Objection! AI Security Mistakes on Trial with Kubeflow and Confidential Computing

Annie Talvasto & Karl Ots

AI & Cybersecurity

TIMELINE IN HISTO

Alan Turing Proposes the Turing Test

1956

annaCry Ransomware Attack

May

2017

AI-Powered Cybersecurity and Attacks

Ongoing

2025

1997

May

IBM's Deep Blue Defeats Chess Champion Garry Kasparov

2022

November

ChatGPT and the Rise of Generative AI

@AnnieTalvasto @KarlOts.com

Annie Talvasto

- CEO at Waovo
- CNCF Ambassador
- Azure and AI Platform MVP

- CloudGossip podcast & TechCraft Show
- Previously: CloudNativeLive & Program Chair for Secure AI Summit

Karl Ots

- Head of Cloud Security at EPAM
- Microsoft Security MVP

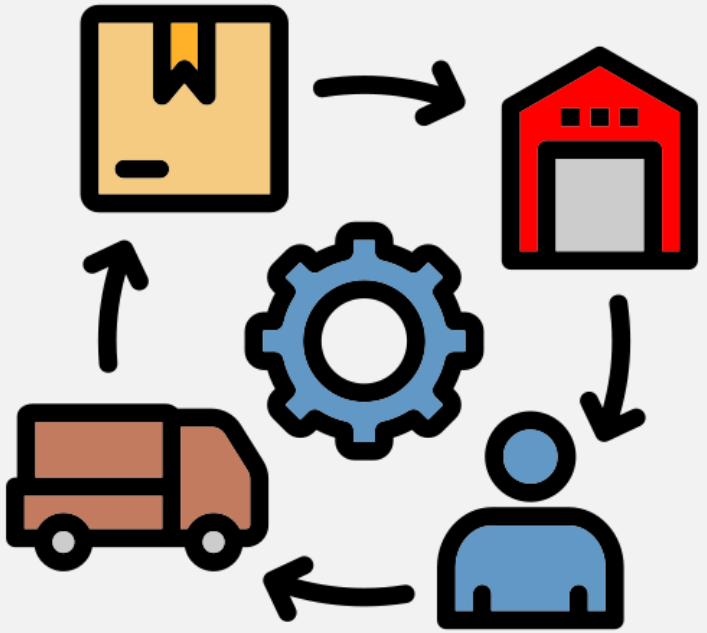
- Author of Securing Microsoft Azure OpenAI (Wiley)
- Cloud Gossip podcast


SUMMARY OF RESULTS
$96 393
TOTAL COMPANY
USA


SUMMARY OF RESULTS
$96 393
TOTAL COMPANY
USA

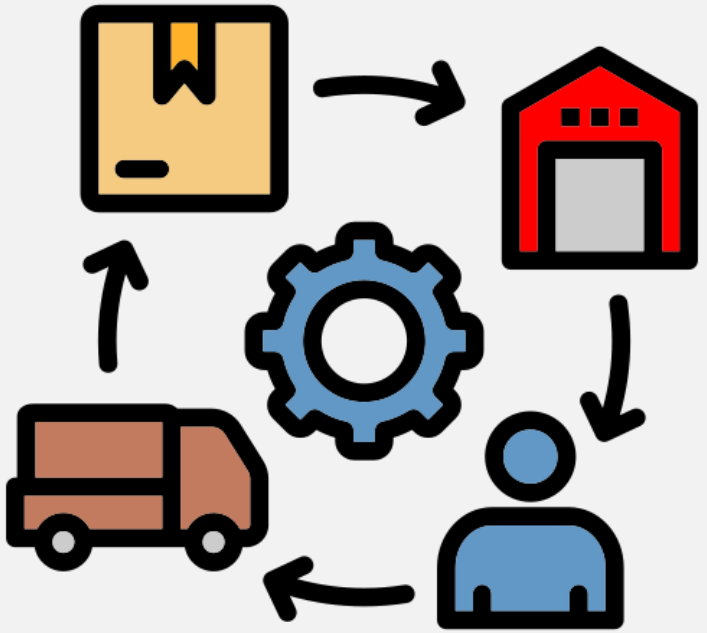

SUMMARY OF RESULTS
$96 393
Kubeflow
CLOUD NATIVE + KUBERNETES
AI DAY
EUROPE
CONFIDENTIAL CONTAINERS
OpenTelemetry
AI

AI Security Mistake #1:
Supply chain attacks



AI Security Mistake #1:
Supply chain attacks

- Prevention: Trust but verify, scan any 3rd party modules, implement DevSecOps

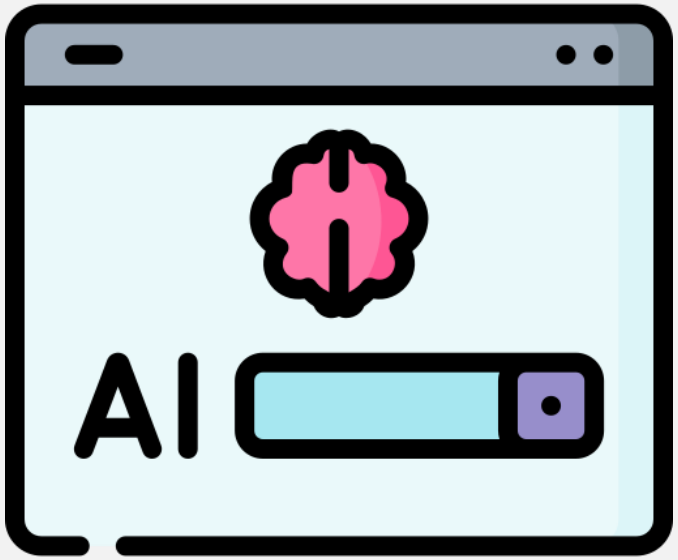


AI Security Mistake #2:
Hallucinations

AI Security Mistake #2: Hallucinations

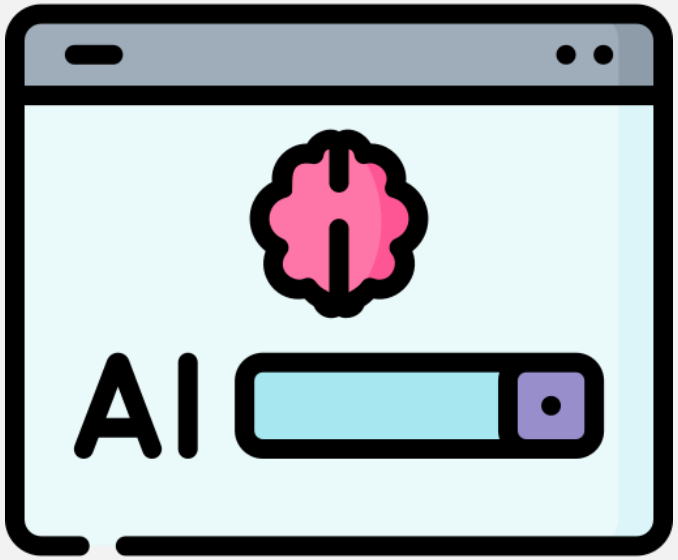
- Prevention: RAG, Orchestrate verification with KubeFlow, Policies, Telemetry

AI Security Mistake #3:
AI platform jacking

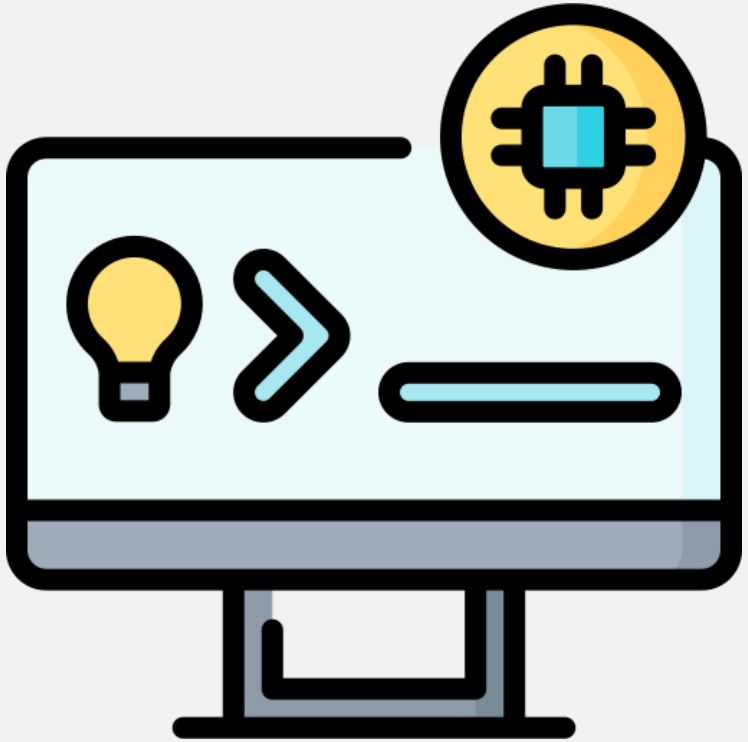


AI Security Mistake #3:
AI platform jacking

- Prevention: No secrets in code, use key vaults, adhere to cloud security best practices

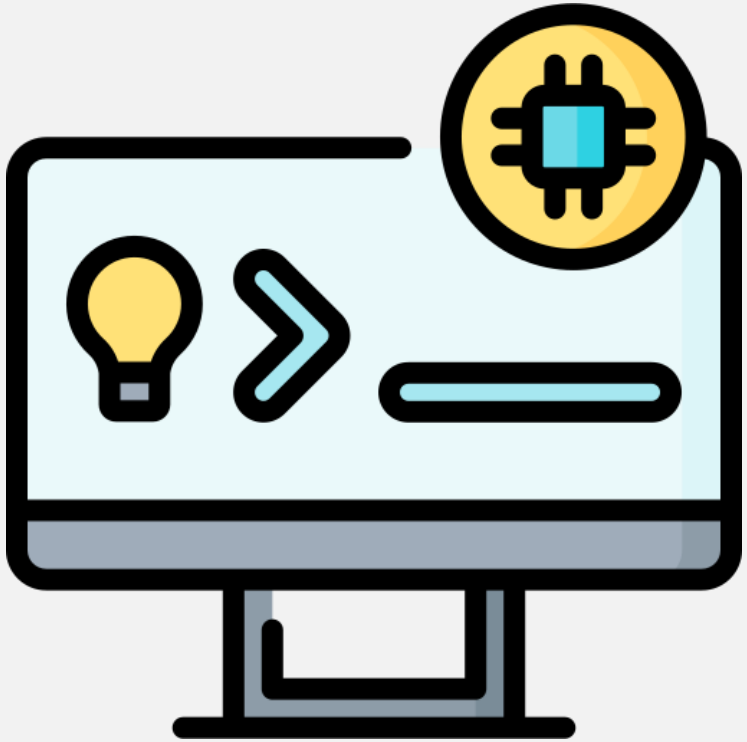


AI Security Mistake #4:
Prompt injection



AI Security Mistake #4:
Prompt injection

- Prevention: Security guardrails for LLMs, Nemo & Llama Guard

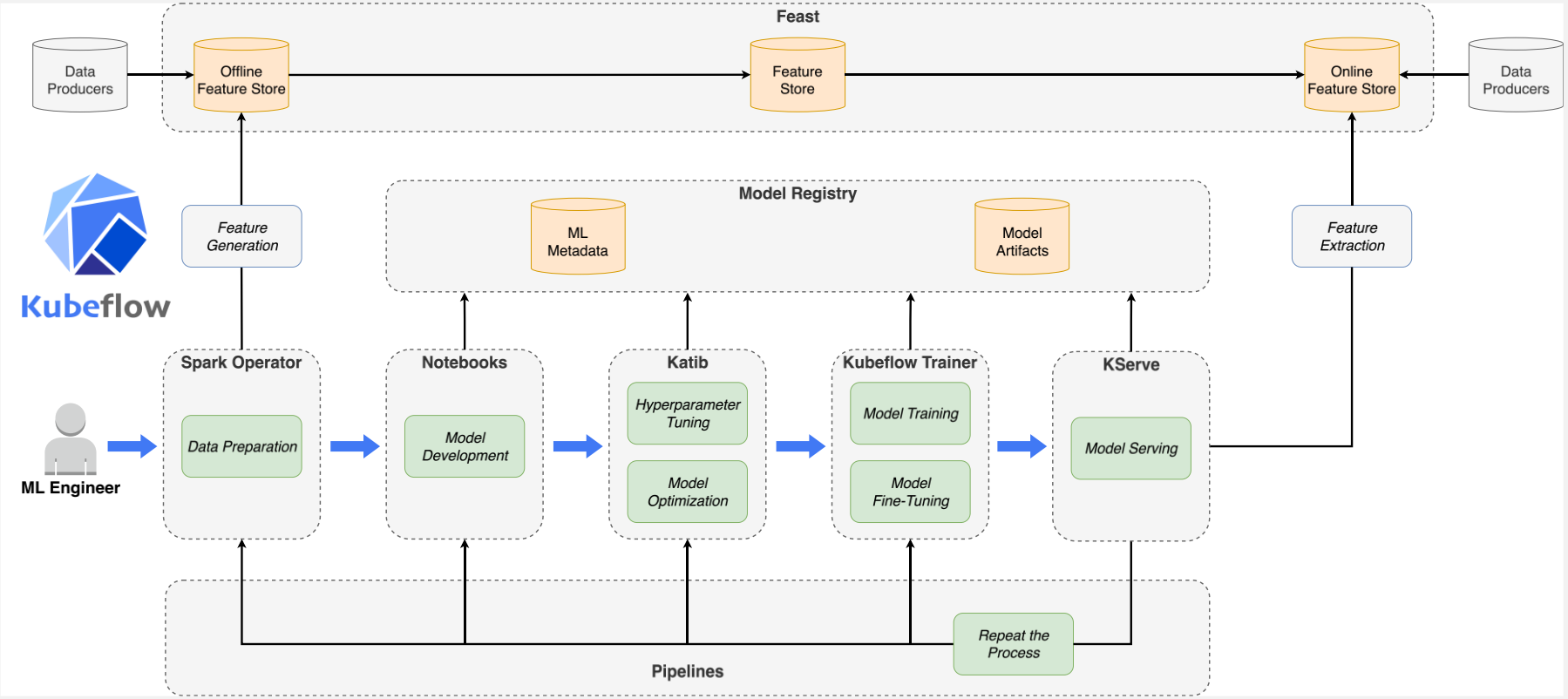


Which one is the worst in your opinion?

Supply chain attacks
Hallucinations
AI platform jacking
Prompt injection

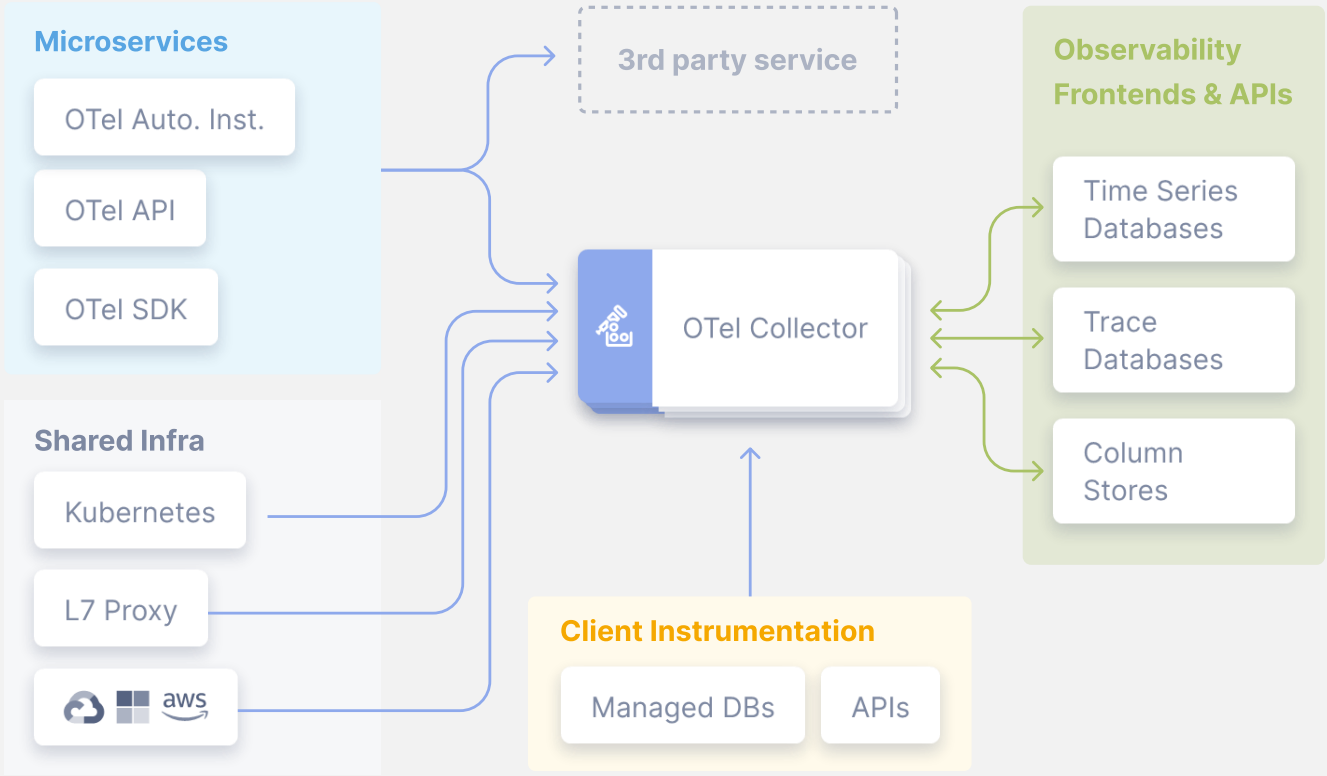
Kubeflow

The Machine Learning Toolkit for Kubernetes

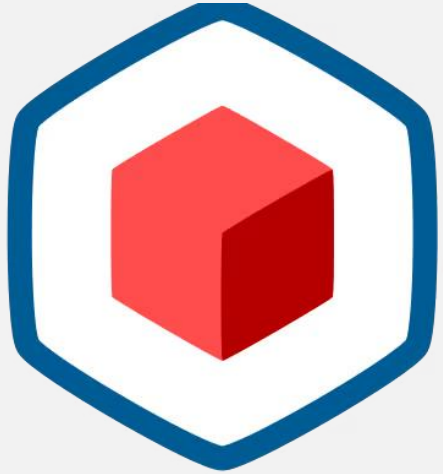


OpenTelemetry

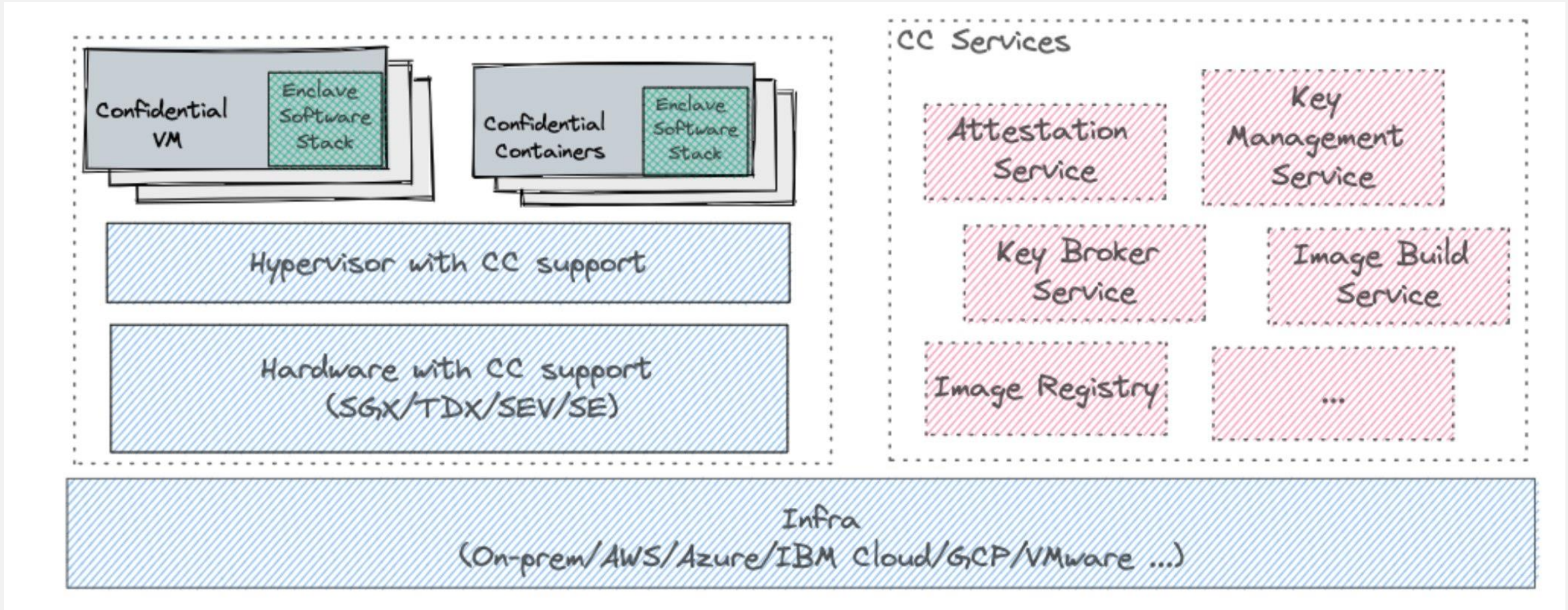
High-quality, ubiquitous, and portable telemetry to enable effective observability



Confidential Containers:
Deploy Cloud Native Applications inside Confidential Enclaves



CONFIDENTIAL CONTAINERS


Confidential VM
Enclave Software Stack
Confidential Containers
Enclave Software Stack
Hypervisor with CC support
Hardware with CC support (SGX/TDX/SEV/SE)
CC Services
Attestation Service
Key Management Service
Key Broker Service
Image Build Service
Image Registry
...
Infra (On-prem/AWS/Azure/IBM Cloud/GCP/VMware ...)

Takeaways:

- Four AI Security Mistakes
  - Supply chain attacks
  - Hallucinations
  - AI platform jacking
  - Prompt injection

- Kubeflow
- Confidential Containers
- OpenTelemetry

Learn more

- Links and slides: github.com/annietalvasto

- Cloud Native AI Working Group & Security TAG:
  https://tag-runtime.cncf.io/wgs/cnaiwg/
  https://tag-security.cncf.io

- OWASP Top 10 For LLM apps v2
  https://owasp.org/www-project-top-10-for-large-language-model-applications/

- Securing Microsoft Azure OpenAI by Karl Ots (Wiley). B0F1969HY8
  https://amzn.eu/d/dy4E6UJ

Thank you!

@AnnieTalvasto @KarlOts.com