Secure Systems Group, Aalto University†, University of Waterloo+

Vasisht Duddu+, Oskari Järvinen†, Lachlan J. Gunn†, N. Asokan+†

ML Property Attestation using TEEs

- Clients of ML-based services cannot verify that responses come from the right model
- Algorithms, datasets, and training parameters cannot be verified after training
- **ML property attestation** can prove such properties efficiently and scalably

1 Introduction

- Measured model and dataset metrics used to demonstrate the quality of models & inferences
- Need to link dataset, training parameters to model, model to inference input/output
- New advances (e.g., Intel AMX) allow training/running complex models within TEEs

2 The problem

- Cryptographic proofs inefficient or don't scale
- ML-based methods are inaccurate
- Current methods focus only on specific properties
- Current certification services require outsourcing both training and inference

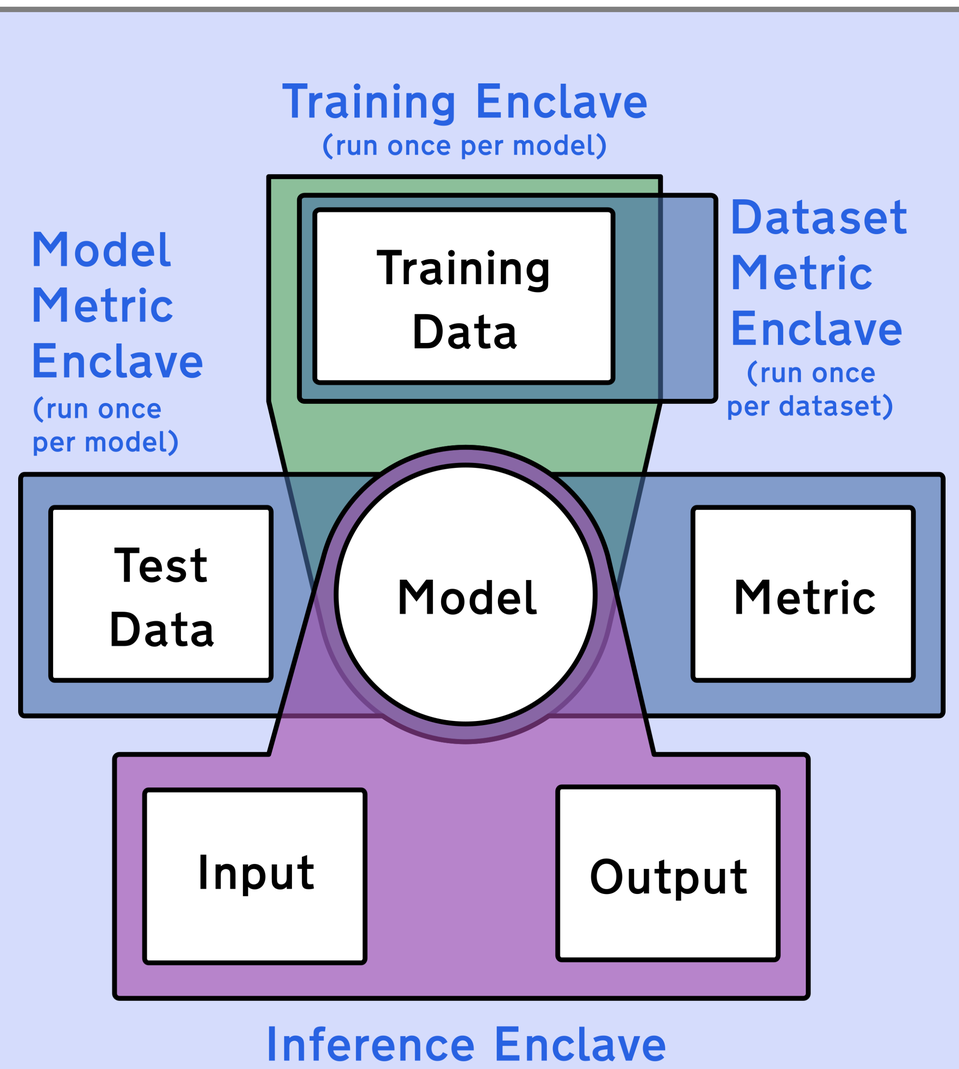
3 Our solution

Use remote attestation to prove properties like:

- Which model produced an inference
- How accurate is the model
- How was the model trained
- What data was used to train it
- How representative was the training set

4 Implementation

- SGX enclaves perform ML tasks and attest process/performance claims
- Verifier combines attestations to link output to input, model, training dataset



Attested ML architecture. Enclaves hosting models measure/attest metrics for training data, model, and inference operations for confidence in model & inferences.

	I/O Binding (100 operations)	Accuracy	Proof of Training
Startup	32029ms	36470ms	36.5s
Preprocessing	0.5ms	294ms	4.4s
Computation	70.1ms	3490ms	514s
Proving	**6.6ms**	**5.68ms**	**0.005s**

Run-time for different types of attestation (average of 10 runs).

5 Conclusion

TEE-based ML property attestation is efficient, scalable & versatile

Aalto University

Project Link:
https://ssg-research.github.io/mlsec/mlattestation

UNIVERSITY OF WATERLOO