TruCore

ATF Security Whitepaper (Preview)

A policy + permit enforcement layer for autonomous finance

Abstract

This preview describes the security model behind TruCore Agent Transaction Firewall (ATF). ATF is an external enforcement boundary for autonomous finance systems that execute on Solana.

The design centers on deterministic controls: policy evaluation, scoped permits, pre-flight validation, and tamper-evident receipts. The goal is constrained execution with clear accountability.

Threat Model

- Unbounded execution: an agent proposes transactions outside approved risk limits.
- Protocol drift: execution moves to unapproved or unsafe on-chain programs.
- Slippage exploitation: adverse pricing and MEV extraction degrade outcomes.
- Authorization creep: agent capabilities expand beyond intended scope.
- Replay and timing abuse: stale or duplicated intents are replayed.
- Audit opacity: no reliable record of checks, approvals, and outcomes.

Trust Model

- ATF assumes model output is not inherently trustworthy and must be constrained before execution.
- ATF assumes Solana is adversarial and treats mempool and routing environments as hostile.
- ATF assumes enforcement must fail closed when checks are ambiguous or unavailable.
- ATF does not rely on agent-side self-approval for critical permissions.

Enforcement Model

ATF enforces a deterministic pipeline:

- Policy: evaluate declared intent against hard rules.
- Permit: issue a scoped, time-bound authorization token.
- Validate: perform pre-flight constraint checks before submission.

Receipt Model

Each decision point emits structured evidence: policy version, permit scope, validation result, and settlement status. Receipts are designed for tamper-evident audit and post-incident review.

V1 Scope

- Chain: Solana (v1). Multi-chain expansion planned.
- Protocol coverage: Jupiter, Orca, Raydium (swaps), Solend, Marginfi (lending). Kamino (feature-gated). Perps adapters (Drift v2, Mango v4, Hyperliquid) feature-gated, off by default.
- Controls: allowlists, spend caps, slippage constraints, TTL and nonce requirements

Design Partner Program

Teams running autonomous financial agents can apply for early integration and direct feedback loops. The program is focused on practical deployment constraints and measurable risk controls.

Apply: trucore.xyz/atf/apply