#2530655 CICI: IPAAI: Multi-Layer Data Provenance and Federated Learning for Securing Scientific AI Pipelines

PI: Wajih Ul Hassan (University of Virginia) Co-PI: Aidong Zhang (University of Virginia)

Goal and Motivation:

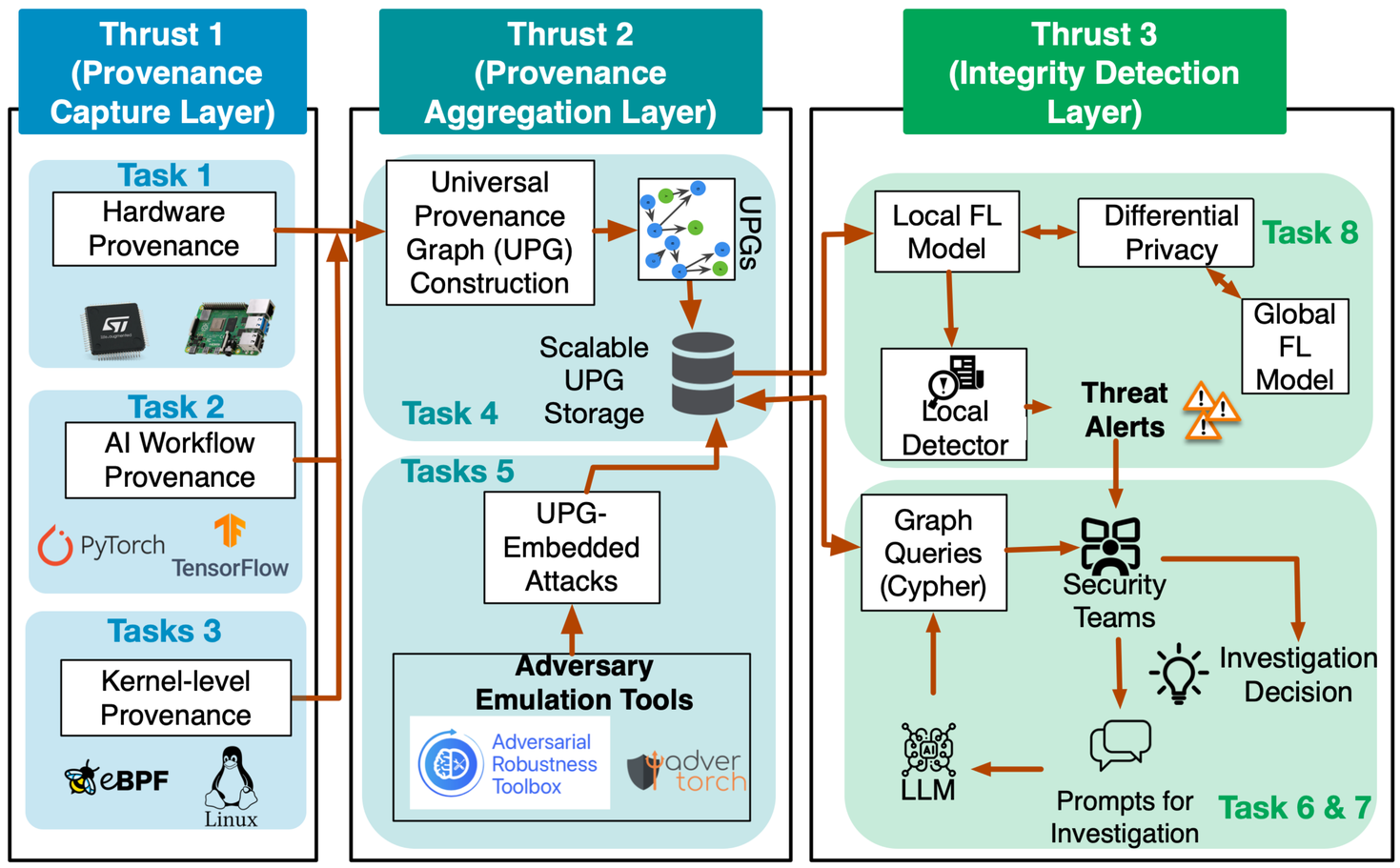
- Ensure the integrity, provenance, and authenticity (IPA) of AI datasets used in scientific workflows
- Tackle threats arising from multi-source, distributed, and unverified scientific data
- Enable traceability across the entire data lifecycle from sensors to AI models
- Promote accountability and public trust in scientific computing

Core Topics:

- Provenance capture across hardware, kernel, and application layers
- Privacy-preserving federated anomaly detection over distributed datasets
- Simulation of data-centric attacks for robustness evaluation
- Open-source forensic tools to support investigation and reproducibility
- Usable interfaces that allow scientists to trace data misuse through natural language

Technical Approach:

- Capture tamper-evident provenance at scientific instruments using hardware-backed cryptographic signing
- Track AI pipelines and OS events to record dataset transformations with minimal overhead
- Fuse hardware, system, and AI events into a Unified Provenance Graph for fast forensic queries
- Simulate realistic dataset attacks to stress-test the system and generate labeled training data
- Detect cross-institution tampering via federated graph learning, with a natural language investigation assistant



Research Areas:

- Secure telemetry and provenance capture
- Scalable graph construction, compression, and querying
- Federated learning for anomaly and manipulation detection
- Simulation of data and provenance attacks for benchmarking
- Auditable AI via provenance traceability and human-in-the-loop investigation

Evaluating and Demonstrating IPA:

- Demonstrate dataset tampering detection in genomics and climate workflows
- Validate provenance graph performance at scientific data scales
- Show domain scientists can investigate integrity issues without specialist expertise

Potential Advances:

- First end-to-end provenance framework from hardware sensors through AI training
- Privacy-preserving cross-institution collaboration
- Open-source testbed and natural language interface that lower the barrier for scientists