DeFi Sidecar: Latents-as-Evidence vs State-as-Evidence

Earlier Attempt — LLM Latents as Evidence

In earlier experiments (ARC-style and synthetic tests), the system tried to use **LLM latents themselves** as the primary evidence source.

- Hidden states at a GPT-2 layer (e.g., tap -9) were warped, whitened, and projected into PCA space.
- Stage-11 warp/detect/denoise was applied directly to these embeddings in an attempt to extract primitive signals.
- Problem: LLM latents are probabilistic shadows of semantics. They do not guarantee alignment with domain truths. Even with WDD, the wells were noisy and improvements over stock GPT-2 were marginal.
- Hallucinations were not eliminated—only nudged—because the evidence came from a fuzzy, non-deterministic manifold.

Tier-2 Proposal — DeFi State as Evidence

The Tier-2 design is fundamentally different. Instead of relying on raw LLM latents, the **DeFi state itself** becomes the evidence source.

- Prompt enters → pooled latent from the sidecar LLM is used only to suggest candidate primitives.
- Candidate primitives are checked against **actual DeFi state traces**: balances, oracle feeds, collateralization ratios, protocol invariants.
- Stage-11 Warp → Detect → Denoise (WDD) runs on these **state-space residual wells**, not on generic LLM embeddings.
- Because the signals come from deterministic protocol math, phantoms can be suppressed, wells converge, and hallucinations are eliminated by design.
- The sidecar thus points to candidates, but **truth is anchored in state-space geometry**, with the verifier enforcing correctness.

Core Difference

- **Earlier approach:** LLM latent = evidence (weak, probabilistic).

- **Tier-2 approach:** DeFi state = evidence (deterministic, math-grounded).

This shift makes Tier-2 integration conceptually much stronger and safer than the earlier “warp tap-9” attempt.