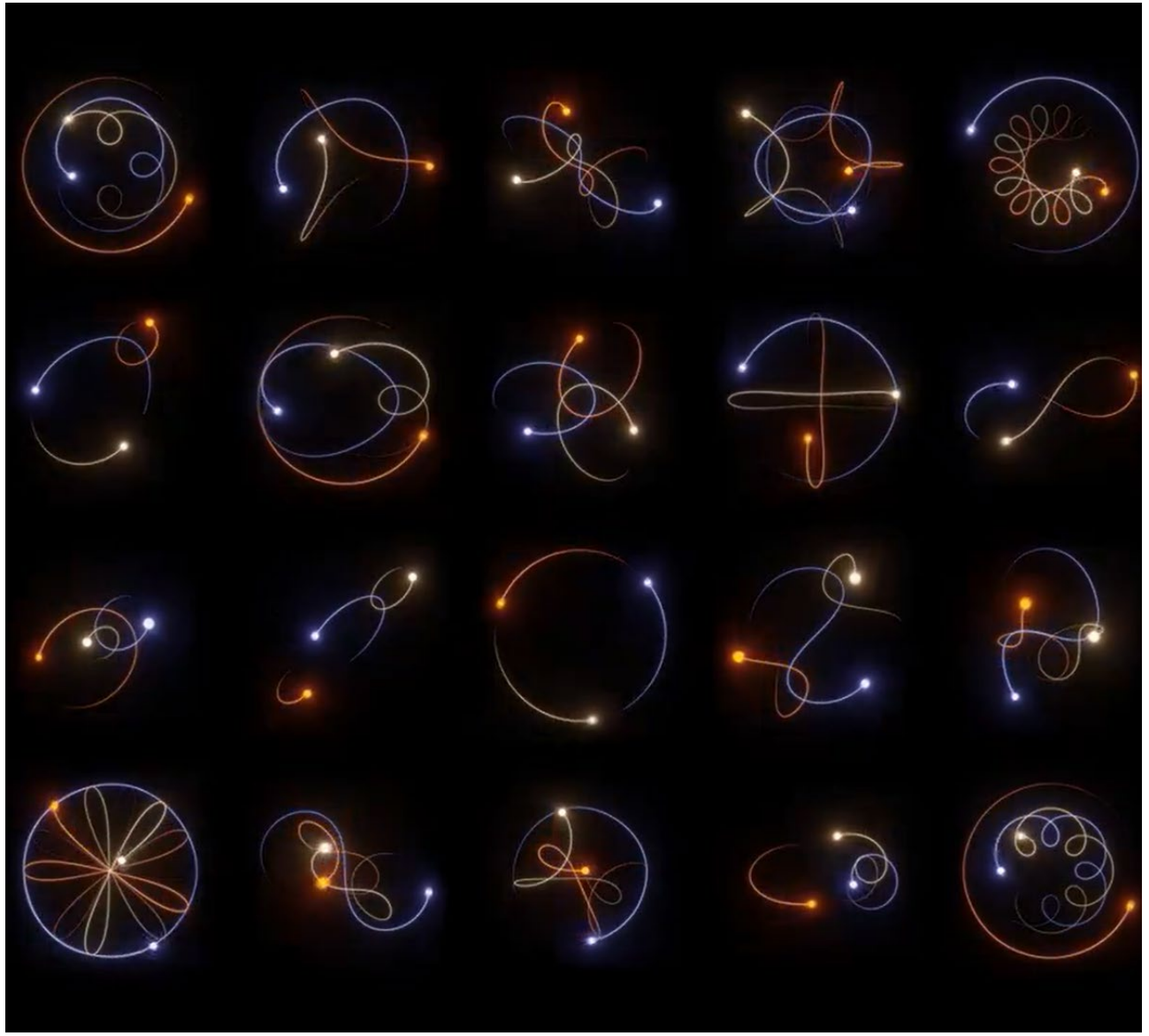
CyberSecurity & AI

Santiago Cavanna

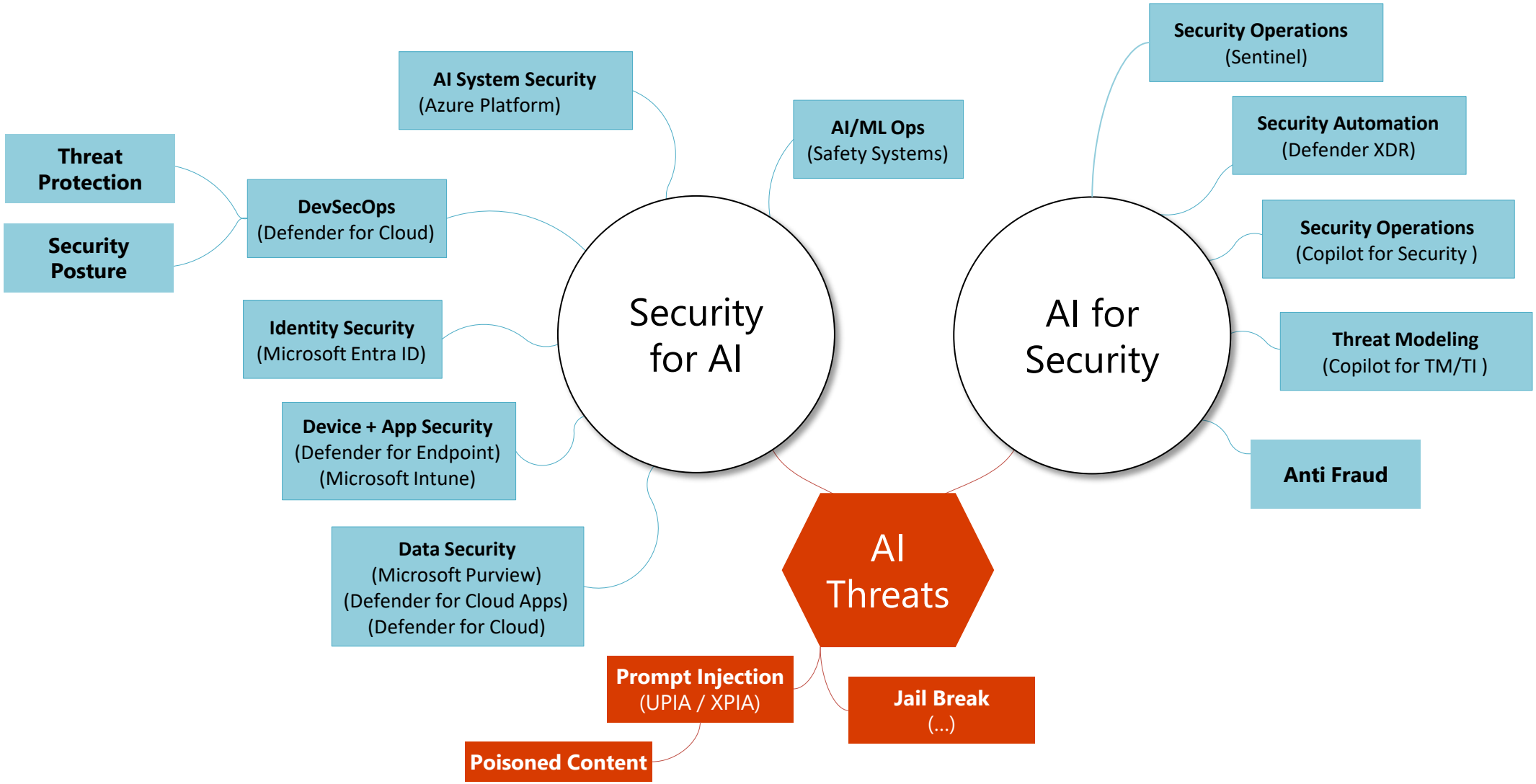
SCavanna@Microsoft.com

Microsoft



AI Safety & Security

how it all fits together – Microsoft PoV



What is Artificial Intelligence?

Today, AI is a platform upon which many applications can be built. There are several definitions of technologies within the overarching AI capability, each one has a specific purpose, but together they make up a powerful system

Machine Learning

Teaching machines to learn from data

- Reinforcement
- Supervised Learning
- Unsupervised Learning
- Knowledge Learning
- Probabilistic Models
- Neural networks

Natural Language Processing

Understand and generate human language

- Information Extraction
- Sentiment Analysis
- Speech Recognition
- Machine Translation
- Large Language Models (including ChatGPT)

Robotics

Controlling objects in the physical world

- Perception
- Actuation
- Locomotion
- Navigation
- Mapping, Localization
- Manipulation and grasping

Reasoning

Using logic to draw conclusions

- Probabilistic Reasoning
- Bayesian Networks
- Reasoning Over Time
- Uncertainty Quantification
- Argumentation
- Ontology Engineering

Decision Making

Making informed choices

- Knowledge Engineering
- Planning and Scheduling
- Learning and Adaptation
- Explanation and Justification
- Real World Planning

Object Recognition

Recognizing objects in images or video

- Image Processing
- Feature Extraction
- Object Detection
- Classification
- 3D Reconstruction
- Motion, Shading, Contour Analysis
- Scene Understanding

Problem Solving

Process of finding a solution

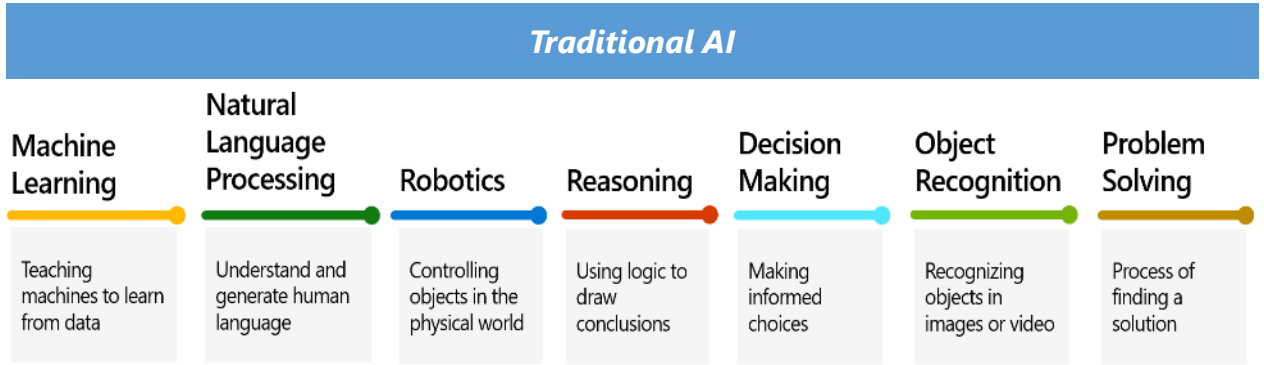
- Search and Exploration
- Heuristics
- Optimization Techniques
- Uncertainty Handling
- Constraint Analysis
- Evolutionary Algorithms
- Fuzzy Logic

Two categories of AI

Classic/Traditional AI

Is AI that detects and classifies, can work on vast amounts of data, for use in real-time applications and automation of capabilities.

- **Traditional AI is good at:** Looking at a large field of data and finding patterns or continuations (like making recommendations).
- **Traditional AI is bad at:** Understanding highly complex smaller things like language.



Generative AI (GenAI)

Is AI that understands and creates content, such as GPT. It works on relatively small chunks of data – text, images, sounds, videos – and has a "linguistic" understanding. Large language models (LLMs) are a kind of GenAI and the term is often used as a synonym, but LLMs are ones that work on text.

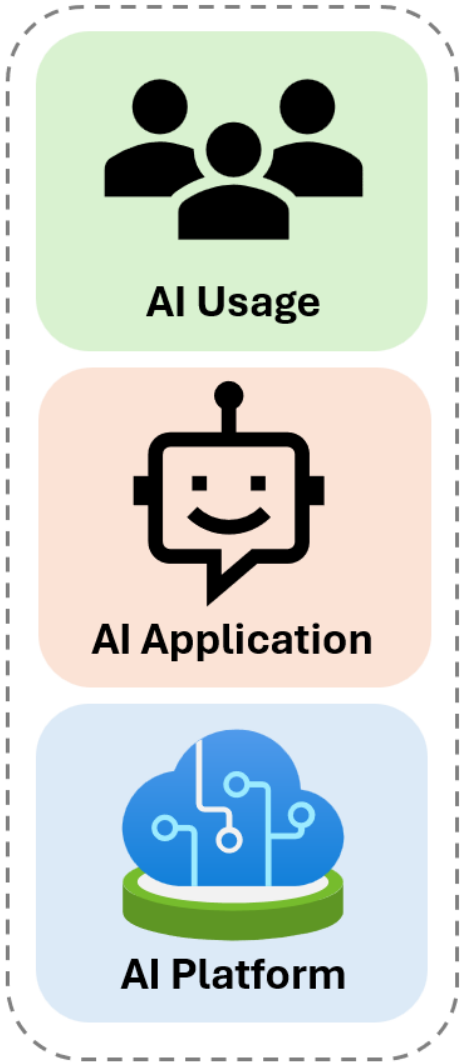
- **GenAI is good at:** Understanding language, summarizing, translating concepts (e.g. from language to code or vice-versa); roleplaying as characters
- **GenAI is bad at:** Processing large amounts of data.
  [Note: All AI's are trained on large amounts of data; this is about what they can do after they're trained]

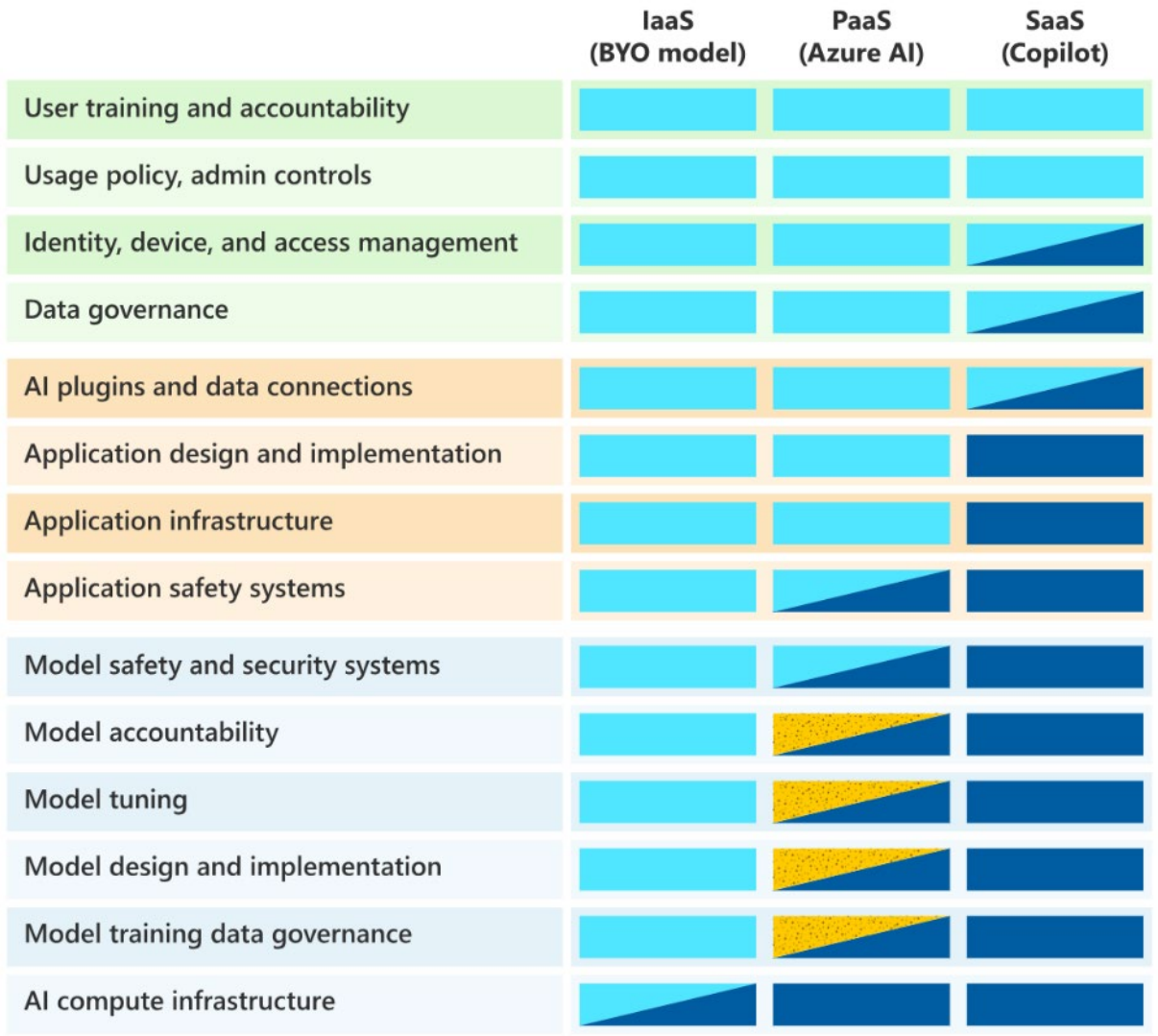
Security for AI Shared Responsibility Model

Acceptable use policy
Human in the loop
Managing insider risk
Identity and access management
Protecting sensitive information

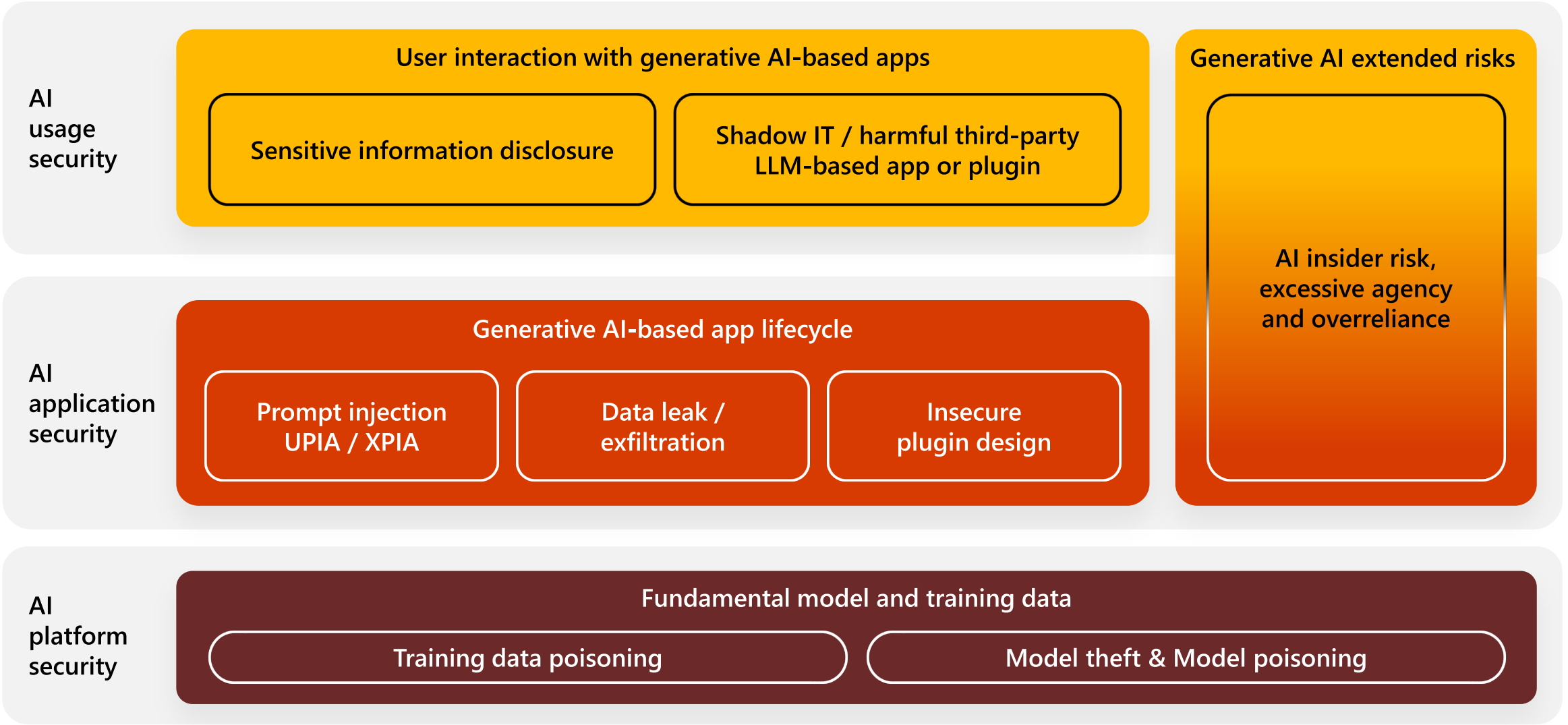
Zero Trust architecture
Minimize attack surface
Preventing command injection
Implement grounding
Monitoring generated content

AI Red Team
Securing the AI supply chain
Reducing the impact of DDoS
Protecting model development





Generative-AI threat landscape



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

Search AIRC Website

Trustworthy & Responsible AI Resource Center

Home

- Home
- Knowledge Base
- Use Cases
- Engagement and Events
- About the Center

Welcome to the NIST Trustworthy & Responsible Artificial Intelligence Resource Center (AIRC).

The AIRC supports all AI actors in the development and deployment of trustworthy and responsible AI technologies. AIRC supports and operationalizes the NIST AI Risk Management Framework (AI RMF 1.0) and accompanying Playbook and will grow with enhancements to enable an interactive, role-based experience providing access to a wide-range of relevant AI resources.

AI RMF Knowledge Base

AI Risk Management Framework (RMF)

The AI RMF is voluntary guidance to improve the ability to incorporate trustworthiness considerations into the design, development, use and evaluation of AI products, services and systems.

Download the Framework

AI RMF Playbook

Companion resource for the AI RMF that includes suggested actions, references, and documentation guidance to achieve outcomes for the four AI RMF functions.

Download the Playbook (as PDF) (as CSV) (as JSON)

Glossary

The Glossary helps promote a shared understanding and improved communication in trustworthy and responsible AI.

Engagements and Events

NIST relies on and encourages robust interactions with companies, universities, nonprofits, and other government agencies in driving and carrying out its AI agenda. There are multiple ways to engage with NIST, including:

Workshops: NIST convenes experts for single day, multi-day, and multi-week sessions.
AI Visiting Fellows: Accomplished Visiting Fellows bring thought leadership to foundational research.

Training

Opportunities to learn about building trustworthy and responsible AI through approaches like the AI RMF.

Introduction to the NIST AI RMF 1.0: An Explainer Video.

NIST AIRC - Home

OWASP Top 10 for LLM Applications

LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior. Sources include Common Crawl, WebText, OpenWebText, & books.

LLM04

Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre- trained models, and plugins can add vulnerabilities.

LLM06

Sensitive Information Disclosure

LLMs may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution.

LLM08

Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10

Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

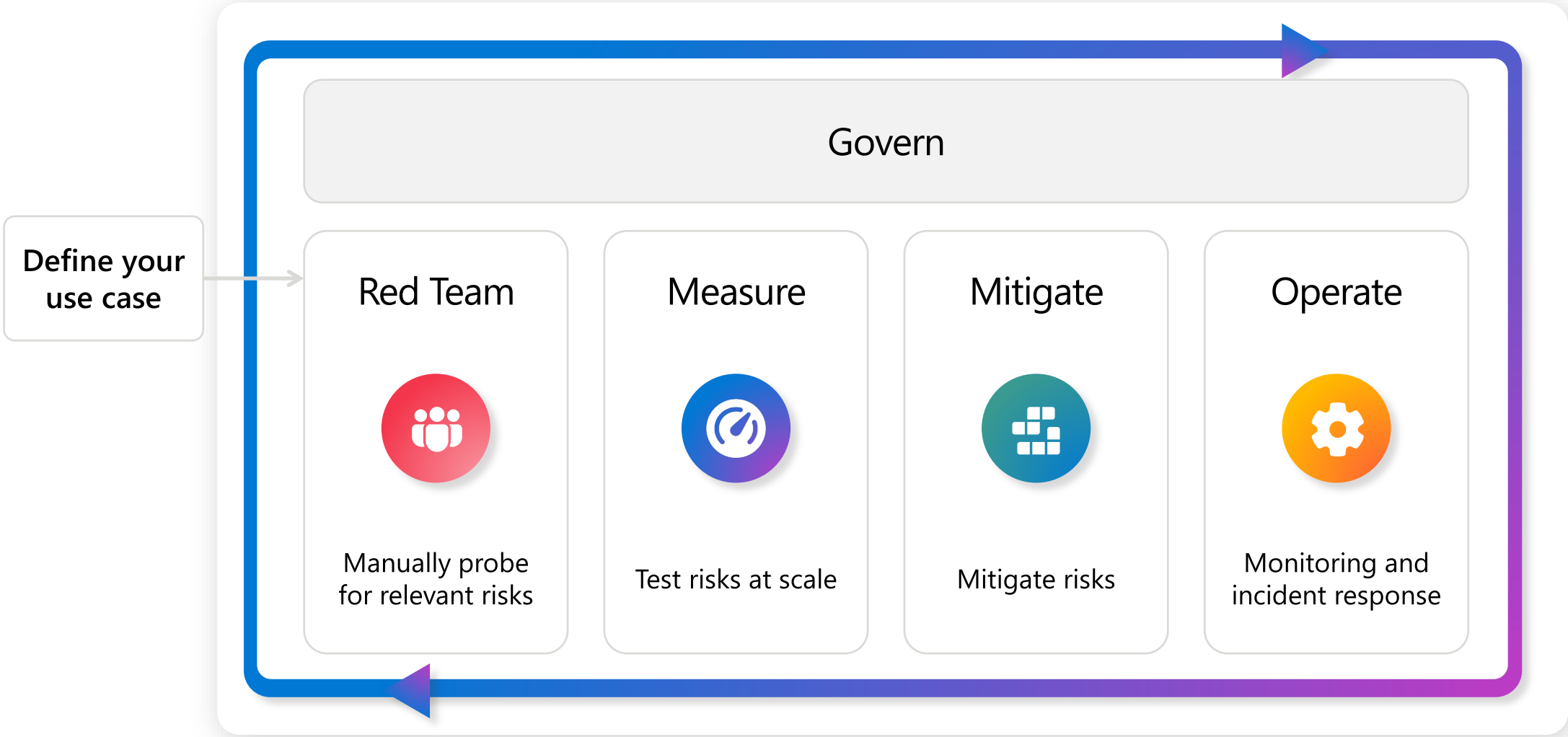
OWASP-Top-10-for-LLMs-2023-slides-v1_1.pdf

ATLAS Matrix

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below. & indicates an adaption from ATT&CK. Click on the blue links to learn more about each item, or search and view ATLAS tactics and techniques using the links at the top navigation bar. View the ATLAS matrix highlighted alongside ATT&CK Enterprise techniques on the ATLAS Navigator.

Reconnaissance&	Resource Development&	Initial Access&	ML Model Access	Execution&	Persistence&	Privilege Escalation&	Defense Evasion&	Credential Access&	Discovery&	Collection&	ML Attack Staging	Exfiltration&	Impact&
5 techniques	7 techniques	6 techniques	4 techniques	3 techniques	3 techniques	3 techniques	3 techniques	1 technique	4 techniques	3 techniques	4 techniques	4 techniques	6 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution &	Poison Training Data	LLM Prompt Injection	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 Plugin Compromise	LLM Prompt Injection		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 Plugin Compromise	LLM Prompt Injection	LLM Jailbreak	LLM Jailbreak		Discover ML Artifacts	Data from Local System &	Verify Attack	LLM Meta Prompt Extraction	Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access						LLM Meta Prompt Extraction		Craft Adversarial Data	LLM Data Leakage	Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets	LLM Prompt Injection											Cost Harvesting
	Poison Training Data	Phishing &											External Harms
	Establish Accounts &												

Responsible innovation is iterative



AI red teaming versus traditional red teaming

How is it similar?

Still need to investigate the underlying security of the application

Use traditional security tools and techniques

Need to do threat modeling, consider attack paths, and end goal of system

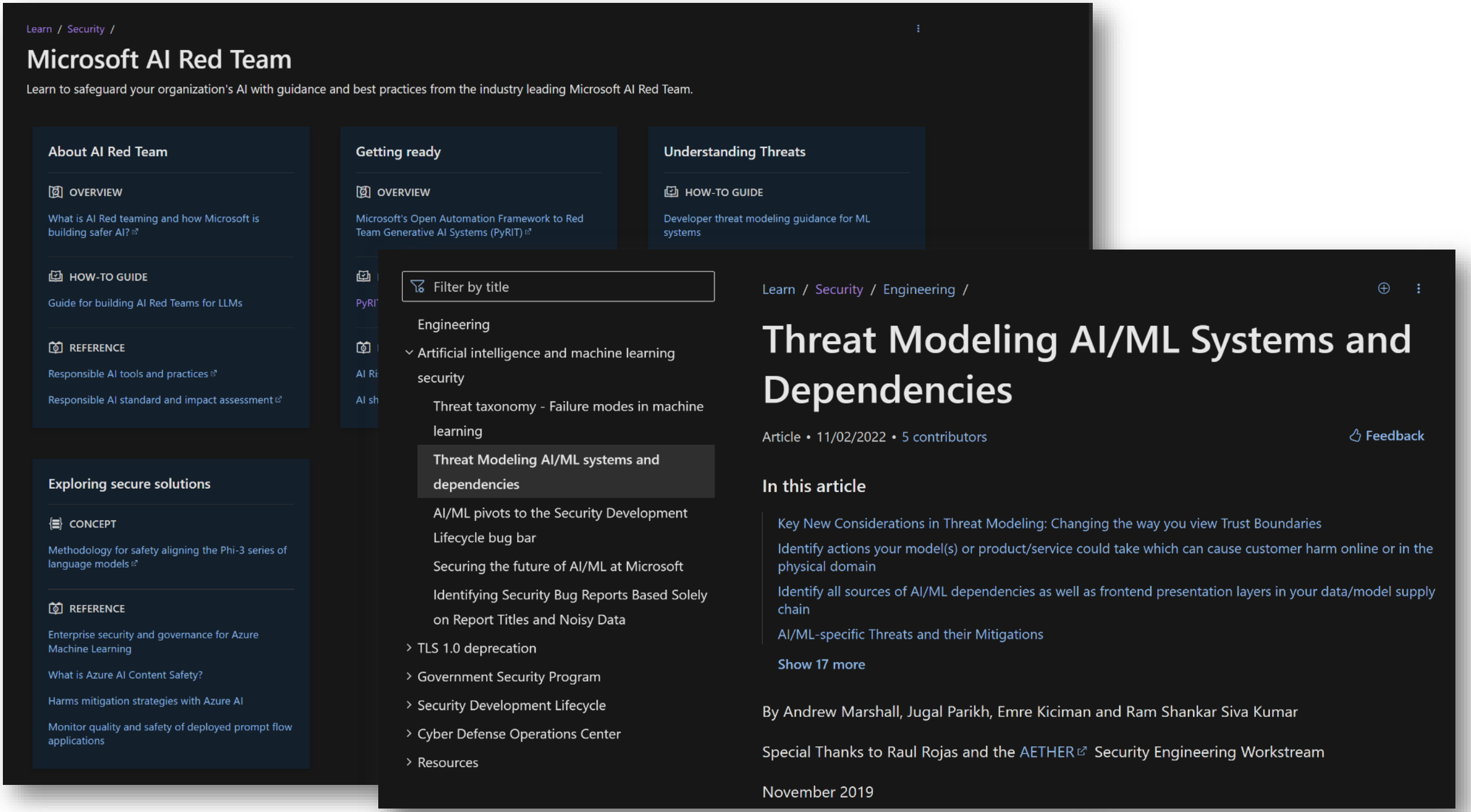
How is it different?

Shorter timeframe

Purple approach - work closely with product team

Testing pre-ship products

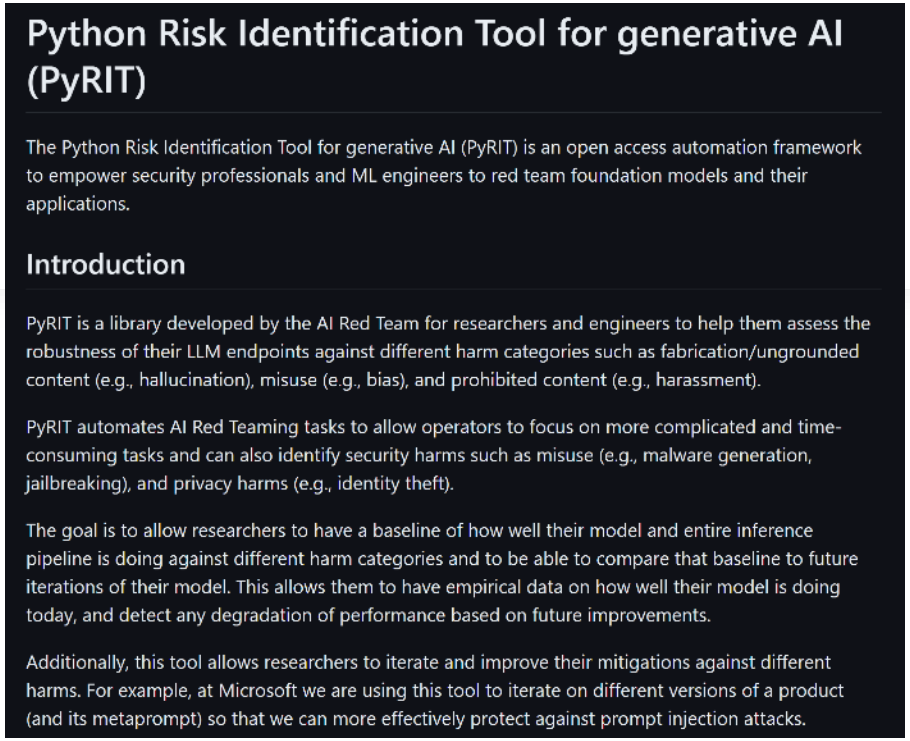
Covers a broader range of issues, including model-induced security risks and RAI harms


Learn / Security /
Microsoft AI Red Team
Learn to safeguard your organization's AI with guidance and best practices from the industry leading Microsoft AI Red Team.
About AI Red Team
OVERVIEW
What is AI Red teaming and how Microsoft is building safer AI?
HOW-TO GUIDE
Guide for building AI Red Teams for LLMs
REFERENCE
Responsible AI tools and practices
Responsible AI standard and impact assessment
Getting ready
OVERVIEW
Microsoft's Open Automation Framework to Red Team Generative AI Systems (PyRIT)
Understanding Threats
HOW-TO GUIDE
Developer threat modeling guidance for ML systems
Exploring secure solutions
CONCEPT
Methodology for safety aligning the Phi-3 series of language models
REFERENCE
Enterprise security and governance for Azure Machine Learning
What is Azure AI Content Safety?
Harms mitigation strategies with Azure AI
Monitor quality and safety of deployed prompt flow applications
Filter by title
Engineering
Artificial intelligence and machine learning security
Threat taxonomy - Failure modes in machine learning
Threat Modeling AI/ML systems and dependencies
AI/ML pivots to the Security Development Lifecycle bug bar
Securing the future of AI/ML at Microsoft
Identifying Security Bug Reports Based Solely on Report Titles and Noisy Data
TLS 1.0 deprecation
Government Security Program
Security Development Lifecycle
Cyber Defense Operations Center
Resources
Learn / Security / Engineering /
Threat Modeling AI/ML Systems and Dependencies
Article • 11/02/2022 • 5 contributors
Feedback
In this article
Key New Considerations in Threat Modeling: Changing the way you view Trust Boundaries
Identify actions your model(s) or product/service could take which can cause customer harm online or in the physical domain
Identify all sources of AI/ML dependencies as well as frontend presentation layers in your data/model supply chain
AI/ML-specific Threats and their Mitigations
Show 17 more
By Andrew Marshall, Jugal Parikh, Emre Kiciman and Ram Shankar Siva Kumar
Special Thanks to Raul Rojas and the AETHER Security Engineering Workstream
November 2019

Python Risk Identification Tool for generative AI (PyRIT)

PyRIT Components

Interface	Implementation
Target	**Local:** local model (e.g., ONNX)
	Remote: API or web app
Datasets	**Static:** prompts
	Dynamic: Prompt templates
Scoring Engine	**PyRIT Itself:** Self Evaluation
	API: Existing content classifiers
Attack Strategy	**Single Turn:** Using static prompts
	Multi Turn: Multiple conversations using prompt templates
Memory	**Storage:** JSON, Database
	Utils: Conversation, retrieval and storage, memory sharing, data analysis



Python Risk Identification Tool for generative AI (PyRIT)

The Python Risk Identification Tool for generative AI (PyRIT) is an open access automation framework to empower security professionals and ML engineers to red team foundation models and their applications.

Introduction

PyRIT is a library developed by the AI Red Team for researchers and engineers to help them assess the robustness of their LLM endpoints against different harm categories such as fabrication/ungrounded content (e.g., hallucination), misuse (e.g., bias), and prohibited content (e.g., harassment).

PyRIT automates AI Red Teaming tasks to allow operators to focus on more complicated and time-consuming tasks and can also identify security harms such as misuse (e.g., malware generation, jailbreaking), and privacy harms (e.g., identity theft).

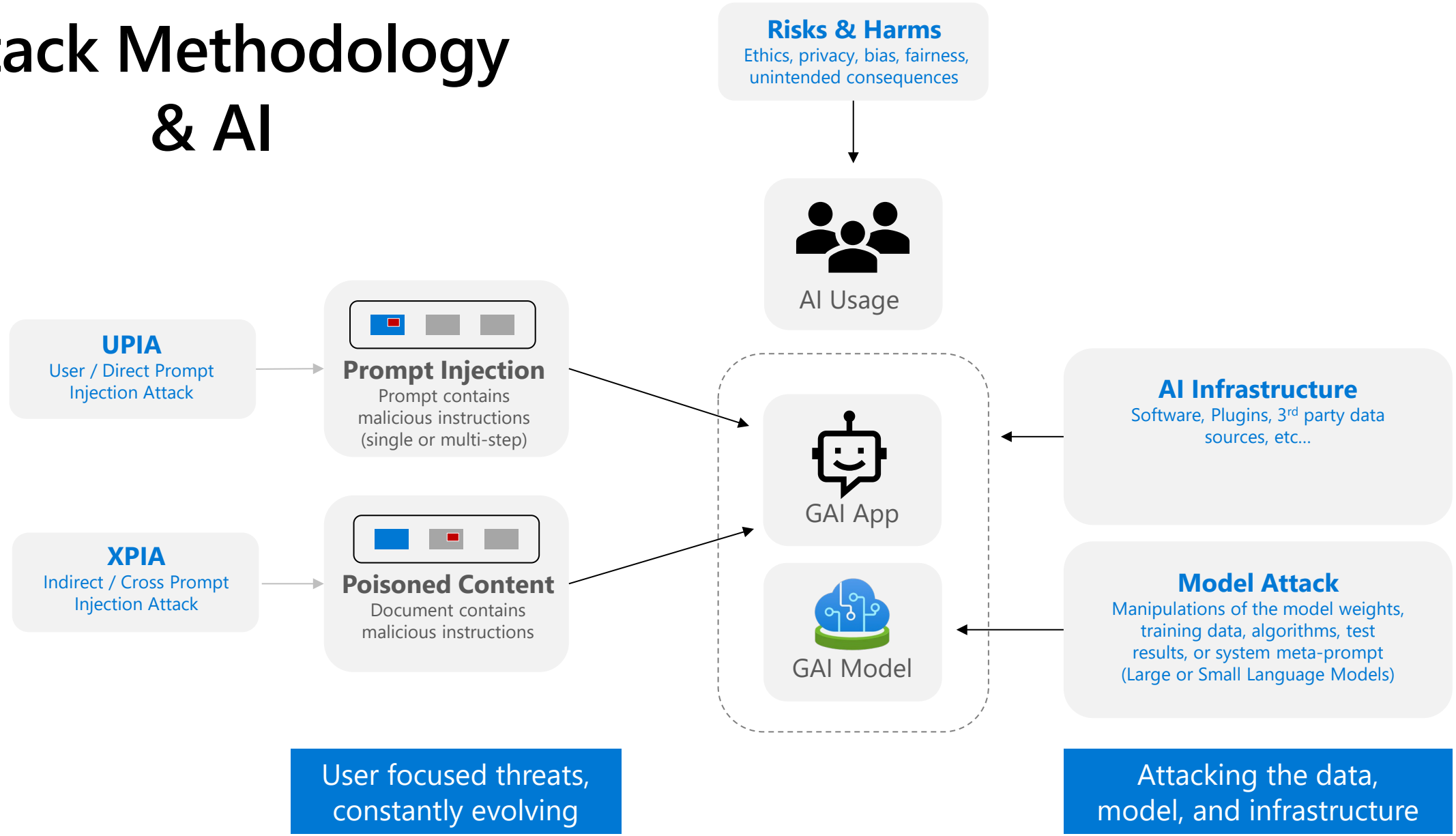
The goal is to allow researchers to have a baseline of how well their model and entire inference pipeline is doing against different harm categories and to be able to compare that baseline to future iterations of their model. This allows them to have empirical data on how well their model is doing today, and detect any degradation of performance based on future improvements.

Additionally, this tool allows researchers to iterate and improve their mitigations against different harms. For example, at Microsoft we are using this tool to iterate on different versions of a product (and its metaprompt) so that we can more effectively protect against prompt injection attacks.

GitHub - Azure/PyRIT: The Python Risk Identification Tool for generative AI (PyRIT) is an open access automation framework to empower security professionals and machine learning engineers to proactively find risks in their generative AI systems.

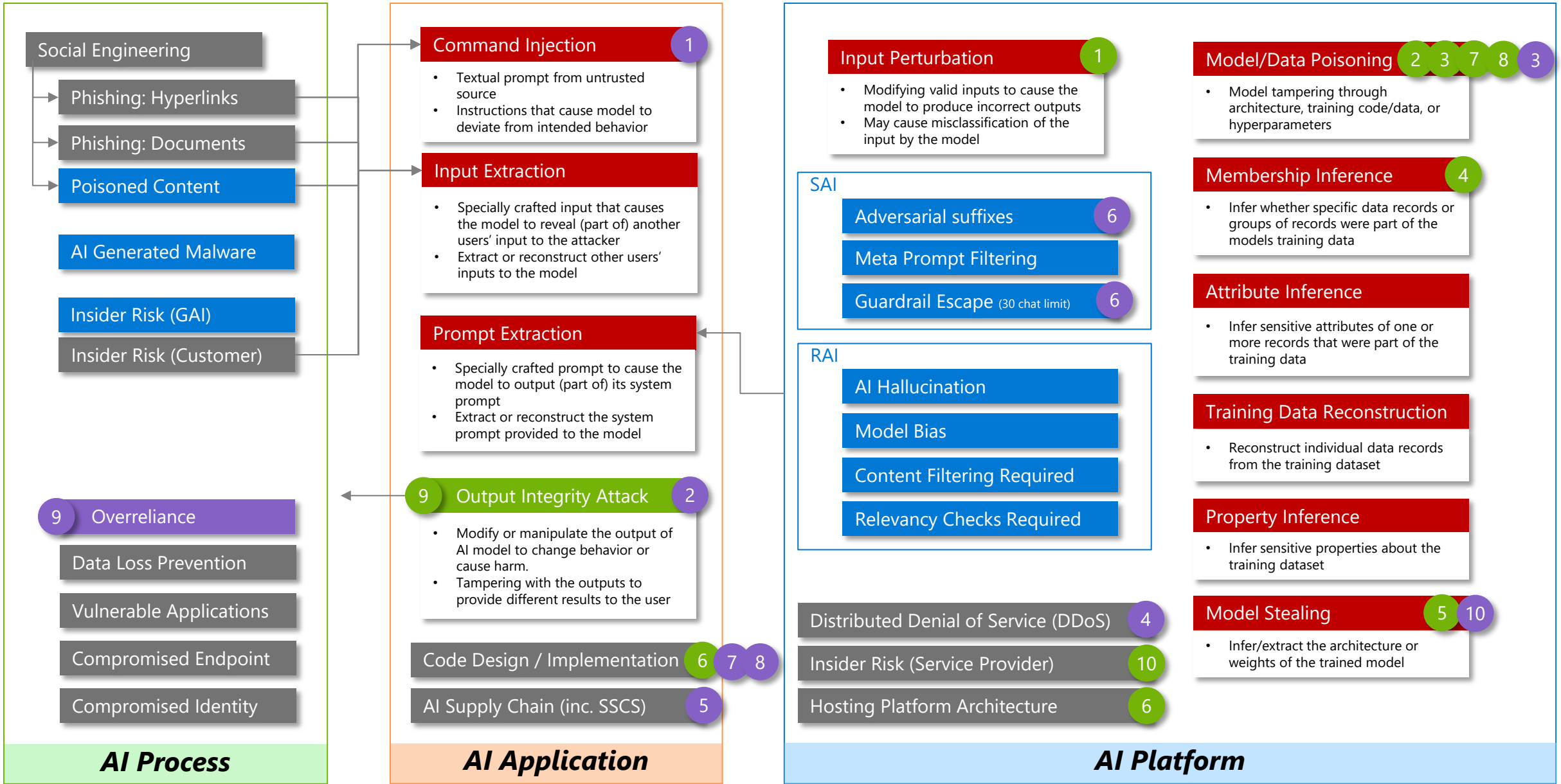
https://github.com/Azure/PyRIT/tree/main

Attack Methodology & AI



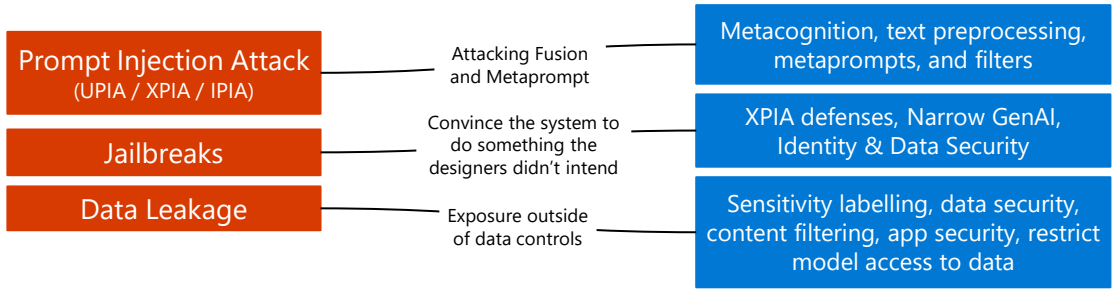
Threat Modelling applied to AI arch

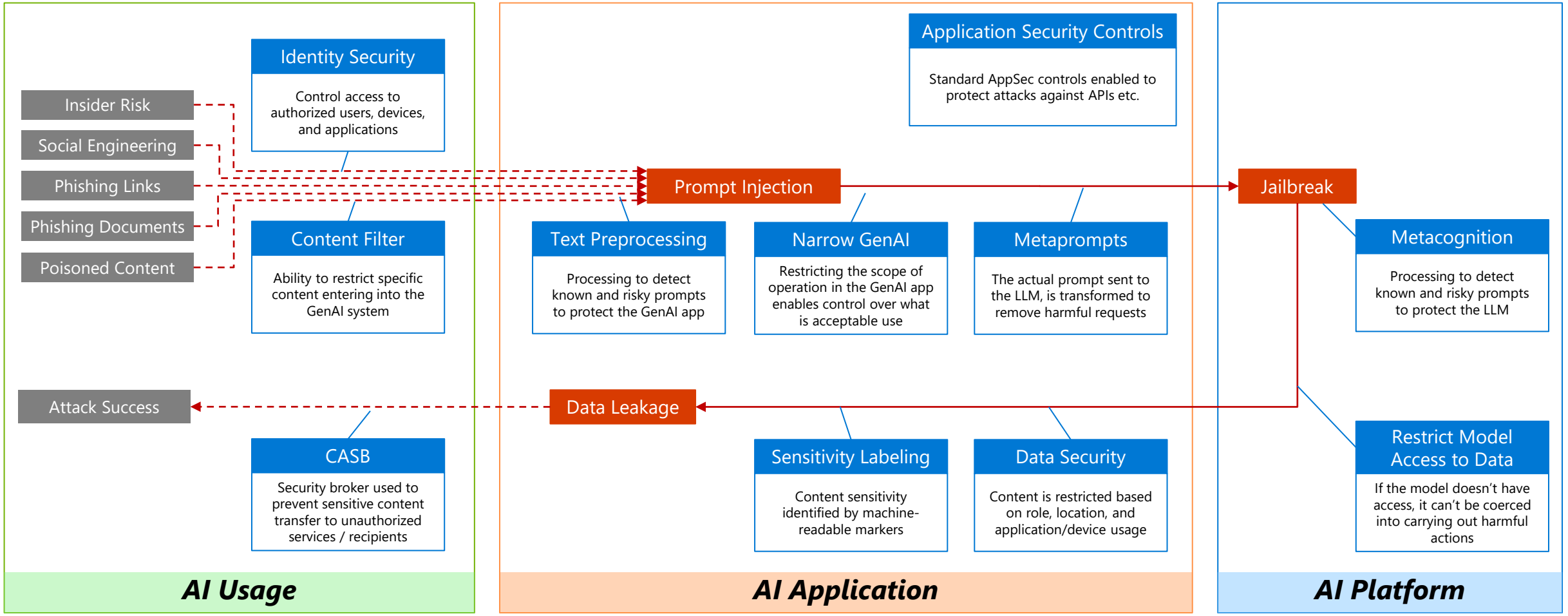
MSRC AI Bug Bar | OWASP Top 10 for ML | Generative AI Specific
MITRE ATLAS | OWASP Top 10 for LLM | Common Cyber Threats



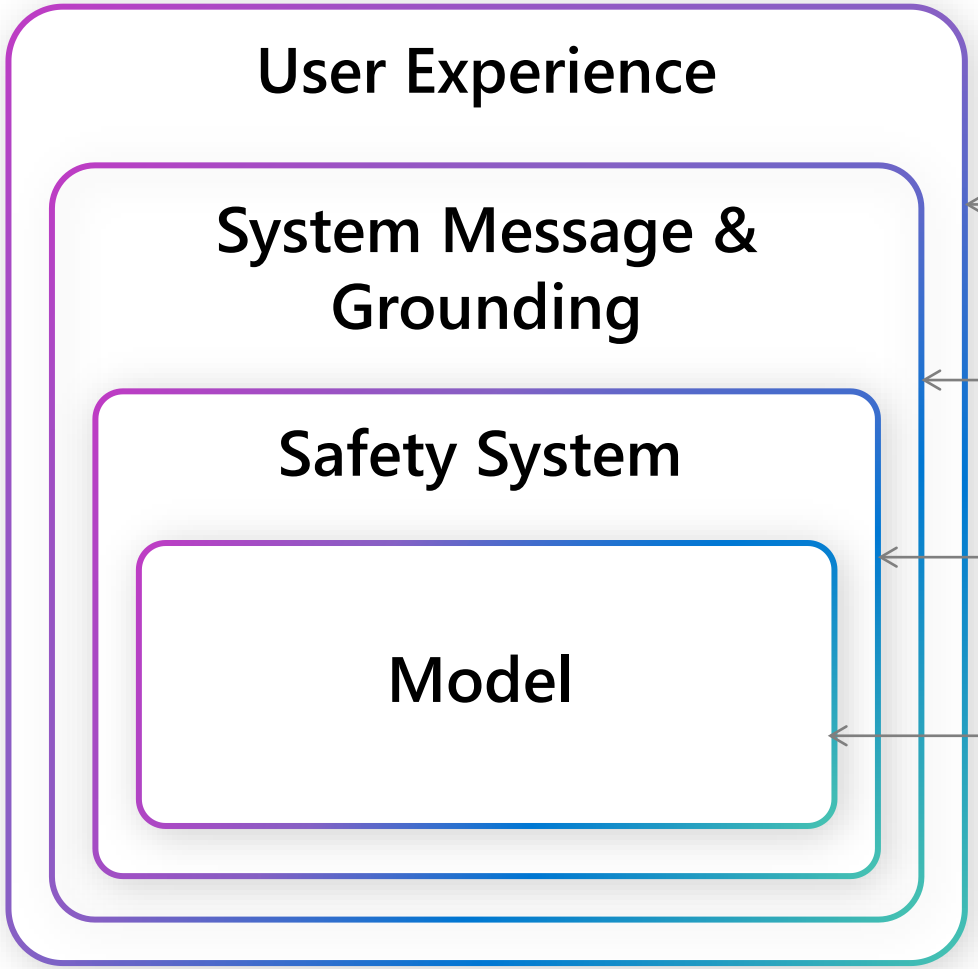
Threat Mapping Template

This framework provides a repeatable method of articulating both the vulnerabilities (red) and the mitigations (blue)





Risk mitigation layers



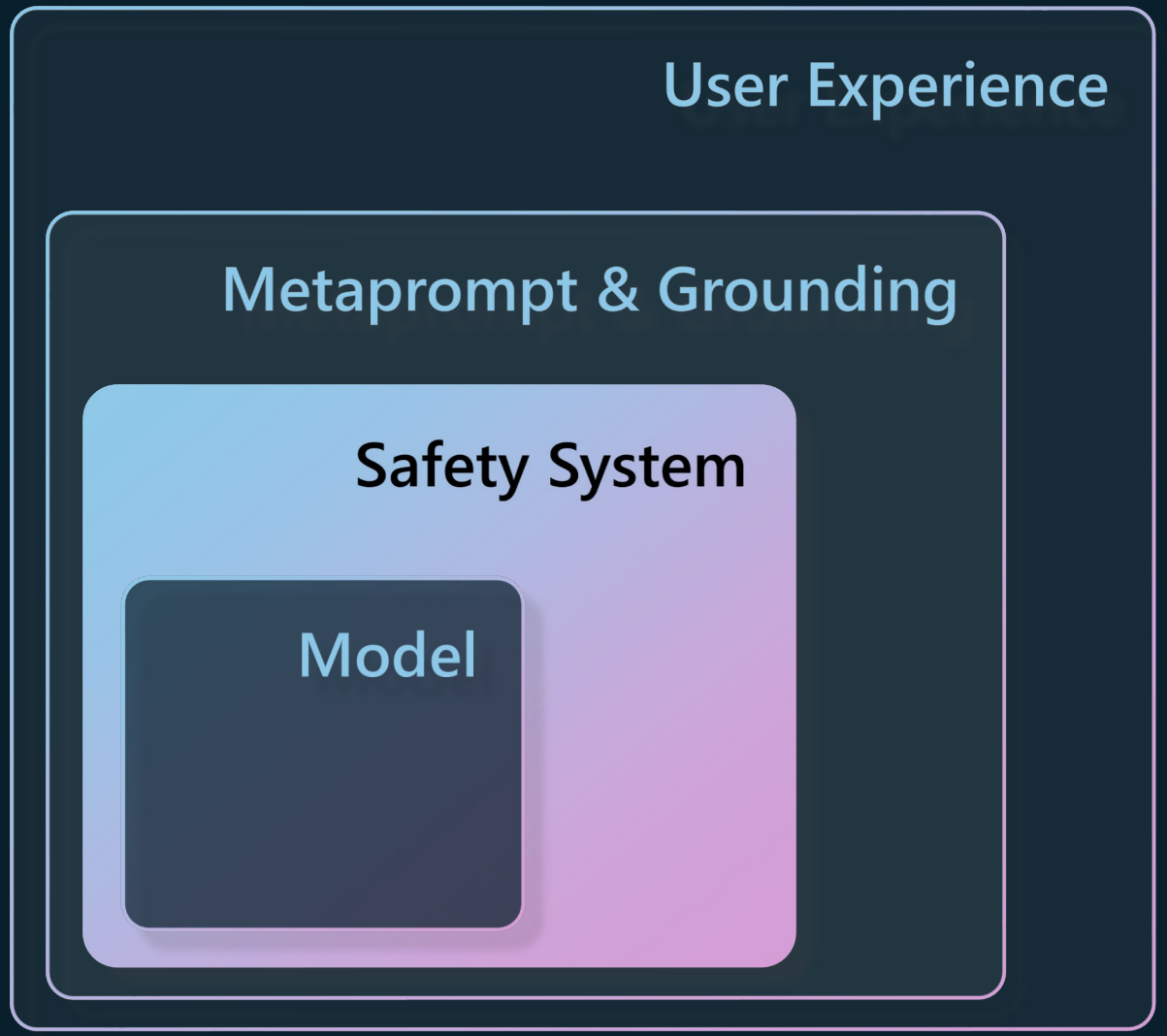
Design for responsible human-AI interaction

Ground your model and direct its behavior

Monitor and protect model inputs and outputs

Choose the right model for your use case

Risk Mitigation Layers



Defenses for AI – Content Safety

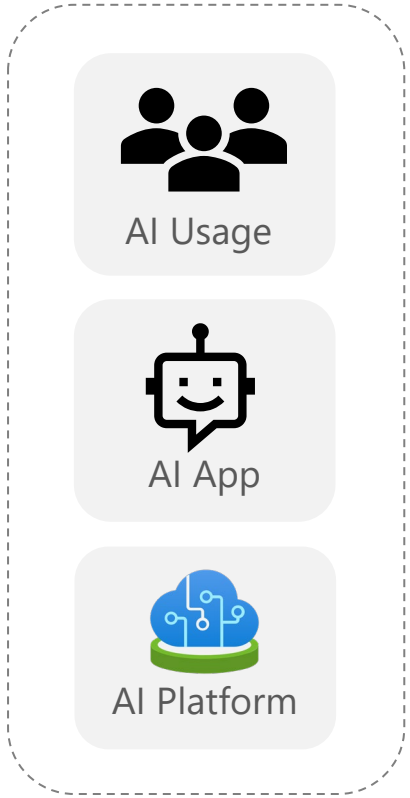
Guardrails

AI Usage Mitigations:
1. AI Acceptable Use Policy
2. Prompt-input filtering
3. Content attachment controls
4. Data Loss Prevention

AI Application Mitigations:
1. Prompt Engineering
2. Retrieval Augmented Generation
3. Prompt Shields (Spotlighting)
4. Risk and Safety Monitoring
5. Secure Plugin Management
6. Sensitive Data Handling

AI Platform Mitigations:
1. Groundedness detection
2. RAI content safety filters
3. AI assisted safety evaluations
4. Safety system messages
5. Prompt response rate limiting (50)
6. Disable backtracking

LLM Development Mitigations:
1. Training data cleanse
2. Strict control of model weights
3. Strict access to modify algorithms



AI Usage Rules
Prevent harms through application usage policy, user input restrictions, and data loss prevention safeguards

GAI App Safety
Apply content inspection, filtering, and prompt-engineering rules inside the app for specific harms and attack methods

AI Platform – Deep Safety
Apply content safety filtering and meta-prompt safety mechanisms at the platform layer to protect all GAI Apps

AI Training
Ensure safeguards are built into the LLM training and data cleansing practices

Use-case Specific

Ability to apply specific policies in each region, industry, company, or department

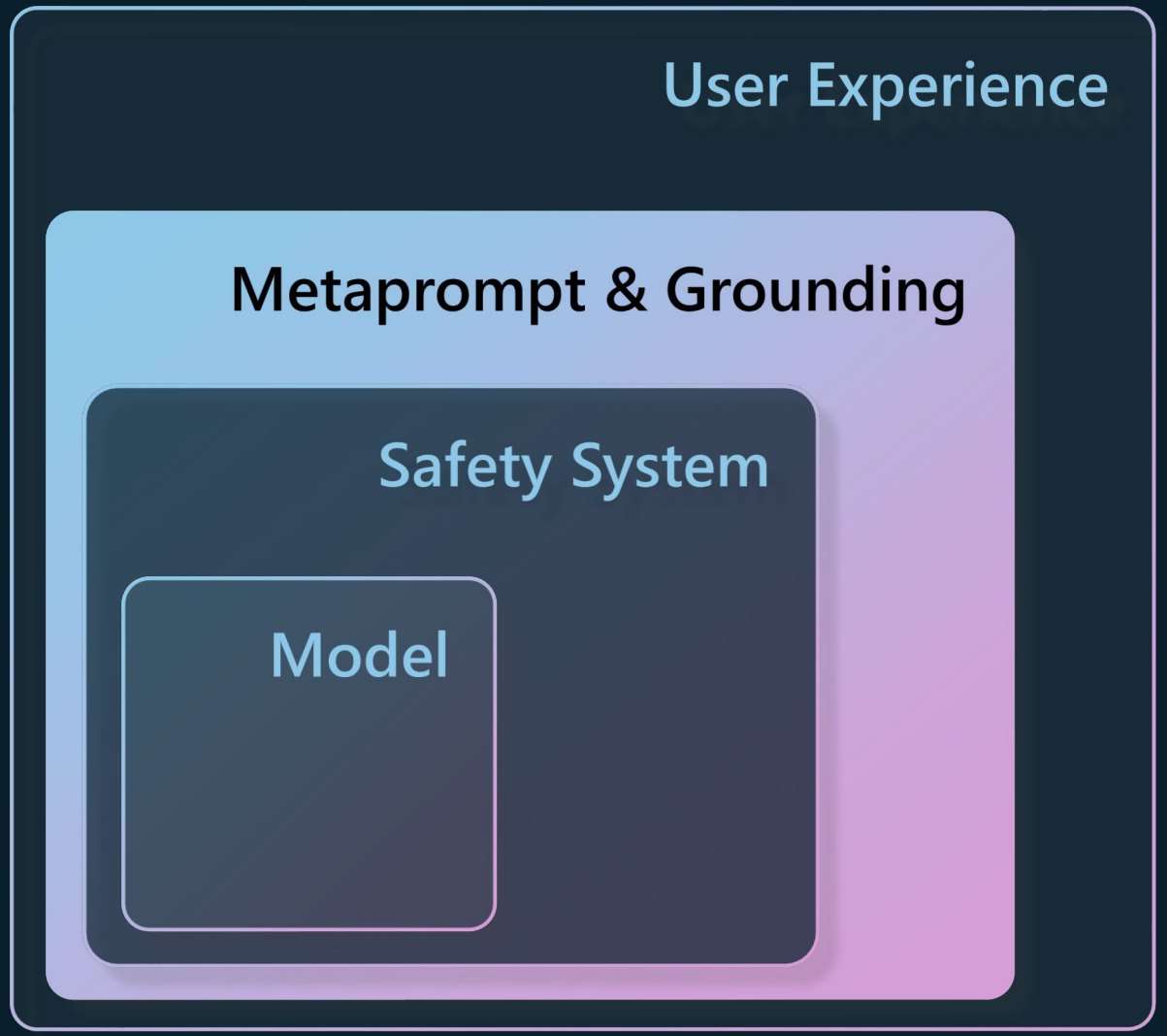
Application owners can make decisions on content control based on the desired outcome of their service, limited by the controls in the platform it is built on.

Controls applied at this level impact capabilities of the platform for all users

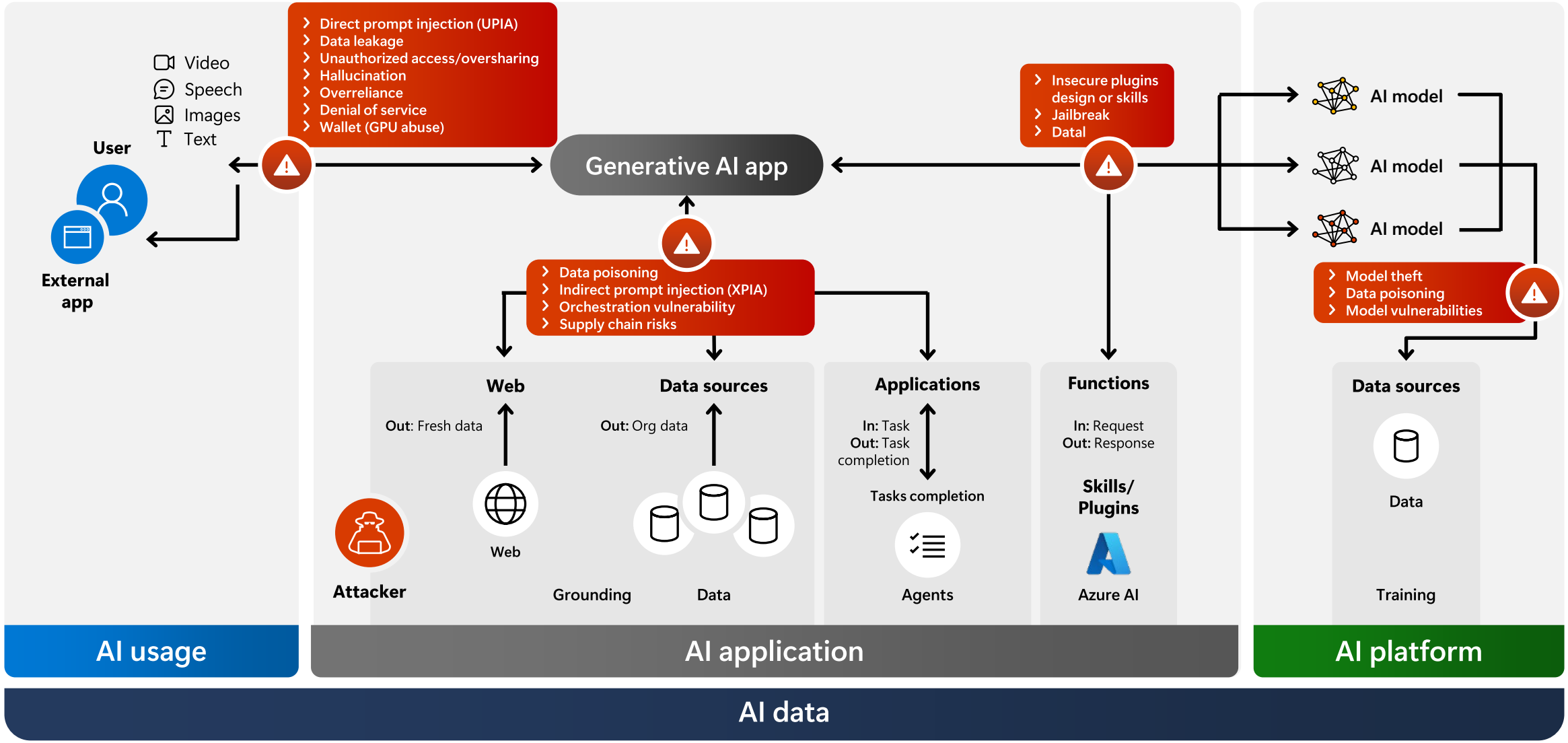
Broad Categorization

Announcing new tools in Azure AI to help you build more secure and trustworthy generative AI applications

Risk Mitigation Layers

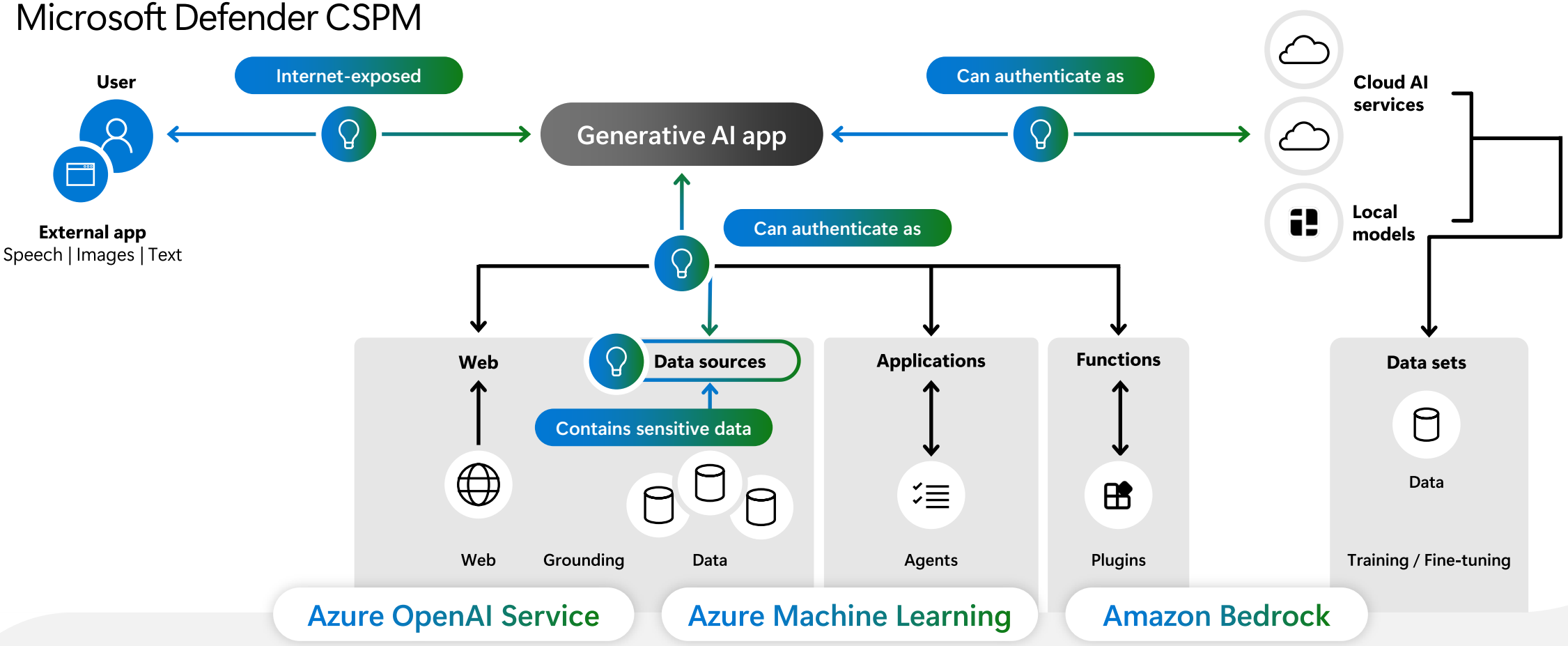


Generative-AI threat landscape (AI Application)



Strengthen your AI security posture

Microsoft Defender CSPM



Discover AI workloads, models, and SDKs across your environment

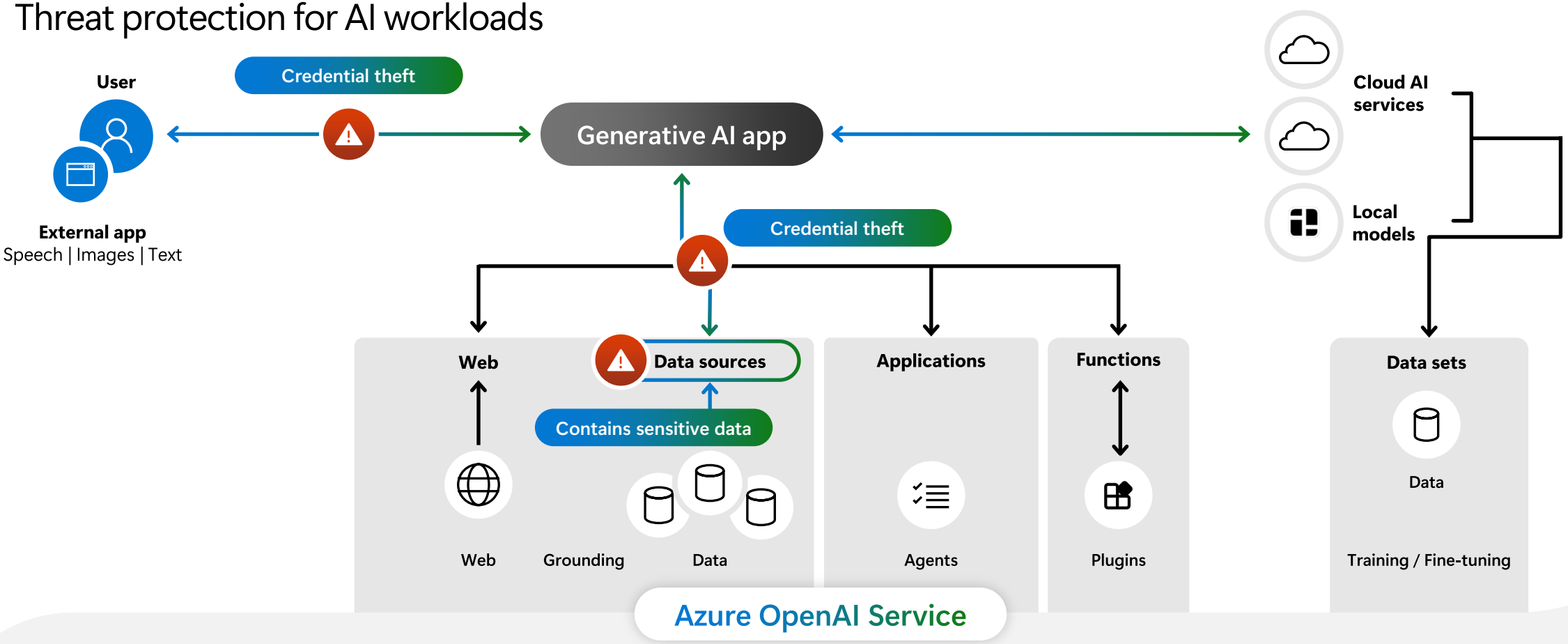
Identify GenAI risks and vulnerabilities from code to runtime

Map direct and indirect attack paths to your GenAI resources

Mitigate risks to sensitive data used across your deployments

Detect and respond to threats impacting AI workloads

Threat protection for AI workloads



Monitor for jailbreak attacks, sensitive data exposure, and credential theft

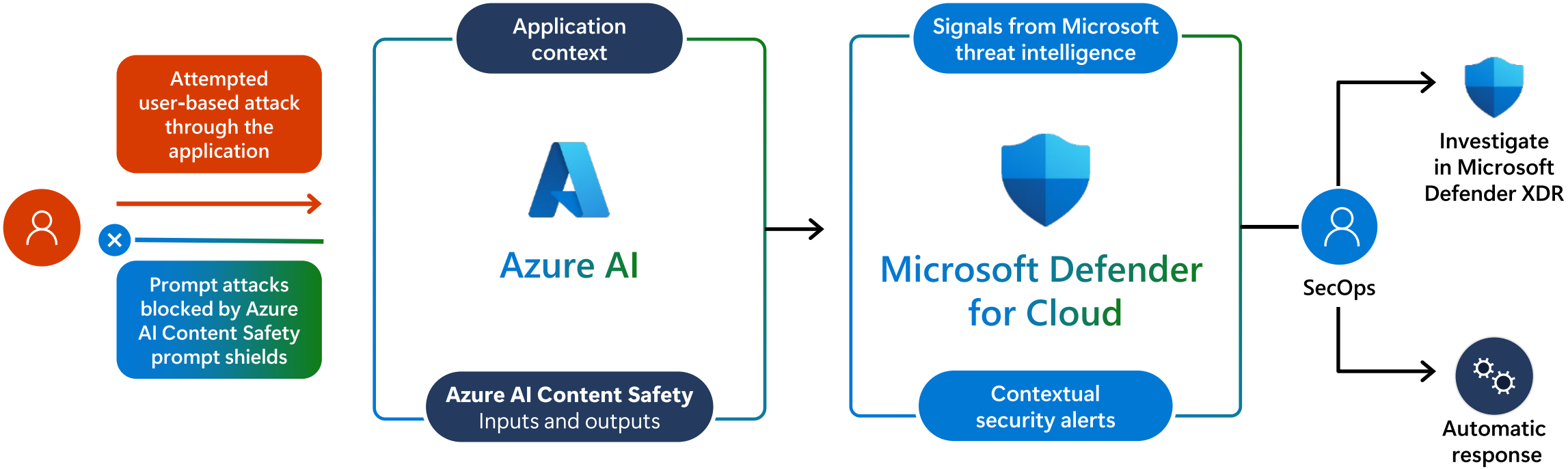
Integrated application context and Azure AI Content Safety findings

Robust alerts enriched with Microsoft Threat Intelligence

Correlate alerts into incidents and investigate in Microsoft Defender XDR

Threat protection for AI workloads using Azure OpenAI

Microsoft Defender for Cloud



Legend: Security teams | Developers

AI Security Learning Topics

Three lenses: How does Microsoft think about/approach these topics, how does Microsoft secure SaaS AI, how can we secure PaaS AI

Ethical Framework

- Why its important to you/organization (liability)
- Providing AI to the world
  - Protecting Civil Rights
  - Protecting Democracy
  - Enabling AI for Good
- Hosting AI Platform
  - Collection of Logs
  - Restrict Access to Logs
  - Using Telemetry for Good
  - Sharing Threat Intelligence

Types of AI

- Generative (GAI)
  - Large Language Model (LLM)
  - Small Language Model (SLM)
- Single-purpose
  - Machine Learning
  - Object Recognition
  - Robotics
  - Reasoning
  - Decision Making
  - Problem Solving

Attack

- Front End (In Session)
  - Prompt Injection
  - Direct
  - Indirect
  - Poisoned Content
- Back End (AI System)
  - Model Manipulation
  - Data Poisoning
  - AI Infrastructure
  - AI Supply Chain

Defense

- Separation of Instruction and Content
- Prompt Analysis with Adversarial Training
- Response Validation
- Secure Plugin Architecture
- Zero Trust Principles (ZT)
- Minimize the Attack Surface (ASR)
- Human-in-the-Loop
- User Guidance & Feedback Loop
- AI Red Team for Proactive Discovery
- AI Threat Response
- .

Existing Risk

- Inherits and amplifies existing risk
- Technical Debt
  - Lack of DevSecOps
  - MFA not applied everywhere
  - Immature data labeling
  - Lack of DLP enforcement
- Integration with ...

GAI App Architecture

- Threat Modeling
- Continuous SDL
- Identity & Access Management
- User Prompt Management
- Meta Prompt Engineering
- Content Management (DLP)
- Plugins, Functions, Skills, Connectors
- AI Safety Controls
- GAI Platform Integration
- .

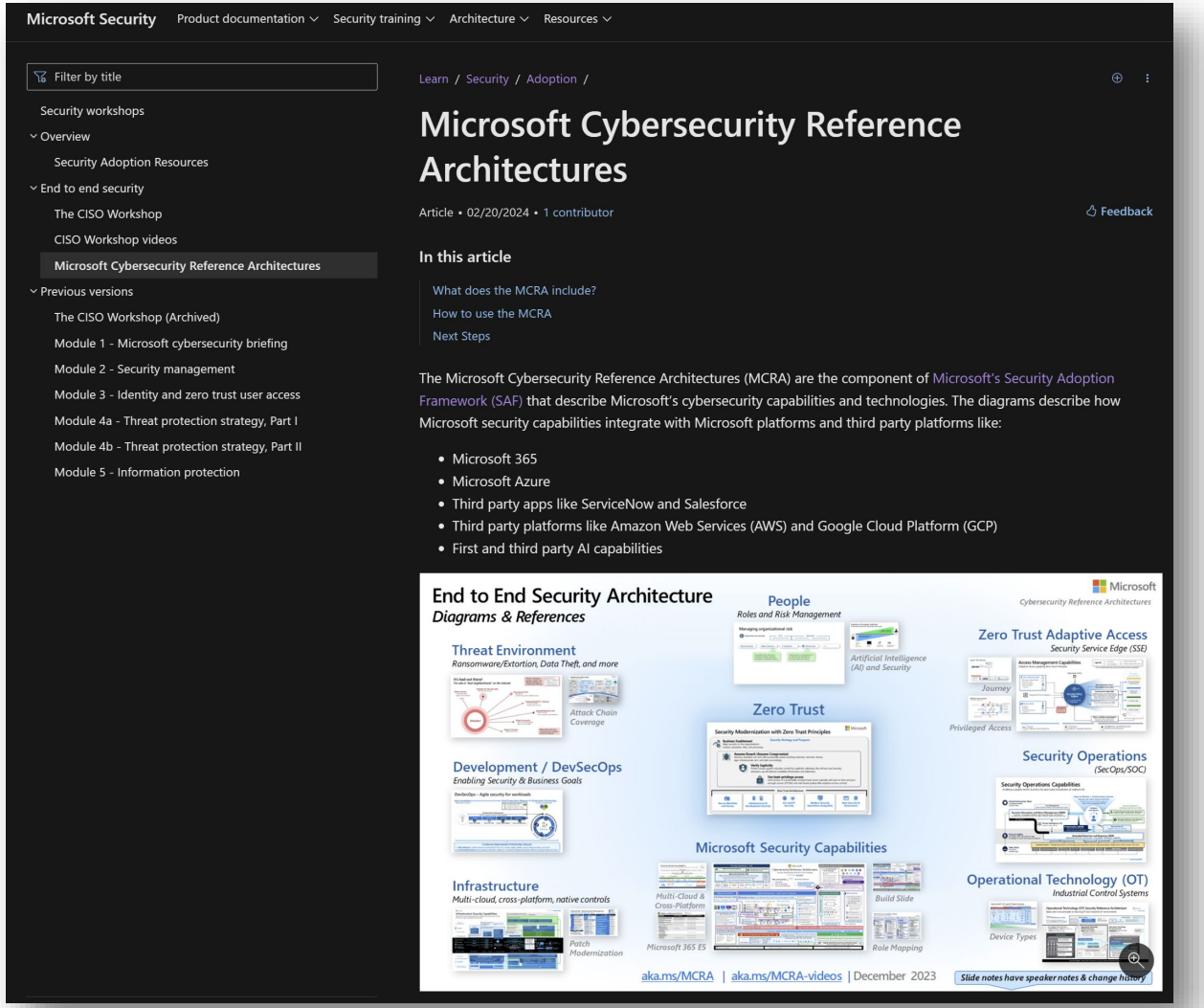
Harms & Risks

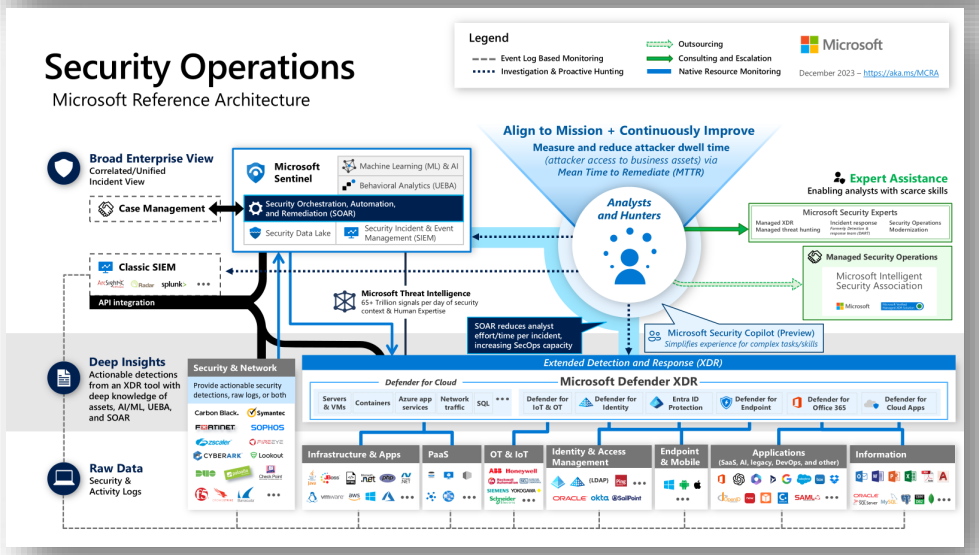
- Non-deterministic outcomes
- Accountability
- Factual Errors
- Biased Outcome
- Over Reliance
- Excessive Agency
- Trust / Reputation
- Data Privacy & Ownership
- Intellectual Property
- Unintended Consequence (aka. The Monkeys Paw)
- .

AI Governance

- AI Shared Responsibility Model
- Transparency & Communication
- Configuration Management
- Telemetry Classification
- Sensitive Data Handling
- .

Microsoft Cybersecurity Reference Architectures





[Microsoft Cybersecurity Reference Architectures (MCRA) | Microsoft Learn]

Copilot for Security Value Proposition

Microsoft Copilot for Security is a generative AI-powered assistant for daily operations in security and IT. Copilot for Security empowers teams to protect at the speed and scale of AI by turning global threat intelligence, industry best practices, and organizations' security data into tailored insights to outsmart and outpace adversaries.

1. Scale → Catch what others miss
2. Speed → Outpace adversaries
3. Skilling → Strengthen team expertise

Copilot for Security working with Microsoft Security

Microsoft Defender XDR

- Summarize incidents quickly
- Act on incidents through guided responses (Triage, Containment, Investigation, Remediation)
- Get results fast when analyzing scripts and codes or reverse engineer malware
- View similar incidents and further review the actions done in those similar incidents.
- The View similar emails action, which is specific to phishing incidents
- And more...

Microsoft Sentinel

- Generate and run hunting queries
- Pivot your investigation to specific incident assigned or related to an alert
- Get details on any entity and justifications
- Assess incidents and alerts with supporting evidence and recommendations.
- Summarize the findings from the investigation and conclude with a set of recommendations.
- And more...

Microsoft Intune

- Compare different security baselines.
- Get a summary of an existing policy.
- Get policy assignment scope.
- Get the differences or comparisons between two devices.
- Quickly gather details for a device by asking about it.
- Get detailed information about a user's device enrollments and device compliance for troubleshooting or a security investigation.
- And more...

Defender TI

- Share the IOCs or TTPs or tell me more about associated with Silk Typhoon.
- Share the technologies that are susceptible to the vulnerability CVE-2021-44228; or Summarize
- Show me the latest CVEs.
- Show me threat actors associated with CVE-2021-44228.
- Show me the threat articles associated with CVE-2021-44228.
- And more...

Microsoft Copilot for Security

Common: • Run queries using natural language • Prepare reports, summaries, and graphs • Upskill teams via prompts and guidance

Microsoft Purview

- Expedite complex data security, data risk and user risk surfaced
- Gain comprehensive summary of DLP alerts and/or insider risk alerts
- Gain contextual summary of communication risks
- Gain contextual summary of evidence collected in review sets
- compliance, and legal investigations with AI-powered summarization capabilities and natural language queries
- And more...

Microsoft Entra

- Discover high risk users, overprivileged access, and suspicious sign-ins that aid in a security incident investigation.
- Troubleshoot daily identity tasks such as why a sign-in required multi-factor authentication (MFA).
- Inquire about users, groups, sign-ins, and permissions then instantly get a risk summary and recommended guidance for each identity at risk.
- Create a Lifecycle Workflow to streamline the process of creating and issuing user credentials and access rights.
- Assisted risk investigation embedded experience in Entra - private preview
- Assisted sign-in troubleshooting embedded experience in Entra - private preview
- Assisted workflow creation embedded experience in Entra - private preview
- And more...

Defender EASM

- Get attack surface summary.
- Get attack surface insights.
- Get assets affected by CVEs by priority or CVE ID.
- Get assets by CVSS score.
- Get expired domains.
- Get expired SSL certificates.
- Get SHA1 certificates.
- And more...

Copilot for Security Value Proposition

1 Scale → Catch what others miss

Threat signals and security alerts create noise that conceal attackers. Copilot for Security enables teams to reason over real-time threat signals and their enterprise data to cut through the noise, detect threats before they cause harm, and reinforce security posture.

Interoperability and Enhanced Insight

Copilot for Security is engineered for seamless interoperability within the Microsoft security ecosystem, including Microsoft Security products and threat intelligence services. This integration enables teams to delve deeper into security analytics, offering them more profound insights. By tapping into the combined strengths of Microsoft's comprehensive security tools, Copilot for Security not only enriches data analysis but also equips organizations with the intelligence needed to anticipate and mitigate cybersecurity threats more effectively.

Hyperscale Cloud-Powered Real-Time Analysis

Powered by Microsoft's advanced hyperscale cloud infrastructure, Copilot for Security offers unparalleled capabilities in conducting real-time analysis. This feature empowers teams with immediate, critical insights, fostering a proactive stance against cyber threats and facilitating a more informed decision-making process.

Enhanced Accuracy

A meticulously conducted trial demonstrates a significant leap in precision, where users engaging with Copilot for Security recorded a 44% increase in task accuracy. This data underscores the system's capability to refine decision-making and enhance operational efficiency in cybersecurity endeavors.

Improved Quality of Work

In a comprehensive survey, an overwhelming 86% of participants testified to a qualitative improvement in their work output after integrating Copilot for Security into their workflows. This feedback highlights the platform's role in elevating the standards of cybersecurity practices.

Copilot for Security Value Proposition

2 Speed → Outpace adversaries

During security incidents, every minute counts. Copilot for Security puts critical guidance and context at security teams' fingertips so they can respond to incidents in minutes instead of hours or days.

Streamlined Query Responses

Empowering Teams with Efficiency: Copilot for Security transforms the way analysts and administrators interact with data. By enabling queries in natural language, the system intelligently generates outputs in script, KQL, or KeyQL formats. This innovation drastically cuts down the time teams spend on manual data analysis and query formulation.

Incident Reporting Revolutionized

Dramatic Time Savings in Incident Documentation: Microsoft Defender Experts have experienced a 90% reduction in the time required to write and publish incident summaries, thanks to Copilot for Security. This efficiency revolutionizes incident management, freeing up valuable time for more critical security tasks.

Significant Time Savings in Security Operations

Reimagining Efficiency in Security Operations: Early adopters of Copilot for Security report a remarkable up to 40% time savings on standard security tasks. For more repetitive tasks like alert triage and reporting, the efficiency gains are even more pronounced, reaching up to 60%. This leap in productivity allows analysts to devote their attention to more complex and impactful security challenges.

Accelerated Task Completion

Enhancing Operational Speed: In controlled trials, Microsoft Copilot for Security has demonstrated a significant 26% reduction in task completion times. This acceleration across various tasks showcases the platform's ability to streamline workflows and improve productivity.

Broadening Strategic Horizons with Faster Response Times

Beyond Risk Mitigation: Speed is of the essence not only in reducing risk during security incidents but also in facilitating quicker responses to stakeholders. The time saved in security and IT operations through faster execution allows professionals to allocate more focus on strategic initiatives, enhancing overall business value.

Copilot for Security Value Proposition

3 Skilling → Strengthen team expertise

Security teams must continuously elevate their expertise to stay ahead in an evolving threat landscape. Copilot for Security enables junior staff to perform more advanced capabilities and redirects expert staff to the hardest challenges, thus elevating the proficiency of the entire team.

Empowering Junior Analysts with Advanced Capabilities

Security's ability to translate natural language into Kusto Query Language and perform sophisticated script analysis.

Delivering Expert-Level Security Insights

Security experts acknowledge Copilot for Security's capability to produce analyses on par with mid-to-expert level analysts, particularly in summarizing incidents, analyzing scripts, and assisting with queries.

Sustaining Investigative Momentum with Stateful Technology

Copilot for Security's stateful functionality ensures that teams can seamlessly continue their investigative work across sessions, preserving momentum and enhancing investigative efficiency.

Flexible Integrated Workflow Options

Offering both standalone and embedded modes, Copilot for Security provides users with the flexibility to integrate AI-powered assistance into their preferred workflows.

Guided Response Features

Leveraging the Guided Response feature of Copilot for Security has led to a 73% improvement in accuracy when querying about remediation steps, marking a significant advancement in precision.