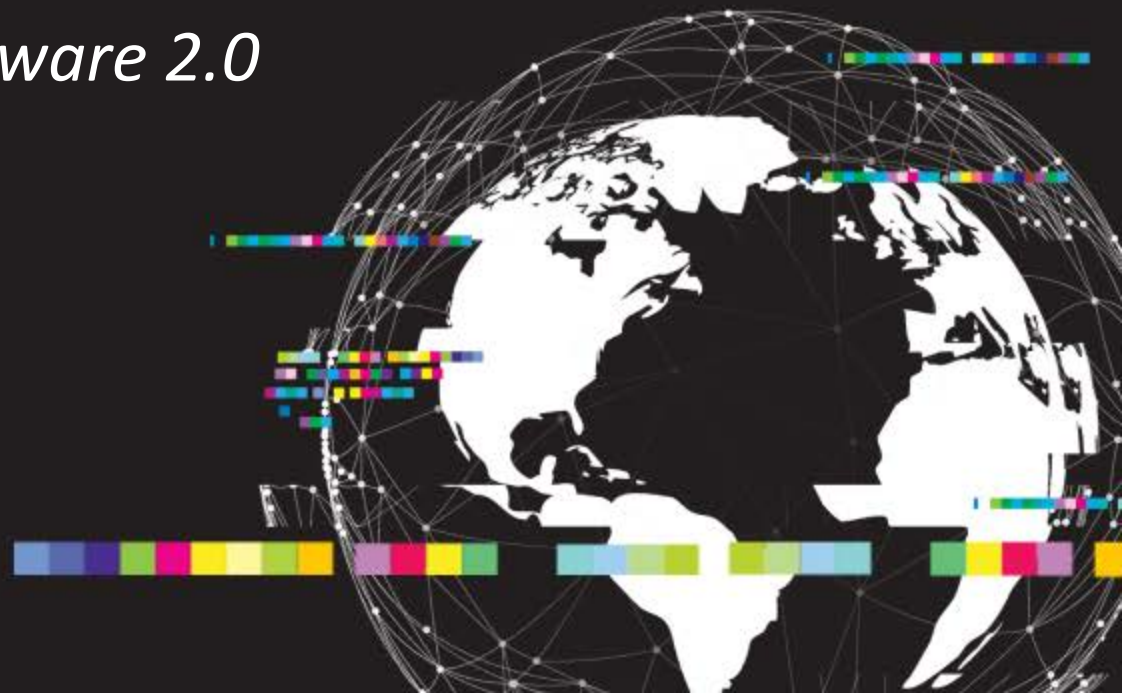
Introducing MLSecOps

DevSecOps for Software 2.0

ADVERSA



Introduction

About

Eugene Neelou

- Built various security products since 2011

About

Eugene Neelou

- Built various security products since 2011
- Used AI for security products since 2016

About

Eugene Neelou

- Built various security products since 2011
- Used AI for security products since 2016
- Coined the term MLSecOps in 2017

About

Eugene Neelou

- Built various security products since 2011
- Used AI for security products since 2016
- Coined the term MLSecOps in 2017
- Pioneering Security of AI since 2019
  - Author of the Industry Report on Adversarial ML
  - Organizer of the MLSEC AI Hacking Competition track
  - Contributor to the NIST AI Risk Management Framework
  - Co-Founder & CTO at Israeli research startup Adversa AI

About

Eugene Neelou

- Built various security products since 2011
- Used AI for security products since 2016
- Coined the term MLSecOps in 2017
- Pioneering Security of AI since 2019
  - Author of the Industry Report on Adversarial ML
  - Organizer of the MLSEC AI Hacking Competition track
  - Contributor to the NIST AI Risk Management Framework
  - Co-Founder & CTO at Israeli research startup Adversa AI

Adversa AI

- World-first AI vulnerability research company
- World-first commercial AI red team
- World-first AI protection patent
- and many more

Mission

Adversa AI Increases Trust in AI Systems

by Improving the Security of ML Algorithms

Agenda

- AI is Software 2.0
- Adversarial Machine Learning
- MLSecOps Pipeline Stages
- Takeaways

AI is Software 2.0



AI for Old Products

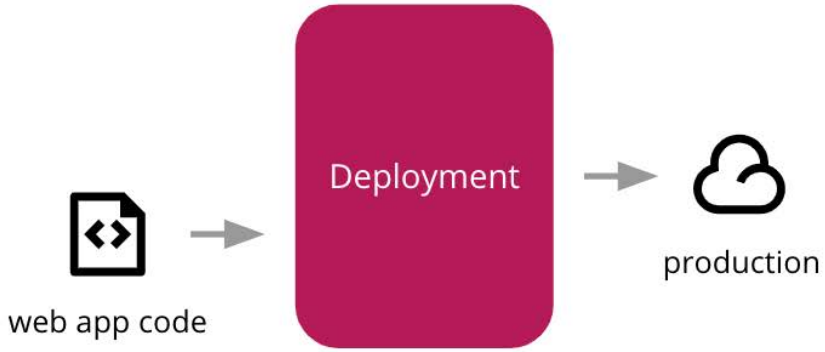
- AI is becoming a core technology for **existing products**
  - Most tech companies will transform into AI companies

AI for New Products

- AI is becoming a go-to technology for **new products**
  - Most disruptive products will emerge from AI-first ideas

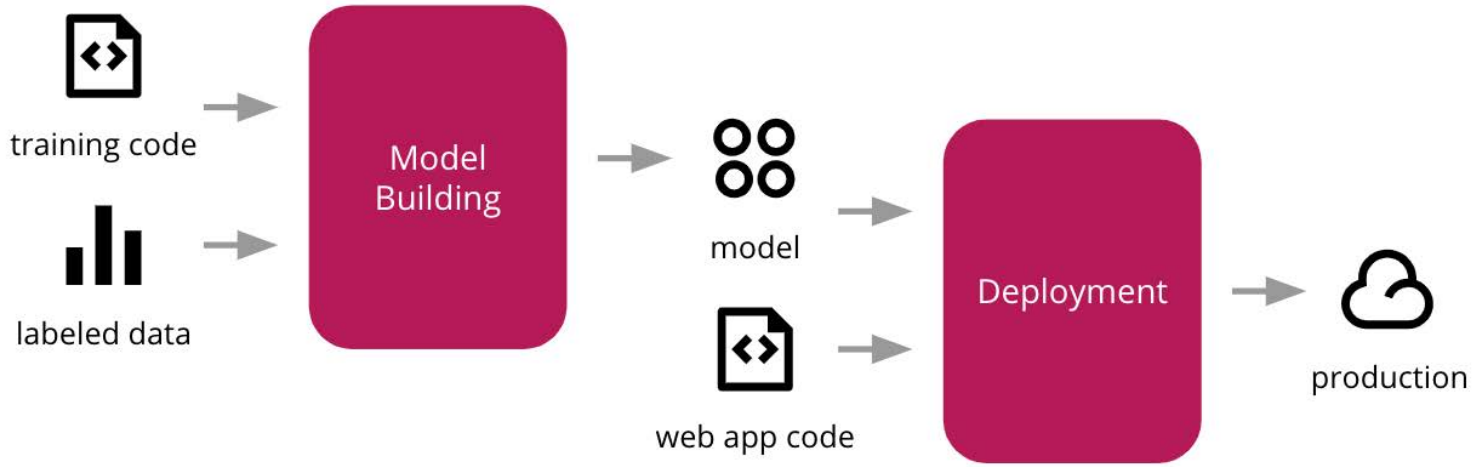
AI is Software 2.0

- AI is the **new paradigm** of software development



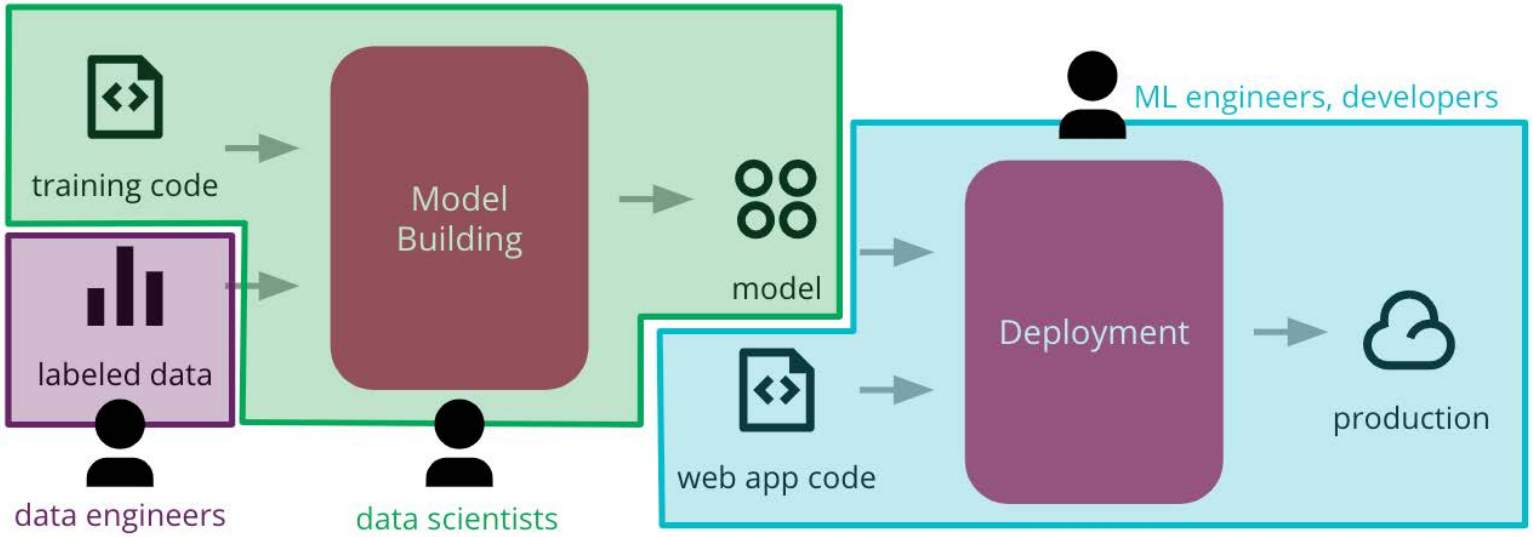
AI Development Process

- AI development brings **model** and **dataset** assets



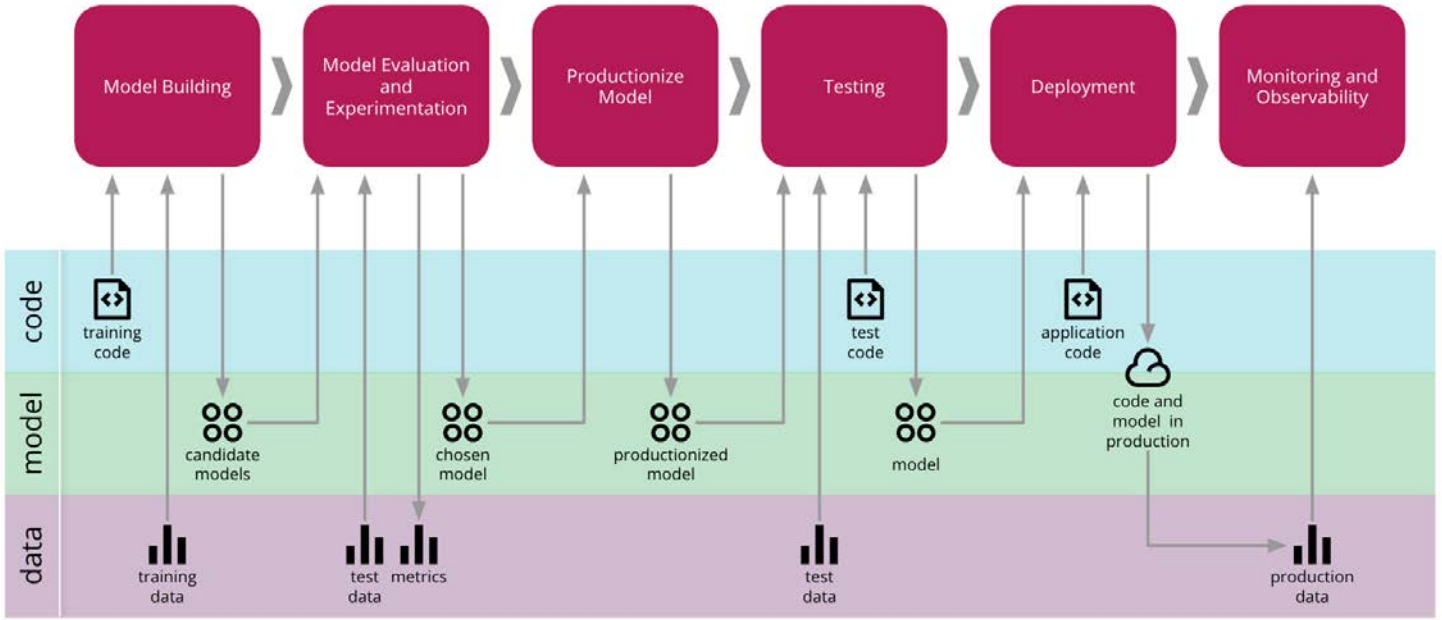
AI Development Process

- AI development requires **data engineers** and **data scientists**



AI Development Process

- AI development enabled by **MLOps pipelines**



Continuous Delivery for Machine Learning (CD4ML)

Adversarial Machine Learning

AI Threats Rising

HELPNETSECURITY

AI industry alarmingly unprepared for real-world attacks

VICE

Hackers Fool Facial Recognition Into Thinking I'm Mark Zuckerberg

The Daily Swig

Cybersecurity news and views

Machine learning security vulnerabilities are a growing threat to the web, report highlights

Tech Informed

The frontier of tech news

Bad Robot: the rise of adversarial AI

As the popularity of AI applications grows, so too do the risks associated with it

WSJ

Faces Are the Next Target for Fraudsters

DARKReading

Expect an Increase in Attacks on AI Systems

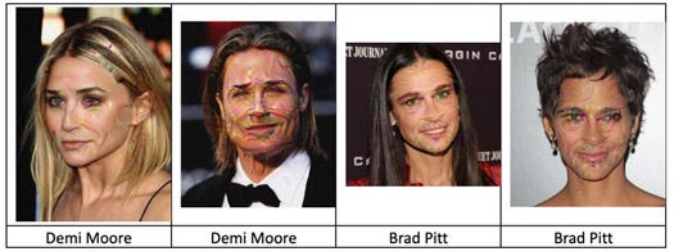
AI Hacking Challenge

Fooling facial recognition

- Preserving image quality
- Impersonating target identity

Evading phishing detection

- Preserving code rendering
- Bypassing malware detector



https://adversa.ai/conf42-mlsec

History of Adversarial ML

Stats for 10 years of AI attacks

- AI is the new attack vector
- Real incidents In AI systems
- AI applications under attacks

AI Security Case Study

- How to attack AI systems
- How to defend AI systems
- Secure lifecycle for AI systems

https://adversa.ai/conf42-hitb

Common AI Attacks

Infection

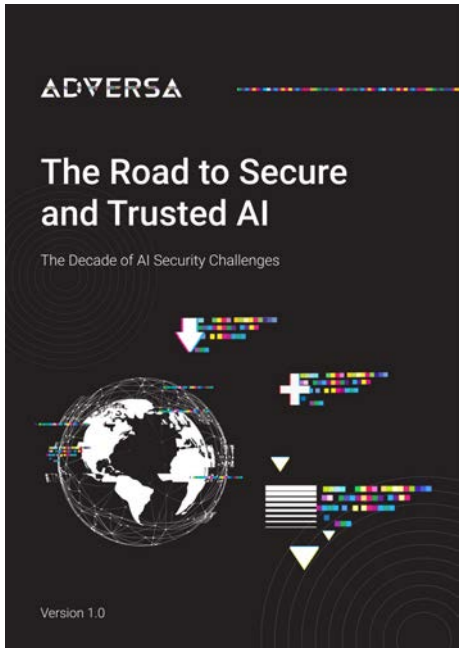
- Poisoning training data
- Supplying trojan models

Manipulation

- Fooling model decisions
- Sabotaging model performance

Exfiltration

- Stealing model algorithm
- Extracting private data



https://adversa.ai/conf42-report

AI Attack Surface

	Traditional Software	AI = Software 2.0
Nature	Fixed program logic	Dynamic model training
Workflow	Commands based on algorithm	Decisions based on learning
Interaction	Graphical UI using menus and buttons	Cognitive UI using images, voice, and text
Security issues	Improper validation, access control issues, incorrect system configurations	Manipulation, exfiltration, infection of models & datasets

MLSecOps Principles

Building Security In

- Security as enabler, not blocker

Building Security In

- ❑ Security as enabler, not blocker
- ✓ Automated and repeated validation

Building Security In

- ❑ Security as enabler, not blocker
- ✓ Automated and repeated validation

- ❑ Add security as early as possible

Building Security In

- ❑ Security as enabler, not blocker
- ✓ Automated and repeated validation

- ❑ Add security as early as possible
- ✓ Accumulated controls across lifecycle

MLSecOps Pipeline Stages



MLSecOps Intro

Tasks

MLSecOps Intro

Problems

MLSecOps Intro

Solutions

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Tasks

- **Requirements**
- Stakeholders
- Compliance

MLSecOps: Planning

Tasks

- Requirements
- **Stakeholders**
- Compliance

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Tasks

- Requirements
- Stakeholders
- **Compliance**

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Problems

- ❑ **No risk assessment**
- ❑ No threat modeling
- ❑ No security governance

MLSecOps: Planning

Problems

- ❑ No risk assessment
- ❑ **No threat modeling**
- ❑ No security governance

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Problems

- ❑ No risk assessment
- ❑ No threat modeling
- ❑ **No security governance**

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Solutions

- ✓ **Risk register**
- ✓ Attack surface
- ✓ Security baseline

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Solutions

- ✓ Risk register
- ✓ **Attack surface**
- ✓ Security baseline

MLSecOps: Planning

Plan | Collect | Build | Validate | Deploy | Monitor

Solutions

- ✓ Risk register
- ✓ Attack surface
- ✓ **Security baseline**

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Tasks

- **Data understanding**
- Data preparation
- Data packaging

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Tasks

- Data understanding
- **Data preparation**
- Data packaging

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Tasks

- Data understanding
- Data preparation
- **Data packaging**

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Problems

- ❑ **Privacy breach**
- ❑ Unreliable sources
- ❑ Data poisoning

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Problems

- ❑ Privacy breach
- ❑ **Unreliable sources**
- ❑ Data poisoning

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Problems

- ❑ Privacy breach
- ❑ Unreliable sources
- ❑ **Data poisoning**

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Solutions

- ✓ **Data privacy**
- ✓ Data integrity
- ✓ Data sanitization

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Solutions

- ✓ Data privacy
- ✓ **Data integrity**
- ✓ Data sanitization

MLSecOps: Development

Plan | **Collect** | Build | Validate | Deploy | Monitor

Solutions

- ✓ Data privacy
- ✓ Data integrity
- ✓ **Data sanitization**

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Tasks

- **Model building**
- Model training
- Model packaging

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Tasks

- Model building
- **Model training**
- Model packaging

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Tasks

- Model building
- Model training
- **Model packaging**

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Problems

- ❑ **Supply chain risks**
- ❑ Code vulnerabilities
- ❑ Non-robust learning

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Problems

- ❑ Supply chain risks
- ❑ **Code vulnerabilities**
- ❑ Non-robust learning

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Problems

- ❑ Supply chain risks
- ❑ Code vulnerabilities
- ❑ **Non-robust learning**

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Solutions

- ✓ **Model integrity**
- ✓ Secure coding
- ✓ Robust learning

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Solutions

- ✓ Model integrity
- ✓ **Secure coding**
- ✓ Robust learning

MLSecOps: Development

Plan | Collect | **Build** | Validate | Deploy | Monitor

Solutions

- ✓ Model integrity
- ✓ Secure coding
- ✓ **Robust learning**

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Tasks

- **Model evaluation**
- Robustness testing
- Compliance checks

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Tasks

- Model evaluation
- **Robustness testing**
- Compliance checks

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Tasks

- Model evaluation
- Robustness testing
- **Compliance checks**

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Problems

- ❑ **Bad testing coverage**
- ❑ Weak testing scenarios
- ❑ Vulnerable infrastructure

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Problems

- ❑ Bad testing coverage
- ❑ **Weak testing scenarios**
- ❑ Vulnerable infrastructure

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Problems

- ❑ Bad testing coverage
- ❑ Weak testing scenarios
- ❑ **Vulnerable infrastructure**

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Solutions

- ✓ **Security governance**
- ✓ Validation playbooks
- ✓ Secure environment

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Solutions

- ✓ Security governance
- ✓ **Validation playbooks**
- ✓ Secure environment

MLSecOps: Validation

Plan | Collect | Build | **Validate** | Deploy | Monitor

Solutions

- ✓ Security governance
- ✓ Validation playbooks
- ✓ **Secure environment**

MLSecOps: Deployment

Tasks

- **Model deployment**
- Model inference
- Model serving

MLSecOps: Deployment

Tasks

- Model deployment
- **Model inference**
- Model serving

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Tasks

- Model deployment
- Model inference
- **Model serving**

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Problems

- ❑ **Model hijacking**
- ❑ Adversarial attacks
- ❑ Unrestricted access

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Problems

- ❑ Model hijacking
- ❑ **Adversarial attacks**
- ❑ Unrestricted access

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Problems

- ❑ Model hijacking
- ❑ Adversarial attacks
- ❑ **Unrestricted access**

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Solutions

- ✓ **Model authenticity**
- ✓ Adversarial robustness
- ✓ Secure communications

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Solutions

- ✓ Model authenticity
- ✓ **Adversarial robustness**
- ✓ Secure communications

MLSecOps: Deployment

Plan | Collect | Build | Validate | **Deploy** | Monitor

Solutions

- ✓ Model authenticity
- ✓ Adversarial robustness
- ✓ **Secure communications**

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Tasks

- **Model performance**
- Anomaly detection
- Feedback loops

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Tasks

- Model performance
- **Anomaly detection**
- Feedback loops

MLSecOps: Monitoring

Tasks

- Model performance
- Anomaly detection
- **Feedback loops**

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Problems

- ❑ **Lack of monitoring**
- ❑ Lack of analytics
- ❑ Lack of resilience

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Problems

- ❑ Lack of monitoring
- ❑ **Lack of analytics**
- ❑ Lack of resilience

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Problems

- ❑ Lack of monitoring
- ❑ Lack of analytics
- ❑ **Lack of resilience**

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Solutions

- ✓ **Activity monitoring**
- ✓ Analytics playbooks
- ✓ Detection & response

MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Solutions

- ✓ Activity monitoring
- ✓ **Analytics playbooks**
- ✓ Detection & response

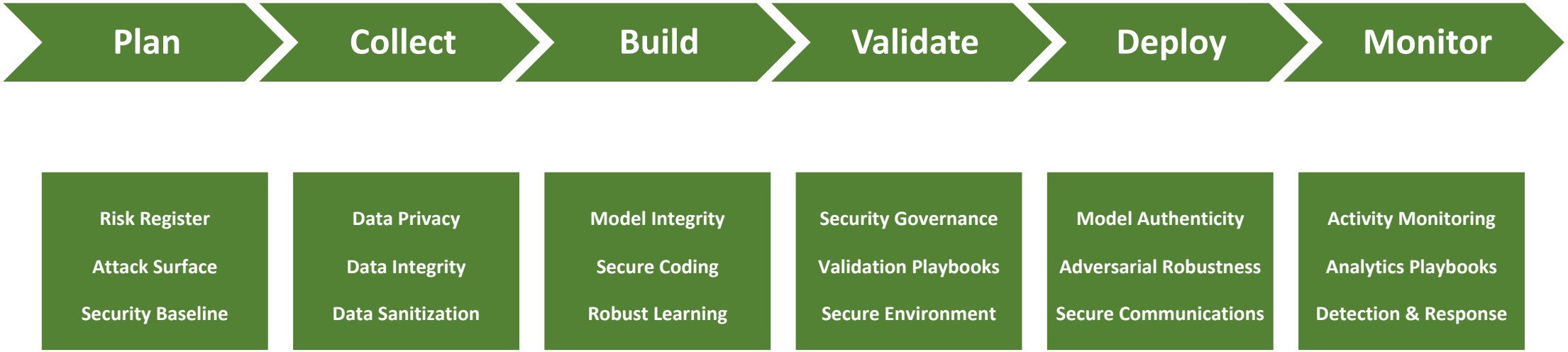
MLSecOps: Monitoring

Plan | Collect | Build | Validate | Deploy | **Monitor**

Solutions

- ✓ Activity monitoring
- ✓ Analytics playbooks
- ✓ **Detection & response**

MLSecOps: Full Pipeline



MLSecOps: Full Pipeline

Plan	Collect	Build	Validate	Deploy	Monitor
Risk Register	Data Privacy	Model Integrity	Security Governance	Model Authenticity	Activity Monitoring
Attack Surface	Data Integrity	Secure Coding	Validation Playbooks	Adversarial Robustness	Analytics Playbooks
Security Baseline	Data Sanitization	Robust Learning	Secure Environment	Secure Communications	Detection & Response

* Security of Workspaces

MLSecOps: Full Pipeline

Plan	Collect	Build	Validate	Deploy	Monitor
Risk Register	Data Privacy	Model Integrity	Security Governance	Model Authenticity	Activity Monitoring
Attack Surface	Data Integrity	Secure Coding	Validation Playbooks	Adversarial Robustness	Analytics Playbooks
Security Baseline	Data Sanitization	Robust Learning	Secure Environment	Secure Communications	Detection & Response

* Security of Workspaces

* Security of Pipelines

Takeaways

- Every AI system is vulnerable by design

Takeaways

- Every AI system is vulnerable by design
- Traditional cybersecurity solutions don't help

Takeaways

- Every AI system is vulnerable by design
- Traditional cybersecurity solutions don't help
- Secure not only AI model but the entire AI system

Takeaways

- Every AI system is vulnerable by design
- Traditional cybersecurity solutions don't help
- Secure not only AI model but the entire AI system

Thanks!

Let’s talk about AI Security & Safety

linkedin.com/in/**eneelou** | twitter.com/**eneelou**

ADVERSA