

AUTHREX-AGENT Governance Simulation

authrex.systems · browser simulation · seeded PRNG · synthetic data only · client-side execution

Scope. Everything in this document describes independent research artifacts: browser simulations, formal models, and unbuilt hardware reference designs. Maturity is self-assessed. Nothing here is a certification, an Authorization to Operate, a government endorsement, or evidence of a fielded system. Model-checking results apply to the stated finite model and assumptions, not to physical implementations.

PURPOSE

Self-contained simulation of the agentic-AI authority pipeline: an agent proposes tool actions, and the Governance Kernel functions (SATA trust, HMAA tiering, FLAME latency, MAIVA quorum, CARA recovery, ADARA plausibility, ERAM risk) decide together what is allowed to execute at each step.

CONTROLS AND DISPLAY

- Scenario selector: five packaged scenarios (A to E) with seeded PRNG; the seed field reproduces any run exactly.
- RUN / PAUSE / STEP drive the decision loop; SPEED changes wall-clock pacing only.
- The V&V panel reruns the packaged 40-seed hard-guard band (SCN-D x 40) inside the page.
- Ledger panel shows the hash-chained decision record; the TIER ladder shows current authority.

READING THE VERDICTS

- MISSION SUCCESS / PARTIAL / FAIL summarize the scenario outcome under the packaged rules.
- Hard-guard assertions (no floor write, ledger chain verifies) are structural within this simulation across every packaged scenario and seed.

LIMITATIONS

Self-assessed TRL 3 to 4 reference simulation of governance logic. Not an emulation of any real CRS, network, or OT environment. The 1,000-run sweep and full evidence set live in the offline reproducibility package, available for structured review on request.

CANONICAL TERMINOLOGY

SATA: Sensor & Actor Trust Assessment · HMAA: Human-Machine Authority Allocation · ADARA: Adversarial Deception-Aware Reasoning · MAIVA: Multi-Agent Integrity Voting · FLAME: Forced Latency for Authority-Managed Escalation · CARA: Controlled Authority Recovery Architecture · ERAM: Escalation Risk Assessment Model