COURSE SYLLABUS

AI Security for Federal Systems Course (Self-Paced)

Learn to govern, secure, and audit federal AI systems using the NIST AI Risk Management Framework, EO 14110, and OMB M-24-10. This self-paced course covers the full lifecycle: GOVERN, MAP, MEASURE, and MANAGE functions, 800-53 Rev 5 controls, adversarial threat identification, vendor evaluation, and continuous monitoring.

Group classes in Live Online and onsite training is available for this course. For more information, email onsite@graduateschool.edu or visit: https://www.graduateschool.edu/courses/ai-security-for-federal-systems-course-online

support@graduateschool.edu •
(888) 744-4723

Course Outline

Module 1: Federal AI Policy Landscape and Governance Obligations

- EO 14110 — Safe, Secure, and Trustworthy AI: agency requirements, timelines, and accountability mechanisms
- OMB M-24-10 — Advancing Governance, Innovation, and Risk Management: AI use case inventory and tiered review requirements
- NIST AI RMF 1.0: four core functions (GOVERN, MAP, MEASURE, MANAGE), subcategories, and relationship to 800-53
- Chief AI Officer (CAIO) role and AI governance board: authority, membership, and required decision processes
- Federal AI use case inventory: required disclosure elements and OMB M-24-10 transparency obligations
- AI use case inventory exercise: catalog AI systems deployed or in development at a sample federal agency, classify each by OMB M-24-10 risk tier, and identify the three use cases requiring the most immediate governance action based on risk tier and current documentation gaps

Module 2: NIST AI RMF — GOVERN, MAP, MEASURE, and MANAGE Functions

- GOVERN: organizational AI risk culture, CAIO authority, cross-functional accountability, and AI impact assessment documentation
- MAP: stakeholder impact analysis, training data provenance, third-party AI risk mapping, and use case categorization
- MEASURE: accuracy, reliability, bias measurement methods, explainability requirements, and confidence threshold standards
- MANAGE: risk treatment selection, model drift monitoring, AI incident response procedures, and system decommissioning protocols
- Connecting AI RMF subcategories to 800-53 control families: where the frameworks overlap and where gaps exist
- AI RMF application exercise: complete GOVERN/MAP exercises for a sample federal ML-based benefits eligibility system — defining governance structure, categorizing the use case, identifying all affected stakeholder groups, and producing a two-page AI impact assessment summary using the NIST AI RMF Playbook template

Module 3: Adversarial AI Threats and Attack Taxonomy

- Training-time attacks: data poisoning, backdoor injection, and label flipping in federal ML pipelines
- Inference-time attacks: adversarial examples, evasion attacks, and model extraction via API queries
- LLM-specific risks: prompt injection (direct and indirect), jailbreaking, hallucination, and RAG attack vectors

- Membership inference and model inversion: extracting training data from deployed federal models
- Supply chain threats: compromised pre-trained models, poisoned datasets, and malicious AI libraries
- Adversarial threat modeling exercise: apply threat modeling to two federal AI scenarios (an ML fraud detection model and an LLM-powered citizen service chatbot), identify the two most likely attack vectors per system, estimate mission impact, and recommend one preventive and one detective control for each attack vector

Module 4: Applying NIST 800-53 Controls to AI System Security Architecture

- SA-8 system security engineering principles applied to AI: SA-8(33) transparency, SA-8(34) governance, SA-8(35) safeguards
- SR family controls applied to AI supply chain: model provenance documentation, vendor assessment, and SBOM extension to AI components
- AT-2(5) AI awareness training: what federal employees who interact with AI systems must understand
- Secure ML pipeline architecture: data ingestion security, training environment isolation, and model registry access controls
- AI inference endpoint security: API authentication, rate limiting, logging requirements, and anomaly detection
- AI security architecture review: review a system architecture diagram for a sample federal AI system, apply relevant 800-53 Rev 5 AI-specific controls to identify five security gaps, and propose specific technical or procedural controls for each gap

Module 5: AI Acquisition Security — Procurement Requirements and Vendor Evaluation

- AI-specific solicitation requirements: OMB M-24-10 mandatory clauses and agency-specific supplements
- SBOM extension to AI: model card requirements, training data documentation standards, and provenance obligations
- EO 14110 safety commitments: what large AI developers must demonstrate and how to evaluate their representations
- Reading an AI vendor model card critically: accuracy claims, bias testing results, red team report scope, and limitations
- FedRAMP for AI: current status, emerging cloud AI service requirements, and how to assess FedRAMP cloud AI risk
- AI procurement review exercise: evaluate a sample vendor proposal for an ML-based federal document review system, assess security representations across five criteria, identify three material gaps, and draft contract language to address each gap before award

Module 6: AI Security Testing and Red Teaming

- AI security testing vs. traditional IT security testing: what's different, why it matters, and what tools exist
- Adversarial robustness testing: automated frameworks (Counterfit, ART) and manual adversarial input techniques
- Bias and fairness testing: disparate impact analysis, demographic parity measurement, and threshold sensitivity analysis
- LLM red teaming methodology: structured adversarial prompting, jailbreak cataloging, and injection vector mapping
- AI Safety Institute (AISI) red teaming framework: federal collaboration opportunities and evaluation standards
- AI red team simulation: conduct a structured red team exercise against a sample LLM-powered federal knowledge management chatbot, attempting three attack classes (prompt injection, jailbreak, sensitive data extraction); document successful attacks, assess operational impact, and recommend technical guardrails to prevent each attack class in production

Module 7: AI System Auditing and Continuous Monitoring

- AI audit objectives: accuracy, fairness, security, transparency, and regulatory compliance — and how each is measured
- GAGAS application to AI auditing: evidence standards, sampling considerations, and reporting requirements
- Evidence collection for AI audits: model documentation packages, decision logs, training data records, and bias test results
- AI continuous monitoring requirements: accuracy drift thresholds, fairness monitoring schedules, and security anomaly detection
- AI incident classification and mandatory reporting under OMB M-24-10 and EO 14110
- AI audit planning exercise: develop an audit plan for a sample federal benefits eligibility ML system, defining five audit objectives, identifying evidence sources and collection methods for each, specifying data analytics procedures for bias analysis, and designing a monitoring dashboard with five KPIs tracking AI trustworthiness over a 12-month cycle

Module 8: Federal AI Risk Governance Program

- AI use case inventory requirements: OMB M-24-10 required elements, review cadence, and public reporting obligations
- AI governance committee: charter structure, membership, tiered review procedures, and escalation authorities

- AI risk tier classification: operational criteria, documentation thresholds, and authorization integration with RMF
- Implementation roadmap: governance program build-out timeline, resource requirements, and quick-win priorities
- Governance capstone: design an AI risk governance program for a sample agency with 12 active AI use cases, including a risk tiering framework, new deployment approval workflow, continuous monitoring plan, AI incident response procedure, and 90-day implementation roadmap