with stringent privacy, security, and compliance requirements, which is particularly critical for agencies handling sensitive public data and operating under complex regulatory frameworks.

This implementation guide from the Data Integration Support Center (DISC) at WestEd provides a structured methodology for establishing secure AI environments within public sector constraints. Based on empirical deployment experience, this brief details the technical, procedural, and organizational steps required for successful implementation. A companion resource, "Maintaining a Secure AI Environment: Resource Considerations," addresses operational sustainability including staffing models, policy frameworks, and total cost of ownership considerations.

PREREQUISITES FOR SECURE AI DEPLOYMENT. Effective AI implementation demands comprehensive organizational readiness across multiple domains: risk assessment and mitigation strategies, governance frameworks aligned with public sector requirements, technical infrastructure evaluation, workforce capability analysis, and clearly defined use case parameters with measurable outcomes. Many of these topics and how WestEd worked on them are outlined in prior briefs. DISC's foundational guidance addresses these prerequisite elements and provides facilitation support for agencies developing AI implementation strategies. Organizations lacking established frameworks in these areas should prioritize foundational work before proceeding with technical deployment.

WestEd's Use Case

ORGANIZATIONAL DRIVERS AND STRATEGIC RESPONSE. Staff requests for GenAI tools created tension between operational efficiency gains and risk management requirements. Leadership identified critical concerns spanning intellectual property protection, data privacy and security, client contractual obligations regarding AI use, data governance (including use and ownership rights), and algorithmic fairness and bias mitigation. Rather than implementing blanket restrictions, the technology team proposed developing a controlled AI environment that would enable tool access while addressing enterprise risk requirements.

IMPLEMENTATION APPROACH AND RESOURCE PLANNING. WestEd's secure AI environment deployment followed a phased implementation methodology requiring cross-functional coordination between information technology (IT) architecture, cybersecurity, legal, and agency operations teams. Each phase involved strategic decision points that directly affected subsequent implementation steps and the overall project timeline. The implementation framework below reflects resource allocation of 10–20 hours per week across participating teams. While specific technical implementations will vary based on existing infrastructure and chosen technology stack, the core implementation phases and decision frameworks remain broadly applicable across similar organizational contexts.

Overview of Effort to Implement a Secure AI Enclave

Implementation Steps

1 Planning and Assessment

DURATION: 2 weeks

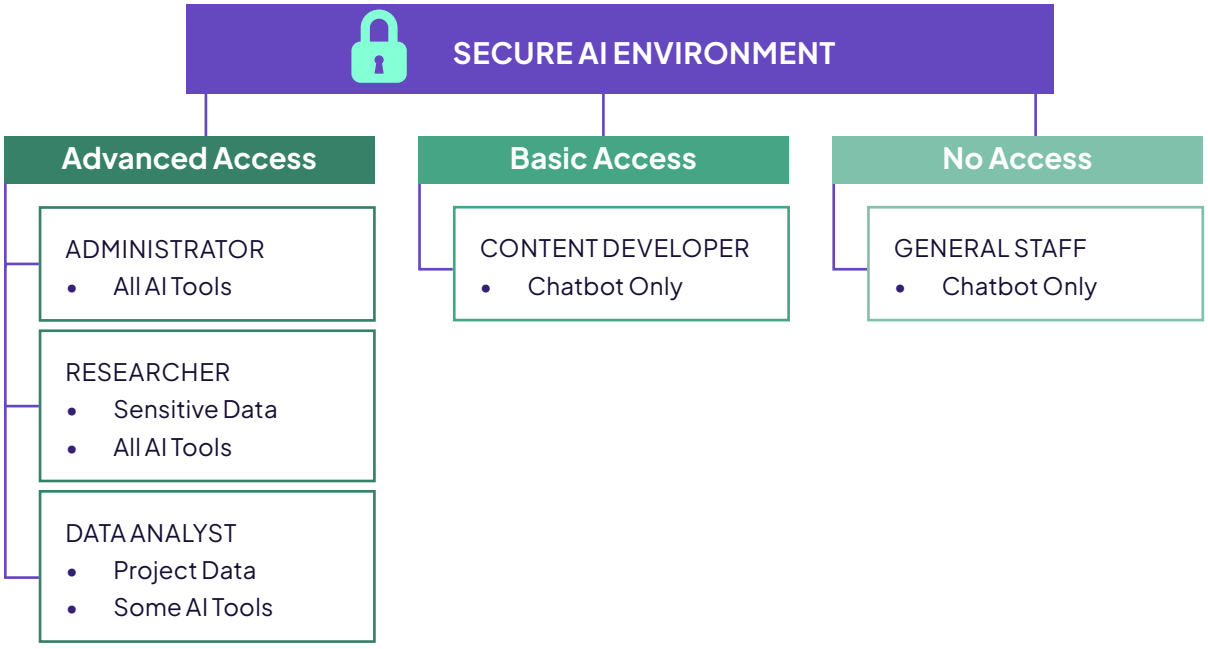
During this first step, organizations should conduct an evaluation that lays the groundwork for all subsequent activities. This process should begin with a thorough assessment of the organization's security requirements and compliance standards to ensure that the AI environment meets all necessary regulatory and institutional requirements. Simultaneously, teams across the organization should identify specific business processes or use cases that could benefit from AI integration. This analysis should include determining which AI models and services will best support the business objectives while maintaining security and compliance. The step should conclude with a detailed assessment of the organization's current IT infrastructure and security posture to help identify any gaps that need to be addressed before proceeding with implementation. This careful planning stage is essential for ensuring that the resulting AI environment will be both secure and aligned with organizational needs.

2 Virtual Desktop Deployment and Security Configuration

DURATION: 2 weeks

This step focuses on creating a controlled, isolated workspace where AI activities can be conducted safely. Organizations should configure their virtual desktop infrastructure with robust security measures, including multifactor authentication for user access and conditional access policies that restrict connections to approved devices and networks. The implementation team needs to establish role-based access controls and configure network security groups that manage traffic flow and network segmentation. This step will ensure appropriate access and maintain strict isolation of the AI environment. Additionally, the team needs to integrate security monitoring and threat detection systems to protect the environment from potential security breaches and attacks. This foundation of secure virtual infrastructure serves as the cornerstone for all subsequent AI implementation steps.

WestEd's Role-Based Access Control Examples



3 AI Model Integration With Emphasis on Security

DURATION: 4 weeks

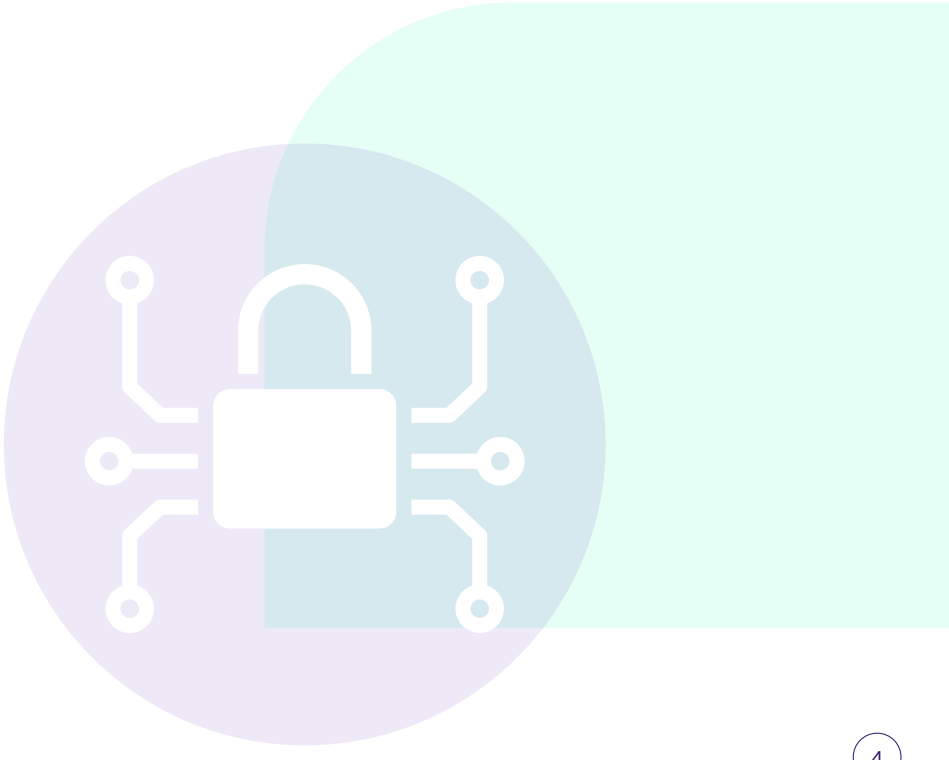
The integration of AI models into the secure environment represents the most technically complex step in the implementation process. This step involves employing multiple layers of security, including encryption for sensitive workloads and data, secure key management systems for protecting sensitive information, and strong access controls for training and deployment. Teams must configure the environment to ensure that all AI interactions are properly encrypted, monitored, and logged. The integration process requires careful attention to data-handling procedures, particularly when training models or processing sensitive information. This step often involves iterative testing and refinement as teams test that the AI models function effectively within the security constraints. Organizations should expect an ongoing effort in this area as models are updated and new security requirements emerge.

4 Responsible Use and User Training

DURATION: 2 weeks

Before granting access to AI tools, organizations should establish clear expectations for responsible use. This step requires developing policies that address ethical considerations, appropriate application of AI, and more. These policies should align with existing regulatory, security, and privacy requirements as well as other organizational procedures. Once these responsible-use frameworks are established, they should be integrated into user training programs.

Effective user trainings are crucial for ensuring that the secure AI environment is used properly. A comprehensive training program addresses both technical aspects and the organization's expectations. Technical training teaches topics such as how to access the environment, proper handling of sensitive data, prompt engineering techniques, and security protocols. Training on organizational expectations educates participants about the policies and procedures for AI use. Regardless of the topic, user trainings can include hands-on practice, open discussion of various scenarios, or simulated exercises done together. Regular refresher training and updates will ensure that users remain knowledgeable about both the technical aspects and responsible AI use, adapting as the AI environment and policies evolve.

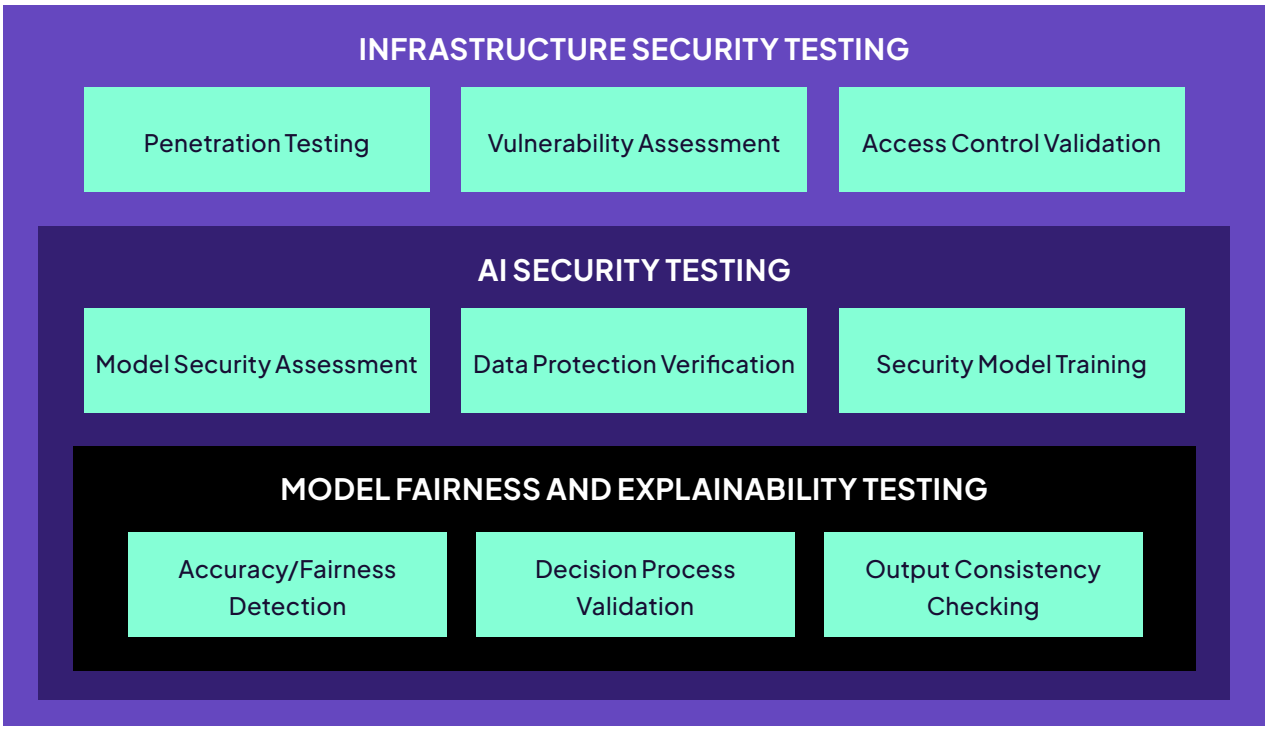


5 Testing and Security Validation

DURATION: 2 weeks

Rigorous testing and security validation is needed to ensure that the AI environment meets all requirements before wider deployment. Organizations can conduct comprehensive security assessments, including penetration testing and vulnerability assessments of both the virtual desktop infrastructure and AI integrations. Security teams should validate implemented measures by running simulated attack scenarios and testing response procedures. This step also includes assessing AI model integrity by both examining technical security and checking for model fairness and explainability. Testing for model fairness includes evaluating outputs and use cases to identify potential biases, while explainability testing ensures that the AI's decision-making process can be understood and audited by humans. The testing process should verify that all security controls, access restrictions, and monitoring systems function as intended. As organizations complete their testing, they should document findings, address any identified vulnerabilities, and establish baselines for ongoing security monitoring.

WestEd's Testing and Security Validation Layers

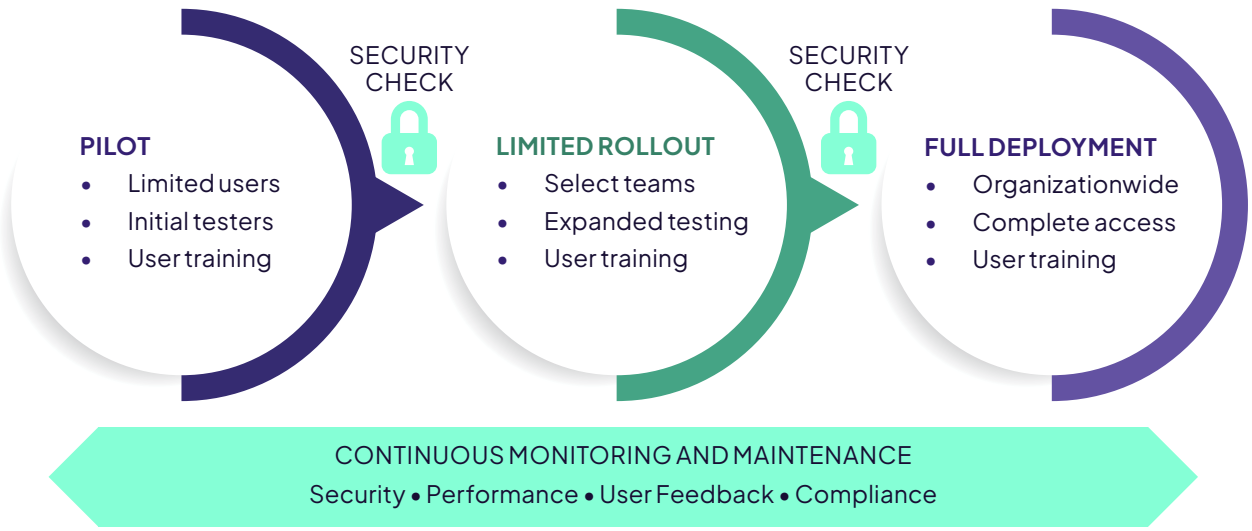


6 Deployment and Go-Live

DURATION: 1 week

The go-live step marks the transition from testing to production use of the secure AI environment. This step requires careful coordination to ensure a smooth migration of users and workloads into the secure virtual environment. Organizations must systematically deploy AI-integrated applications while maintaining compliance with data privacy regulations throughout the process. During this step, teams should implement a phased rollout strategy, beginning with a pilot group of users before expanding to the entire organization. Close monitoring during the initial deployment period is crucial for quickly identifying and addressing any security concerns or user access issues. The deployment team should maintain detailed documentation of the production environment configuration and establish clear communication channels for user support during the transition.

WestEd Deployment Phases



What Is Next?

A carefully crafted implementation plan is critical to the success of a safe, secure AI environment, but the resources involved and the ongoing effort to maintain the environment must also be considered. For practical insights on staffing considerations, cost factors, and policy requirements for system sustainability, read the companion brief, "Maintaining a Secure AI Environment: Resource Considerations."

How DISC Can Help

DISC at WestEd can facilitate productive conversations and assist in the development of an AI strategy for an organization's integrated data system. DISC offers technical assistance to public agencies free of cost, including the following services:

- training agency leadership and/or staff in GenAI best practices
- providing expertise on the scope and variety of AI tools to support data integration efforts
- developing use cases that leverage AI tools and aligning those use cases to the state's strategic priorities
- conducting neutral, external assessments of the maturity of the integrated data system and its capacity to support and scale AI technologies
- facilitating structured discussions to develop a foundation for the use of AI in the integrated data system

Suggested citation: Rodriguez, B., Tiderman, L., & Kock, S. (2025, June). *Building a secure AI environment: Implementation plan.* WestEd.