Threat Modeling in the Age of AI

•••

Disesdi Susanna Cox

OWASP AI Exchange

OWASP Global AppSec 2024
San Francisco, CA, USA

About Me - Susanna aka Disesdi

Red teamer -> Data scientist -> MLE -> AI architect -> Chief Data Officer

Tenth generation security professional

Author, “Securing AIML Systems in the Age of Information Warfare”

OWASP AI Exchange Core Team

Daughter of the American Revolution

Survivalist

Owned by 3 pugs

Threat Modeling in the Age of AI

Two quotes:

“An ounce of prevention is worth a pound of cure.” - Unknown

“Don’t make the perfect the enemy of the good.” - Voltaire

Introduction: Threat modeling 101

- Rising threat: The 2023 World Economic Forum's (WEF) Global Risks Report ranks cybersecurity as a top threat, interlinked with privacy, digital inequality, infrastructure breakdown, terrorism, & more
- Interconnectedness of systems increases attack surface & amplifies failure modes
- AIML interconnected systems are increasingly embedded in mission-critical aspects of daily human life

System Resilience is the Goal

- The goal: Cyber Resilience
- Proactive vs Reactive security
- We can’t prevent every cyber incident, but we can build resilience into our systems

From Reactive to Proactive: Modeling Threats

- How do we move from being reactive to proactive in our security stance?
- Proactive security starts with understanding the threat landscape

But,,,

- If we had to learn every threat to every system, it would take forever!
- Understanding threats could quickly become prohibitively difficult
- **Every system is unique – time to use a model**

What Threat Modeling Is & What It Isn’t

“Threat modeling is analyzing representations of a system to highlight concerns about security and privacy characteristics.”

“At the highest levels, when we threat model, we ask four key questions:

- *What are we working on?*
- *What can go wrong?*
- *What are we going to do about it?*
- *Did we do a good enough job?”*

Source: Threat Modeling Manifesto

What Threat Modeling *Is*

- A structured, systematized approach
- Clearly articulated
- Understood by stakeholders
- Consistently applied (with adaptations documented)

What Threat Modeling *Is*

- A tool to contextualize risks
- “Cyberattacks” (big quotes) are a risk–but how?
- What vectors?
- How likely is each attack?
- And what might the effects be?

What Threat Modeling *Is*

- **A means of preparing & documenting mitigations**
- This may be one of the more overlooked aspects of threat modeling
- Once we know what are threats are, we can begin to prepare our response

What Threat Modeling *Is*

- The original "purple team" technique ^_^
- How do we get defenders to think like attackers?
- Threat modeling!
- AI systems **require** a purple team approach

Threat Modeling **Isn't**

Threat modeling anti-patterns (via *Threat Modeling Manifesto*)

- **Hero Threat Modeler:** Threat modeling does not depend on one's innate ability or unique mindset; everyone can and should do it.
- **Admiration for the Problem:** Go beyond just analyzing the problem; reach for practical and relevant solutions.
- **Tendency to Overfocus:** Do not lose sight of the big picture, as parts of a model may be interdependent. Avoid exaggerating attention on adversaries, assets, or techniques.
- **Perfect Representation:** It is better to create multiple threat modeling representations because there is no single ideal view, and additional representations may illuminate different problems.

The New AI Landscape: Data is the Vector

- 2012: Data is the “new oil”
- 2024: Data is the new attack vector

A Quick Intro To AI/Mlops

- Why are Ops so important to AIML systems?

  Because our goal is ***inference*** *at* ***scale*** .
- **Inference** :

  Any system that works with a gradient can be tricked

  This can be thought of as a cognitive bias
- **Scale** :

  Massive data requirements mean a new frontier in data provenance

  AIML deployments require a suite of new of Ops techniques for monitoring in production

A Quick Intro To AI/Mlops

DevOps vs AI/MLOps: familiar concepts, new systems

- Familiar: Communication, integration of expertise, & operationalization
- New: Continuous monitoring of data, AIML-specific systems
- Goal: De-siloing development expertise across fields to integrate for continuous deployment

A Quick Intro To AI/Mlops

Continuous deployment in AIML means:

- Continuous data processing
- Acquisition, validation, pipelines, oh my!
- Don’t forget: Your data scientists are probably running experiments - generating data, using data
- This is your IP - where is it going, where does it live, what processes are in place to make sure your R&D is secure?

A Quick Intro To AI/Mlops

Continuous deployment in AIML means:

- Continuous monitoring
- Model output can degrade - would you know if it did? How?
- Data quality can degrade - population changes, concept drift, malicious activity - would you know if it did? How?

A Quick Intro To AI/Mlops

Continuous deployment in AIML means:

- De-siloing: who are the actors?
    - Developers
    - DevOps
    - Data scientists
    - MLEs

A Quick Intro To AI/Mlops

What is MLSecOps?

- Machine Learning Security Operations - the integration of AIML-specific security mitigations into the MLOps pipeline
- Production-grade AIML at scale is **impossible** without MLSecOps

MLSecOps Architecture

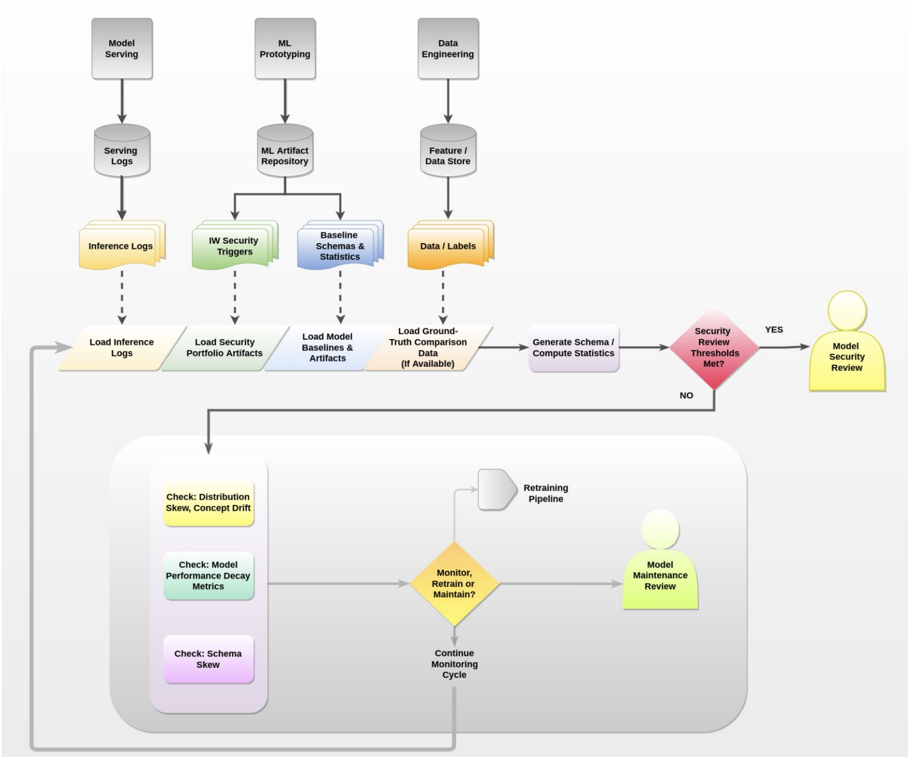
Model Serving
ML Prototyping
Data Engineering
Serving Logs
ML Artifact Repository
Feature / Data Store
Inference Logs
IW Security Triggers
Baseline Schemas & Statistics
Data / Labels
Load Inference Logs
Load Security Portfolio Artifacts
Load Model Baselines & Artifacts
Load Ground-Truth Comparison Data (If Available)
Generate Schema / Compute Statistics
Security Review Thresholds Met?
YES
Model Security Review
NO
Check: Distribution Skew, Concept Drift
Check: Model Performance Decay Metrics
Check: Schema Skew
Retraining Pipeline
Monitor, Retrain or Maintain?
Model Maintenance Review
Continue Monitoring Cycle

Fig 6: Continuous monitoring pipeline with model security checkpoint

How AI Systems Differ From Traditional Deployments

- Inference **at** scale necessitates operationalizing **to** scale
- Begin with the end in mind - how will this data product scale? What about the pipelines?
- It’s not just the pipelines–this understanding begins with the product itself.

Threat Modeling Your AIML Systems

First 3 questions:

1. Is it secure?
2. Can we operationalize?
3. Does it scale?

Threat Modeling Your AIML Systems

Is it secure?

- Gets developers, data scientists, & MLEs in the security headspace
- Review AIML-specific systems
- Do you *really* know all your AIML dependencies?

Threat Modeling Your AIML Systems

Can we operationalize?

- Because the goal of any AI system is inference at scale, operationalization is key
- If you don’t have a plan to operationalize monitoring of your systems, outputs, and data, you very likely have security problems

Threat Modeling Your AIML Systems

Does it scale?

- If you have a plan to operationalize, but it’s unclear how it scales....you probably have security problems

Mapping the attack surface

First 3 steps to understanding your AIML system attack surfaces

1. Know your data flows
2. Know your data provenance
3. Know your data governance

Mapping the attack surface

Utilyze the OWASP AI Exchange

- Threats through use
- Development-time threats
- Run-time application security threats

Thank you!

Resources:

- OWASP AI Exchange
- Threat Modeling Manifesto
- Threat Modeling Capabilities
- Securing AIML Systems in the Age of Information Warfare