ARTIFICIAL INTELLIGENCE (AI): CYBERSECURITY

JOHN HARRISON, CYBERSECURITY STATE COORDINATOR, CISA

AUGUST 15, 2024

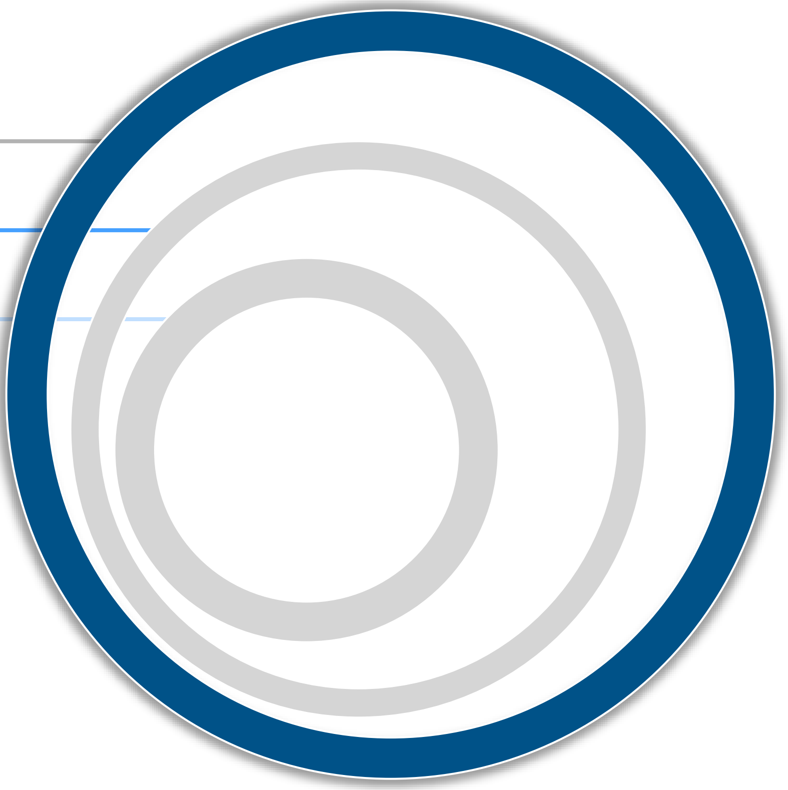
Artificial Intelligence: Overview

The term artificial intelligence (AI) is challenging to define due to constantly evolving technology.

Artificial Intelligence

Definition: Artificial Intelligence is a [machine-based] system that can, for a given set of [human-defined] objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to*:

- Perceive real and virtual environments;
- Abstract such perceptions into models through analysis in an automated manner; and
- Use model inference to formulate options for information or action.



Machine Learning

Large Language Models

Subsets of AI

*National AI Initiative (WH/OSTP)

Machine Learning: Overview

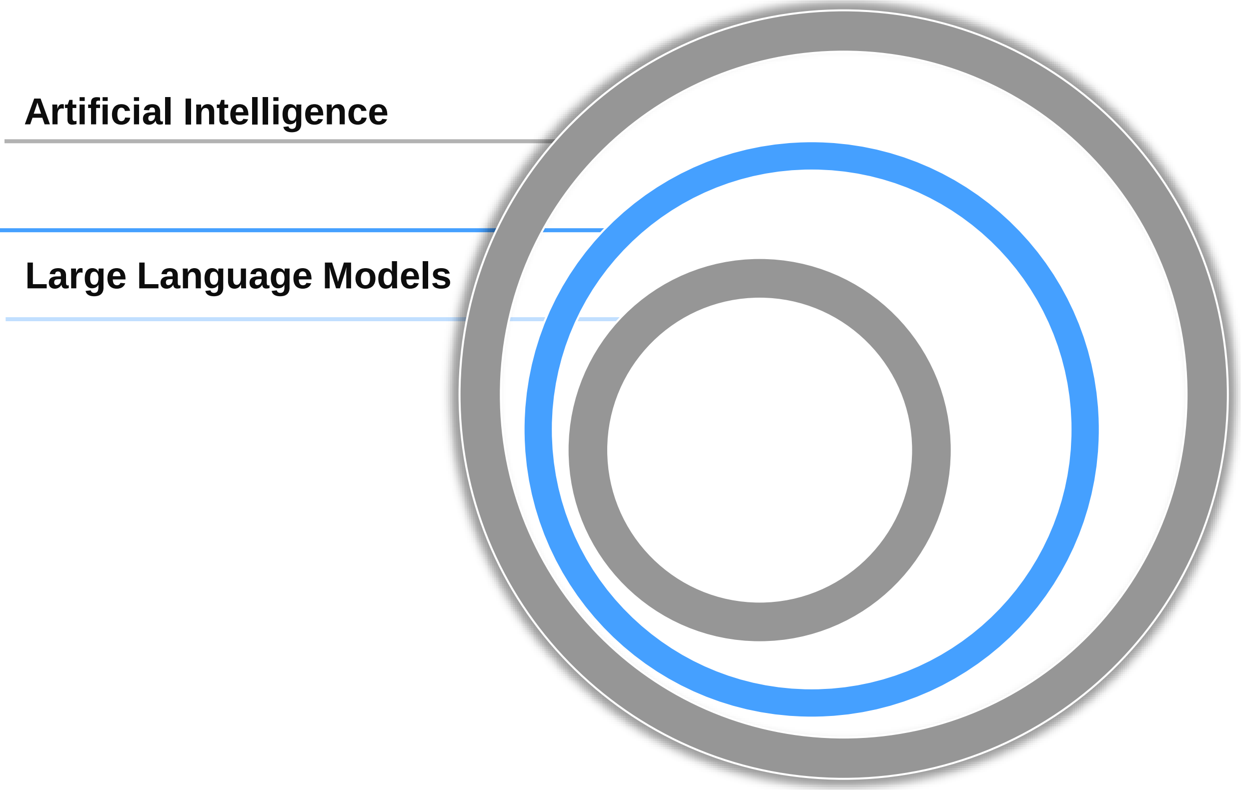
When individuals today are discussing AI, they are often discussing machine learning (ML).

Machine Learning

Training a computer model to understand a representation of data, rather than explicitly incorporating instructions into programming.

Machine Learning is:

- A subset of AI
- Typically does not refer to traditional statistics models, though it may leverage them.



Subsets of AI

Large Language Models: Overview

Large language models (LLMs) are the key to human-AI interaction as their text-based prompts provide human interoperability to other models.

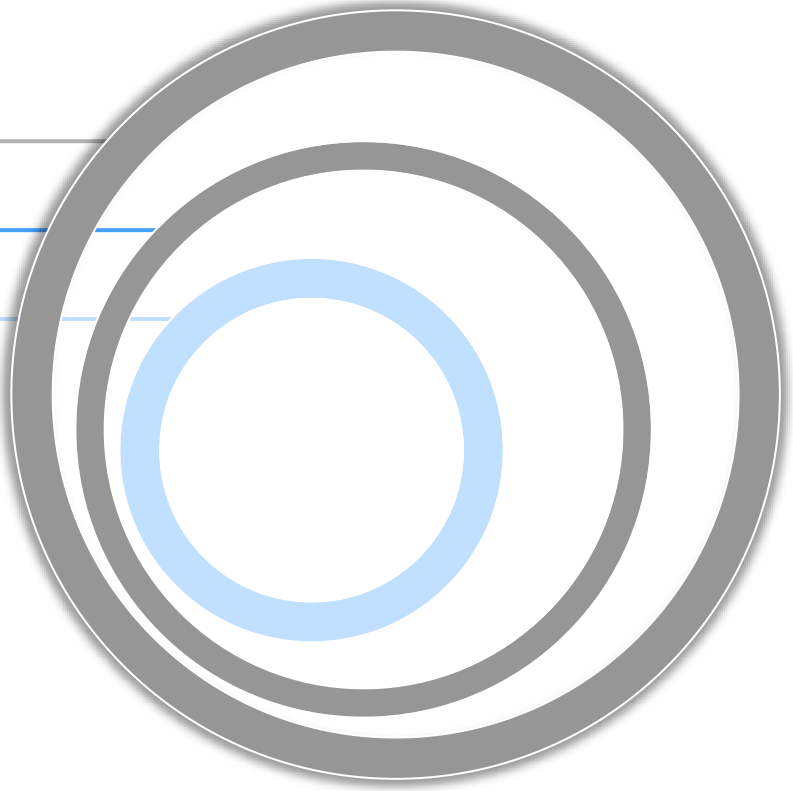
Artificial Intelligence

Machine Learning

Large Language Models

LLMs: A class of language models that use deep learning algorithms and are trained on extremely large textual datasets that can be multiple terabytes in size*.

- Successful LLMs require **fine-tuning**. ChatGPT was fine tuned using a large number of human interactions.
- The **growth curve for LLMs has been exponential** and far exceeds previous technology adoption curves. Change management across USG will be a challenge due to this growth curve.



Subsets of AI

*https://doi.org/10.6028/NIST.AI.100-3

LLM Adoption: Environment

The disruptive LLM environment is rapidly changing.

Exponential Growth: The computational power of LLMs are currently doubling every 6 months*. While this will be an S-Curve**, if the exponential phase extends for a single decade, this will be paradigm shift that society has rarely, if ever, seen. Technology S-Curves can have long lifespans in the exponential phase.

Immeasurable Novelty: The full capabilities of LLMs are currently unknown and new use cases are being discovered daily. Chaining different LLMs as well as multi-modal LLMs is an active area of research that will further upend the existing paradigm.

Practical Usability Challenges: These technologies are rooted in our primary form of communication and thus deceptively easy to learn and use. However, they have confounding aspects for users to master and require deep domain expertise to evaluate the applicability and accuracy of the model's responses.

Leverageable Capabilities: Existing NLP represents a powerful class of tools that can be leveraged to meet an exceptionally broad range of agency requirements.

**https://arxiv.org/abs/2202.05924*
***https://dl.acm.org/doi/pdf/10.1145/3467017*

US AI Policy & Strategy

The U.S. has previously published policy that underlies the AI Security position.

Notable Policy Foundations

AI highlighted as an Administration and congressional priority; CISA will fulfill important coordinating role.

National AI Initiative (NAII) Act of 2020: Coordinated complementary AI R&D, demonstration activities among FCEB, DOD, IC.

AI in Government Act of 2020: Established the AI Center of Excellence within GSA.

EO 13859: Maintaining American Leadership in AI: Established federal principles and strategies to strengthen the nation's capabilities in AI.

EO 13960: Promoting the Use of Trustworthy AI in the Federal Gov't: Required Agencies to inventory and share AI use cases.

Focused Strategy Development

Articulation of DHS/CISA and broader FCEB strategic needs, as a way of operationalizing on the need for AI.

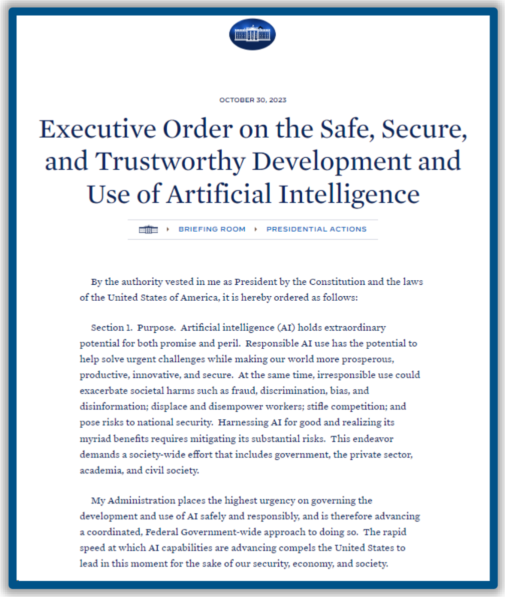
DHS AI Strategy: Strategic vison for DHS role in policy development, governance, use of AI, and risk mitigation.

Implementation Plan for a National AI Research Resource: Memorializes findings of National Artificial Intelligence Research Resource (NAIRR) Task Force on national AI research infrastructure.

OGA Strategies: FDA, Nuclear Regulatory Commission, and others have released AI strategies tailored to specific mission areas.

National Priorities for AI RFI: OSTP RFI on key themes to inform Administration's updated National AI Strategy.

Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI



OCTOBER 30, 2023

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Purpose. Artificial intelligence (AI) holds extraordinary potential for both promise and peril. Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security. Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society.

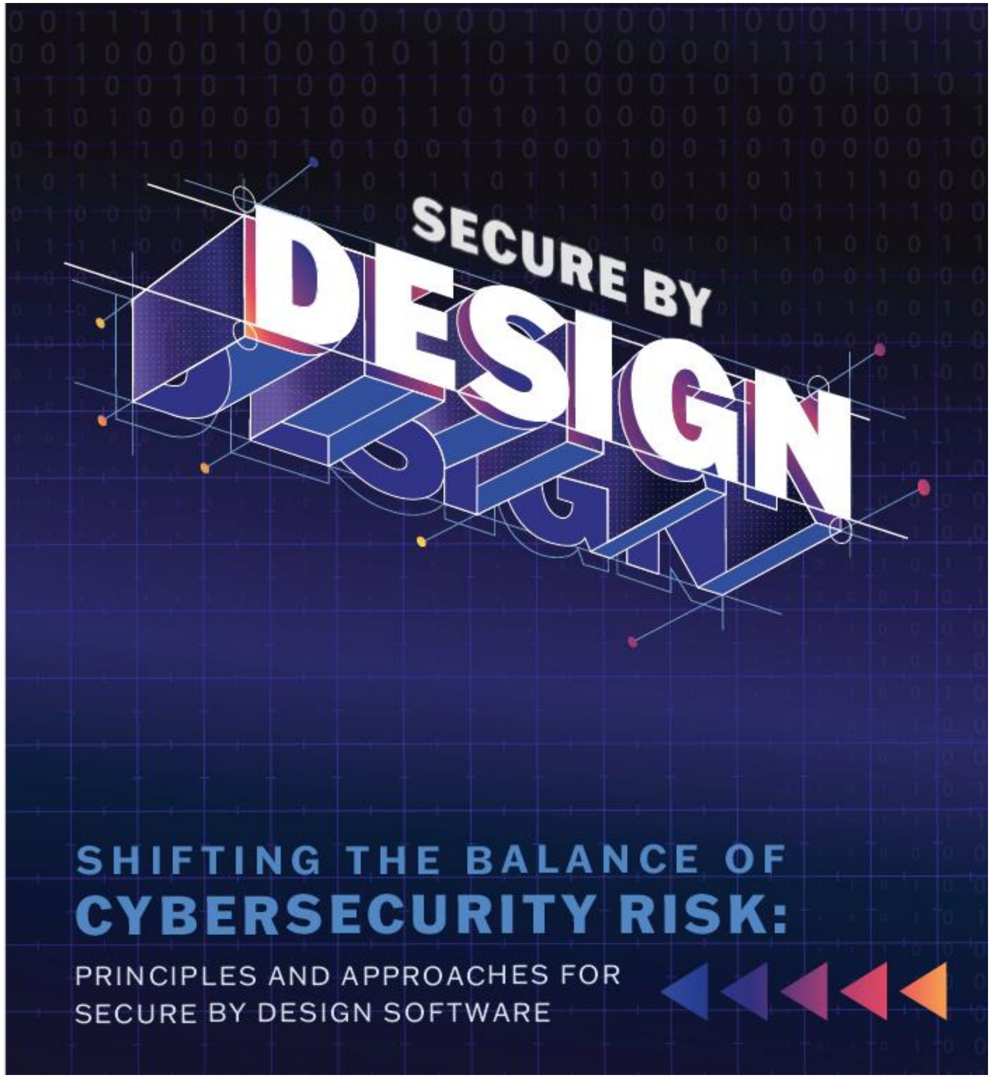
My Administration places the highest urgency on governing the development and use of AI safely and responsibly, and is therefore advancing a coordinated, Federal Government-wide approach to doing so. The rapid speed at which AI capabilities are advancing compels the United States to lead in this moment for the sake of our security, economy, and society.

CISA’s Mission

- To defend against today’s threats and collaborate to build a more secure and resilient infrastructure for the future

- To serve as the operational lead for federal cybersecurity and the national coordinator for critical infrastructure security and resilience

AI security risks echo traditional software risks

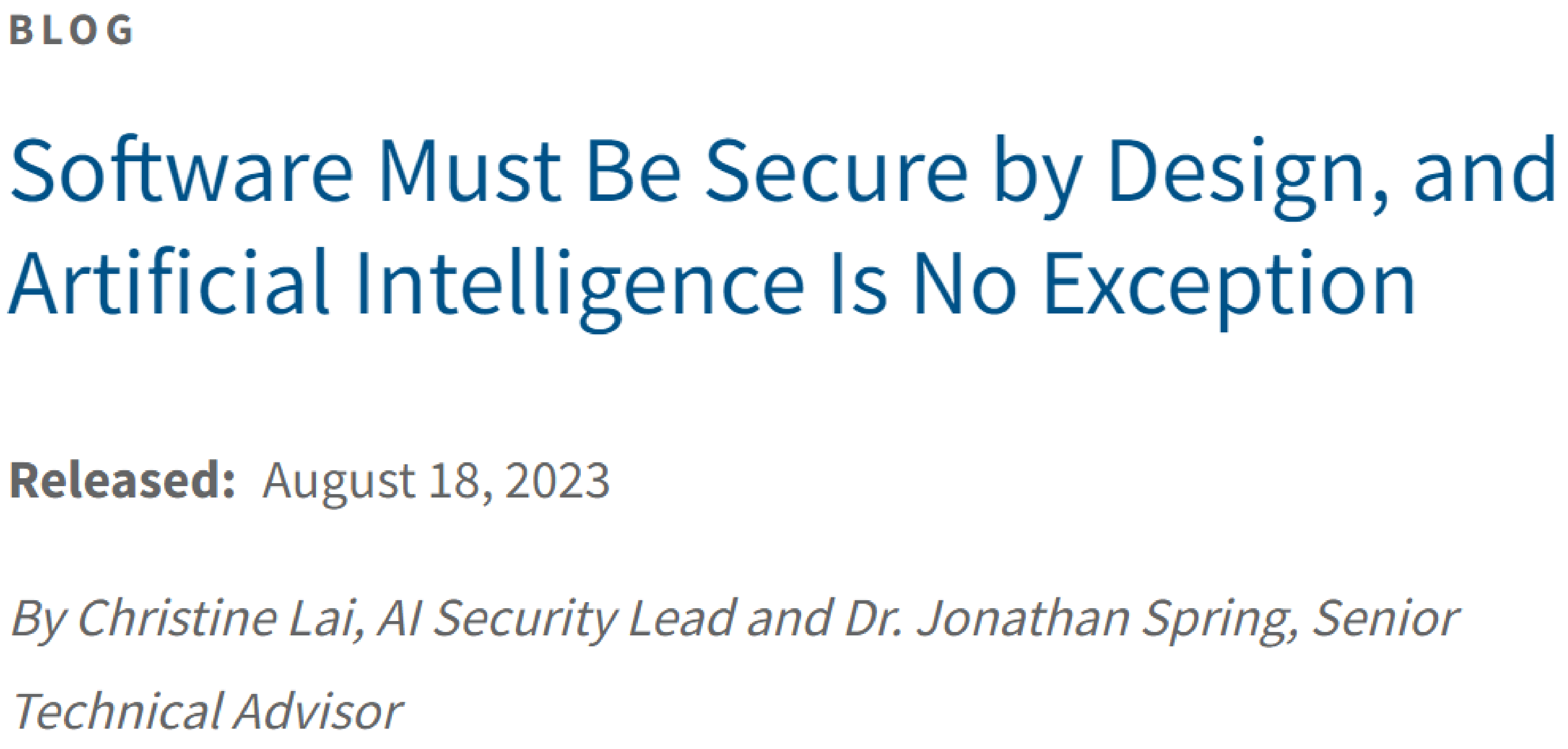


SECURE BY DESIGN

SHIFTING THE BALANCE OF CYBERSECURITY RISK:

PRINCIPLES AND APPROACHES FOR
SECURE BY DESIGN SOFTWARE

AI is Powerful Software



BLOG

Software Must Be Secure by Design, and Artificial Intelligence Is No Exception

Released: August 18, 2023

By Christine Lai, AI Security Lead and Dr. Jonathan Spring, Senior Technical Advisor

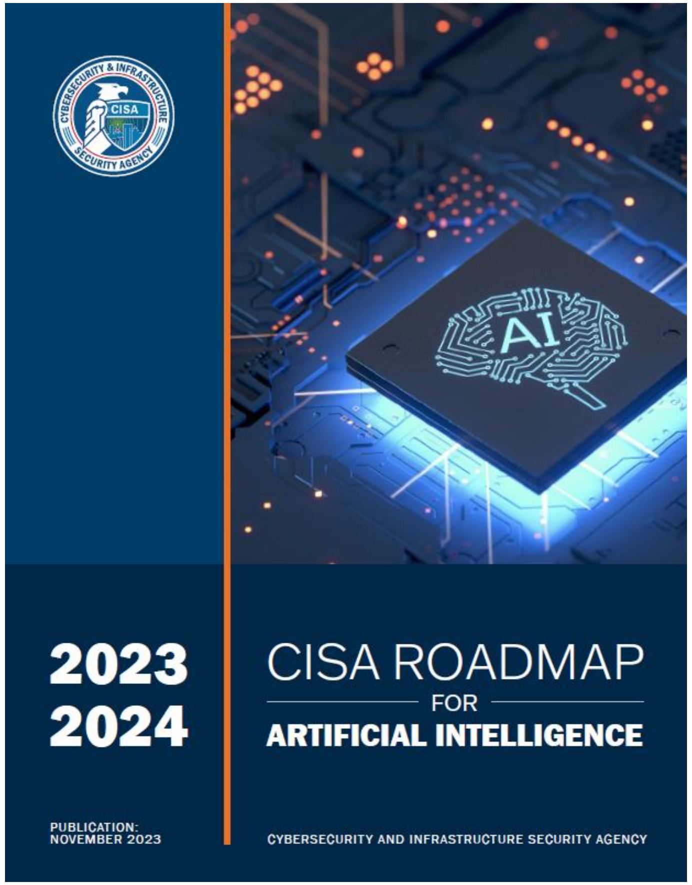
CISA Roadmap for AI

Purpose

CISA's AI Roadmap is a whole-of-agency plan aligned with national AI strategy to align our cross-agency efforts and communicate our role in AI safety and security.

Areas of Focus

1. Promote the beneficial uses of AI to **enhance cybersecurity capabilities**.
2. Ensure **AI systems are protected from cyber-based threats**.
3. **Deter the malicious use of AI capabilities** to threaten the critical infrastructure Americans rely on every day.



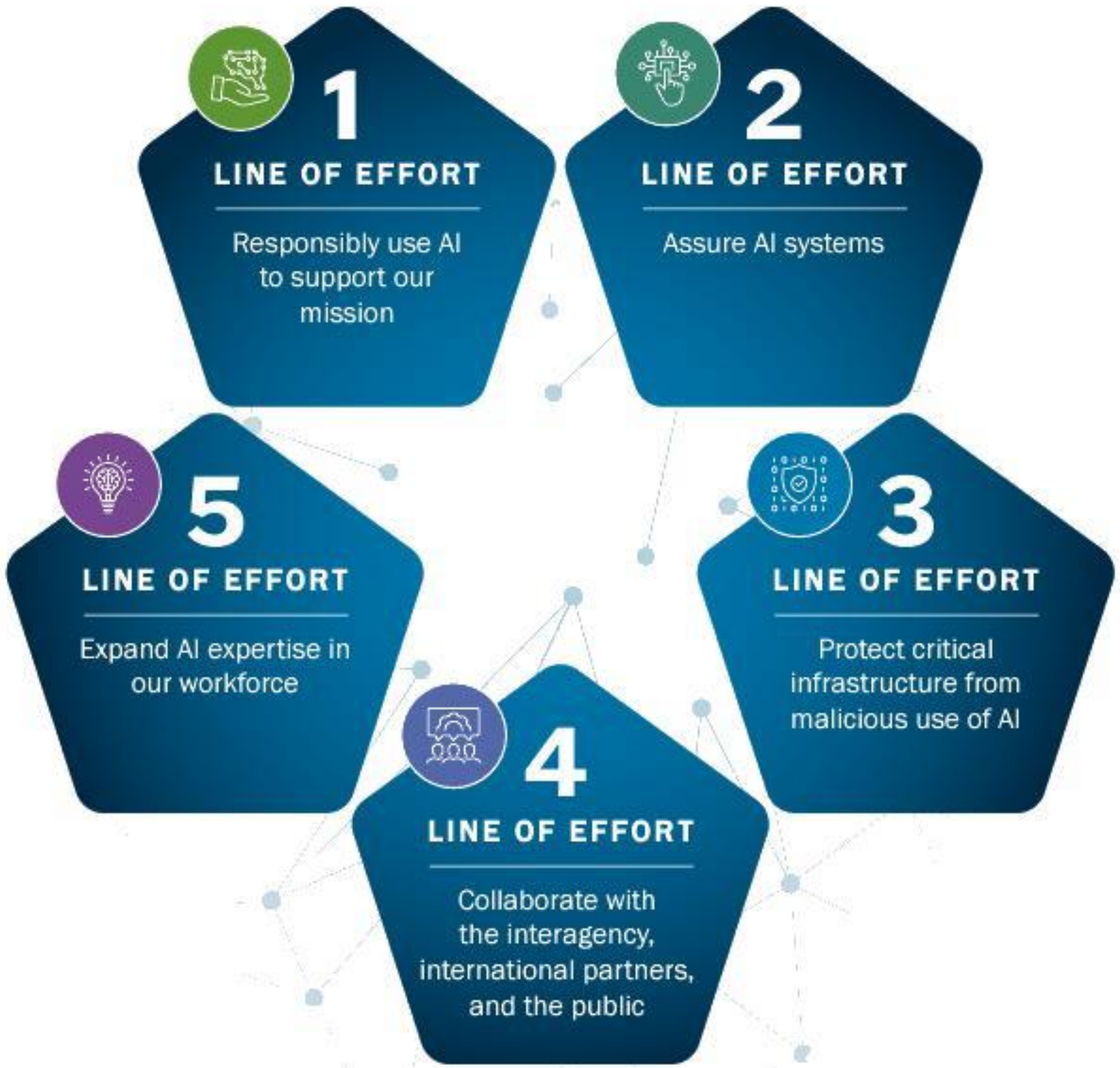
2023
2024

CISA ROADMAP
FOR
ARTIFICIAL INTELLIGENCE

PUBLICATION:
NOVEMBER 2023

CYBERSECURITY AND INFRASTRUCTURE SECURITY AGENCY

CISA AI Lines of Effort



1
LINE OF EFFORT
Responsibly use AI to support our mission

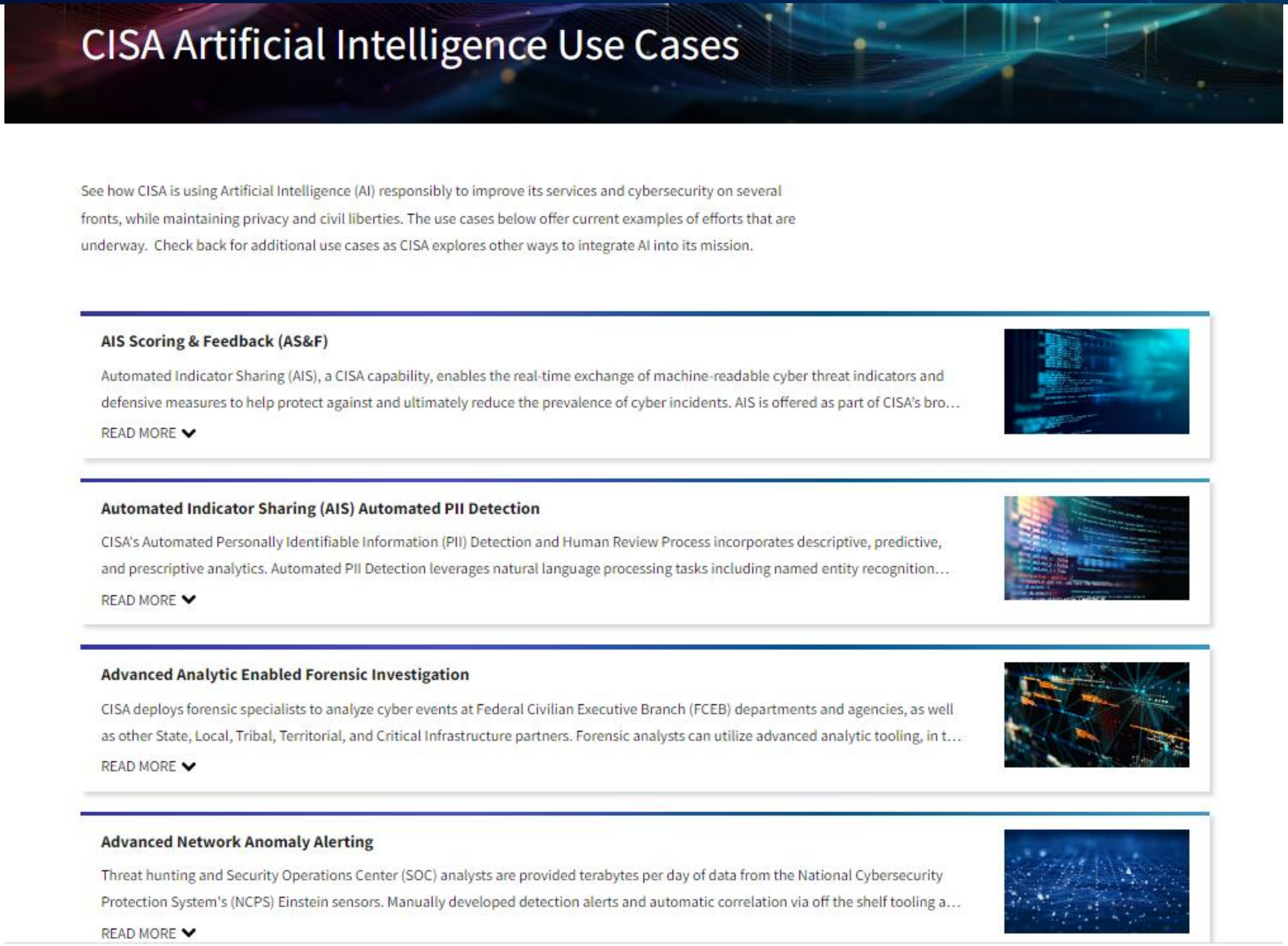
2
LINE OF EFFORT
Assure AI systems

3
LINE OF EFFORT
Protect critical infrastructure from malicious use of AI

4
LINE OF EFFORT
Collaborate with the interagency, international partners, and the public

5
LINE OF EFFORT
Expand AI expertise in our workforce

AI Use Cases



CISA Artificial Intelligence Use Cases

See how CISA is using Artificial Intelligence (AI) responsibly to improve its services and cybersecurity on several fronts, while maintaining privacy and civil liberties. The use cases below offer current examples of efforts that are underway. Check back for additional use cases as CISA explores other ways to integrate AI into its mission.

AIS Scoring & Feedback (AS&F)

Automated Indicator Sharing (AIS), a CISA capability, enables the real-time exchange of machine-readable cyber threat indicators and defensive measures to help protect against and ultimately reduce the prevalence of cyber incidents. AIS is offered as part of CISA's bro…

READ MORE

Automated Indicator Sharing (AIS) Automated PII Detection

CISA's Automated Personally Identifiable Information (PII) Detection and Human Review Process incorporates descriptive, predictive, and prescriptive analytics. Automated PII Detection leverages natural language processing tasks including named entity recognition…

READ MORE

Advanced Analytic Enabled Forensic Investigation

CISA deploys forensic specialists to analyze cyber events at Federal Civilian Executive Branch (FCEB) departments and agencies, as well as other State, Local, Tribal, Territorial, and Critical Infrastructure partners. Forensic analysts can utilize advanced analytic tooling, in t…

READ MORE

Advanced Network Anomaly Alerting

Threat hunting and Security Operations Center (SOC) analysts are provided terabytes per day of data from the National Cybersecurity Protection System's (NCPS) Einstein sensors. Manually developed detection alerts and automatic correlation via off the shelf tooling a…

READ MORE

AI Security: Risks and Threats

There are a variety of threats that are actively being identified in the wild – these recent examples signify the relevance of the current discussion.

Term	Description	Examples
Confidentiality	Risks associated with data privacy and security, including the potential for sensitive information to be inadvertently shared or used inappropriately.	*OpenAI: ChatGPT payment data leak caused by open-source bug (bleepingcomputer.com)*
Supply Chain	Risks associated with reliance on third-party providers for AI systems and dependencies.	*Compromised PyTorch-nightly dependency chain between December 25th and December 30th, 2022. | PyTorch*
Adversarial Use of AI	Risks associated with threat actors leveraging AI to enhance the sophistication of their operations.	*ChatGPT Powered Malware Bypasses EDR | by David Merian | Mar, 2023 | System Weakness* *Disinformation Researchers Raise Alarms About A.I. Chatbots - The New York Times (nytimes.com)*
Adversarial Machine Learning (AML)	Providing deceptive inputs to a machine learned model to cause it to behave in an unexpected fashion	Prompt Injection Attack on GPT-4

AI in Critical Infrastructure

Critical Infrastructure Sector	Relevant AI Enabled Technology
Chemical	Plant Automation
Commercial Facilities	Facial Recognition
Communications	Satellite Tracking
Critical Manufacturing	Supply Chain Analysis
Dams	Flow Controls
Defense Industrial Base	Nonproliferation Monitoring
Emergency Services	Response Routing
Energy	Grid Stabilization
Financial Services	Inflation Prediction
Food and Agriculture	Crop Yield Models
Government Facilities	Access Control
Healthcare and Public Health	Medical Diagnostics
Information Technology	Intrusion Prevention
Nuclear Reactors, Materials, and Waste	Nuclear Waste Monitoring
Transportation Systems	Air Traffic Control
Water and Wastewater Systems	Wastewater Treatment

Interagency and International Collaboration

DHS AI Task Force

Coordination across government

AI security guidance with international partners

A few recent developments

Voluntary Commitments

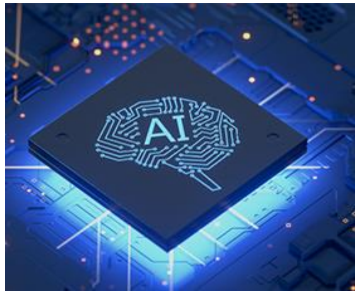
July 2023

Executive Order 14110

October 2023

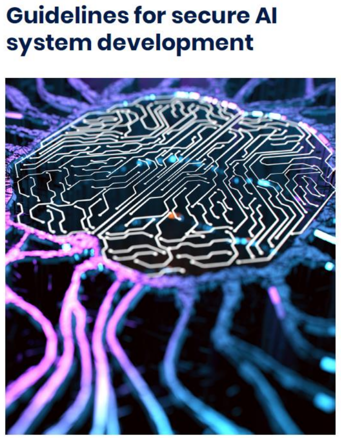
UK AI Safety Summit

November 2023



CISA AI Roadmap

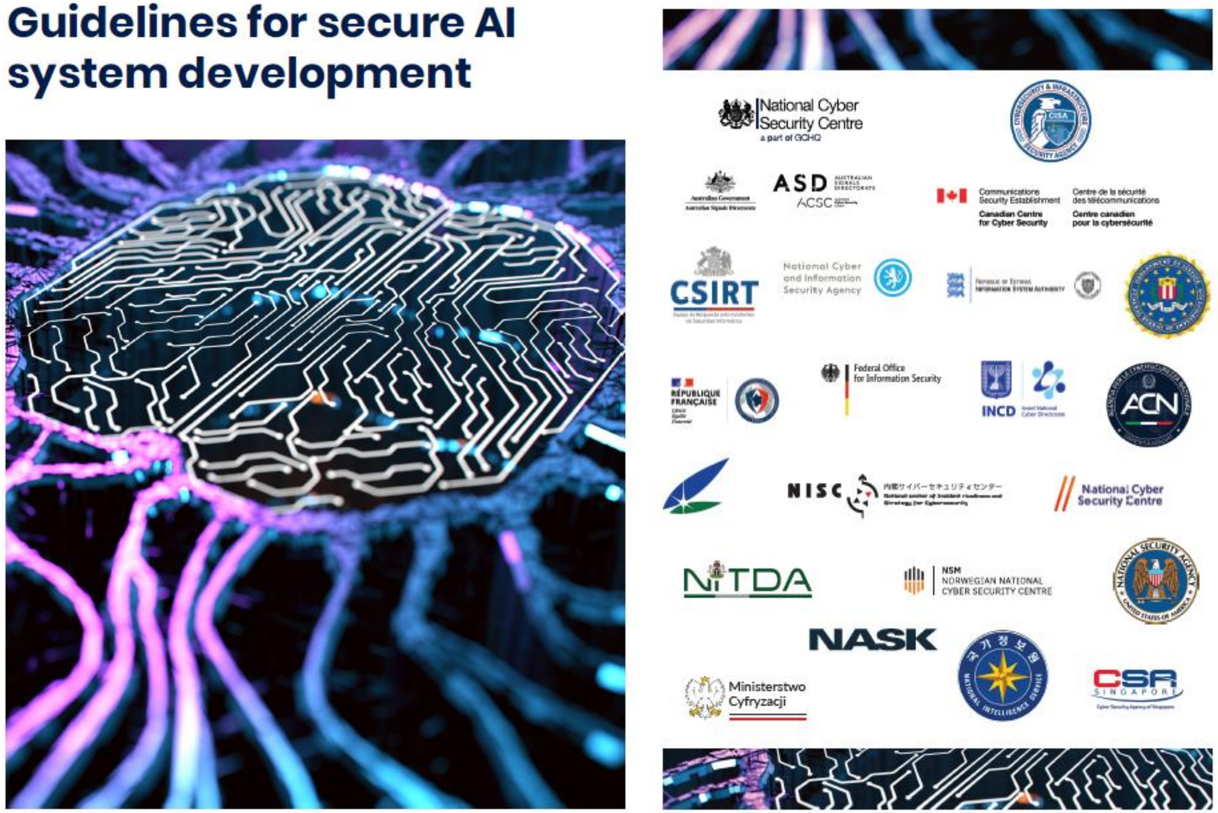
November 2023


Guidelines for secure AI system development

Guidelines for Secure AI System Development
November 2023

Guidelines for Secure AI System Development

- Co-authored with UK NCSC
- Co-sealed with 21 additional international agencies from 18 countries including all of the G7
- Developed in collaboration with industry
- Broken into four key areas:
  - Secure design
  - Secure development
  - Secure deployment
  - Secure operation and maintenance



Guidelines for secure AI system development

National Cyber Security Centre
a part of GCHQ
ASD
ACSC
Communications Security Establishment
Centre de la sécurité des télécommunications
Canadian Centre for Cyber Security
Centre canadien pour la cybersécurité
CSIRT
National Cyber and Information Security Agency
RÉPUBLIQUE FRANÇAISE
Federal Office for Information Security
INCD
ACN
NISC
内閣サイバーセキュリティセンター
National Cyber Security Centre
NITDA
NSM
NORWEGIAN NATIONAL CYBER SECURITY CENTRE
NASK
Ministerstwo Cyfryzacji
CSR
SINGAPORE

Looking Forward

CISA Opportunities

- CISA partnership opportunities https://www.cisa.gov/doing-business-cisa
- Check out DHS Acquisition Planning Forecast System (APFS) https://apfs-cloud.dhs.gov/ to learn about CISA upcoming requirements and email APFS-inquiries@cisa.dhs.gov for questions.
- July 29, 2024: CISA Pilot for Artificial Intelligence Enabled Vulnerability Detection
  - From late 2023 to early 2024, CISA performed an operational pilot to examine whether current vulnerability detection software products that use AI, including large language models (LLMs), are more effective at detecting vulnerabilities than those that do not use AI.
- Check out MITRE's Atlas™ AI Security 101
  - https://atlas.mitre.org/resources/ai-security-101

Questions?

Contact Information:

- John.Harrison@cisa.dhs.gov
- CISA.IOD.REGION.R03_Cyber_Security@cisa.dhs.gov

CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY

CISA