

Autonomous Cyber-Defense Patch Authority Console

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

Simulates AUTHREX governing an autonomous patching agent for critical-infrastructure controllers: each proposed patch is scored on its evidence and routed to AUTHORIZE, HUMAN-IN-THE-LOOP, or DENY, so that autonomous action is bounded by a safety floor and a hard envelope rather than left to the agent alone.

CONTROLS AND DISPLAY

- Proposal stream arrives continuously; click a proposal to inspect its evidence.
- Scenario and seed controls reproduce runs; the V&V panel reruns packaged assertions.
- Status bar tracks the two structural properties: the floor (no autonomous patch to a safety-critical, human-life asset) and the bounds (no authorized action outside its envelope).

READING THE VERDICTS

- SYSTEM NOMINAL / DEGRADED-CONTAINED / GUARANTEE BREACH are simulation state labels for those two properties.
- A patch to a live controller cannot execute in this simulation without a human in the loop and a pre-armed rollback, as asserted across the packaged scenarios.

LIMITATIONS

Self-assessed TRL 3 to 4 reference architecture demonstration. Aligned conceptually with Five Eyes agentic-AI guidance and the AlxCC lineage; alignment statements are research positioning, not endorsements.

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