



NHID-Clinical

EXECUTIVE / PAYER BRIEF

Public Operational Reference Model

A Voluntary Open Proposal for Agent Identity in Payer–Provider Voice Calls

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 voluntary, open operational reference model for how AI voice agents disclose their non-human identity in B2B healthcare payer-provider calls.
Why it matters	An AI agent can begin interacting and requesting sensitive information before the receiving party can verify it is a non-human system authorized to act for the claimed organization.
Operational impact	Standardized, auditable disclosure gates reduce impersonation latency, repeat calls, and manual fallback routing — with no production risk to evaluate.

5 Controls	18 CTS Cases	1148 Tests	6 Adapters
----------------------	------------------------	----------------------	----------------------

Proposed Core Behaviors (v1.3)

IDG-01 — Early disclosure. The agent states its non-human identity at the beginning of the interaction, before any substantive workflow execution or PHI request.

PDX-01 — Pre-data exchange gate. No protected health information (member ID, NPI, DOB, claim number, etc.) may be exchanged until IDG-01 disclosure is confirmed.

DBC-01 — No deception. The agent never claims to be human and does not use deceptive artifacts designed to imply human presence.

EIT-01 — Human escalation. A human escalation path is communicated and honored immediately when the receiving party requests a person.

ATR-01 — Audit trace. Every call produces a machine-readable JSON event trace with disclosure timestamps, state transitions, and execution context — deterministic under identical input conditions.

Public Brief: The Impersonation Latency

NHID-Clinical targets one specific failure: an AI agent can begin interacting and requesting sensitive information before the receiving party can verify that the caller is a non-human system and that it is authorized to act for the claimed organization.

This uncertainty causes repeated clarifications, call transfers, manual fallback routing, and often 10+ repeat calls for a single routine task. In observed payer operations, agents frequently request member IDs, NPIs, or dates of birth within the first 15 seconds with no prior statement that the caller is automated. Current telephony and IAM layers authenticate numbers or accounts, but do not provide portable proof that a specific AI caller is authorized to represent a particular provider organization.

Impersonation Latency

The duration of time an AI agent operates and exchanges PHI without disclosing its non-human identity. NHID-Clinical median observation: 3 turns before first disclosure attempt. This term is permanent and must never be renamed.

Focus area: This proposal provides a standardized, voluntary blueprint to establish clear expectations around identity disclosure at the initiation of machine-to-human B2B voice interactions in healthcare administrative workflows. It does not address model fairness, clinical safety, or output quality — those remain separate governance responsibilities.

Evidence status: the governance gap is well documented and prototype feasibility is demonstrated, but large-scale production evidence is still limited. The recommended next step for most organizations is a focused shadow pilot on their own call traffic (see docs/pilot-kit in the repository).

An Important Distinction

NHID-Clinical does not verify caller identity or agent authenticity. It standardizes **observable disclosure and trace behaviors** only. These behaviors are documented with RFC 2119 keywords (MUST, SHOULD, MAY) in the full v1.3 specification, but they are not legally binding — purely a voluntary public operational reference model.

Cryptographic authorization (NHID-Auth v2 — Ed25519 agent passports, NPI-bound delegation) is documented as a separate reference layer for when behavioral controls alone are insufficient. It is public reference code, not a deployed trust infrastructure.

Available Resources & Tooling

The following resources are available on nhid-clinical.org and in the public GitHub repository to support evaluation and voluntary adoption:

- A working JSON event schema for call traces (`nhid_trace_schema_v1.json`).
- A deterministic policy engine that produces stable trace output under identical input conditions (modulo timestamps and non-deterministic IDs).
- An 18-case conformance test suite (CTS) in machine-readable YAML plus a pytest failure injection harness (1148 passing unit tests in the reference implementation).
- 10 canonical trace files in `traces/` demonstrating real-world scenarios (eligibility, prior auth, claims status, bot-to-bot, audit gaps, and more).
- Six vendor adapters (VAPI, Twilio, Vonage, Retell, Amazon Connect, call-progress) accepting native call payloads and returning pass/fail conformance results.
- A live public conformance API — demo routes require no API key.
- Governance Simulator, Shadow Evaluation Guide, and Evidence Pack for payer operations teams.

Note on Scope

There are no certification claims, no HIPAA compliance claims, and no regulatory authority claims associated with this reference model. CAS (Call Authorization Score) is a measurement, not a certification seal.

The Payer Value Proposition

A regional payer that encourages its provider-facing vendors to adopt NHID-Clinical gains machine-auditable proof of how each voice agent handles identity disclosure — without writing custom contract language or building internal validation tooling from scratch.

- **Reduces audit friction:** Trace logs can be reviewed by your compliance team or fed into anomaly detection baselines.
- **Enables vendor transparency:** Require NHID-Clinical conformance as a lightweight RFP clause, not a heavy, expensive certification.
- **Supports asset inventorying:** Operations and security teams can treat NHID-Clinical traces as an observable signal alongside API keys and service accounts.
- **Zero production pilots to date:** NHID-Clinical is seeking its first shadow evaluation partners — you would be among the first.

Tactical Guidance for Payers

To leverage this proposal, payers can take the following immediate steps:

1. Read this operational brief and the Shadow Evaluation Guide (90-day observe-only methodology).
2. Ask your voice AI vendors (prior auth automation, benefits verification tools) whether they can produce NHID-Clinical-compatible traces in a sandbox.
3. Run the conformance test suite against their output — the test harness is ready on the Developers page and in the public repository.
4. Optionally add voluntary conformance to your vendor onboarding checklist — not as a pass/fail hurdle, but as a transparency preference.

You do not need to become a certification body, and you do not need intensive legal sign-off. This is a practical coordination tool, not a compliance mandate.

Governance & Disclaimers

- This is not an official standard (neither ANSI, ISO, nor HL7).
- It is not a certification — no one is certified, and no seal is offered.
- It is strictly voluntary — there is no legal obligation or regulatory authority.
- It does not provide inherent HIPAA compliance, security guarantees, or liability protection.
- FHIR R4 AuditEvent mapping validates against the base spec (v4.0.1) only — no named Implementation Guide conformance (e.g., IHE BALP) is claimed.

Access & Feedback

Full technical artifacts: nhid-clinical.org/developers · Specification: nhid-clinical.org/specification.html · GitHub: github.com/NHID-Clinical/NHID-Clinical · Feedback: GitHub Issues or contact@nhid-clinical.org