Safe-State Authority Governance 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

Safe-state authority governance gate for the safing lifecycle of an autonomous system: the authority to enter a safe state on a demanded fault, to hold and deepen it while the fault persists, and to exit it only under an authorized re-arm with human approval. A fail-safe governor adjudicates every safing command, the safe-state floor is never applied as a grant of authority, and the run replays deterministically against five published reference roots.

CONTROLS AND DISPLAY

- INITIALIZE, RUN, STEP +60s, and RUN TO END drive the deterministic run; APPROVE and DENY resolve reviewer holds through the simulated reviewer adapter.
- VERIFY CHAIN confirms the SHA-256 hash-chained ledger with its entry count; TAMPER TEST demonstrates detection while the live chain stays intact.
- CHECK REFERENCE ROOT reproduces the published root for the loaded scenario (five scenarios, seed 20260610); SCN_D is a PARTIAL run by construction.
- EXPORT REPORT and EXPORT TELEMETRY write local files; the console is fully client-side and makes no network calls.

READING THE VERDICTS

- Each safing command is adjudicated across the lifecycle: SAFE-ENTRY, SAFE-HOLD, SAFE-DEEPEN, and a re-arm gated SAFE-EXIT, with DENIED as the refusal outcome.
- The safe-state floor is never applied as an authority grant, human approval precedes any recovery-owned application, and a faulty authorization is always overridden; results apply to the modeled conditions only.

LIMITATIONS

AUTHREX-SAFING is a deterministic synthetic research prototype. It does not model physical dynamics, real actuation, certified hardware, operational identity infrastructure, or an accredited deployment environment. Demonstrated in a synthetic simulation environment; no formal TRL determination has been performed. Governance and assurance only, with no dynamics, no actuation, and no targeting scope.

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

AUTHREX Systems · authrex.systems · Burak Oktenli · July 10, 2026 · Site documentation, CC BY 4.0 where stated; artifact licenses vary by record · Independent single-author research; self-assessed maturity; not a government publication