Got Trust Issues? The Power of TEEs

Pierre Le Guen

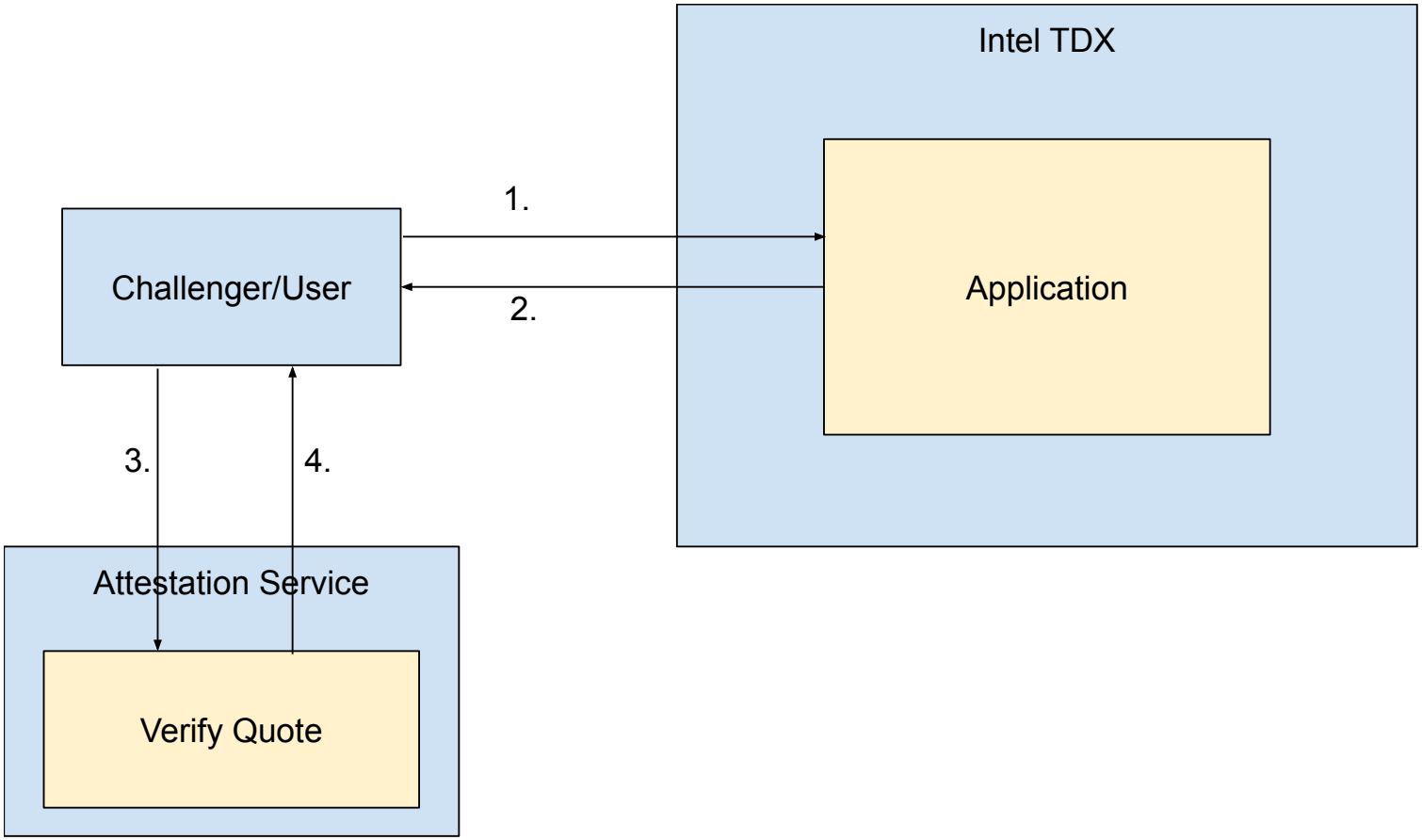
Head of Engineering @ NEAR AI

July 3, 2025

What are Trusted Execution Environments?

- Hardware-based isolation
- Secure computation without data exposure
- Key examples:
  - Intel TDX, AMD SEV, ARM CCA
  - NVIDIA H100 GPUs
- Ensures CIA – Confidentiality, Integrity, Attestation

What are Trusted Execution Environments?



Attestation flow[1]

[1] https://www.intel.com/content/www/us/en/developer/tools/software-guard-extensions/attestation-services.html

TEE Use Cases

- **Healthcare:** Privacy-preserving diagnostics using sensitive medical data.
- **Finance:** Risk models trained on confidential datasets.
- **Decentralized Agents:** AI agents capable of executing financial agreements.
- **Artificial Intelligence:** Protecting inference, securing AI model weights

AI and Confidential Computing: The Core Problem

- Frontier AI models are extremely valuable and create new security vectors.
- **Key Risks During Inference:**
  - **Model Theft:** Protecting proprietary model weights, the core IP.
  - **User Data Leakage:** Preventing exposure of sensitive user data (e.g., financial, medical).
  - **Execution Tampering:** Ensuring the correct model is running without modification.
- **The Solution:** Confidential inference using TEEs to mitigate these risks.

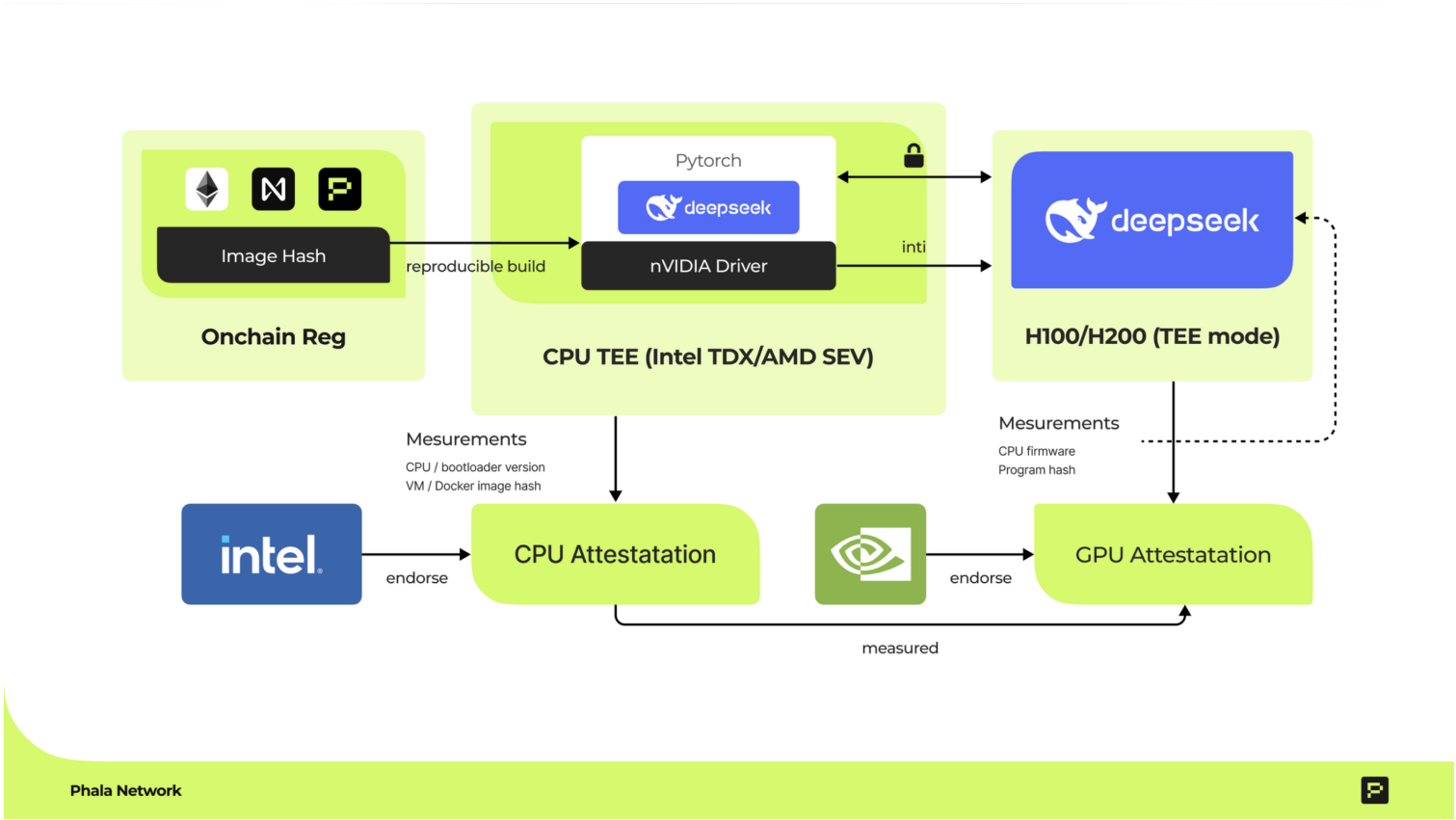
NEAR AI's Decentralized Confidential ML (DCML)

- **Private Inference**: User data confidentiality
- **Verifiable Open-Source Models**: Transparent computations
- **Decentralized Monetizable Models**: Encrypted distribution, blockchain monetization
- **Community-Owned Models**: Collaborative and trustless
- DCML Paper

NEAR AI Private ML SDK

- **Open-source SDK**: github.com/nearai/private-ml-sdk
- Enables confidential and verifiable inference
- Pre-dates Anthropic's confidential inference paper
- Easy integration with existing ML workflows

NEAR AI Private ML SDK



GPU Attestation flow[2]

[2] https://docs.phala.network/overview/phala-network/gpu-tee

RAND's Threat Model

- Importance of securing AI weights (core capabilities)
- 38 identified attack vectors
- Operational Capacity Categories (OC1-OC5)
- Five incremental Security Levels (SL1-SL5)
- RAND Report

Anthropic's Confidential Inference Principles

- User and model confidentiality during inference
- Attestation and cryptographic assurance
- Threat Landscape:
  - Systemic risks (hardware vulnerabilities)
  - Introduced risks (misconfiguration)
- Recommendations: Proactive risk management
- Anthropic Paper

Vision for Confidential AI

- NEAR AI's decentralized and monetizable infrastructure
- RAND's structured incremental security model
- Anthropic's proactive risk approach
- Need: Secure, decentralized, monetizable, confidential AI

Implementation Insights

- TEE overhead minimal (1.5%–12%)
- Blockchain-based monetization methods
- Challenges: Hardware trust, decentralized training scalability

Model	Total Token Throughput (tokens/s)			Requests Throughput (req/s)		
	CVM	Bare metal	Overhead	CVM	Bare metal	Overhead
Mistral-24B	2382.29	2476.27	3.79%	3.57	3.71	3.77%
Qwen-32B	1832.29	1861.78	1.58%	2.61	2.65	1.5%
DeepSeek-R1-70B	1250.06	1421.99	12.09%	1.09	1.24	12.09%

Performance comparison of TEE-on and TEE-off modes for various models in terms of TPS and QPS.[3]

[3] DCML Paper: https://raw.githubusercontent.com/nearai/por/refs/heads/main/DecentralizedConfidentialM

Closing & Call to Action

- TEEs essential, not optional
- Incremental, structured security (RAND benchmarks)
- Cross-sector collaboration necessary
- Join NEAR AI to shape user-owned AI

Questions & Answers

Thank you! Questions?

linktr.ee/pierre_lg