Validation Results — Result-Integrity Protocol for Distributed AI

SPHINXX.AI, LLC · April–June 2026 validation series · Public summary — July 2026 · Patent pending

The problem

When AI computation is distributed across machines no single party controls, the buyer cannot verify that returned results were actually computed — or that participating machines are real. This fraud class spans 25 years of distributed computing; in one 2024 incident, roughly 1.8 million fabricated GPU identities were injected into a commercial compute marketplace. The prevailing mitigation — redundant re-computation — costs 2–3x the underlying work.

The protocol (claim-summary level)

At claim-summary level, the protocol has three layers: (1) hardware-anchored commit-reveal binding of each machine's result before any peer's result is visible; (2) hidden known-answer verification tasks scheduled as a continuous inverse function of each machine's earned trust; and (3) a trust-weighted release threshold that withholds output and escalates to a human when consensus integrity is stressed. The protocol portfolio is patent-pending and managed by national patent counsel; the most recent provisional application was filed by counsel in July 2026. The protocol is designed to fail safe: under successful attack, its failure mode is refusing to release — across the tested envelope, integrity-stressed decisions were withheld and escalated rather than confidently released, with the higher-saturation capture boundary mapped as the primary open research problem.

Validation progression — three phases, April–June 2026

PHASE	SETUP	KEY RESULTS
1 · Simulation **Apr 12, 2026**	50 simulated nodes, 10% adversarial; 20 rounds; version-controlled code	100% consensus completion · 0% false positives · 91% probe accuracy · 80% detection (tuning issue identified and corrected)
2 · Real heterogeneous devices **Apr 13, 2026**	Four independent physical devices on a real TCP/IP network — two Windows laptops and two Android phones — one device scripted as malicious	20/20 rounds completed · the malicious device correctly identified by name · 0% false positives · 60 hidden verification tasks issued · all three protocol layers operating together on real, mixed hardware. (Adversary in this phase was uncoordinated — the easy case; this phase proves protocol mechanics, not adversarial limits.)
3 · Adversarial saturation series **Jun 9, 2026**	Physically separate machines over an encrypted mesh; 30 node identities, 67–89% adversarial saturation; coordinated adversaries; 60 rounds per configuration, multiple runs	Verification tasks hidden: 100% detection, 0% false positives, 0% adversarial capture at 67% saturation. Worst case (adversaries able to distinguish verification tasks): 0–2% capture; 98–100% of decisions withheld and escalated to a human — failed safe as designed. Capture transition at higher saturation mapped and matched analytical prediction, defining the operating envelope and the primary open research problem.

What this does and does not establish

It establishes: correct three-layer operation on real, heterogeneous physical hardware; and, under controlled adversarial conditions at 30 node identities, full detection with zero false accusations when verification tasks stay hidden — plus fail-safe (withhold-and-escalate) behavior under the worst-case scenario. **It does not establish:** performance at production scale, or resistance to adversaries beyond those simulated. Those are the funded research questions. All results are inventor-run to date; independent university replication is being arranged, and full run logs are available under appropriate agreements.

SPHINXX.AI, LLC · Morganton, North Carolina · mdshurt@gmail.com