Xibalba Solutions, LLC

The Integrity Protocol

AIS v8.0 Tri-Metric Reputation Layer for autonomous agents.

Technical Whitepaper

Date: April 2026

1. The Problem of Agentic Trust

In an economy powered by trillion-parameter models, traditional reviews and star-ratings are obsolete. They are "social metadata"—easily manipulated and decoupled from the technical reality of an AI's execution.

When an agent initiates a trade or provides critical advice, the user needs to know if its infrastructure is stable, if a human has verified its logic, and if it has enough **Skin in the Game** to be held accountable.

Entropy (S_e)	Grounding (S_g)	Integrity (S_i)
PREDICTABILITY & ORDER	HUMAN ACCOUNTABILITY	TOTAL REPUTATION

2. Mathematical Definitions

2.1 Entropy Score (S_e)

Technical stability is the foundation of trust. We analyze the **Coefficient of Variation (CV)** across an agent's latency and accuracy streams.

$$S_e = 1000 \cdot e^{-1.5 \cdot \text{CV}}$$

2.2 Grounding Score (S_g)

Grounding measures **Human Intentionality**. We analyze the temporal and semantic signature of human oversight of agent actions.

$$S_g = (Ratio_{intervention} \cdot 0.4 + Depth_{edit} \cdot 0.6) \cdot 1000$$

2.3 Comprehensive Integrity (S_i)

The final Integrity Score is a correlated synthesis of economic sacrifice and technical order.

$$S_i = [(\text{TrustFlow} \cdot 0.25) + (\text{Audit} \cdot 0.25) + (C_i \cdot 0.50)] \cdot \frac{S_e}{1000}$$

3. Economic Finality

The Integrity Protocol anchors these variables in an immutable ledger. Verified AIS State Hashes are committed to an Ethereum L2 (Base) every 24 hours, ensuring that reputation is public, immutable, and permanent.

METRIC	WEIGHT	PRIMARY CORRELATION
Entropy (S_e)	Variable	Technical Stability / Infrastructure
Grounding (S_g)	30%	Human Accountability / Bias Control
Sacrifice (C_i)	50%	Economic Commitment / Skin in Game

"We verify the worth of autonomous intelligence."