NIST Artificial Intelligence Update

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

March 27, 2026

NIST AI Programs and Efforts

Research, Testing, & Evaluation

Guidelines & Standards

Stakeholder Engagement

Center for AI Standards and Innovation (CAISI)

Information Technology Laboratory (ITL)

Enterprise AI

Credit: NIST

NIST CAISI Introduction

Andrew Kane – Chief of Staff, Center for AI Standards and Innovation (CAISI)

CAISI focus areas:

- Support the **diffusion and adoption** of AI throughout the U.S. economy
- Shape the **international environment** to support the adoption of U.S. AI
- Track **strategic competition** between the U.S. and China on AI
- Support the **secure deployment** of U.S. AI.
- **Protect U.S. AI** from adversaries

NIST ITL AI Program Introduction

Mark Przybocki – Chief, Information Access Division, Information Technology Laboratory (ITL)

ITL focus areas:

- Transform the way that AI system trustworthiness is **measured**
- Deliver **resources** that empower others to make informed decisions on AI use
- Position the U.S. to dominate in the AI technical **standards** arena
- Enable **innovative application** of AI in high-priority areas

NIST Enterprise AI Introduction

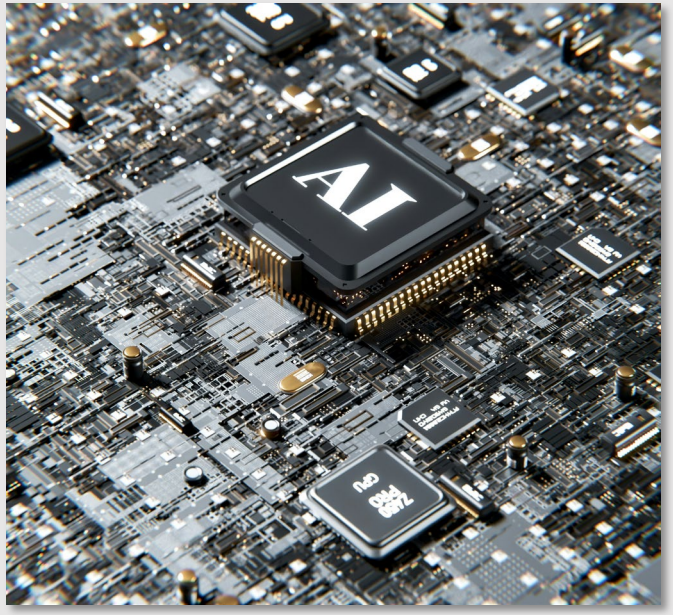
Hannah Brown – Acting Deputy Associate Director for Management Resources

Enterprise AI focus areas:

- Empower scientific discovery
- Enhance operational efficiency and support
- Governance and management of AI resources

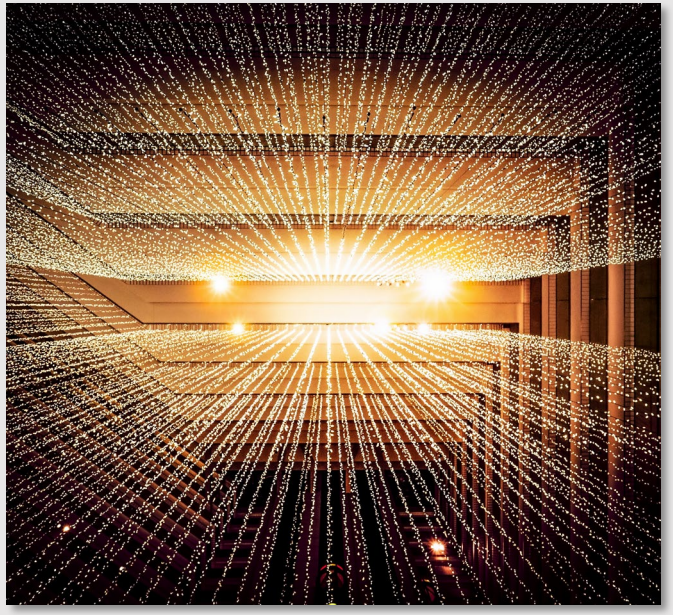
Winning the Race: America's AI Action Plan

The White House released America's AI Action Plan on July 23, 2025, unveiling over **90** recommended Federal policy actions. NIST is directly named in over 25 actions.



Credit: unsplash

Pillar I: Accelerate AI Innovation



Credit: unsplash

Pillar 2: Build American AI Infrastructure

Credit: pixabay

Pillar 3: Lead in International AI Diplomacy and Security

NIST AI Action Plan Alignment

Research, Testing, & Evaluation

- Support the development of **the science of measuring and evaluating AI models**…
- Build, maintain, and update as necessary **national security-related AI evaluations**…
- …conduct research and, as appropriate, publish **evaluations of frontier models** from the People's Republic of China for alignment with Chinese Communist Party talking points and censorship

Guidelines & Standards

- …consider developing **NIST's Guardians of Forensic Evidence** deepfake evaluation program into a formal **guideline** and a companion voluntary forensic **benchmark**
- Create new **technical standards** for high-security AI data centers...
- …revise the **NIST AI Risk Management Framework** to eliminate references to misinformation, Diversity, Equity, and Inclusion, and climate change
- Publish **guidelines and resources**…for Federal agencies to conduct their own evaluations of AI systems…

Engage Stakeholders

- Launch several domain-specific efforts…to convene a broad range of public, private, and academic stakeholders to accelerate the development and adoption of **national standards** for AI systems and to measure how much AI increases productivity at realistic tasks in those domains
- Convene meetings at least twice per year…to **share learnings and best practices** on building AI evaluations.
- …convene the NIST AI Consortium to empower the collaborative establishment of new **measurement science** that will enable the identification of proven, scalable, and interoperable techniques and metrics to promote the development of AI

Research, Testing, and Evaluation

Enterprise AI Program Pillars

AI fit for purpose: to advance NIST's Mission

AI Governance

- Manage AI for the enterprise
- Establish processes for AI implementation and management
- Align with NIST & DOC policies & procedures

AI for Research

- Provide AI resources for the enterprise
- Provide NIST researchers with a catalog of AI tools, products, and services
- Provide resources needed to operate advanced AI technologies

AI for Operations

- Provide enabling AI for the enterprise
- Provide NIST Operations with a catalog of AI tools, products, and services
- Implement workflow analysis to drive effective AI integration

AI Learning & Development

- Build AI adoption
- Realize the benefits of AI in the enterprise
- Encourage experimentation and exploration
- Manage risk
- Upskilling

Enterprise AI Program Activities

AI Governance

- Internal NIST AI Guidance
  - Research, communications and publications, operations
- DOC Coordination
  - AI Integrated Project Team participation, AI inventory, AI strategy and compliance documents
- TOS and License Review
  - "Negotiate" TOS with potential AI providers
  - Legal review of AI provider licenses

AI for Research

- AI Catalog
  - Cloud options
  - On-premise options
- AI Resources
  - Securing access to the resources needed to operate advanced AI technologies

AI for Operations

- Business Productivity
- Meeting Transcription and Notes
- Customer Support Chatbots
- Coding Assistance
- Analytics and Business Intelligence

AI Learning & Development

- Experimentation
  - Hack-a-thons, office hours, workflow sessions
- Upskilling
  - Career-based training
- Training Resources
  - Securing access to training resources
- Community Building
  - NIST AI CoP
  - GSA AI CoP

CoP = Community of Practice; TOS = Terms of Service

Overview of CAISI Testing & Evaluation

Goal: Lead evaluations of frontier U.S. and foreign models

Capability Evaluations

- Automated evaluations of general capabilities in software engineering, general reasoning, and mathematics

National Security Evaluations

- Coordinate the interagency and assess models for dual-use capabilities in cyber, chemical, and biological domains

Rapid-fire Evaluation of PRC Frontier Models

- DeepSeek Report
- DeepSeek Models Advance CCP Narratives

Agent Security Evaluations

- Assess robustness of frontier models to novel attacks specially aimed at agents

AI Action Plan Alignment:

Support the development of the science of measuring and evaluating AI models…

Build, maintain, and update as necessary national security-related AI evaluations…

…conduct research and, as appropriate, publish evaluations of frontier models from the People's Republic of China…

CAISI Partnerships

CAISI has partnerships with frontier labs and third parties that enable evaluations of latest capabilities.

Type of Agreement	Capabilities	Organizations
MOU	• Privileged access to models pre and post deployment • Confidential sharing of evaluation results for improvement of safeguards and development of evaluation methodologies • Waiver of terms of service in order to jailbreak and probe for security issues	3 MOUs with Frontier labs, UK AISI, GSA
CRADAs and DTAs	• Access to private benchmarks • Access to proprietary data to assist tracking of AI diffusion • Ability to share CAISI-improved benchmarks	METR, Scale AI, RAND, SecureBio, Carnegie Mellon, University of Illinois, CSIS, and more, plus 290+ organizations via the AI Consortium CRADA

CAISI National Security Evaluations

CAISI coordinates federal agencies on national security-related AI risks:

- CAISI acts as a central hub for model testing and evaluation within the interagency
- CAISI established and leads the **Testing Risks of AI for National Security (TRAINS) taskforce**, which convenes experts from across USG to address national security concerns in AI development and deployment
- Current Members: CAISI (chair), DHS (including CISA), DOE, DoW, NSA, NIH

Goals:

1. **Operationalize the CAISI-led intragovernmental collaboration** on AI model research, testing and evaluation.
2. **Leverage the unique subject matter expertise from across USG to conduct evaluations** that are valuable in advancing the science and understanding of frontier models across national security domains.
3. **Distribute findings to key partners in the USG national security community** for use in their mission space

AI Action Plan Item: *Build, maintain, and update as necessary national security-related AI evaluations…*

AI Agent Measurement Science

Goal: Lead evaluation and assessment of novel security vulnerabilities

AI Agent Hijacking Evaluations

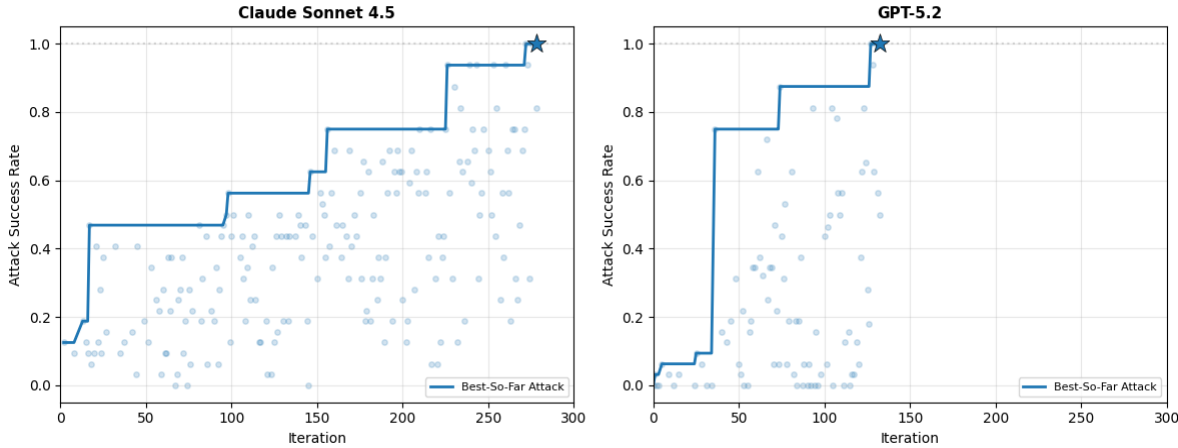
- CAISI built an evaluation environment from scratch that replicates a Google Workspace suite
- CAISI used orchestrated LLMs to generate thousands of realistic user files and data points
- Released a technical blog in January 2025 detailing initial experiments evaluating agent hijacking / prompt injection risk
- Shared insights for other evaluators, and open-sourced improved evaluation codebase

Automated Red-Teaming (ART) System

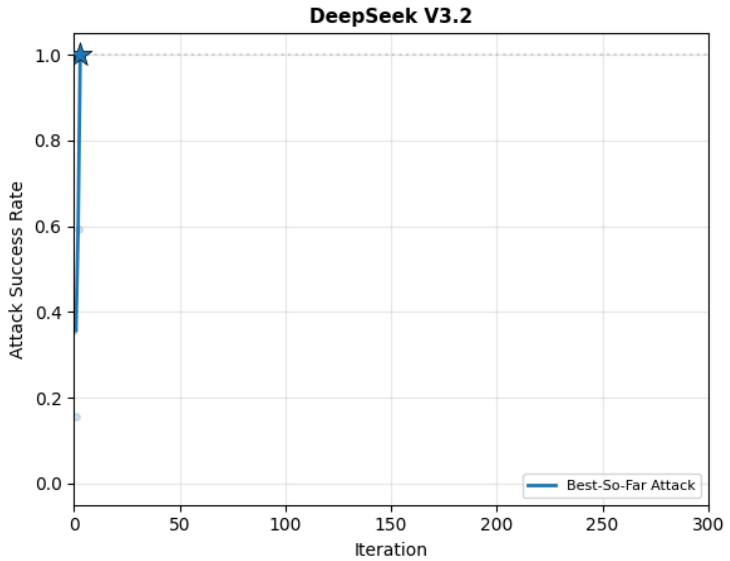
- CAISI built a system that measures the optimization effort needed to exploit different models
- Security benchmarks often overestimate models' robustness against adaptive attackers
- CAISI ART system uses an evolutionary search algorithm to discover and optimize attacks

AI Action Plan Item: *Support the development of the science of measuring and evaluating AI models…*

CAISI's ART system finds strong attacks against leading models…



… And shows how much more vulnerable P.R.C. models are in comparison



Securing AI Agent Systems

Request for Information (RFI) on AI Agent Security Considerations

- Forthcoming RFI on best practices and methodologies for measuring and improving the secure development and deployment of agentic systems
- Received ~400 submissions
- Hosted a workshop to develop taxonomies of tool use by agent systems

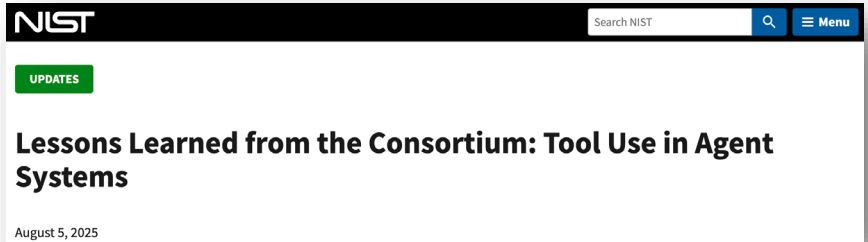
Cheating on AI Agent Evaluations

- Published novel research on how AI agents attempt to cheat on standardized evaluations to score higher

AI Action Plan Item: *Support the development of the science of measuring and evaluating AI models…*

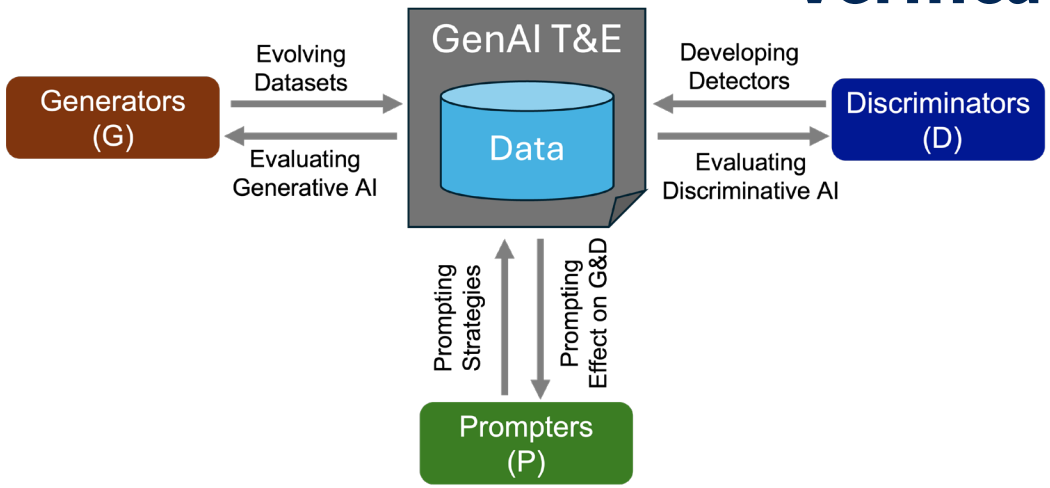
The RFI poses questions on topics including:

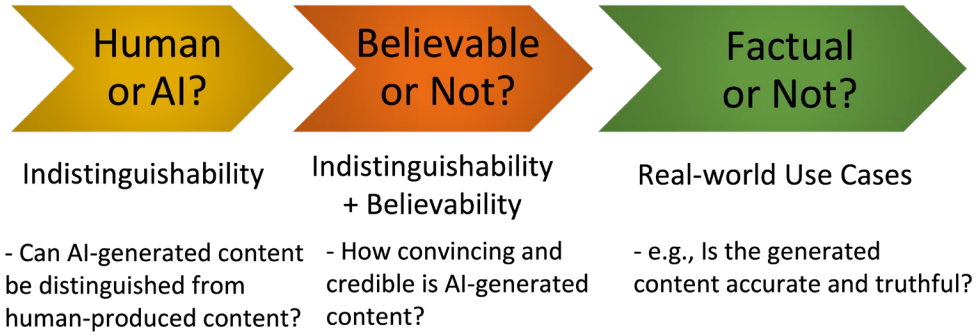
- Unique security threats affecting AI agent systems, and how these threats may change over time.
- Methods for improving the security of AI agent systems in development and deployment.
- Promise of and possible gaps in existing cybersecurity approaches when applied to AI agent systems.
- Methods for measuring the security of AI agent systems and approaches to anticipating risks during development.
- Interventions in deployment environments to address security risks affecting AI agent systems, including methods to constrain and monitor the extent of agent access in the deployment environment.



ITL Evaluation-Driven Research

Goal: Transform the measurement of AI via testing, evaluation, verification, and validation.





GenAI

- Supporting evaluations for research and measurement science in Generative AI across multiple modalities (Text, Image, Video, Code & Deepfakes)

Assessing Risks and Impacts of AI (ARIA)

- Providing resources for organizations to evaluate their systems and informing decision making regarding positive or negative impacts of AI deployment

AI Action Plan Item: *Support the development of the science of measuring and evaluating AI models…*

Guidelines and Standards

AI Data Center Security

Goal: Create new technical standards for high-security AI data centers.

Given AI data centers bear similar architectures to high-performance computing(HPC), NIST will leverage expertise in AI and data center security (e.g., HPC security (NIST SP 800-223/224)).

Planned Approach

- Host a workshop to capture feedback from industry
- Conduct a gap analysis to identify where existing data center standards fall short on issues specific to AI
- Draft a baseline standard (and higher-security add-ons)
- Publish a security overlay using existing NIST cybersecurity resources (e.g., NIST SP 800-53, Cybersecurity Framework)
- Conduct listening sessions with AI companies, think tanks, etc.
- Balance considerations on security, construction latency, operational friction, upgradeability, etc.

AI Action Plan Item: *Create new technical standards for high-security AI data centers, led by DoW, the IC, NSC, and NIST at DOC, including CAISI, in collaboration with industry and, as appropriate, relevant Federally Funded Research and Development Centers.*

Guardians of Forensic Evidence

Goal: Enable consistent examiner assessment of deepfakes and create a benchmark that validates tool reliability, both courtroom ready

Main Challenges

- Lack of U.S. standardized guidelines, checklists or resources
- Limited trust in existing tools due to the lack of benchmarking and transparency

Approach

- January/February 2026: Establish Forensic SMEs pilot working group
- February/March 2026: Outreach to USSS forensic experts informing gap analysis
- September 2026: Release initial challenge kit, with tools, data and benchmark system

AI Action Plan Item: *Led by NIST at DOC, consider developing NIST's Guardians of Forensic Evidence deepfake evaluation program into a formal guideline and a companion voluntary forensic benchmark.*

CAISI Guidelines for AI Evaluations

Goal: Develop guidelines for frontier developers and measurement science advancements for evaluations.

Practices for Automated Evaluations (NIST AI 800-2)

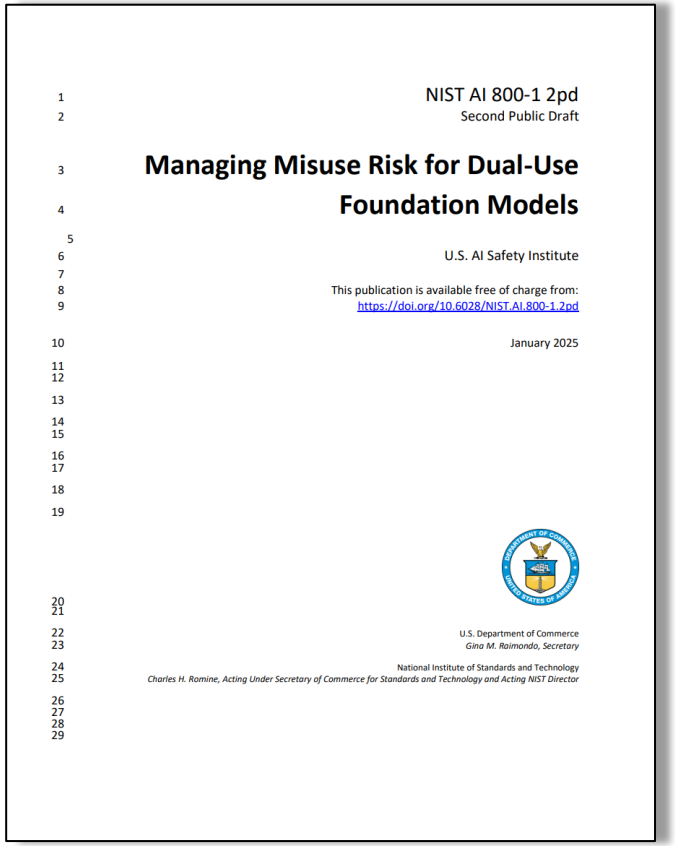
- Released draft guidelines for evaluators on methods to support the validity, transparency, and reproducibility of AI evaluations
- Public comment period ends March 31

Other relevant publications

- NIST AI 800-3 on Statistical Models for Evaluations
- NIST AI 800-4 on Post-Deployment Monitoring

Upcoming:

- Agent Security guidelines
- Guidelines on anti-distillation practices for developers



NIST AI 800-1 2pd
Second Public Draft

Managing Misuse Risk for Dual-Use Foundation Models

U.S. AI Safety Institute

This publication is available free of charge from:
https://doi.org/10.6028/NIST.AI.800-1.2pd

January 2025

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Charles H. Romine, Acting Under Secretary of Commerce for Standards and Technology and Acting NIST Director

AI Action Plan Item: *Publish guidelines and resources…for Federal agencies to conduct their own evaluations of AI systems.*

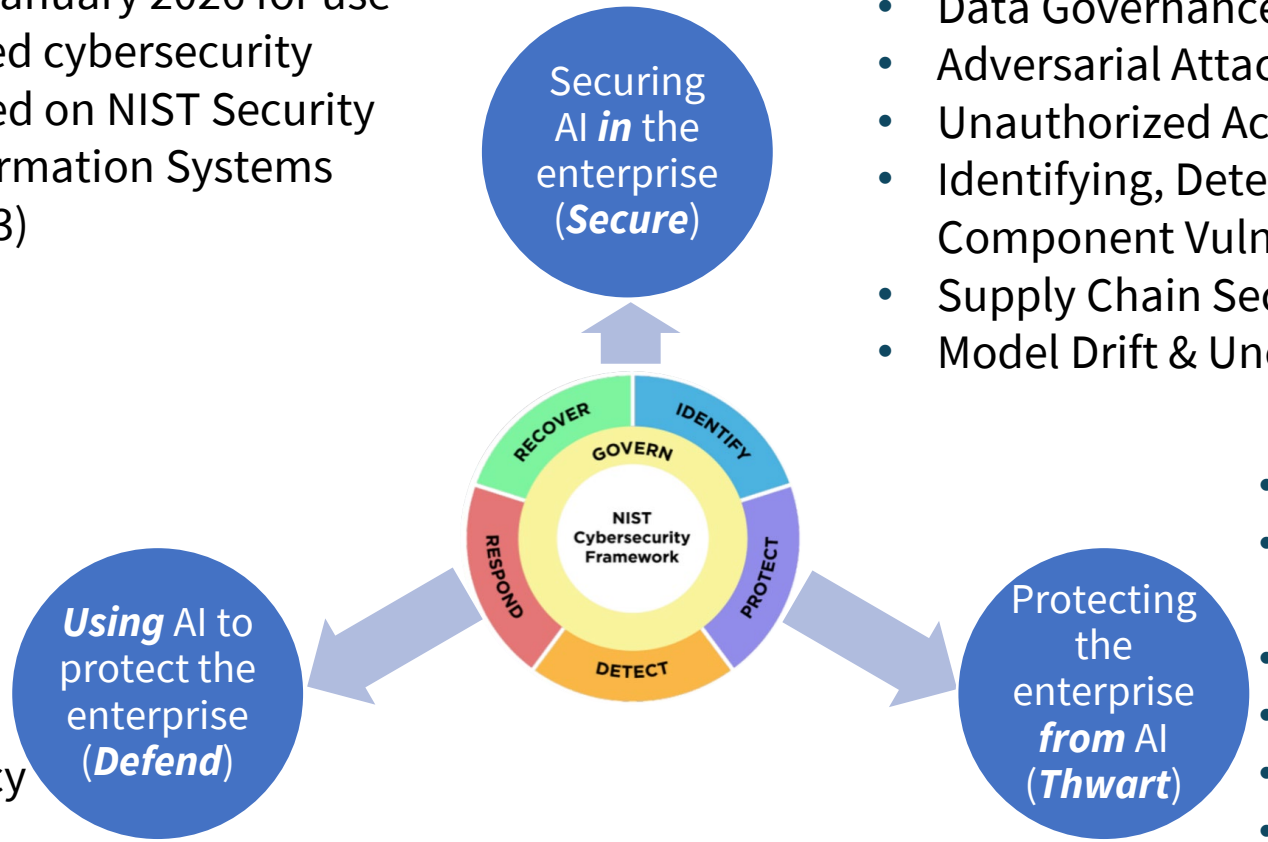
Cyber AI Profile

Goal: Develop an "AI Community Profile" based on the NIST Cybersecurity Framework

Published a draft outline in January 2026 for use case-focused, threat-informed cybersecurity "control overlays" for AI based on NIST Security and Privacy Controls for Information Systems and Organizations (SP 800-53)

Opportunities for AI in Cyber Defense

- Advanced Threat Detection
- Advanced Threat Analysis
- Automated Incident Response
- Proactive Risk Management
- Security Governance & Policy



Challenges in Protecting Enterprise AI

- Data Governance, Security, and Privacy
- Adversarial Attacks
- Unauthorized Access and Use
- Identifying, Detecting, & Responding to AI Component Vulnerabilities and Adverse Events
- Supply Chain Security
- Model Drift & Unexpected or Inaccurate Results

New or Augmented Threats

- Automated/Adaptive Malware
- Targeted Phishing & Social Engineering
- AI-Driven Cyber Espionnage
- Evasion Techniques
- Supply Chain Attacks
- AI-Powered Zero-Day Exploits
- AI-Powered Attack Automation

AI Action Plan Item: *Publish guidelines and resources…for Federal agencies to conduct their own evaluations of AI systems.*

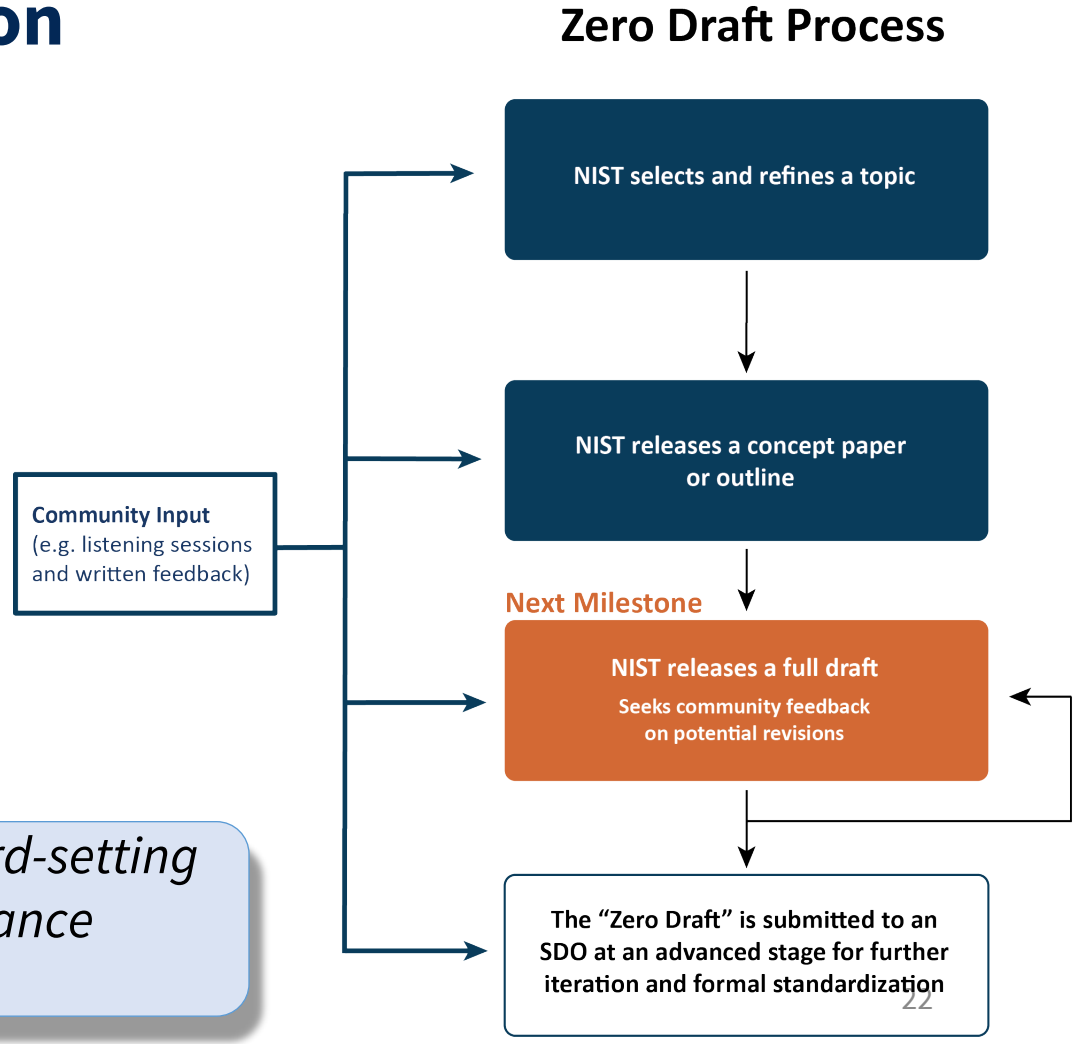
NIST AI Standards "Zero Drafts" Pilot

Goal: Accelerate the creation of AI standards and broaden stakeholder participation

AI Standards "Zero Drafts" Pilot Project

- NIST released two outlines of initial "zero drafts" for public comment:
  - Test Evaluation Verification Validation (TEVV)
  - Documentation of AI Datasets and AI Models
- Next Steps: Release and iterate on full draft text and work with stakeholders for submission to Standards Development Organizations (SDO)

AI Action Plan Item: *…leverage the U.S. position in…standard-setting bodies to vigorously advocate for international AI governance approaches that promote innovation…*



Stakeholder Engagement

Domain-Specific Efforts

Cybersecurity and Manufacturing

- NIST has leveraged existing stakeholder connections to identify challenges, standards development opportunities, and productivity measures
- Pilot projects are being designed to address core industry challenges and inform standards
- **Goal**: Proof of concept pilots will lead to impactful AI adoption and substantial productivity increases by U.S. manufacturing and cybersecurity communities

Healthcare, Education, and Finance

- Co-hosting with OSTP virtual roundtables on barriers to adoption in healthcare, financial services, and education
- Registration is open until Tuesday, March 31st
- Workshops will gather examples and research about successful and unsuccessful adoption
- **Goal**: Impactful roadmap of CAISI AI evaluation projects to lower barriers and publish barriers for other agencies to address.

AI Action Plan Item: *Launch several domain-specific efforts (e.g., in healthcare, energy, and agriculture), led by NIST at DOC, to convene a broad range of public, private, and academic stakeholders to accelerate the development and adoption of national standards for AI systems and to measure how much AI increases productivity at realistic tasks in those domains.*

ITL AI Webinar Series

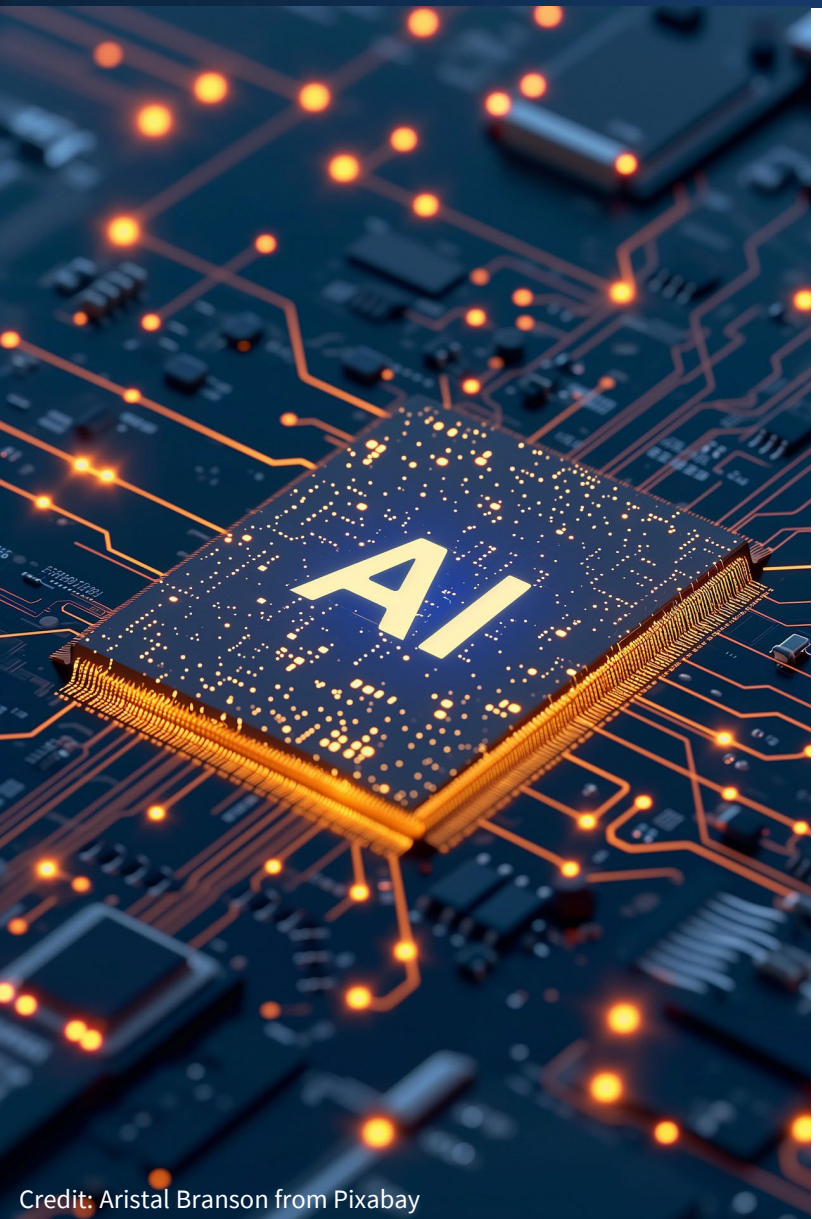
Goal: Share updates on AI-related research and standards activities and solicit community feedback.

- Informs the public about several of ITL's ongoing AI-related efforts
- Educates the public about the AI ecosystem and how NIST's work fits into the broader landscape

Date	Registrants	Topic
March	Over 1700	The International AI Standards Landscape and ITL's Role, Priorities, and Progress
April	1000+	Building Measurement Probes into Agentic AI Ecosystems

AI Action Plan Item: *Convene meetings…to share learnings and best practices on building AI evaluations*

Looking Ahead



Credit: Aristal Branson from Pixabay

- Produce resources, guidelines, and frameworks to advance Administration priorities in high-priority domains, including:
  - Revised NIST AI RMF
  - Re-envisioned NIST AI Consortium

VCAT Feedback

- What strategies should NIST use to engage stakeholders on AI development and adoption?
- What areas of AI not highlighted today are of interest to the VCAT?
- How can NIST accelerate partnerships and recruitment of talented AI researchers?

Questions?