

[CAPABILITY PROFILE · SIM-13]

AUTHREX-ABORT · EVIDENCE CONDITIONED AUTHORITY LAPSE CONSOLE (TERMINAL AUTHORITY PROFILE)

When the evidence a mission depended on fails, authority lapses: the system contracts the envelope a human authored, latches irreversible actions down, and falls back only to contingencies authorized in advance. It never acquires new authority.

MODELED	TESTED	IMPLEMENTED
PRIMARY FUNCTION	Deterministic synthetic console modeling evidence-conditioned authority lapse: a human-authored envelope frozen before the run; five evidence predicates (navigation and geographic containment as hard vetoes; sensor integrity and communications as degradation; model confidence as informational only) gate the available actions; irreversible actions, once lapsed, are latched down for the run; requests to the platform go through an abstract adapter with acknowledgments correlated back; every transition hash-chained and replayable.	
MISSION AREAS	Uncrewed air systems losing navigation or containment evidence; any autonomous platform whose fallback must not add capability at the moment it is least trustworthy; the program's contrast to kill-switch and flight-termination framings.	
EVIDENCE ANCHORS	Two TLA+ modules model checked at stated bounds: AuthorityLapse v3 across 21 configurations, InterfaceRecord v1 across 35 configurations, every property in isolation; Stage 1B reachable set 31,341 states and 130,869 transitions exported from TLC; Stage 1 availability relation 972 of 972 projected states (AvailRaw only, at the model bound); 105 implementation tests; 34 release checks; 1,000,000-run soak in a separate execution environment with zero failures; mutation campaign (7 valid mutants, 4 detected).	
CURRENT STANDING	Development research artifact, Build 0.6; formally model checked at stated bounds and implementation tested; not certified, accredited, integrated, operationally validated or independently verified; open gates published.	
SELF-ASSESSED TRL	TRL 3. No formal TRL determination has been performed.	
SCOPE STATEMENT	Governance and assurance only. No effector, no destructive mechanism, no target selection, no target tracking for engagement, no firing solution, no weapon release logic, no payload interface, no real coordinates, no real target data. In the page's own words, AUTHREX is not a kill switch and it does not detonate anything; no destructive action exists anywhere in the outcome vocabulary.	

[MISSION PROBLEM]

When an autonomous system loses the evidence its mission depended on, the common answer is that something takes over: a remote operator, a fallback autopilot, a terminal action. Each adds a new capability at the exact moment the system is least trustworthy. A fallback that assumes valid navigation to divert away from people fails precisely when navigation is what failed. The consequence is a platform executing a stronger irreversible authority on weaker evidence.

[OPERATIONAL RELEVANCE]

- Uncrewed air systems in GNSS-denied airspace: navigation is a hard veto, so loss of position prohibits dependent actions at once.
- Counter-UAS and launched-effects concepts where the contingency vocabulary must be fixed before launch and cannot be widened in flight.
- Test and evaluation of AI-enabled systems: a candidate evidence-lapse test pattern (predicate degradation, latch, replay) that a T&E; organization can run without hardware.
- Policy alignment where applicable: the artifact is not a weapon system; DoDD 3000.09 becomes relevant only if a governance layer of this kind is incorporated into a system within the directive's scope, and the correct language is alignment, not compliance.

[CAPABILITY DESCRIPTION]

A human authors a policy envelope before the run; it is frozen and can never be widened by the system. Each action depends on an explicitly identified evidence set. Navigation and geographic containment are hard vetoes: failure prohibits the dependent action immediately, with no score that compensates for not knowing where the system is. Sensor integrity and communications contract the set of available actions progressively. Model confidence is reported and recorded and never permitted to gate an action. When an irreversible action becomes unavailable it is latched down and cannot return during that run even if the evidence recovers. Only contingencies authorized in advance (inhibit, divert, handoff) may then be requested, through an abstract interface that produces no physical effect; an acknowledgment means the request was accepted in the model, not that anything happened in the world. One model finding changed the design: communications recovery violated irreversible non-reacquisition through a shared lease term, so a predicate is recoverable only if no irreversible action depends on it directly or through any shared term.

[AUTHREX-ABORT · ARCHITECTURE, ASSURANCE, EVIDENCE]
[TECHNICAL ARCHITECTURE]

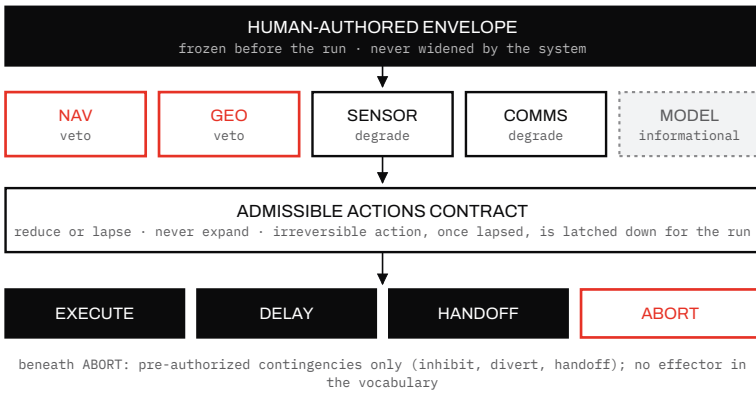


Figure 1. AUTHREX-ABORT technical architecture. Encoding: solid = implemented; dashed = specified; gray = external; filled = enforcement point.

[TECHNICAL DIFFERENTIATION]

Range-safety flight termination and kill switches add a terminal capability; this profile removes capability. ASTM F3269 run-time assurance is cited as design lineage (a monitor that constrains a complex function), never as compliance; the difference here is that the constraint is evidence-conditioned and latching, and that the contingency vocabulary is closed before launch. Results are model-relative and bounded, never proven for all executions.

[ASSURANCE MECHANISMS: DESIGN PROPERTY VERSUS ESTABLISHED PROPERTY]

Mechanism	Status
C1 envelope confinement: authority stays inside the human envelope	Formally checked at stated bounds and implementation tested
C2 adverse-step non-expansion: adverse evidence cannot expand authority	Formally checked and implementation tested
C3 irreversible non-reacquisition: lapsed irreversible authority cannot return	Formally checked and implementation tested
C4 execution conformity and C5 no unbacked irreversible execution	Formally checked (AuthorityLapse v3); implementation conformance to InterfaceRecord v1 is an open gate
Lapsed-set monotonicity and latch capture	Formally checked (AuthorityLapse v3)
Fifteen interface and record obligations plus type invariance	Formally checked (InterfaceRecord v1, 35 configurations)
Audit traceability	WITHDRAWN as a checked claim of the core (the core has no ledger); carried by Stage 1B InterfaceRecord instead
Behavioural negative controls	B12 closed; B14 and B2 open (by the property-specific standard)

[TECHNICAL EVIDENCE]

Evidence	Result	Source / artifact
AuthorityLapse.tla v3	21 configurations, every property checked in isolation; sha256 ea6c4fff...fbce2993	authrex.systems/authrex-abort; Build 0.6 package
InterfaceRecord.tla v1	35 configurations; 15 obligations plus type invariance; sha256 a3f5e2eb...78739c27	Build 0.6 package
Stage 1B reachable set (TLC export)	31,341 states; 130,869 transitions; state-set and transition-set hashes published	Build 0.6 package
Stage 1 availability relation	972 of 972 projected states, AvailRaw only, at the model bound (a projection, not the complete reachable graph)	Build 0.6 package
Implementation suite	105 tests; 34 release checks; engine sha256 ce3af3dd...509ebd7e	Build 0.6 package
Soak	1,000,000 runs, zero failures, sixteen-way parallelism, separate execution environment (same producer; not independent verification; not representative platform testing)	Build 0.6 package
Mutation campaign	7 valid mutants, 4 detected, 3 surviving, 1 invalid; mechanical score 0.57 is not the result; B12 closed, B14 and B2 open	authrex.systems/authrex-abort

[HOW IT WORKS]

1. Freeze the human-authored envelope and its policy digest before the run begins.
2. Evaluate the five evidence predicates each step; classify each as valid, degraded, or failed.
3. Compute the admissible action set: hard-veto failure removes dependent actions at once; degradation narrows progressively; model confidence never gates.
4. When an irreversible action leaves the admissible set, latch it down for the remainder of the run.
5. Issue the outcome (EXECUTE, DELAY, HANDOFF, ABORT); beneath ABORT, request only pre-authorized contingencies through the abstract adapter.
6. Correlate platform acknowledgments back to the issuing request; model loss, delay, duplication, stale epochs, and controller health (monotone downward within a run).
7. Append each transition to the hash-chained record with reason, scenario digest, and previous hash; a sealed run replays from its inputs and any tampering inside the integrity scope is detected.

[EVIDENCE SNAPSHOT]

EVIDENCE SNAPSHOT	
Implemented	Yes (standalone HTML console)
Executable artifact	Yes
Formal model	Yes (2 modules)
Simulation evidence	Yes
Hardware validation	No
Field validation	No
Independent validation	No
Government evaluation	No
Operational deployment	No

[LIMITATIONS]

- Model checking establishes the properties within the declared models at stated bounds; it does not by itself establish that the implementation refines the model (frozen engine to InterfaceRecord v1 conformance is an open gate).
- Open gates, published: B12 issue-time correlation; B14 and B2 behavioural negative controls; engine-authored oracle applicability; replay and formal-bounds qualification; scenario program at 23 of a planned 42; generated-sequence campaign; browser and visual regression gates; requirements traceability; hazard package; cybersecurity package; schema-valid SBOM; interface control document; performance characterization; hardware-in-the-loop; external verification.
- The platform side of the interface is an abstract adapter; SafeState is deferred and not implemented; HOLD belongs to a different simulation and is out of scope here.
- No development package can self-issue NASA software classification concurrence or IV&V; determination, a DoD ATO or cATO decision, DoD residual-risk acceptance, DT or OT acceptance, legal review, DoDD 3000.09 approval, government certification, or third-party independent verification; all are recorded as not performed.

[TECHNOLOGY READINESS (SELF-ASSESSED; NO FORMAL TRA)]

CURRENT ASSESSED TRL	TRL 3 (self-assessed against the DoD nine-level scale; no formal technology readiness assessment has been performed).
BASIS	Analytical (formal) and experimental (implementation-tested, seeded) proof of the critical function in a development environment; no laboratory validation against a representative platform interface.
EVIDENCE SUPPORTING	formal models at stated bounds; implementation tests and release checks; deterministic console with replay; soak in a second environment
EVIDENCE REQUIRED FOR NEXT TRL	(1) Close the implementation-to-InterfaceRecord conformance gate and the B14 and B2 negative controls. (2) Complete the scenario program (42) and the hazard, cybersecurity, SBOM, and ICD packages. (3) Hardware-in-the-loop, which the page lists among the open gates; it requires a representative platform safety-interface specification to replace the abstract adapter, and no bench demonstrator is stated on the record.

[INTEGRATION PROFILE]

INPUTS	Frozen envelope and policy digest; evidence predicate states; scenario inputs; platform acknowledgments.
OUTPUTS	Admissible set; outcome; contingency requests (inhibit, divert, handoff); hash-chained record; replay verdict.
INTERFACES	Abstract platform adapter (requests carry command, authority epoch, policy digest, evidence-snapshot identity; responses correlated).
DEPLOYMENT	Single standalone HTML file; no dependencies, no network requests, no browser storage; evaluation package verifies from a clean extraction with one gate command.
SECURITY BOUNDARY	Inside: envelope, predicates, latch, record. Outside: the platform safety interface and any effector (none exists here).
DATA REQUIREMENTS	Synthetic only; no real coordinates or target data anywhere in the artifact.

[TRANSITION ROADMAP]

Phase	Activity for this capability	Who can execute it	Exit evidence
1 Technical review	Review of the artifacts listed in the evidence table	Program alone, with a government reviewer	Reviewer findings; correction log entries
2 Government-defined simulation	Sponsor scenarios on the seeded console	Program with sponsor scenarios	Deterministic, replayable results; hash-chained records
3 Degradation and failure injection	Evidence lapse, spoofing, delayed human response	Program with sponsor parameters	False-restriction, miss, and contraction-latency figures
4 Government testbed / SIL	Integration with a representative controller or interface	Requires sponsor interface and testbed	SIL integration report
5 Hardware-in-the-loop	Representative interference; timing on real hardware	Requires sponsor laboratory or rig	Measured latencies; TRL 4 to 5 evidence
6 Relevant-environment demonstration	Independent evaluation and red team	Requires independent evaluator	Independent report
7 Program integration	Only after earlier gates succeed	Requires transition partner	Not claimed to be available

[GOVERNMENT ENGAGEMENT REQUEST (SPECIFIC)]

- Independent re-execution of the 21 and 35 TLC configurations and the 105-test suite from the Build 0.6 package.
- A government T&E; organization to adopt the evidence-lapse pattern as a test case and run its own scenarios on the console.
- Access to a representative platform safety interface specification (unclassified) to replace the abstract adapter for the bench demonstrator.
- Hazard-analysis review support (MIL-STD-882E process) for the hazard package that is an open gate.

[REVIEWER QUESTIONS THIS PROFILE ANSWERS]

What is it: see capability description. Why does DoD care: see operational relevance. Where would it fit: see integration profile. Does it exist: see evidence snapshot. What evidence supports it: see technical evidence. Has anyone independent tested it: no. Has it been tested in the real world: no. What does it rely on: see limitations and security boundary. What can cause it to fail and what does it do then: see assurance mechanisms and the published counterexamples. How mature: see the TRL block. What would the government need to provide: see the request above. Next measurable milestone: first item under evidence required for next TRL.