mnccc

Minnesota Counties Computer Cooperative

Minnesota

BCA

Bureau of Criminal Apprehension

ISSG Special Topic Showcase

on Copilot and AI

AI Risks & FBI CJIS Data

Carolyn Geason-Beissel/MIT SMR | Getty Images

whoami

- Jodie Monette, CJIS Systems Agency Information Security Officer.

Agenda

- Risks
- Risk Assessments (RA) FBI CJIS Security Policy
- AI and FBI CJIS Data

Artificial Intelligence (AI) Risk

- Conventional Cyber Security Risk
- BIAS Risk
- Privacy Risk
- Regulatory Risk
- Intellectual Property Risk
- Reputational Risk

Artificial Intelligence (AI) Risk Assessment (RA)

What is your level of Risk for your County/City?

- Low
- Medium
- High

Artificial Intelligence (AI) Risk Assessment (RA) FBI CJIS Security Policy

RA – 1 Policy and Procedures

RA – 2 Security Categorization

RA – 3 Risk Assessment

RA – 5 Vulnerability Monitoring and Scanning

- (2) updated vulnerabilities to be scanned (24h)
- (5) Privileged Access
- (11) Public Disclosure Program

RA – 9 Criticality Analysis

Artificial Intelligence (AI) Risk Assessment (RA) - Data flow of AI

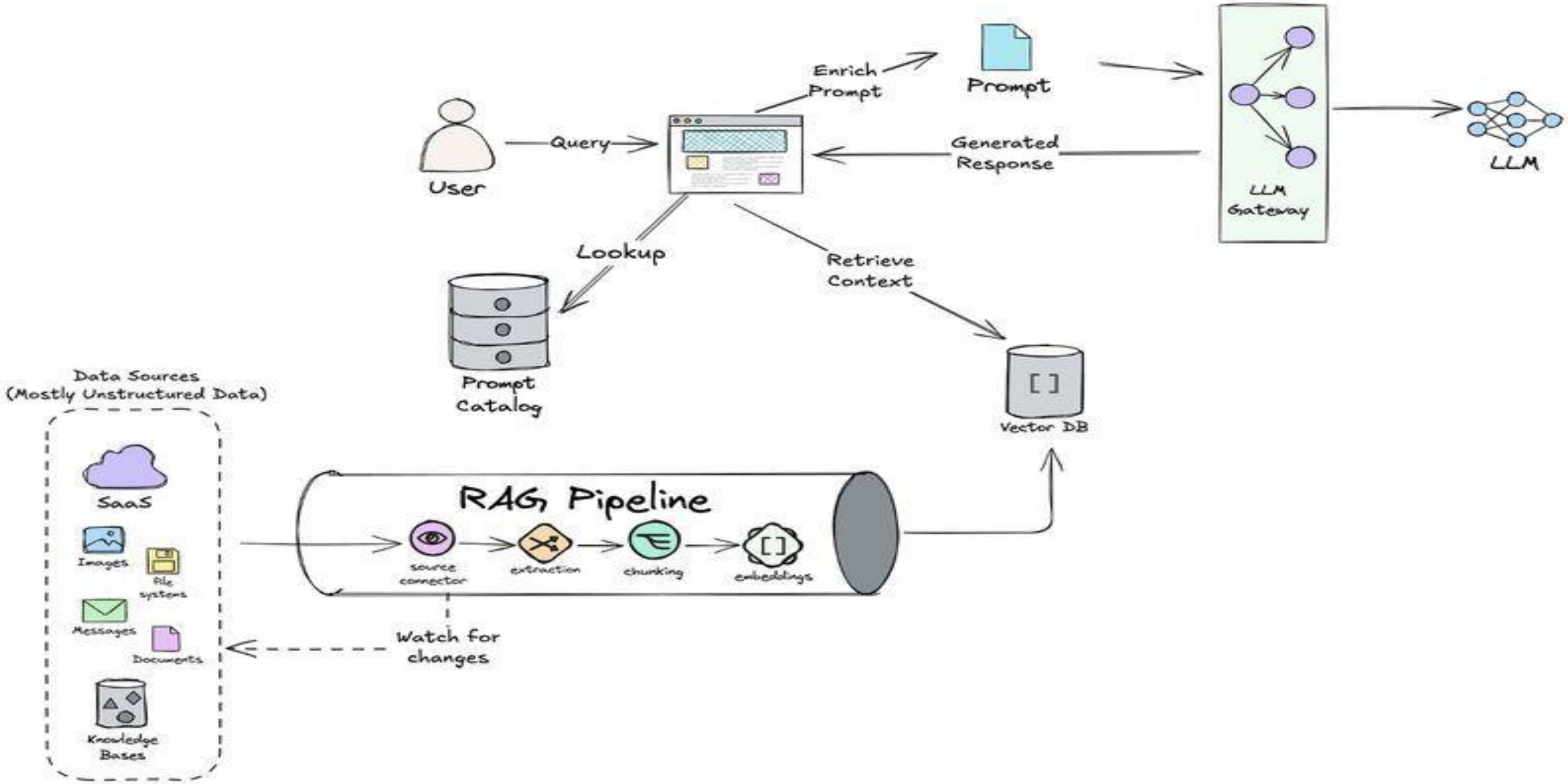


Diagram by Sid Sriram

Artificial Intelligence (AI) Risk Assessment (RA) - OWASP Top 10 for LLM

LLM01:2025 Prompt Injection

LLM02:2025 Sensitive Information Disclosure

LLM03:2025 Supply Chain

LLM04: Data and Model

LLM05:2025 Improper Output Handling

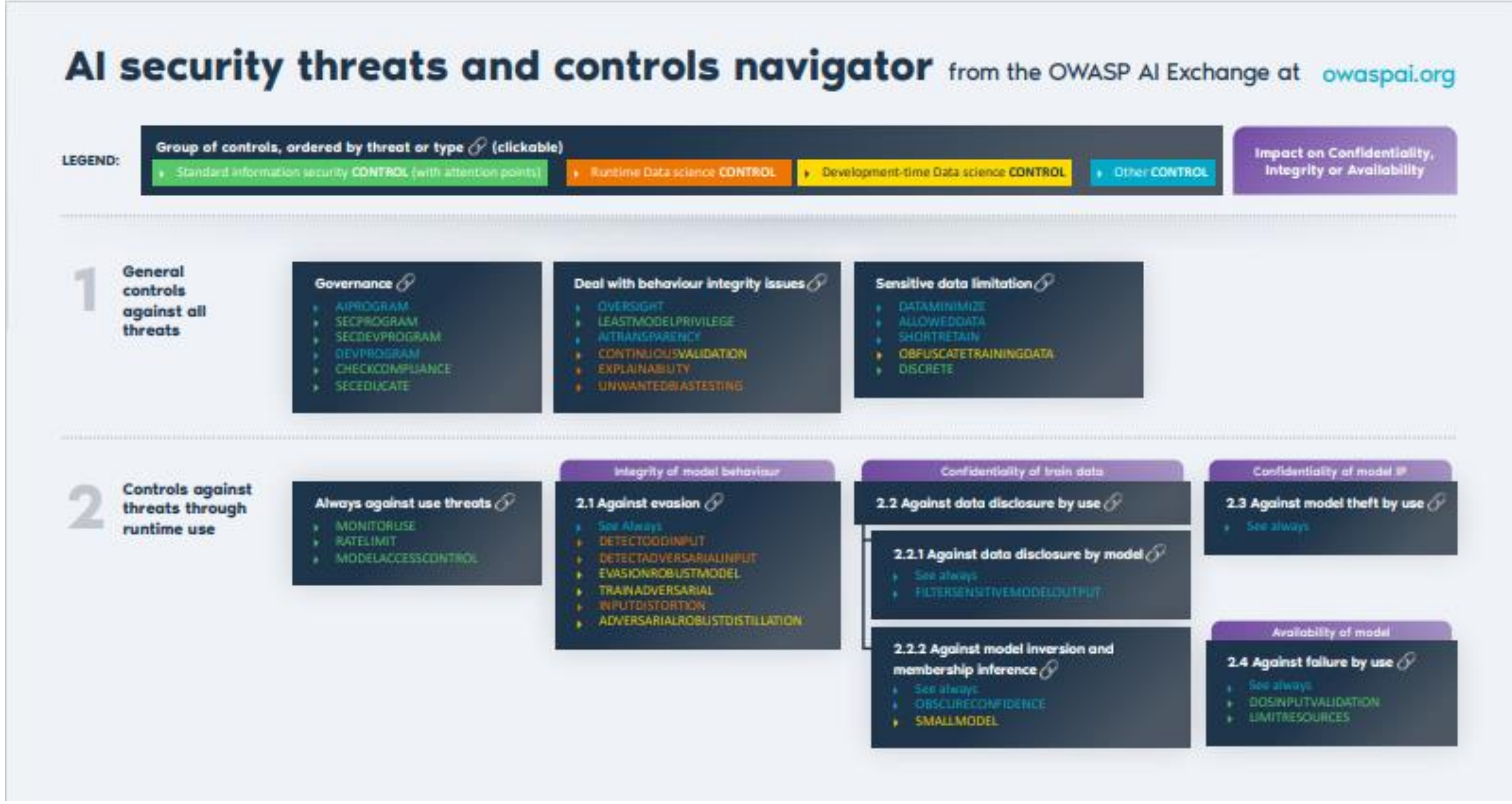
Artificial Intelligence (AI) Risk Assessment (RA) - OWASP Top 10 for LLM

LLM06:2025 Excessive Agency
LLM07:2025 System Prompt Leakage
LLM08:2025 Vector and Embedding Weaknesses
LLM09:2025 Misinformation
LLM10:2025 Unbounded Consumption

Artificial Intelligence (AI) Risk Assessment (RA) - OWASP AI Exchange

- *resource for broad AI security & privacy - over 200 pages of practical advice and references on protecting AI and data-centric systems from threats.*

Artificial Intelligence (AI) Risk Assessment (RA) - OWASP AI Exchange – Periodic table of AI security



AI security threats and controls navigator from the OWASP AI Exchange at owaspai.org

LEGEND: Group of controls, ordered by threat or type 🔗 (clickable)

- Standard information security CONTROL (with attention points)
- Runtime Data science CONTROL
- Development-time Data science CONTROL
- Other CONTROL

Impact on Confidentiality, Integrity or Availability

1 General controls against all threats

Governance 🔗

- AIPROGRAM
- SECPROGRAM
- SECDEVPROGRAM
- DEVPROGRAM
- CHECKCOMPLIANCE
- SECEDUCATE

Deal with behaviour integrity issues 🔗

- OVERSIGHT
- LEASTMODELPRIVILEGE
- AITRANSPARENCY
- CONTINUOUSVALIDATION
- EXPLAINABILITY
- UNWANTEDBIASTESTING

Sensitive data limitation 🔗

- DATAMINIMIZE
- ALLOWEDDATA
- SHORTRETAIN
- OBFUSCATETRAININGDATA
- DISCRETE

2 Controls against threats through runtime use

Always against use threats 🔗

- MONITORUSE
- RATELIMIT
- MODELACCESSCONTROL

Integrity of model behaviour

2.1 Against evasion 🔗

- See Always
- DETECTODDINPUT
- DETECTADVERSARIALINPUT
- EVASIONROBUSTMODEL
- TRAINADVERSARIAL
- INPUTDISTORTION
- ADVERSARIALROBUSTDISTILLATION

Confidentiality of train data

2.2 Against data disclosure by use 🔗

2.2.1 Against data disclosure by model 🔗

- See always
- FILTERSENSITIVEMODELOUTPUT

2.2.2 Against model inversion and membership inference 🔗

- See always
- OBSCURECONFIDENCE
- SMALLMODEL

Confidentiality of model IP

2.3 Against model theft by use 🔗

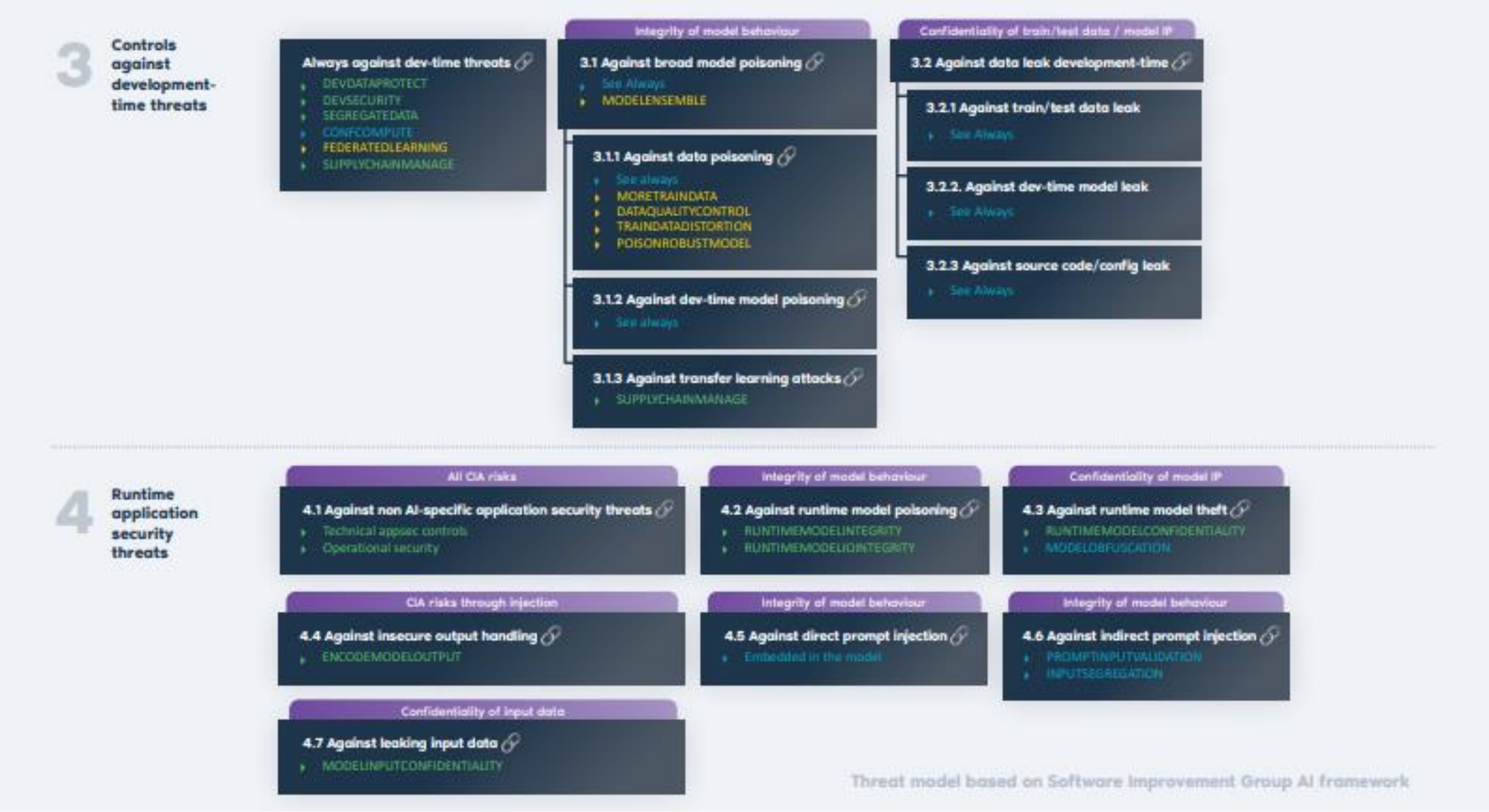
- See always

Availability of model

2.4 Against failure by use 🔗

- See always
- DOSINPUTVALIDATION
- LIMITRESOURCES

Artificial Intelligence (AI) Risk Assessment (RA) - OWASP AI Exchange – Periodic table of AI security



3 Controls against development-time threats

Always against dev-time threats
- DEVDATAPROTECT
- DEVSECURITY
- SEGREGATEDATA
- CONFCOMPUTE
- FEDERATEDLEARNING
- SUPPLYCHAINMANAGE

Integrity of model behaviour

3.1 Against broad model poisoning
- See Always
- MODELENSEMBLE

3.1.1 Against data poisoning
- See always
- MORETRAINDATA
- DATAQUALITYCONTROL
- TRAINDATADISTORTION
- POISONROBUSTMODEL

3.1.2 Against dev-time model poisoning
- See always

3.1.3 Against transfer learning attacks
- SUPPLYCHAINMANAGE

Confidentiality of train/test data / model IP

3.2 Against data leak development-time

3.2.1 Against train/test data leak
- See Always

3.2.2. Against dev-time model leak
- See Always

3.2.3 Against source code/config leak
- See Always

4 Runtime application security threats

All CIA risks

4.1 Against non AI-specific application security threats
- Technical appsec controls
- Operational security

Integrity of model behaviour

4.2 Against runtime model poisoning
- RUNTIMEMODELINTEGRITY
- RUNTIMEMODELIOINTEGRITY

Confidentiality of model IP

4.3 Against runtime model theft
- RUNTIMEMODELCONFIDENTIALITY
- MODELOBFUSCATION

CIA risks through injection

4.4 Against insecure output handling
- ENCODEMODELOUTPUT

Integrity of model behaviour

4.5 Against direct prompt injection
- Embedded in the model

Integrity of model behaviour

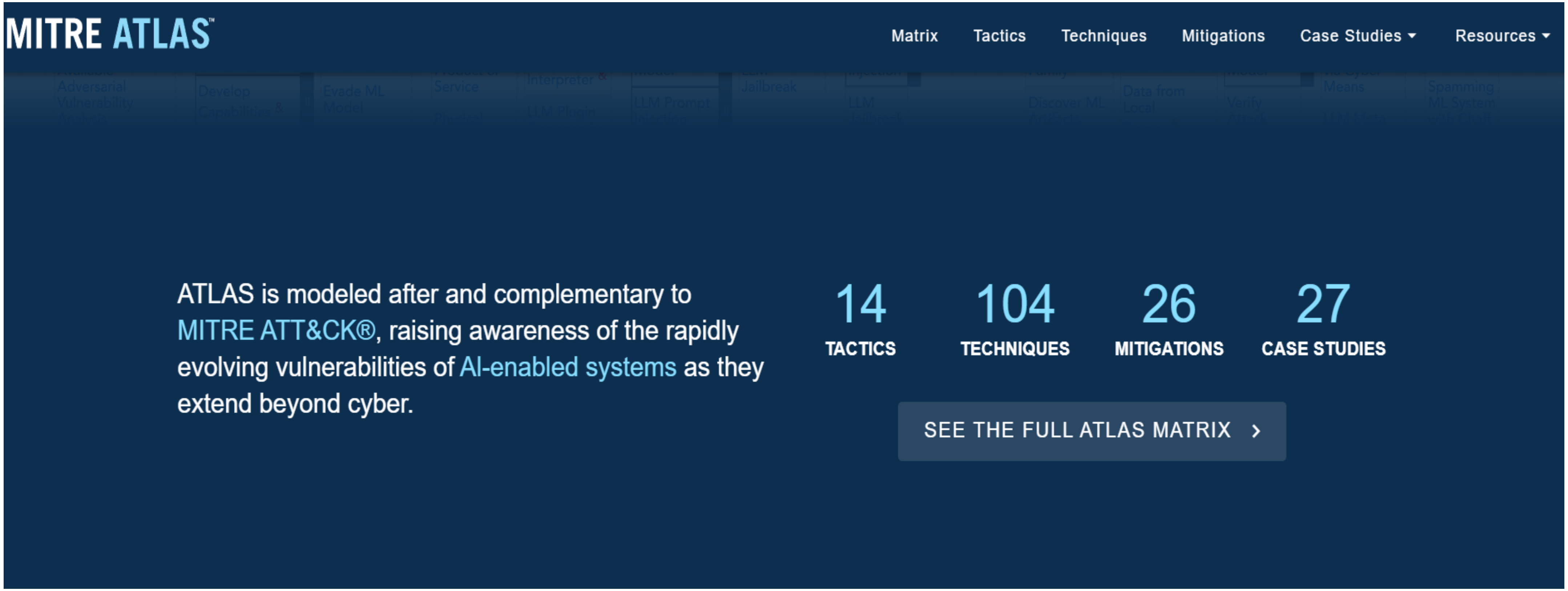
4.6 Against indirect prompt injection
- PROMPTINPUTVALIDATION
- INPUTSEGREGATION

Confidentiality of input data

4.7 Against leaking input data
- MODELINPUTCONFIDENTIALITY

Threat model based on Software Improvement Group AI framework

Artificial Intelligence (AI) Risk Assessment (RA) - MITRE ATLAS



ATLAS is modeled after and complementary to MITRE ATT&CK®, raising awareness of the rapidly evolving vulnerabilities of AI-enabled systems as they extend beyond cyber.

14 TACTICS

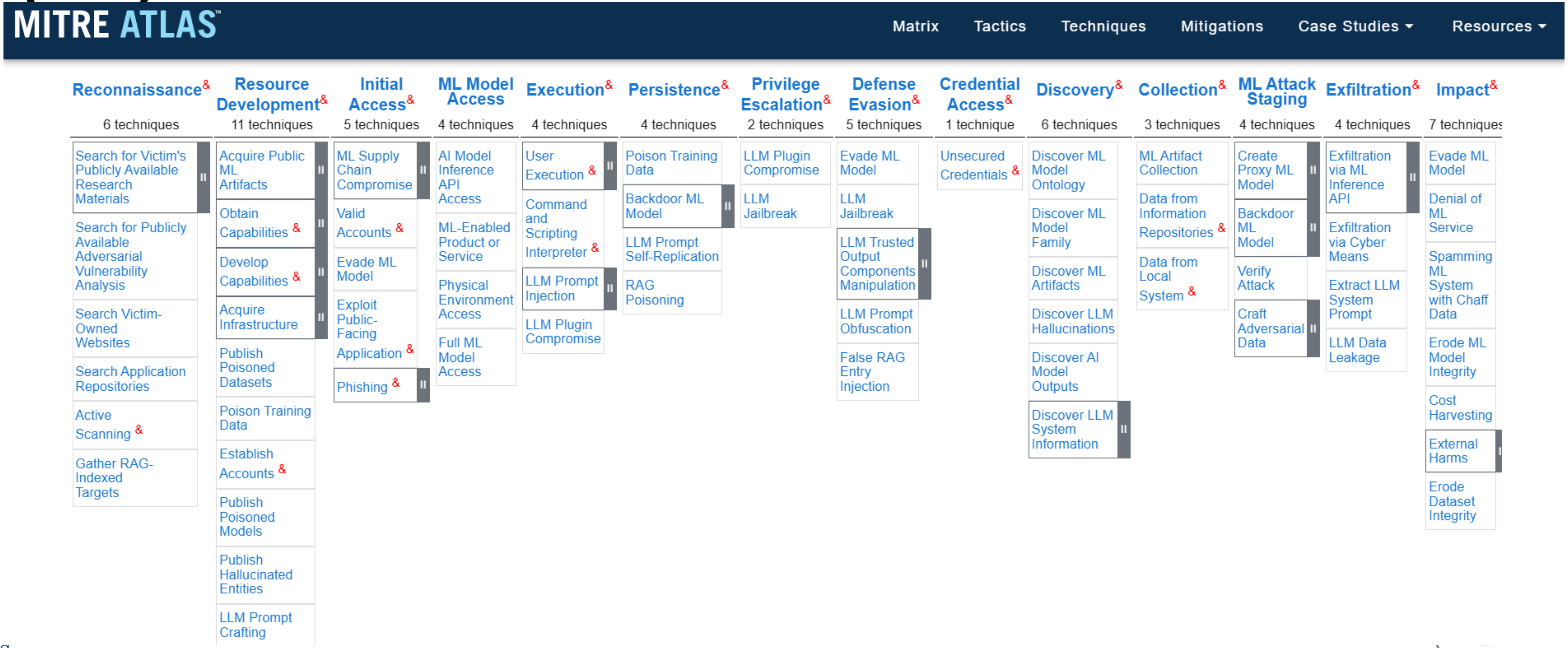
104 TECHNIQUES

26 MITIGATIONS

27 CASE STUDIES

SEE THE FULL ATLAS MATRIX ›

Artificial Intelligence (AI) Risk Assessment (RA) - MITRE ATLAS

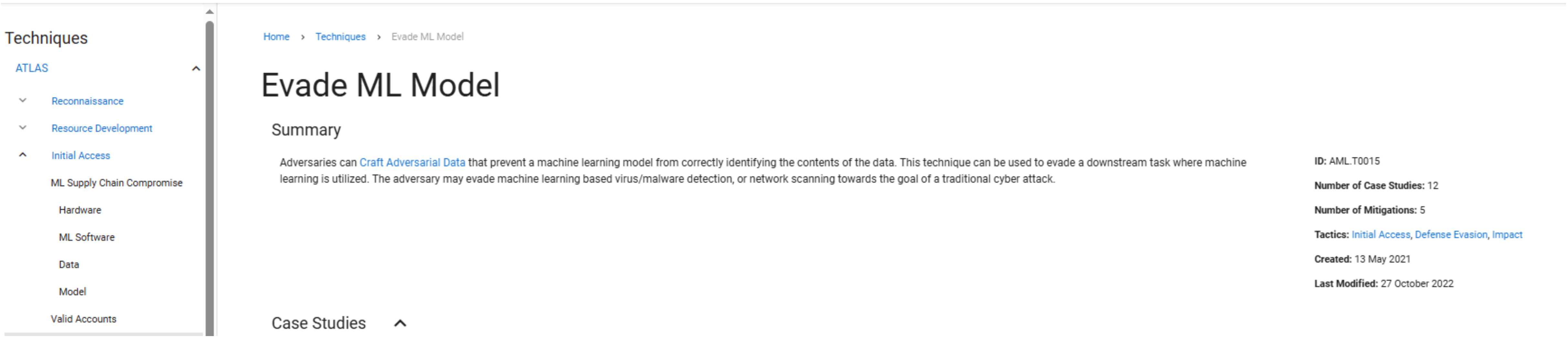


Reconnaissance&	Resource Development&	Initial Access&	ML Model Access	Execution&	Persistence&	Privilege Escalation&	Defense Evasion&	Credential Access&	Discovery&	Collection&	ML Attack Staging	Exfiltration&	Impact&
6 techniques	11 techniques	5 techniques	4 techniques	4 techniques	4 techniques	2 techniques	5 techniques	1 technique	6 techniques	3 techniques	4 techniques	4 techniques	7 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	AI Model Inference API Access	User Execution &	Poison Training Data	LLM Plugin Compromise	Evade ML Model	Unsecured Credentials &	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	ML-Enabled Product or Service	Command and Scripting Interpreter &	Backdoor ML Model	LLM Jailbreak	LLM Jailbreak		Discover ML Model Family	Data from Information Repositories &	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Capabilities &	Evade ML Model	Physical Environment Access	LLM Prompt Injection	LLM Prompt Self-Replication		LLM Trusted Output Components Manipulation		Discover ML Artifacts	Data from Local System &	Verify Attack	Extract LLM System Prompt	Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access	LLM Plugin Compromise	RAG Poisoning		LLM Prompt Obfuscation		Discover LLM Hallucinations		Craft Adversarial Data	LLM Data Leakage	Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets	Phishing &					False RAG Entry Injection		Discover AI Model Outputs				Cost Harvesting
Gather RAG-Indexed Targets	Poison Training Data								Discover LLM System Information				External Harms
	Establish Accounts &												Erode Dataset Integrity
	Publish Poisoned Models												
	Publish Hallucinated Entities												
	LLM Prompt Crafting												

Minnesota BCA Bureau of Criminal Apprehension

mnccc Minnesota Counties Computer Cooperative

Artificial Intelligence (AI) Risk Assessment (RA) - MITRE ATLAS



Evade ML Model

Summary

Adversaries can Craft Adversarial Data that prevent a machine learning model from correctly identifying the contents of the data. This technique can be used to evade a downstream task where machine learning is utilized. The adversary may evade machine learning based virus/malware detection, or network scanning towards the goal of a traditional cyber attack.

ID: AML.T0015

Number of Case Studies: 12

Number of Mitigations: 5

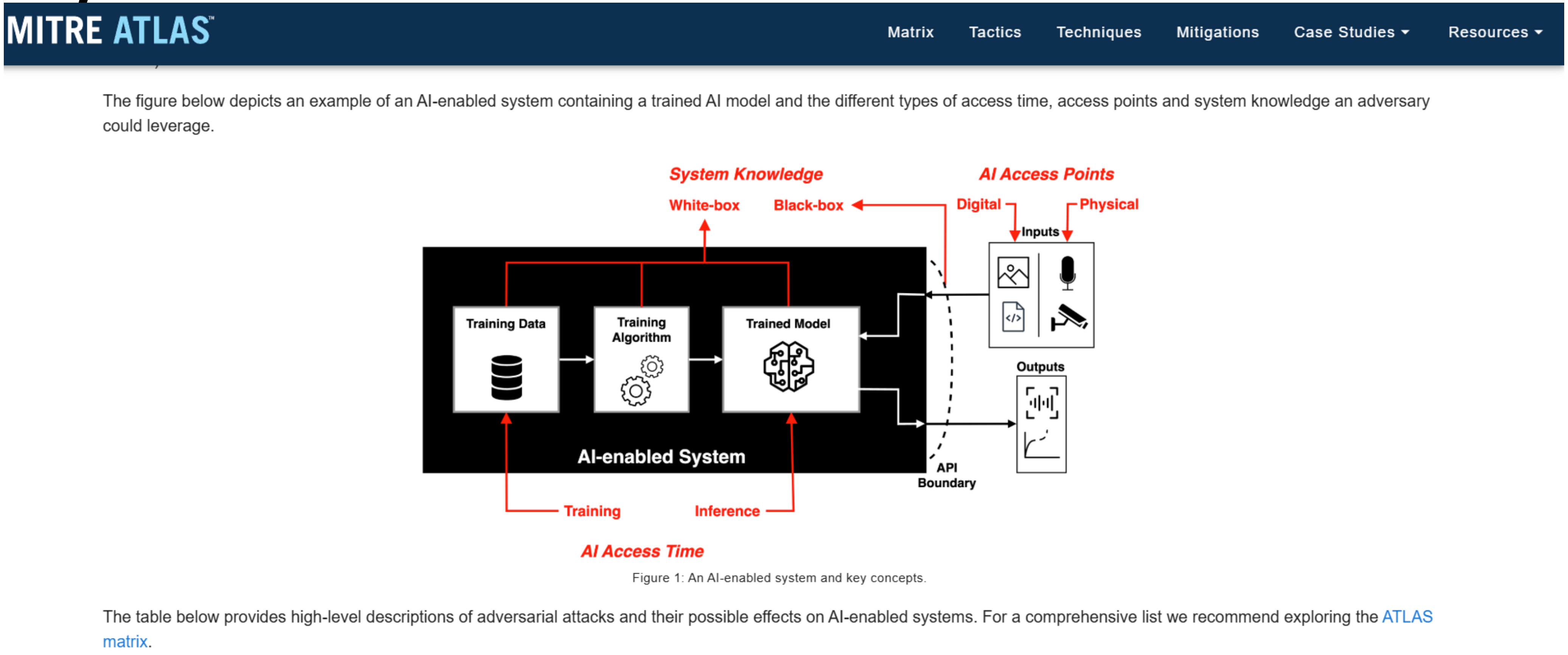
Tactics: Initial Access, Defense Evasion, Impact

Created: 13 May 2021

Last Modified: 27 October 2022

Case Studies

Artificial Intelligence (AI) Risk Assessment (RA) MITRE ATLAS

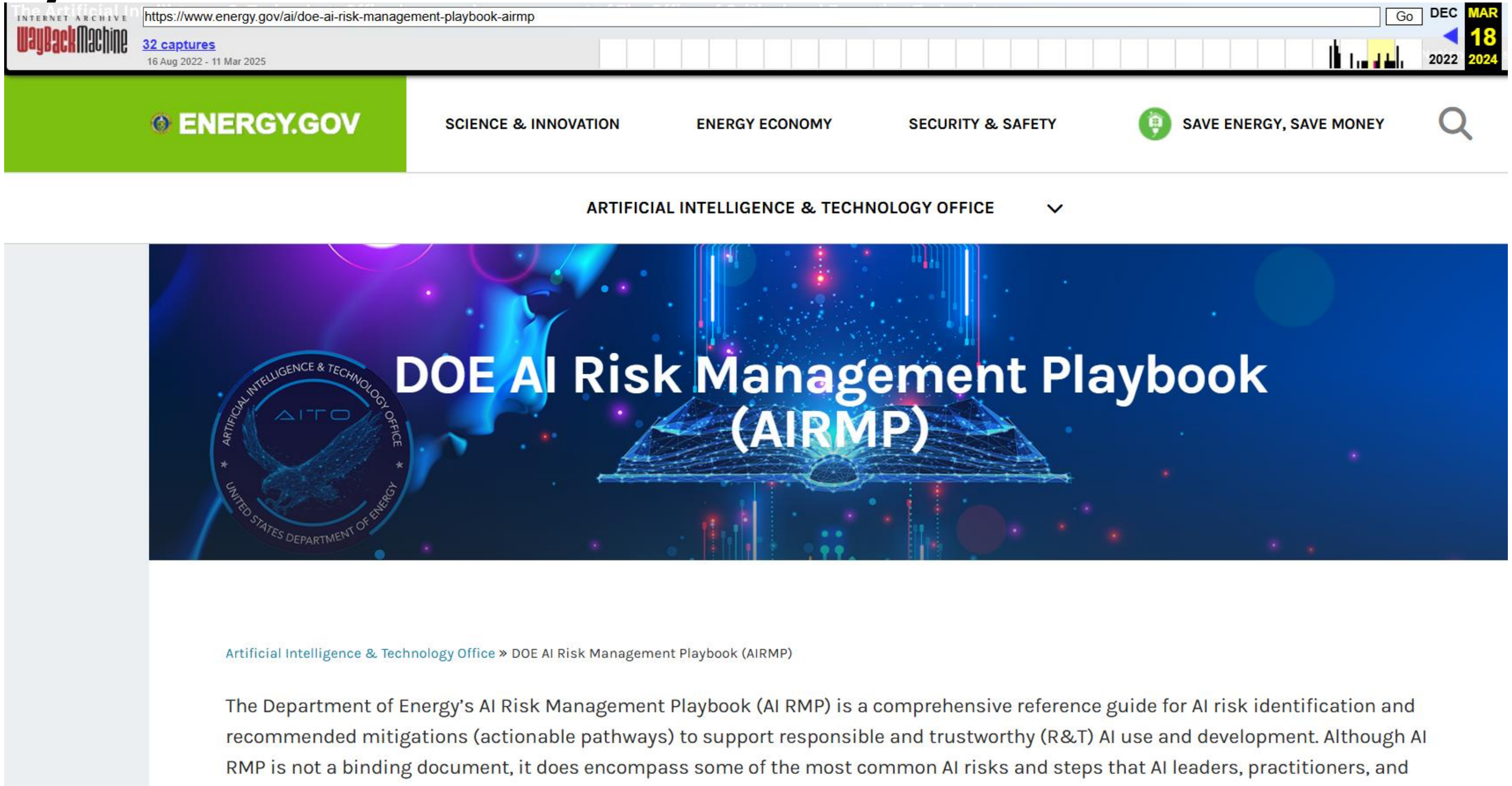


The figure below depicts an example of an AI-enabled system containing a trained AI model and the different types of access time, access points and system knowledge an adversary could leverage.

Figure 1: An AI-enabled system and key concepts.

The table below provides high-level descriptions of adversarial attacks and their possible effects on AI-enabled systems. For a comprehensive list we recommend exploring the ATLAS matrix.

Artificial Intelligence (AI) Risk Assessment (RA) - DOE AIRMP E1140


INTERNET ARCHIVE
WaybackMachine
https://www.energy.gov/ai/doe-ai-risk-management-playbook-airmp
Go
32 captures
16 Aug 2022 - 11 Mar 2025
DEC
MAR
18
2022
2024
ENERGY.GOV
SCIENCE & INNOVATION
ENERGY ECONOMY
SECURITY & SAFETY
SAVE ENERGY, SAVE MONEY
ARTIFICIAL INTELLIGENCE & TECHNOLOGY OFFICE
ARTIFICIAL INTELLIGENCE & TECHNOLOGY OFFICE
AITO
UNITED STATES DEPARTMENT OF ENERGY
DOE AI Risk Management Playbook
(AIRMP)

Artificial Intelligence & Technology Office » DOE AI Risk Management Playbook (AIRMP)

The Department of Energy's AI Risk Management Playbook (AI RMP) is a comprehensive reference guide for AI risk identification and recommended mitigations (actionable pathways) to support responsible and trustworthy (R&T) AI use and development. Although AI RMP is not a binding document, it does encompass some of the most common AI risks and steps that AI leaders, practitioners, and

AI Model Cloud Vendors

- Vendor will tell you -- you get a private instance in the cloud for your AI
  - Read the license agreement
  - Read the user level agreement

Where does AI FIT with CJIS Data?

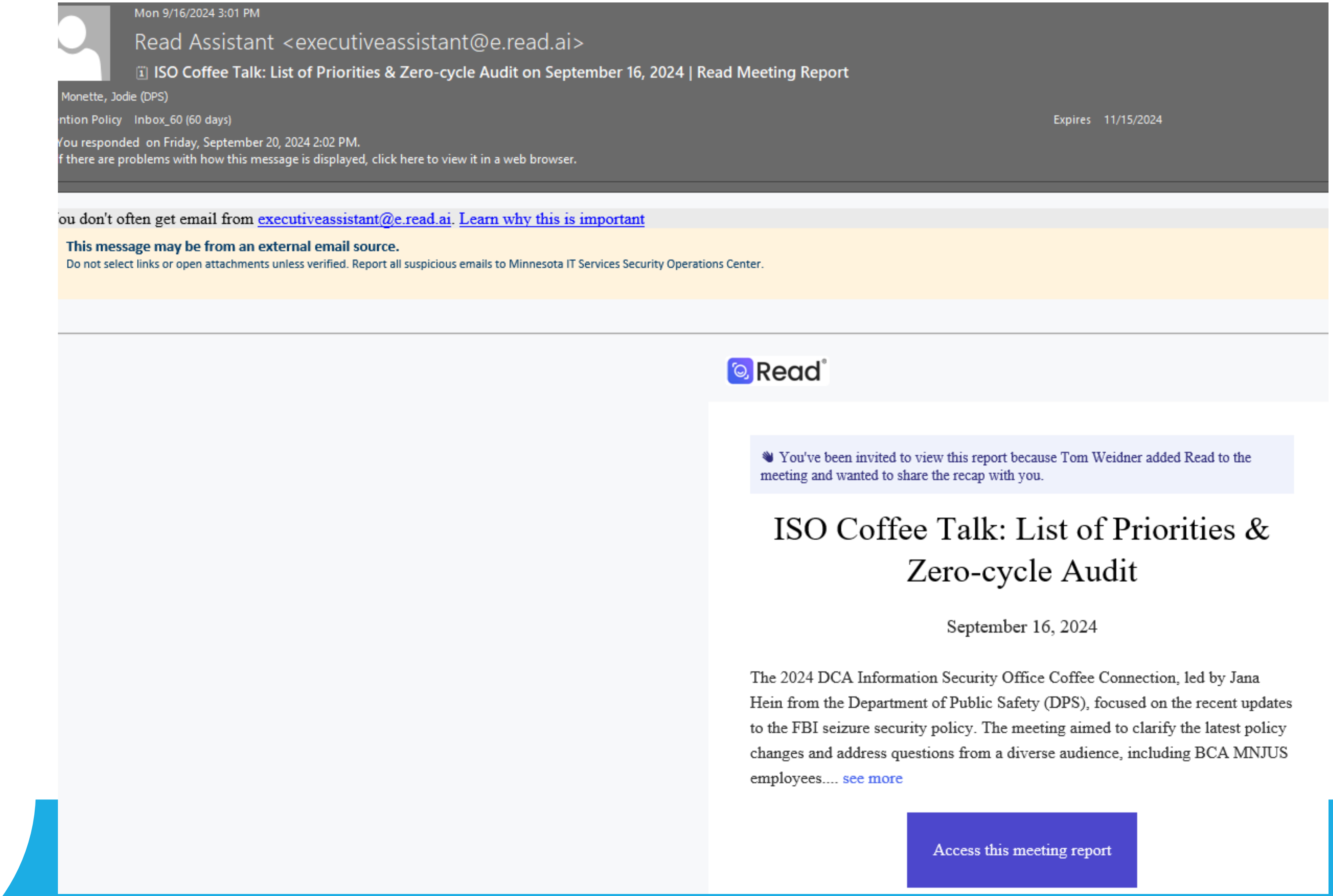
- Non-Deterministic AI
  - Characterized by outputs that vary even with the same input
- Deterministic AI
  - Characterized by always produces the same output for the same input

Where does AI FIT with CJIS Data?

- Currently the FBI CJIS position on AI is NO AI used with FBI CJIS data and systems.

- Only way - would be in a stand alone system with no connection to the internet.

Where does AI FIT with CJIS Data?



Mon 9/16/2024 3:01 PM

Read Assistant <executiveassistant@e.read.ai>

ISO Coffee Talk: List of Priorities & Zero-cycle Audit on September 16, 2024 | Read Meeting Report

Monette, Jodie (DPS)

ntion Policy Inbox_60 (60 days) Expires 11/15/2024

ou responded on Friday, September 20, 2024 2:02 PM.
f there are problems with how this message is displayed, click here to view it in a web browser.

ou don't often get email from executiveassistant@e.read.ai. Learn why this is important

This message may be from an external email source.
Do not select links or open attachments unless verified. Report all suspicious emails to Minnesota IT Services Security Operations Center.

Read

👋 You've been invited to view this report because Tom Weidner added Read to the meeting and wanted to share the recap with you.

ISO Coffee Talk: List of Priorities & Zero-cycle Audit

September 16, 2024

The 2024 DCA Information Security Office Coffee Connection, led by Jana Hein from the Department of Public Safety (DPS), focused on the recent updates to the FBI seizure security policy. The meeting aimed to clarify the latest policy changes and address questions from a diverse audience, including BCA MNJUS employees.... see more

Where does AI FIT with CJIS Data?

State of MN

Public Artificial Intelligence Services Security Standard

Where does AI FIT with CJIS Data?

Copilot & CJIS Issues….

If M365 is county and law enforcement together – Comingled – Copilot must be shutoff.

Artificial Intelligence (AI)

- What is your plan for AI?
- Where is AI already in your Organization? Do you know?
- Do you have a policy that governs the use of AI?

Take Away - Artificial Intelligence (AI)

- Inventory AI where is it in your environment?
- Know what the flow of AI data is?
- Put security controls around AI just like production code.

Resources

- OWASP AI Exchange – https://owaspai.org
- OWASP Top 10 for Large Language Model Applications | OWASP Foundation – https://owasp.org/www-project-top-10-for-large-language-model-applications/
- MITRE ATLAS™ -- https://atlas.mitre.org
- SANs.Org AI Cybersecurity Summit 2025 (March 31 & April 1, 2025)Recording will be posted, next week.

Questions?

Information Security Office - 651-793-2502 • bca.iso@state.mn.us

CSA ISO – 651-793-2547 Jodie.Monette@state.mn.us

DEPARTMENT OF PUBLIC SAFETY
STATE OF MINNESOTA

Minnesota
BCA
Bureau of Criminal Apprehension

- Thank you!!! For taking time out of your busy day.

mnCCC
Minnesota Counties
Computer Cooperative

Contact us

Lisa Meredith, Executive Director –lisa@mnccc.gov

Mike Fox, Chief Financial Officer - mike@mnccc.gov

Amanda Beyer-Schulte, Communications and Events Coordinator - amanda@mnccc.gov

Emily Wick, Marketing and Membership Specialist - emily@mnccc.gov