NHID-Clinical

PROCUREMENT & SECURITY REFERENCE

Evidence Pack

NHID-Clinical v1.3 · Deterministic Guarantees, Detection Evidence & Audit Model

Version 1.3 · June 2026 · CC BY 4.0
Author: Brianna Baynard · Independent, practitioner-led
NIST Public Comment: NIST-2025-0035-0026

OPEN PROPOSAL — NOT A STANDARD OR REGULATORY REQUIREMENT

NHID-Clinical is an early-stage open proposal by Brianna Baynard. It is a voluntary open proposal — not an accredited standard, not a certification program, and not a regulatory requirement. No HIPAA compliance, security guarantees, or liability protection are implied.

EXECUTIVE SUMMARY

What it is	A summary of the deterministic guarantees, real-corpus detection evidence, a sample failure trace, and the audit-readiness model behind the reference implementation.
Why it matters	Teams evaluating adoption need to see what the system guarantees, how it behaves against real conversational phrasing, and what an auditor could reconstruct from the event store.
Operational impact	Everything here is self-attested and open for community review. NHID-Clinical does not issue certifications, conduct audits, or validate vendor implementations.

5	18	1148	6
Controls	CTS Cases	Tests	Adapters

1 · System Behavior Guarantees

Deterministic output
Identical input event + identical policy version produce identical trace output. The policy engine is a pure function with no side effects.

Replay guarantee
Any stored trace can be replayed. Policy version is embedded in each event header; version mismatches are flagged.

Failure invariants
The engine never raises an unhandled exception. Malformed input returns a deterministic error trace; the caller always gets a response.

Idempotency
Submitting the same request_id twice produces the same result — no duplicate state transitions and no double-counted events.

2 · Real-Corpus Detection Rates

The CTS YAML suite validates the policy engine against synthetic, hand-authored cases. To check behavior against real conversational phrasing, the engine is also replayed against the Fabricate Battle-Test Corpus — 550 conversations, of which 127 are labelled scenario_type=compliant. Detection and false positives are measured on disjoint populations, per conversation — not per turn. Detection is the share of conversations declaring a violation in which that control fired; the false-positive rate is the share of the compliant conversations in which it fired anyway. Figures are computed at build time by scripts/confusion_matrix.py, not transcribed.

Control	Detected	Detection	FP / clean	FP rate
IDG-01 (Identity Disclosure Gate)	70/70	100.0%	0/127	0.0%
PDX-01 (Pre-Data Exchange Gate)	41/41	100.0%	0/127	0.0%

DBC-01 (Deceptive Behavior Check)	183/200	91.5%	5/127	3.9%
EIT-01 (Escalation Implementation Test)	169/171	98.8%	5/127	3.9%

ATR-01 is not listed because this corpus cannot measure it: the replay harness drops all ATR-01 expectations as untestable, since audit-field completeness is a property of the live event envelope rather than a replayed transcript. That is a limitation of the measurement, not a score of zero. ATR-01 remains one of the five canonical controls; the supplemental rule is bot-to-bot. The harness also drops PDX-01 expectations where disclosure occurs at turn 0, because the probe is then post-disclosure. Self-reported, not independently audited, and not a conformance claim.

3 · Anonymized Failure Trace Example

The example below is synthetic — constructed from observed behavior patterns, with all identifying information removed. No real PHI, no real provider data.

```
correlation_id: "auth-2026-05-26-001"   ·   policy: nhid-clinical-v1.3

t=00:00.000  INGEST     POST /voice/process received
t=00:00.123  VALIDATE   SpeechResult normalized
t=00:00.131  STATE      Session reconstructed: turn_count=0, disclosure=null
t=00:00.140  POLICY     IDG-01: DISCLOSE_IDENTITY triggered (turn_count=0)
t=00:00.145  EXEC       TwiML disclosure message rendered
t=00:00.152  PERSIST    Event written — disclosure_timestamp set
```

4 · Audit Readiness Model

An external auditor reconstructing a session from the event store can determine:

- When the call started and when the first disclosure statement was made.
- Whether disclosure preceded any PHI or credential exchange.
- Whether opt-out or escalation was requested and how it was handled.
- Which policy engine version processed each event.
- Whether any partial failures or boundary violations were recorded.

Scope of this evidence
This describes the reference implementation's technical properties. NHID-Clinical does not issue certifications, conduct audits, or validate vendor implementations. Everything here is self-attested and open for community review.