

The Hardware Truth Landscape — Pirates, Traders, and Anchors

A Comparative Analysis of the 2026 TEE Ecosystem (Oasis, Phala, Flashbots, Super Protocol, and Citadel)

Author: Theo Ezell **Date:** February 2026 **Publication Target:** Zenodo.org (Preprint)

Abstract

The transition of artificial intelligence from a generative curiosity to an agentic workforce has precipitated a fundamental crisis in digital governance. This paper argues that the prevailing "Software Governance" paradigm—which relies on system prompts and application-layer logic—is insufficient for securing high-stakes autonomous agents.

To this end, we analyze the "Hardware Renaissance" of 2026, a shift characterized by the rapid migration of agentic infrastructure toward Trusted Execution Environments (TEEs). While the underlying silicon—ranging from Intel SGX/TDX and NVIDIA H100 to IBM Z—is often shared across the industry, the architectural frameworks built atop these foundations are designed to solve radically different problems.

Consequently, this taxonomic survey examines five distinct architectural archetypes: The Privacy Box (Oasis Network), The Pirate Ship (Phala Network), The Dark Pool (Flashbots SUAVE), The IP Vault (Super Protocol), and The Governance Anchor (The Citadel Protocol). We provide a detailed technical breakdown of their respective attestation flows, memory isolation mechanisms, and key management strategies. Ultimately, we conclude that while decentralized frameworks optimize for privacy or censorship resistance, the "Governance Anchor" model establishes the requisite standard for enterprise compliance (ISO 42001).

Keywords: Trusted Execution Environments (TEE), AI Governance, Confidential Computing, ISO 42001, MEV, Privacy-Preserving AI, Hardware Root of Trust.

1. Introduction — The Physics of Trust

1.1 The Collapse of "Trust Me"

Historically, the cloud computing paradigm has operated on a presumption of benevolence, where the "Trust Me" model serves as the foundational agreement between tenant and provider. In this standard environment, an AI agent exists merely as a software process,

leaving it inherently vulnerable to the superior privileges of the host operating system and hypervisor. Consequently, at any point during execution, the "Truth" of the agent's operation can be manipulated without detection: memory can be inspected, weights can be copied, and decision logs can be deleted. As a result, as autonomous agents increasingly manage high-stakes assets, reliance on this legacy architecture has evolved from a manageable technical risk into an existential liability.

1.2 The Hardware Root of Trust

To address these systemic vulnerabilities, the industry is now shifting toward anchoring agent identity to a Hardware Root of Trust (RoT). As argued in "Anchoring Agentic AI Governance to a Hardware Root of Trust" [1], this paradigm shift moves the locus of trust from the administrator to the silicon itself. By leveraging this architecture, an external observer can cryptographically verify three critical properties:

- **Identity Attestation:** The code running is exactly what it claims to be (measured via hash).
- **Environmental Integrity:** The code is running inside a genuine, compromised-free TEE.
- **State Isolation:** The memory has not been tampered with by the host OS or a system administrator.

2.1 Architecture — Runtime Offchain Logic Framework (ROFL)

Oasis Network was founded on the premise that for blockchain to transcend simple financial transactions, it must support confidential data processing. Motivated by the inherent limitations of transparent ledgers where all state is public, Oasis developed a modular architecture that fundamentally separates consensus from computation [2]. This design choice enables the Runtime Offchain Logic Framework (ROFL), a system that allows heavy, privacy-preserving compute to run off-chain in TEEs (Intel SGX/TDX) while securely settling results on the Sapphire confidential EVM [3].

Technical Implementation:

- **Isolation Mechanism:** ROFL apps run as binaries inside Intel SGX/TDX enclaves, effectively isolated from the host OS. Within this environment, a light client runs inside the enclave, which allows the app to independently validate the blockchain state without trusting the host node.
- **Attestation Flow:** For verification, the framework uses Intel DCAP (Data Center Attestation Primitives). Specifically, the application generates a Quote (binding its binary hash, MRENCLAVE, to a signing key). This Quote is then verified on-chain by the Oasis registry before the app is allowed to interact with the Sapphire network.
- **Key Management:** Oasis utilizes a Key Manager committee where keys are generated and stored within TEEs and shared via a threshold cryptography scheme. To ensure

forward secrecy, ephemeral keys are derived securely within the enclave for each session.

2.2 The Gap — The "Black Box"

While Oasis excels at Private DeFi (see Glossary) and Alpha Preservation, it functions primarily as a "Black Box." In this context, it protects the secret but does not enforce the standard of the agent's behavior. Consequently, a malicious agent inside an Oasis TEE is just as protected as a compliant one.

Verdict: Use Oasis for **Alpha Preservation**.

3. The Pirate Ship — Phala Network

3.1 Architecture — Decentralized Compute (DePin)

Phala Network emerged from the Polkadot ecosystem with the goal of creating a "Trustless Cloud". Motivated by the risks of centralized cloud providers – who can censor or de-platform applications at will – Phala built a Decentralized Physical Infrastructure Network (DePin) of TEE workers [4]. Recently, this mission has expanded to support the AI Agent economy, integrating with the Eliza framework (partnering with ai16z) to host "Unstoppable Agents" that operate beyond the reach of traditional regulatory controls.

Technical Implementation:

- **Isolation Mechanism:** Phala primarily utilizes Intel SGX and is currently transitioning to TDX. Architecturally, it employs a hybrid model where the TEE ("pRuntime") executes the contract logic off-chain, while the blockchain handles consensus and ordering.
- **The "dstack" SDK:** To streamline adoption, Phala introduced the dstack SDK, allowing developers to deploy standard Docker containers into TEEs. This abstraction simplifies the deployment of complex AI agents by handling low-level TEE initialization and key generation ("invisible private keys").
- **Key Management (KMS):** A critical differentiator for Phala is its Key Rotation mechanism [5]. Unlike centralized models, the Master Key is not held by a single entity but is distributed across the DePin network using a threshold signature scheme. Furthermore, the KMS constantly rotates management keys, reducing the window of vulnerability if a single node is compromised.

3.2 The Gap — Compliance

Phala explicitly optimizes for censorship resistance. However, its permissionless nature and focus on "unstoppable" code, while ideal for DAOs, present a challenge for regulated

enterprises that require a centralized "kill switch."

Verdict: Use Phala for **Sovereignty & Censorship Resistance**.

4. The Dark Pool — Flashbots (SUAVE)

4.1 Architecture — The Kettle and MEVM

Flashbots originally emerged to illuminate the "Dark Forest" of Ethereum MEV (Maximal Extractable Value), a predatory environment where bots exploited transaction visibility to front-run users. Motivated by the need to decentralize the block-building process and protect user intent, Flashbots developed SUAVE (Single Unifying Auction for Value Expression). This architecture creates a separate domain designed to unbundle the transaction supply chain, utilizing TEEs known as "Kettles" to function as a "Trusted Auctioneer." In this environment, bids can be evaluated privately without revealing sensitive data to the network before execution [6].

Technical Implementation:

- **Execution Environment:** The core of the Kettle is the MEVM (Modified EVM) running inside an Intel SGX enclave. To bridge the gap between privacy and verification, it includes specialized pre-compiles (e.g., `sgxattest_run`) that allow smart contracts to request attestation and access off-chain data securely [7].
- **Gramine OS:** To facilitate compatibility, SUAVE relies on Gramine, a library OS that allows unmodified Linux binaries (like the MEVM) to run inside SGX enclaves. By exposing attestation via pseudo-files (`/dev/attestation/quote`), Gramine simplifies the developer experience, removing the need for custom TEE code rewriting.
- **Confidential Data Store:** Additionally, a specialized key-value store allows users to upload private data (bids/intents). Access to this store is governed by "Peekers"—smart contracts granted specific permissions to view the data only during the auction process, ensuring that sensitive trade information remains encrypted until the exact moment of execution.

4.2 The Gap — Specialization

Despite its robust privacy guarantees, SUAVE is hyper-specialized for atomic transaction ordering and MEV protection. Consequently, while it effectively solves the "Dark Pool" problem for finance, it is not designed to host persistent, long-running agent identities or general-purpose AI workloads.

Verdict: Use Flashbots for **Transaction Physics (MEV)**.

5. The IP Vault — Super Protocol

5.1 Architecture — Confidential AI Cloud

As the value of generative AI models skyrocketed, a new threat vector emerged: the theft of proprietary model weights by malicious cloud administrators or co-tenants. Motivated by the need to secure the "crown jewels" of modern AI companies, Super Protocol evolved to focus on "Heavy Iron" computing. Unlike protocols designed for lightweight transaction logic, Super Protocol utilizes NVIDIA H100 Confidential Computing to create a secure environment specifically optimized for high-performance model inference and fine-tuning [8].

Technical Implementation:

- **Hardware Isolation:** To achieve this, it leverages the NVIDIA Hopper architecture, which extends the TEE boundary from the CPU to the GPU [9]. Crucially, it utilizes an encrypted bounce buffer to transfer data between the CPU TEE (e.g., AMD SEV-SNP) and the GPU memory (HBM). This mechanism prevents physical attackers from snooping on the PCIe bus to steal model weights during transfer.
- **TEE-Agnostic Design:** Furthermore, unlike architectures strictly tied to Intel SGX, Super Protocol is designed to be TEE-agnostic. It supports both Intel TDX and AMD SEV-SNP as the host environment for the GPU orchestration, offering greater deployment flexibility.
- **Model Protection:** The primary technical innovation, therefore, is the ability to run inference and fine-tuning where the model weights are encrypted at rest and only decrypted inside the GPU silicon. This ensures the IP remains protected from the cloud provider at all times.

5.2 The Gap — Liability

Super Protocol effectively protects the Asset (the Model) from theft. However, it does not inherently govern the Liability (the Agent's decisions). While the model is safe, its output is not strictly constrained by an external governance policy.

Verdict: Use Super Protocol for **IP Protection**.

6. The Governance Anchor — The Citadel Protocol

6.1 Architecture — Hub-and-Spoke Governance

While the aforementioned protocols prioritize privacy, decentralization, or transaction speed, The Citadel Protocol (distinct from the Dusk Network identity protocol of the same name [10], and fully defined in the Reference Architecture [12]) focuses on the requirements of enterprise risk management. Designed to facilitate ISO 42001 compliance for autonomous workforce agents, Citadel represents the "Governance Anchor" archetype. Rather than relying on decentralized consensus for state changes, it utilizes a Hub (IBM Z/LinuxONE) and Spoke

(Cloud TEEs) architecture to bind Agent Identity strictly to verified hardware policies.

Technical Implementation:

- **The Hub (Citadel):** The core of the system is anchored on IBM Z Secure Execution [11]. This environment is selected specifically to protect against the "Admin Threat." Because of this architecture, even a root user on the mainframe cannot access the memory of the Secure Execution guest (the Citadel core), as it is protected by hardware-level memory encryption keys managed by an ultra-secure Crypto Express HSM (FIPS 140-2 Level 4).
- **The Spoke:** The agent itself runs in a commodity cloud TEE (e.g., Azure Confidential VM with Intel TDX), functioning as the execution arm of the system.
- **Attestation Handshake:** To establish trust, the system uses a rigorous handshake:
 1. The Citadel Hub challenges the Spoke.
 2. The Spoke generates an attestation report (e.g., via Intel DCAP).
 3. The Citadel verifies the report against the ISO 42001 policy hash.
 4. **Connection Collapse:** If verification fails (e.g., policy deviation), the Hub's HSM refuses to sign the transaction or release the session keys. Consequently, the agent is effectively lobotomized, unable to operate outside its governance constraints.

6.2 The Gap — Cost & Complexity

The reliance on mainframe-grade hardware (IBM Z) creates a significantly higher barrier to entry compared to commodity SGX nodes. Thus, it is suitable primarily for enterprise and regulatory use cases rather than permissionless innovation.

Verdict: Use Citadel for **Corporate Governance & Compliance**.

7. Comparative Technical Analysis

To synthesize these architectural differences, the following table provides a taxonomic breakdown of the five frameworks. While they share a reliance on the "Hardware Root of Trust," their divergence in isolation models and key management strategies dictates their specific utility—ranging from open censorship resistance to closed enterprise compliance.

Feature	Oasis (Privacy Box)	Phala (Pirate Ship)	Flashbots (Dark Pool)	Super Protocol (IP Vault)	Citadel (Governance Anchor)
Primary Hardware	Intel SGX / TDX	Intel SGX / TDX (Hybrid)	Intel SGX (Gramine)	NVIDIA H100 + AMD SEV	IBM Z + Cloud TEEs
Isolation Model	Enclave + Light Client	dstack Containers	MEVM Enclave	Encrypted Bounce Buffer	Secure Execution (Admin-proof)
Attestation	Intel DCAP (On-chain)	Remote Attestation (RA-TLS)	Pre-compile (sgxattest)	GPU-to-C PU Attestation	Hub-and-Spo ke Handshake
Key Management	Threshold Committee	Key Rotation (DePin)	In-Enclave Generation	Session Keys	HSM (FIPS 140-2 L4)
Primary Gap	"Black Box" (No behavior check)	"Permissionless" (No kill switch)	"Ephemeral" (No long-term identity)	"Passive" (Asset protection only)	Complexity (Enterprise focus)

8. Conclusion — Matching Architecture to Threat Model

The "Hardware Renaissance" of 2026 demonstrates that the Trusted Execution Environment (TEE) is no longer a monolithic commodity. While the underlying silicon—Intel SGX/TDX, NVIDIA H100, and IBM Z—provides the raw material for a Hardware Root of Trust, the architectural frameworks built atop them serve fundamentally different purposes.

As this analysis highlights, there is no single "best" TEE architecture; rather, there are optimized tools for distinct threat models:

- **For Alpha Preservation:** Oasis Network provides the necessary "Privacy Box" to shield proprietary logic.
- **For Censorship Resistance:** Phala Network's "Pirate Ship" model ensures agents remain unstoppable in a permissionless environment.
- **For Market Fairness:** Flashbots SUAVE creates a "Dark Pool" to solve specific transaction ordering physics.
- **For IP Protection:** Super Protocol acts as an "IP Vault" for high-value model weights.
- **For Regulatory Compliance:** The Citadel Protocol serves as a "Governance Anchor" for enterprises bound by strict liability standards.

The transition from software-based governance to hardware-anchored truth requires architects to look beyond the spec sheet of the processor. The critical decision is not merely which chip to use, but which governance topology—decentralized, atomic, or federated—best aligns with the agent's intended role in the digital economy.

Conflict of Interest *The author of this paper, Theo Ezell, is the creator and architect of The Citadel Protocol. The comparative analysis presented herein reflects the architectural goals of that project alongside distinct market alternatives.*

9. Glossary of Terms

- **Alpha Preservation:** The protection of a proprietary trading strategy ("Alpha") from being observed, reverse-engineered, or front-run by competitors or market infrastructure.
- **Attestation:** A cryptographic process where a hardware device (TEE) proves to a remote party that it is the correct hardware running the correct software.
- **Confidential Computing:** A security paradigm that focuses on protecting data in use (during processing) by performing computations within a hardware-based Trusted Execution Environment.
- **DePin (Decentralized Physical Infrastructure Network):** A blockchain-based network that incentivizes providers to contribute physical hardware resources (like GPUs or CPUs) to a decentralized network.
- **Encrypted Bounce Buffer:** A secure memory region used to transfer data between a CPU TEE and a GPU TEE (like NVIDIA H100) to prevent the host system from reading the data on the PCIe bus.
- **Governance Anchor:** A high-security hardware node (like an IBM Z mainframe) that holds master keys and policy logic, serving as the ultimate root of trust for a distributed network of agents.
- **Hardware Root of Trust (RoT):** A set of functions within a cryptographic module (usually silicon) that is always trusted by the computer's operating system.
- **ISO 42001:** The international standard for Artificial Intelligence Management Systems (AIMS), specifying requirements for establishing, implementing, maintaining, and continually improving an AI management system.
- **MEV (Maximal Extractable Value):** The maximum value that can be extracted from block production in excess of the standard block reward and gas fees by including, excluding, and changing the order of transactions in a block.
- **Private DeFi:** Decentralized Finance applications that use privacy-preserving technologies (like TEEs or Zero-Knowledge Proofs) to hide transaction details (amounts, strategies) from the public blockchain.
- **TEE (Trusted Execution Environment):** A secure area of a main processor that guarantees code and data loaded inside to be protected with respect to confidentiality and integrity. Examples include Intel SGX, Intel TDX, and AMD SEV.

10. References

1. **Ezell, T.** (2025). "Anchoring Agentic AI Governance to a Hardware Root of Trust." WebMethodMan.com.
<https://www.webmethodman.com/p/anchoring-agentic-ai-governance-to-a-hardware-root-of-trust>
2. **Oasis Network.** (2025). "ROFL: Unlocking Secure Off-Chain Computation with Oasis Network." Oasis Protocol Foundation.
<https://dev.to/caerlower/rofl-unlocking-secure-off-chain-computation-with-oasis-network-3ien>
3. **Oasis Network.** (2025). "Inside ROFL: A Deep Technical Dive into Oasis Protocol's Runtime Offchain Logic Framework." Oasis Protocol Foundation.
<https://medium.com/@caerlower/inside-rofl-a-deep-technical-dive-into-oasis-protocols-runtime-offchain-logic-framework-330c9c97559e>
4. **Phala Network.** (2025). "Build Trustworthy Fintech AI Agents With TEE." Phala Blog.
<https://phala.com/posts/Build-Trustworthy-Fintech-AI-Agents-With-TEE>
5. **Phala Network.** (2025). "Detailed Analysis of Phala Cloud's Decentralized Root of Trust, KMS Protocol, and ZKP Enhancement." Phala Blog.
<https://phala.com/posts/detailed-analysis-of-phala-clouds-decentralized-root-of-trust-kms-protocol-and-zkp-enhancement>
6. **Flashbots.** (2025). "SUAVE Kettle Architecture Technical Specification." Flashbots Github. <https://github.com/flashbots/suave-specs/blob/main/specs/rigil/kettle.md>
7. **Flashbots.** (2024). "Demystifying Remote Attestation by Taking it On-Chain." Flashbots Collective.
<https://collective.flashbots.net/t/demystifying-remote-attestation-by-taking-it-on-chain/2629>
8. **Super Protocol.** (2024). "Exploring the Case of Super Protocol with Self-Sovereign AI and NVIDIA Confidential Computing." NVIDIA Developer Blog.
<https://developer.nvidia.com/blog/exploring-the-case-of-super-protocol-with-self-sovereign-ai-and-nvidia-confidential-computing/>
9. **NVIDIA.** (2023). "Confidential Computing on H100 GPUs for Secure and Trustworthy AI." NVIDIA Technical Blog.
<https://developer.nvidia.com/blog/confidential-computing-on-h100-gpus-for-secure-and-trustworthy-ai/>
10. **Dusk Network.** (2023). "Citadel: Self-Sovereign Identities on Dusk Network." arXiv.
<https://arxiv.org/abs/2301.09378>
11. **IBM.** (2024). "IBM Secure Execution for Linux." IBM Documentation.
<https://www.ibm.com/docs/en/linux-on-systems?topic=management-secure-execution>
12. **Ezell, T.** (2026). "The Citadel Protocol: A Reference Architecture for Hardware-Enforced Agentic Governance." Zenodo. <https://doi.org/10.5281/zenodo.18472859>