Module 9

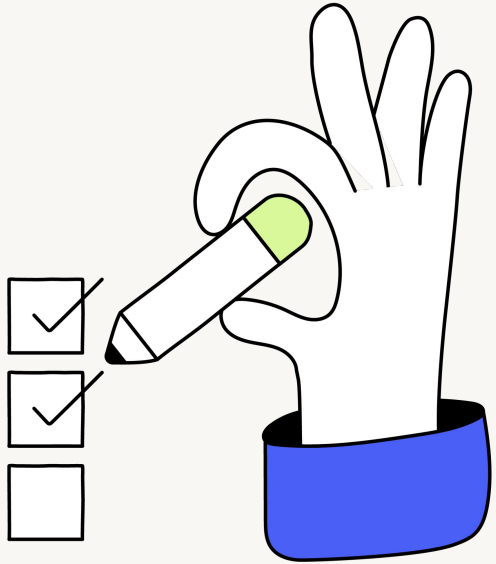
Workshop 1: Designing secure ML systems: A threat modeling and risk mitigation lab

Icebreaker

Security blind spots: What gets missed?

What’s one security risk you think often gets overlooked in ML systems?

Drop your answers in the chat.



Module 9

Workshop 1: Designing secure ML systems: A threat modeling and risk mitigation lab

Agenda

- **Review:** Recap key concepts from async units 1 and 2.
- **Demo:** A guided security walkthrough.
- **Practice:** Securing the PackSure ML system.
- **Closing:** Key takeaways and next steps.

Learning objectives

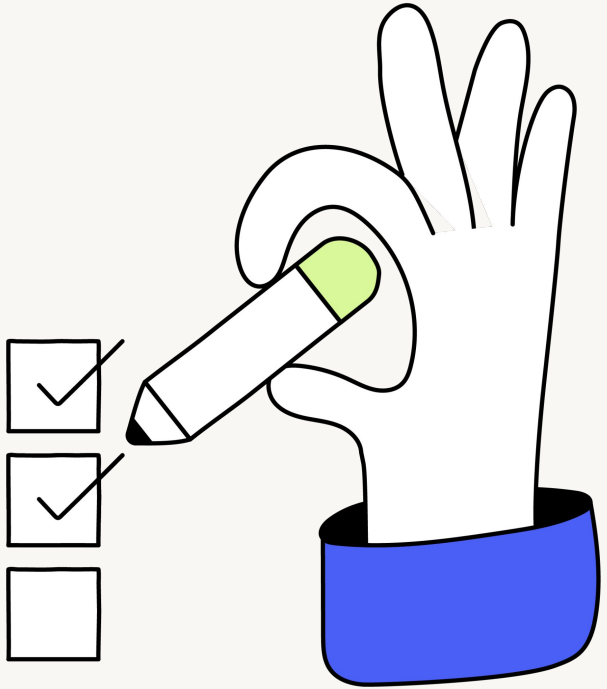
- **Analyse vulnerabilities** across the ML system lifecycle, mapping them to core security principles.
- **Design secure ML workflows** that address identified risks, including infrastructure, API, and model-level threats.
- **Apply threat modeling techniques** to a real-world ML scenario, identifying mitigation strategies at each stage.
- **Develop basic team protocols** for monitoring, incident response, and continuous security assessment.

Async review: Unit 1

Go to My Multiverse, Module 9
Workshop 1

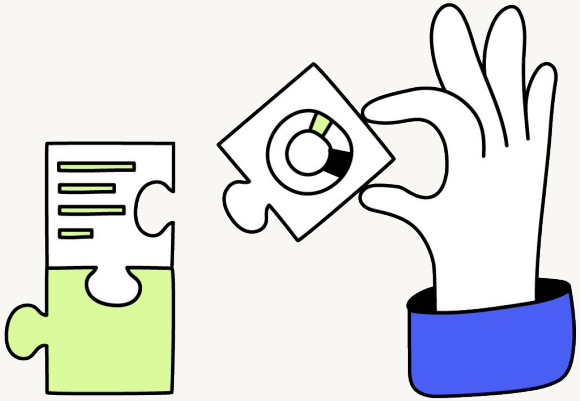
Navigate to the **“Async review”** page

Review the summary of Unit 1



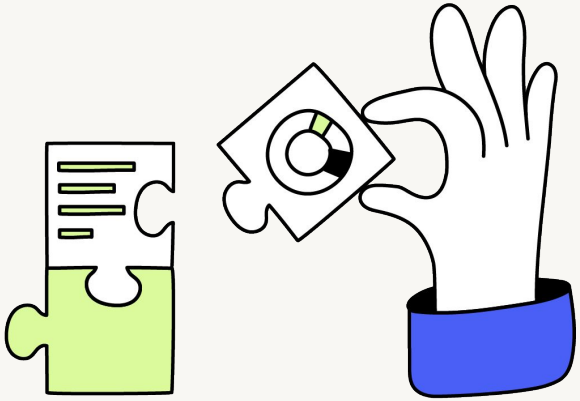
Security Fundamentals in Machine Learning

- **Core security principles –** Confidentiality, integrity, authentication, non-repudiation, and service integrity in the context of ML workflows.
- **Lifecycle threats –** Common vulnerabilities across the ML lifecycle, including data poisoning, model inversion, insecure APIs, and audit gaps.
- **Infrastructure decisions** – Trade-offs between local, cloud, and hybrid setups in terms of control, compliance, scalability, and risk exposure.



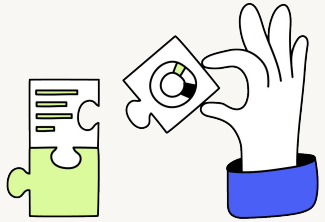
Security Fundamentals in Machine Learning

- **Secure workflow design –** Techniques such as threat modeling, least privilege, encryption, and monitoring embedded into ML pipelines.
- **Team protocols and culture –** Role-based responsibilities, incident response practices, and a culture of continuous security improvement.



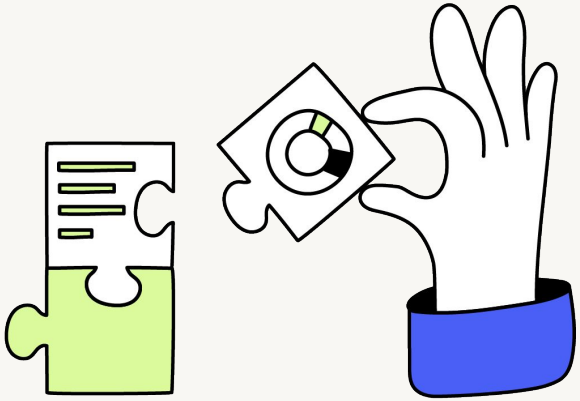
Five principles shape every stage of a secure ML system

Principle	Focus area	ML example
Confidentiality	Preventing unauthorised access to data/models	Encrypting training data and protecting model outputs
Integrity	Ensuring correctness of data and models	Detecting tampering in training labels or model weights
Authentication	Verifying user/system identity	Using API keys and MFA to restrict model access
Non-repudiation	Enabling traceability of actions	Logging model updates and access events
Service integrity	Defending runtime model behavior	Blocking adversarial inputs and monitoring for drift



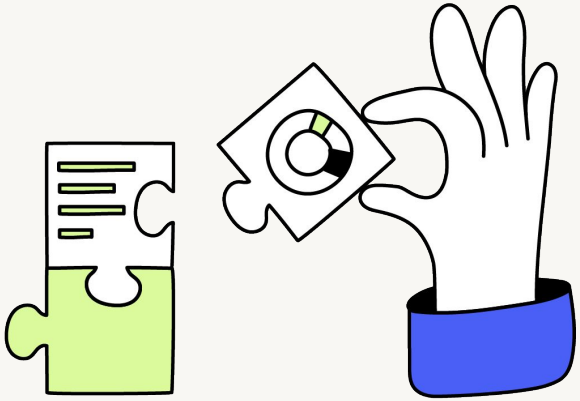
Threats across the ML lifecycle: Where can things go wrong?

- **Data poisoning:** Malicious input during training corrupts model behavior.
- **Model inversion:** Attackers reconstruct training data from outputs.
- **Adversarial inputs:** Carefully crafted inputs trick deployed models.
- **API misuse:** Open or unauthenticated endpoints expose model logic or data.



Secure ML practices

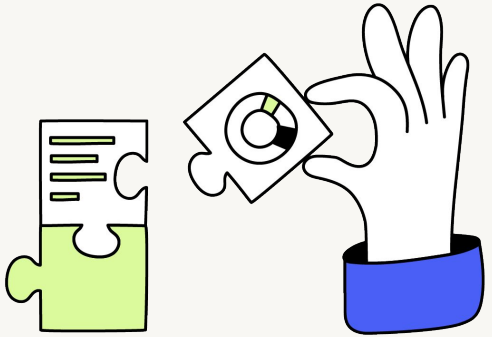
- Encrypt sensitive assets.
- Apply role-based access controls.
- Use secure APIs with rate limiting.
- Log access and model events.



Who protects what in your ML system?

Security in ML systems is a shared responsibility across roles:

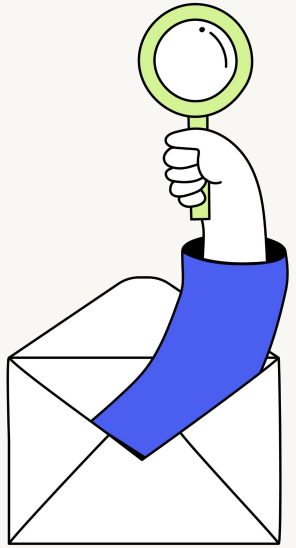
- **Data scientists:** Monitor model behavior and validate training data.
- **Engineers:** Secure data pipelines and configure access controls.
- **Platform admins:** Manage infrastructure and enforce CI/CD security.



Skills application: Security walkthrough

Skills application introduction

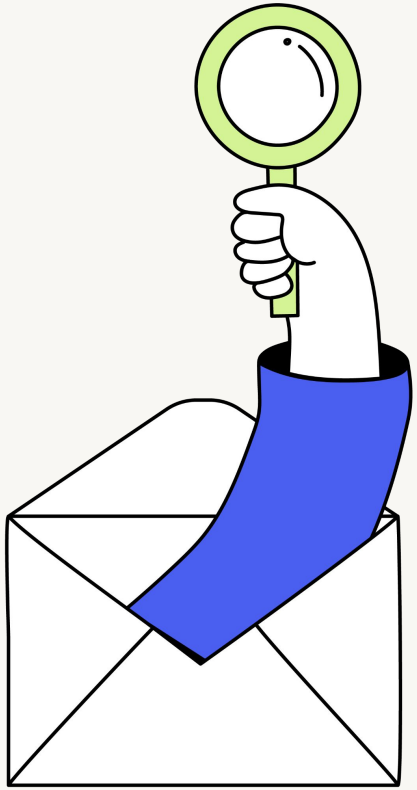
- **Objective:** In this skills application demo, we'll explore the end-to-end threat model for LogiFleet's ML system, which predicts delivery times using geolocation, traffic, and package data.
- **We'll show you how to:**
  - Spot vulnerabilities across data ingestion, training and deployment.
  - Map each risk to core security principles and practical mitigations.



Skills application: Security walkthrough

Skills application

- Work in breakout rooms in small groups.
- Use the Jupyter Notebook and provided dataset.
- The challenge is designed to be solved individually, but collaboration is encouraged.
- Find all instructions and resources on Ariel.

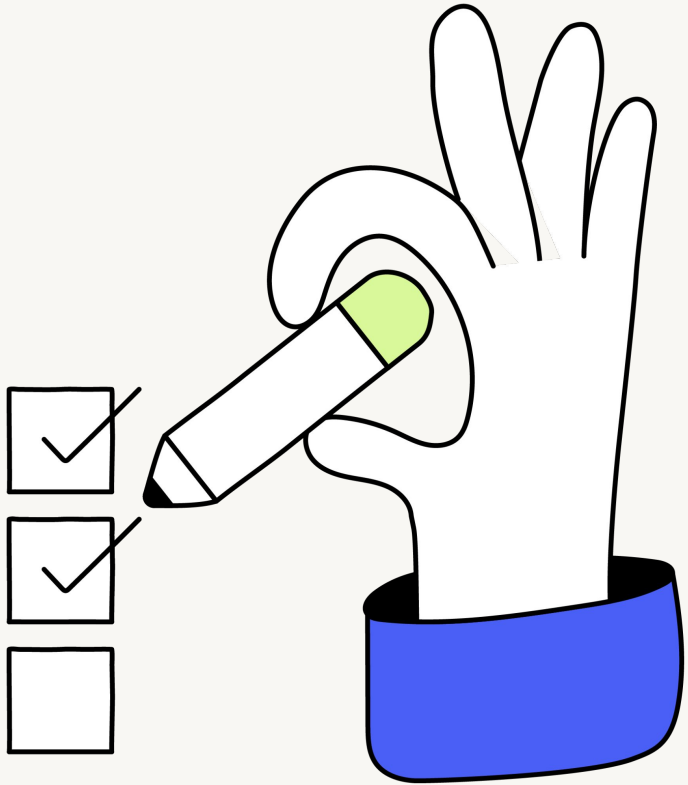


Access the skills application

Go to My Multiverse, Module 9 Workshop 1

Go to the Skills application page and follow the instructions provided

Be prepared to discuss your learnings when we come back together



Skills application debrief

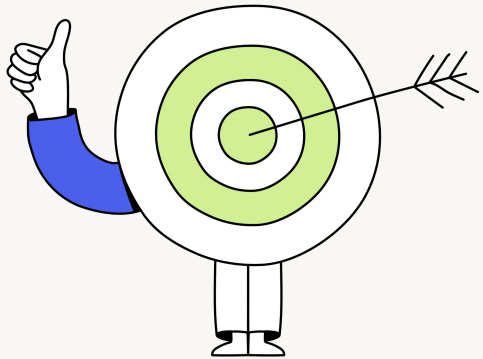
Skills application debrief

What risks did you identify and what steps did you take to secure them?

Key takeaways

1. **Security isn’t a feature—it’s a design principle:** Effective ML security starts at the planning stage. Building with threat modeling, encryption, and access control from the ground up is far stronger than patching holes later.

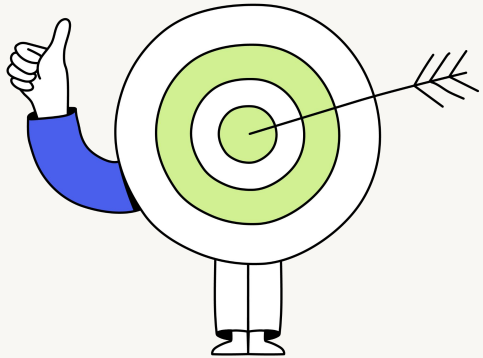
2. **Secure systems need clear roles and protocols:** A good security plan includes not just tools, but people. Define who’s responsible for monitoring, incident response, and escalation—then make sure everyone’s aligned and prepared.



Key takeaways

3 **Threats evolve—so should your security:**

The best defence is a living one. Continuous monitoring, regular audits, and a culture of proactive improvement are key to keeping your ML systems safe over time.



Applying key takeaways to your organisation

Applying key takeaways

1. How does your team currently handle ML system security, and where might gaps exist across the lifecycle?

2. What practical steps could you take to strengthen collaboration around security responsibilities?

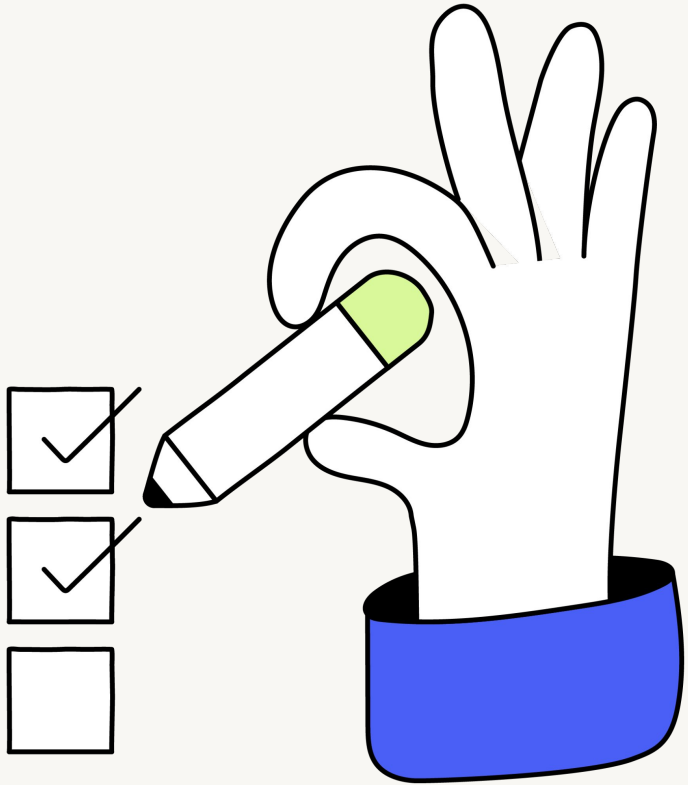
Share your plan to apply new skills

Go to My Multiverse, Module 9
Workshop 1

On the Apply new skills to your role page, find the discussion activity

Take 2 minutes to share how you will apply this skill in your role.

Take time this week to read what others have written and get new ideas!



Closing out

- **Next workshop:** Module 9 Workshop 2
- **Have questions?**
  - Chat with Atlas
  - Reach out to your Cohort Coach
  - Sign up for Tutoring
- **Log your OTJ hours**
- **Take the Workshop Survey** on My Multiverse

multiverse