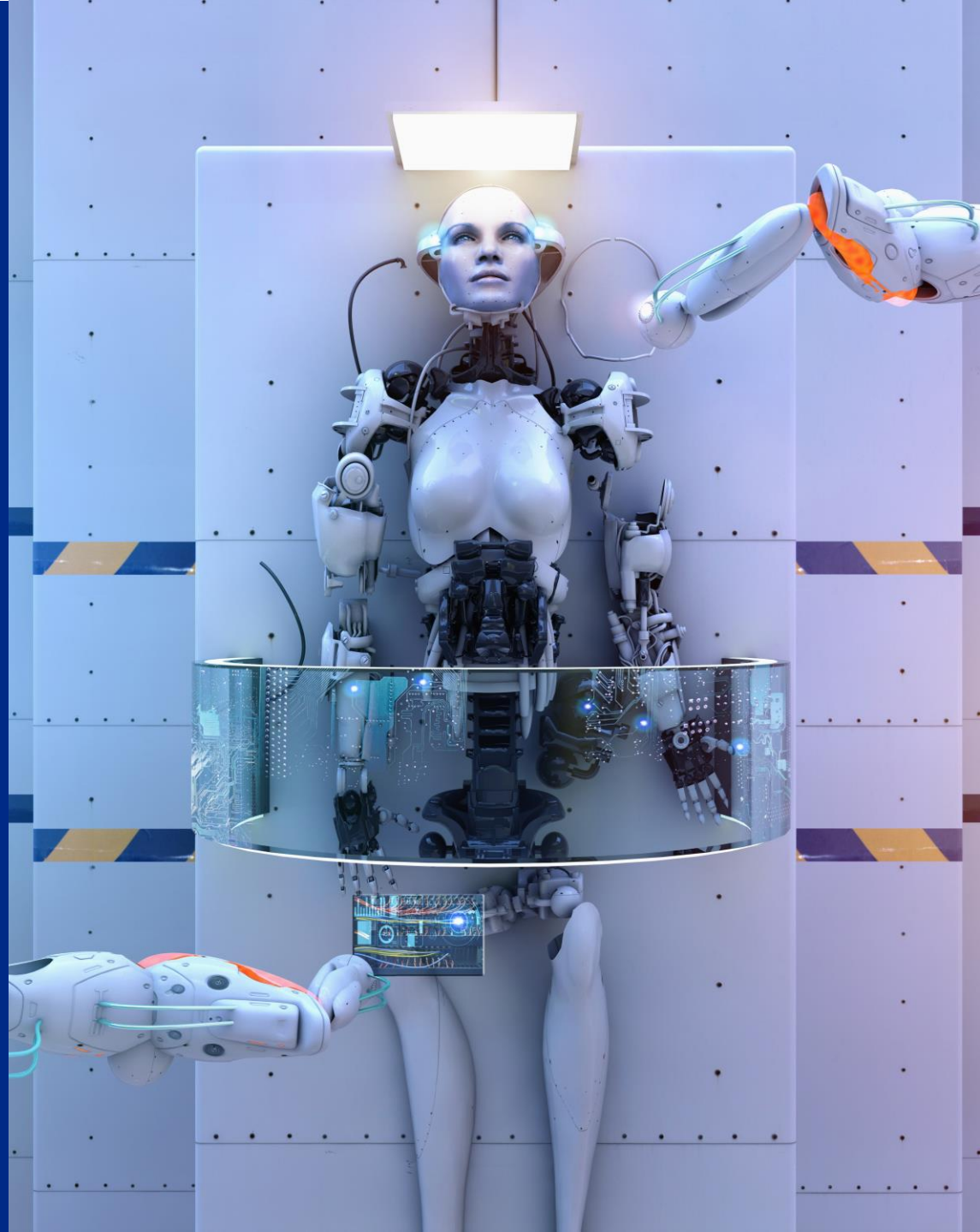
AI Risk Assessment through Threat Modeling and use cases for AI-BOM automation

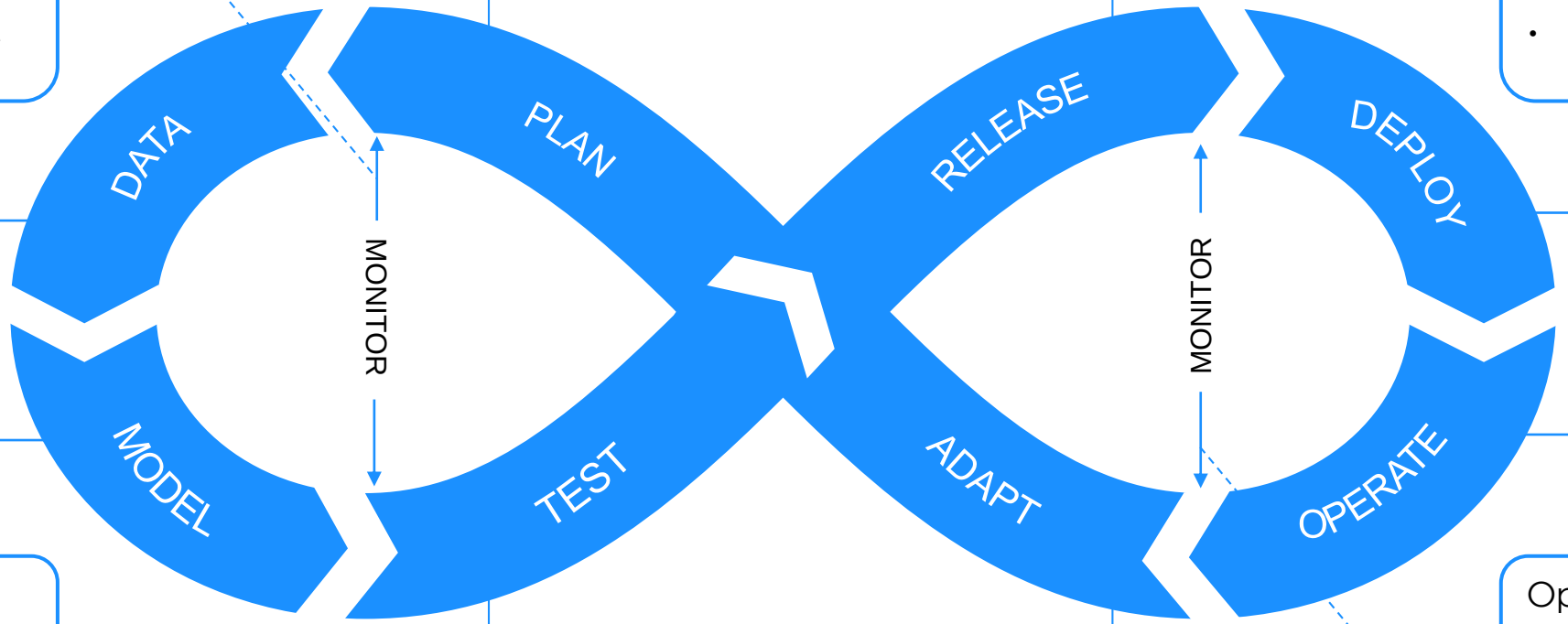
Helen Oakley, CISSP, GPCS, GSTRT
Director of Secure Software Supply Chains & Secure Development, SAP
/in/helen-oakley

May 7th, 2024

PUBLIC



ML-SEC-OPS



Monitor:
- Logging and Monitoring
- Data/Model Drift or Decay

Plan:
- Open Source Strategy
- Threat Modelling
- Risk Identification /Management

Release:
- Model Acceptance
- Model Signing
- Code Signing
- AI/ML BOM Generation

Data Engineering:
- Data Sourcing and Collection
- Regulatory Compliance
- Secure Data Pipelines

Deploy
- KPI/Benchmarking verification
- AI Explainability

Model:
- Model use case verification
- Bias Testing

Testing & pre-Release:
- Model Scans
- Pen Testing
- Verification of security and privacy controls

Adapt
- Model Performance Adaption
- Use Case Verification

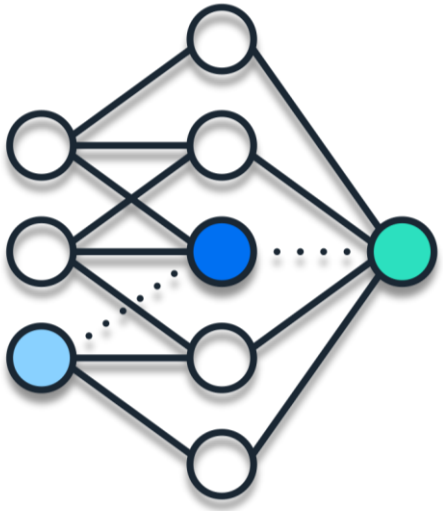
Operate
- Regulating Model Performance
- Auditing Edge Scenarios

Monitor:
- Logging and Monitoring
- Data/Model Drift or Decay

AI Software transparency

MLSECOPS

- DATA OPS
- MODEL OPS
- DEV OPS



AI/ML Bill of Materials

- AI-BOMs: describes the purpose
- AI-BOMs are part of SBOMs

AI Risk Assessment

- AI-BOMs lifecycle & management
- AI-BOMs for Threat Modeling

Threat Modeling: What Can Go Wrong?

Property: informationAboutTraining

- training data used to train the AI model, along with any relevant details about its source, quality, and pre-processing steps;
- specific training algorithms employed, including stochastic gradient descent, backpropagation, and reinforcement learning.
- specific training techniques used to improve the performance or accuracy of the AI model, such as transfer learning, fine-tuning, or active learning; and
- any evaluation metrics used to assess the performance of the AI model during the training process, including accuracy, precision, recall, and F1 score.

https://github.com/spdx/spdx-3-model/blob/main/model/AI/Properties/informationAboutTraining.md

STRIDE

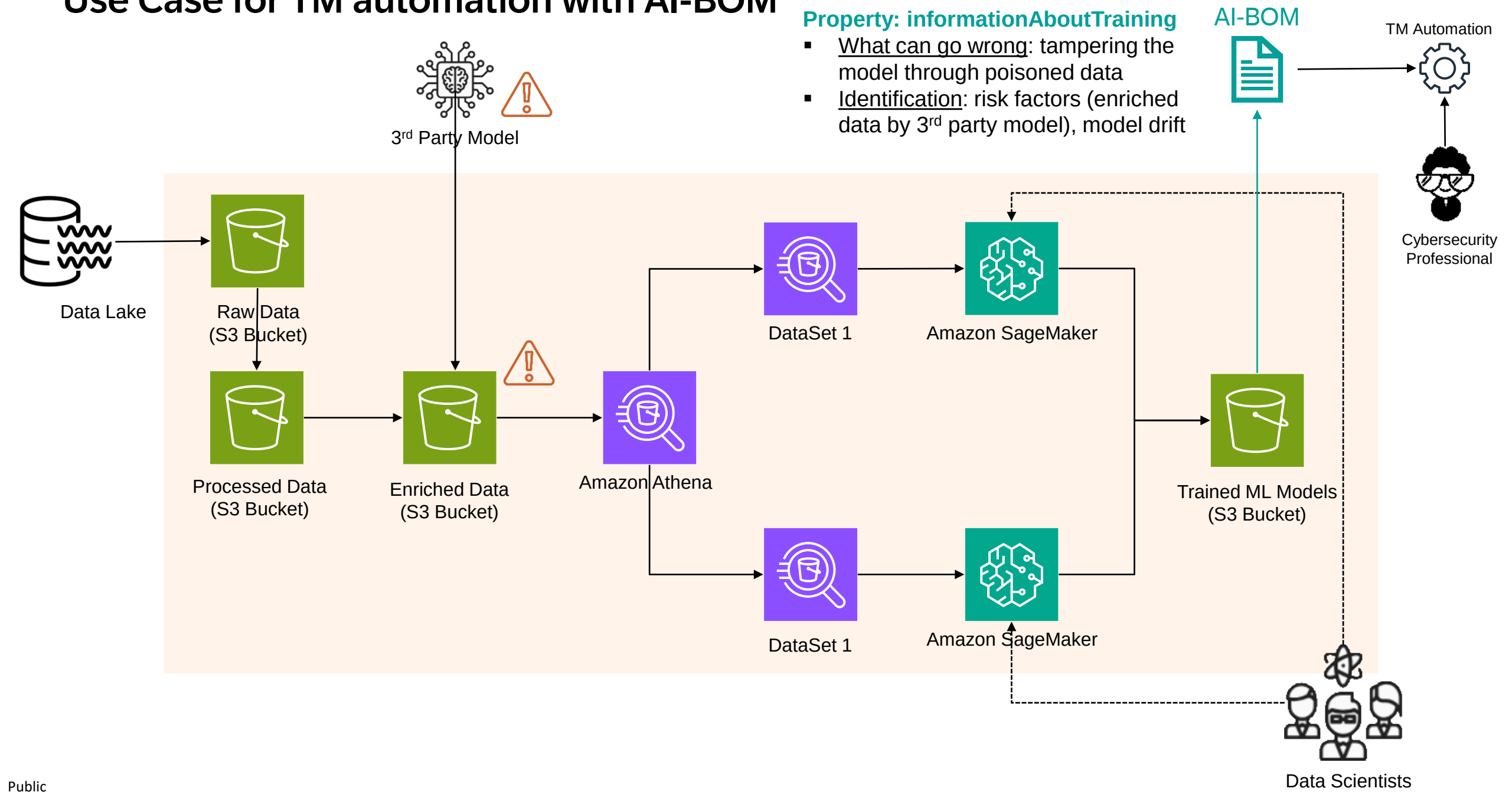
- Spoofing
- Tampering
- Repudiation
- Information Disclosure
- Denial of Service (DoS)
- Elevation of Privilege

https://owasp.org/www-community/Threat_Modeling_Process#stride

Use Case for TM automation with AI-BOM

Property: informationAboutTraining

- What can go wrong: tampering the model through poisoned data
- Identification: risk factors (enriched data by 3rd party model), model drift



Summary: automating processes around AI-BOM

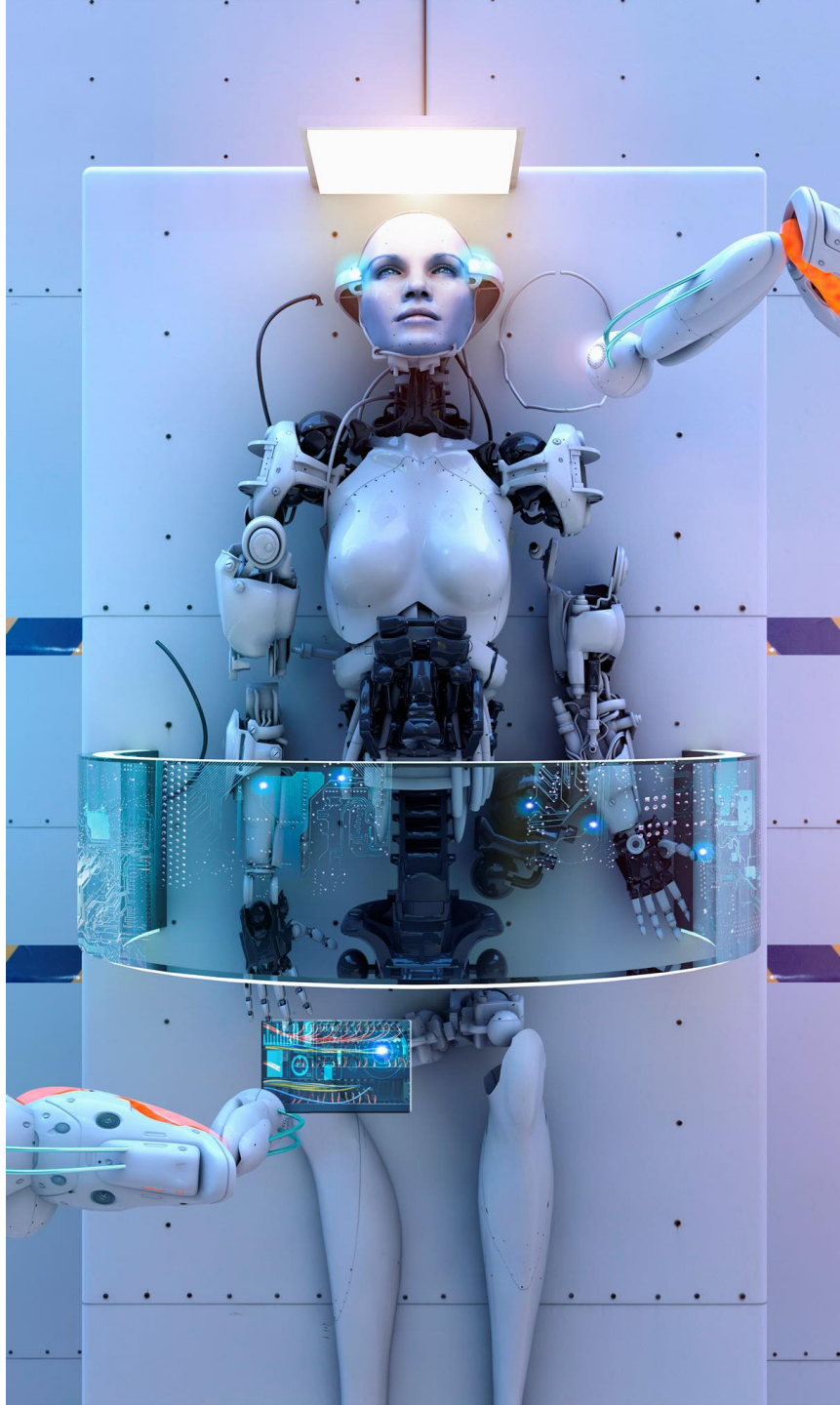
Automating Threat Modeling requires defining your risk or acceptable boundaries and ability collect the data from your sources. AIBOM can help facilitating the collection of data about the model in a machine readable and consistent format.

Challenges

- Rapidly evolving technology
- Lack of framework standardization for MLSecOps
- Runtime SBOM (incl. AI-BOM) concept for dynamic data collection about AI/ML

What do we do?

- This AI-BOM workshop is the first step bringing the AI community together to define best practices across the industry



Thank you!

Helen Oakley, CISSP, GPCS, GSTRT
Director of Secure Software Supply Chains & Secure Development, SAP
/in/helen-oakley