FORMAL VERIFICATION AS REGULATORY VALIDATION FOR AI ACROSS THE CLINICAL TRIAL LIFECYCLE

Steven Thompson · NexTrial.ai · Los Angeles, CA, USA / São Paulo, Brazil

Gate 1 · Jurisdiction Verification — Gate 2 · Formal Proof Certificate — Gate 3 · Human Oversight

Disclosure: S. Thompson is Founder and CEO of NexTrial.ai

BACKGROUND & OBJECTIVE

The Validation Gap

- **FDA-EMA: high concordance on protocol-compliance findings, with frequent divergence on documentation** (Sellers et al., 2022; PMID 35610469) — yet the AI layer generating that compliance evidence has no deterministic verification standard.
- **Statistical performance metrics** describe output-level certainty, not output-level accuracy of this specific output.
- The same system may still be wrong on a specific output for this patient, this site, this jurisdiction, this trial.
- Unified verification standard enables compliance evidence across **FDA, ANVISA, CDSCO, EMA and the EU AI Act.**
- **The FDA Real-Time Clinical Trials (RCTC) initiative moves oversight toward continuous, near-real-time monitoring - making deterministic verification an operational necessity (Docket FDA-2026-N-4390).**
- No industry standard exists for deterministic, auditable AI compliance evidence applicable across **FDA, ANVISA, and CDSCO.**

167 days

Median site activation time — nearly 2× the 90-day target (AACI benchmark: NCI target). Activation is the most controllable driver of the trial timeline

~80%

Trials that fail to complete within their planned timeline — median delay 12.2 months (JAMA Netw Open, 2023; cross-sectional analysis, n=2,542)

0

Unified multi-jurisdiction AI verification frameworks requiring deterministic, jurisdiction-specific compliance evidence — published prior to this work

OBJECTIVE

To propose and validate a three-gate regulatory verification architecture for AI systems across the clinical trial lifecycle that provides:

- Deterministic, auditable compliance evidence through formal mathematical proof (Gate 2)
- Jurisdiction-specific regulatory verification covering FDA, ANVISA, and CDSCO (Gate 1)
- Mandatory human oversight with electronic audit trail per ICH E6(R3) / 21 CFR Part 11 (Gate 3)

FOUNDATIONAL CLAIM

Proof certificates represent fundamentally different evidence than confidence intervals they are used in addition to confidence intervals, never as a replacement.

The system that generates outputs must be architecturally independent from the system that verifies them.

GLOBAL BY DESIGN, COVERING:

- **FDA 21 CFR Parts** 11, 50, 54, 56, 312, 314
- **ANVISA Lei** 14.874/2024; RDC 945/2024; RDC 751/657/848; CFM 2.454/2026 (published Feb 2026, effective Aug 2026)
- **CDSCO NDCTR 2019**; Seventh Schedule; DPDP Act 2023
- **EU AI Act** Reg 2024/1689; Annex III; CTR Reg 536/2014
- **ICH E6(R3)** 2025

Architecture is global by construction, with NMPA and MHRA reachable via adapter

METHODS & ARCHITECTURE

THE VERIFICATION FLOW

Generated output — produced, not approved → **Gate 1 · CFM-1** — regulatory semantics → **Gate 2 · Lean4** — structural proof— designed → **Gate 3 · Human** — final decision

CRITICAL failure → return for correction (self-healing loop)

Immutable audit trail · bi-temporal · 21 CFR Part 11
every output, verification, and decision — including failures

The component that generates an output is architecturally separate from the systems that verify it — no component both produces and approves its own work.

FOUR-LAYER VERIFICATION STACK

Risk-Based Quality Management (RBQM) · ICH E6(R3) · EU AI Act Art. 9
Identifies high-risk AI outputs requiring elevated scrutiny before Gate 1. Classifies output risk by trial phase and regulatory consequence.

Gate 1 - CFM-1 Jurisdiction-Specific Verification
LoRA adapter per jurisdiction — one adapter per jurisdiction, one shared base model. Violations: Critical / Warning / Info. Verification surface: Column 3.

Gate 2 - Lean4 Formal Proof Certificates
Formal theorem prover — generates binary PASS/FAIL. Six structural properties: three shipping (MVP), three Phase 2. Full specification: Column 4.

Gate 3 - Mandatory Human Oversight
Mandatory. PI attestation level configured by sponsor and PI at site activation. 21 CFR Part 11 audit trail. Full governance: Column 5.

How a Regulation Becomes a Deterministic Check

CFM-1 does not interpret regulations at decision time. Each jurisdiction's requirements are encoded into a verification adapter through training against a deterministic verifier — the regulatory verifier itself is the reward signal. This is what separates encoded, validated compliance from a model interpreting rules on the fly.

Stage	Mechanism	
1.	Encode requirements	Each jurisdiction's regulatory requirements rendered into structured verification pairs — the regulation is the ground truth, not generated data
2.	Adapter training (SFT)	LoRA Rank-16 adapter imprints baseline regulatory compliance
3.	RLVR loop (PPO)	is designed to route adapter output → deterministic verifier → reward → policy update. +reward for correct verification, penalty for a missed CRITICAL violation
4.	Freeze	The trained adapter is frozen; the encoding no longer changes at inference

A golden evaluation dataset — held-out, jurisdiction-specific — is the artifact a CI/CD gate runs on every change to an adapter. It is not training data; it is the regression floor that proves an updated adapter has not degraded before it can ship. The distinction between a validation strategy and a validated system

Multi-Jurisdictional Adapter Architecture

FDA Adapter
21 CFR 11, 50, 54
56, 312, 314

ANVISA Adapter
Lei 14.874/2024
RDC 945/2024
RDC 751 · 657 · 848
CFM 2.454/2026*

CDSCO Adapter
NDCTR 2019
Seventh Schedule
DPDP Act 2023

EU AI Act
Reg. 2024/1689
CTR 536/2014
Annex III

CFM nº 2.454/2026: published Feb 2026; effective Aug 2026

PER-SITE, PER-JURISDICTION ISOLATION

Most platforms promise data residency in a contract. This architecture enforces it in the routing

	US site	Brazil site	India site
Routed to	FDA adapter	ANVISA adapter	CDSCO adapter
Runs in region	us-central1	southamerica-east1	asia-south1
Protocol data leaves region?	No	No	No
Residency basis	21 CFR Part 11	RDC 945 · LGPD	DPDP Act 2023

IT DOES NOT MOVE

GATE 1: JURISDICTION VERIFICATION

What Gate 1 Verifies: A Multi-Way Match, Not a Single Check

Gate 1 answers one question for every output: does this satisfy the regulatory requirements of the target jurisdiction? But requirements are rarely a single authority. Once the protocol is verified, it becomes a verification authority in its own right — a document that contradicts the protocol fails Gate 1 as surely as one that violates a regulation.

THE MULTI-WAY MATCH

Output Verified	Checked Against	Match
Protocol	Jurisdiction regulatory requirements	Two-way
SOP	Regulation + the protocol	Three-way
Clinical Trial Agreement	Regulation + the protocol	Three-way
Informed Consent Form	Regulation + the protocol	Three-way
Community / context inputs	Regulation + the protocol	Three-way
Execution-phase artifacts (TMF, monitoring, closure)	Regulation + the protocol	Three-way · architecture-tense

Protocol, SOP, CTA, and ICF verification at trial startup is the framework's present, shipping focus.Extension across execution-phase artifacts is the architecture's designed reach — general by construction.

DOCUMENTS AND DECISIONS BOTH CARRY CERTIFICATES

A proof certificate is designed to be generated not only when the system produces a document, but when the trial records a decision. Verifying an ICF is not only confirming the form is compliant — it is verifying the consent event.

The Decision	What Gate 1 Verifies About the Event
Informed consent	Which ICF version was signed · that it was the current approved version · who obtained consent · when
Eligibility / screening	The criteria applied · the source values · the jurisdiction snapshot at decision time
Enrollment	The authorities the decision was checked against · what the human reserved (Property 4)

A decision that cannot produce a clean certificate is a decision that gets escalated — not waved through

WHAT EACH JURISDICTION ADAPTER ENCODES

Jurisdiction	Representative Encoded Checks	CRITICAL Example — Blocks Progression
United States FDA	21 CFR 11 electronic-records integrity · 21 CFR 50.25 informed-consent elements incl. exculpatory-language prohibition · 21 CFR 56.103 IRB approval · 21 CFR 312.23 IND completeness · 21 CFR 54 investigator financial disclosure	Exculpatory language in consent → blocked (21 CFR 50.25)
Brazil ANVISA	Trial-registration proof matching submitted protocol (RDC 945) · DDCM/DEEC dossier completeness · post-trial-access planning (Lei 14.874/2024) · CEP ethics approval · LGPD lawful basis · SaMD framework: RDC 751 / 657 / 848	Trial-registration proof absent at DEEC submission → blocked (RDC 945)
India CDSCO	Registered Ethics Committee approval before initiation (NDCTR 2019) · Seventh Schedule subject-compensation framework · CTRI registration · data-fiduciary obligations (DPDP Act 2023)	Ethics Committee approval missing before initiation → blocked (NDCTR 2019)

The adapter encodes each jurisdiction's full regulatory tradition — including Brazil's Software-as-a-Medical-Device framework, which few competing approaches have addressed. The three adapters are validated reference instances spanning structurally distinct regulatory traditions; the mechanism is general.

THREE SEVERITIES, ONE SELF-HEALING LOOP

CRITICAL	Blocks progression. Output cannot advance.
WARNING	Flagged for human review with specific remediation guidance
INFO	Logged for audit-trail completeness; does not impede progression.

Generate → verify → correct → re-verify. A Critical failure returns the output to the generating subagent with the specific violation; it is corrected and re-enters Gate 1 — a bounded, logged loop, not open-ended retry. Only output that passes Gate 1 advances to Gate 2, which is designed to emit a Lean4 structural proof. Every attempt is recorded.

CFM-1 owns semantics; Lean4 owns structure. Whether the content of a field is regulatory-accurate is Gate 1's responsibility. Gate 2 proves the document is well-formed; it does not prove the regulation was chosen correctly. That gap — who certifies the encoding — is named in Open Question VI, not hidden

GATE 2: FORMAL PROOF

What Gate 2 Proves: Structure, Not Probability

Gate 2 is where proof certificate stops being a metaphor and becomes a mathematical fact. A formal proof has no probabilistic middle ground — it either succeeds or it fails. This is the gate that makes the framework's central claim literal: a proof certificate is a different class of evidence than a confidence interval, because it is produced by proving, not by estimating.

THE SIX STRUCTURAL PROPERTIES

Property	What Is Proven	
Field Presence	Every required field exists for the document type and jurisdiction	
Version Consistency	Every version reference points to the current document version across the set under verification	
Reference Resolution	Every cross-reference resolves to an existing document in the regulatory package	
Completeness	Every required section is present per document type, jurisdiction, and protocol	
Non-Contradiction	No logical conflict exists between documents in the set under verification	
Temporal Ordering	Every event, approval, and dependency occurs in a logically valid time sequence	

Gate 2 is designed to verify these properties; the three MVP properties define the minimum proof set, with completeness, non-contradiction, and temporal ordering scoped to Phase 2.

WHY A PROOF IS NOT A SCORE

Binary	Succeeds or fails. No confidence score, no statistical estimate, no threshold to tune.
Deterministic	Same input + same proof definition → same result, always.
Independently checkable	Any party with a Lean4 proof checker can re-verify the certificate — without trusting the originating system, the model, or the original run.
Per-output	The proof addresses this document, not aggregate performance across many.

A confidence interval describes how often a system is right across a population. A Lean4 proof addresses whether this output satisfied every structural constraint that governed it. In addition to statistical validation — never instead of it.

Each Lean4 verification is designed to generate a proof certificate: the properties proven, any failures with specifics, a verification identifier for audit-trail linkage, and a record sufficient for independent re-checking. That certificate carries the eight properties specified in the certificate schema — retained in the immutable, bi-temporal audit trail as the per-decision credibility record.

PHYSICS-INFORMED ELIGIBILITY PREDICTION

87.3%

Eligibility prediction accuracy · AL amyloidosis · n=312 retrospective validation cohort

AUC 0.821 Enrollment Eligibility

AUC 0.794 Trial Completion

AUC 0.846 Toxicity Prediction

- PINNs embed governing differential equations directly in the loss function: $dx/dt = f(x,\theta,g) + \varepsilon$ — patients modeled as trajectories through constrained state spaces, not static data points
- Equivalent prediction accuracy with 5–10x less training data vs. purely data-driven approaches (Raissi et al., 2019)
- Every prediction is designed to produce a Gate 2 proof certificate — eligibility verified, not estimated. Three gates applied before any output enters the clinical record
- Scoped result: AL amyloidosis, n=312. Prospective clinical validation ongoing. No breadth claim.

RISK CLASSIFICATION AS ARCHITECTURAL PRIMITIVE

Risk Class	Substitution	Reverification	Gate 3 Min.
CRITICAL	Prohibited	Every cert	Level 2
HIGH	Prohibited	Daily	Level 2
MODERATE	Authorized	Weekly	Level 1
LOW	Authorized	Monthly	Level 1

Risk class is assigned and frozen at the moment of certificate generation. Working session polls 3, 4, and 5 each independently resolved to risk-based approache

GATE 3: HUMAN OVERSIGHT

What Gate 3 Enforces: The Human Always Decides

Gate 3 is the gate that looks soft and is the most structurally enforced of the three. No AI-generated output reaches a decision-maker without it. The system presents the verification evidence — the Gate 1 result, the Gate 2 certificate — and a qualified human makes the final determination. Passing Gates 1 and 2 does not reduce the PI's oversight obligation. The certificate is evidence the human consumes, never a substitute for the human's own verification

THREE ATTESTATION LEVELS

Level 1 Review	Reviewed AI output and its verification evidence (Gate 1 result + Gate 2 certificate)
Level 2 Independent Verification	Independently verified the underlying clinical data against source documents
Level 3 Clinical Judgment	Exercised independent clinical judgment that may supersede the AI recommendation, with documented rationale

The required level is configured by the sponsor and PI at site activation — not by the AI system. Level 3 explicitly authorizes the human to override the AI with documented rationale. Human authority is configured above the system, not by it

RISK CLASS SETS THE FLOOR

CRITICAL	Level 2	Every Certificate
HIGH	Level 2	Daily
MODERATE	Level 1 Permitted	Weekly
LOW	Level 1 Sufficient	Monthly

Risk class is assigned and frozen at decision time. For CRITICAL and HIGH, substitution is prohibited by construction, not by policy

ESCALATION IS STRUCTURAL, NOT A FOOTNOTE

Escalation Property	What the Record Captures
Override invoked	Whether the human overrode the AI, with cited rationale
SLA	Service-level agreement invoked; start and deadline timestamps
Outcome	RESOLVED / EXPIRED / AUTO-PAUSED; adjudicating party; final disposition

A decision that cannot produce a clean record is escalated, not waved through. The enforcement mechanism is an automatic pause, not a request for one.
Every output, every Gate 1 result, every Gate 2 certificate, and every human decision is logged to the immutable, bi-temporal audit trail — timestamped, tamper-evident, 21 CFR Part 11.

OVERRIDE PROTOCOL — SLA ARCHITECTURE

- **Real-time sponsor SLA with auto-pause** — primary mechanism (Poll 13: 7-3 over 24h cooling-off)
- Escalation holds under activation pressure: 9-2 (Poll 11). **Structural, not procedural**
- SLA breach triggers automatic CAPA-equivalent documentation. Arbitration standards: open co-development question (Column 6, Q.V)

FRAMEWORK CRITIQUE — HONEST ACCOUNTING

- **3 gaps acknowledged and addressed:** risk management (Art. 9) → RBQM; data governance (Art. 10) → risk primitive + provenance; rubber-stamp → attestation architecture with invariant PI obligation
- Named honestly: **Level 2 attestation adds estimated 15–20 min** per eligibility determination. Equitable implementation for community, rural, and underserved sites is an open design question

WORKING SESSION — MAY 14, 2026

14 practitioners from US, Brazil, India, and EU — 13 anonymous Mentimeter polls. Four questions verified. Two new architectural questions surfaced.

CFM Nº 2.454/2026 — GATE MAPPING (EFFECTIVE AUG 2026)

Mandatory human mediation — AI may not independently communicate diagnoses or therapeutic recommendations	Art. 5	Gate 3 — mandatory human oversight; no autonomous output
Integral physician responsibility — physician bears full accountability for AI-supported acts	Art. 7	Gate 3 — human attestation; liability attaches to the person, never the system
Complete auditability — reasoning trails and decision lineage	Res. req.	Gate 1 + audit trail — verification surface + bi-temporal immutable record
Institutional risk classification	Res. req.	Risk primitive — CRITICAL/ HIGH/MODERATE/LOW, frozen at decision time
Patient transparency — informed consent for AI use in care	Res. req.	Gate 3 + ICF verification — consent event verified, AI use disclosed

CONCLUSIONS & OPEN QUESTIONS

CORE CONCLUSION

Proof certificates represent fundamentally different evidence than confidence intervals — in addition to them, never instead of them.

AI verification is EVIDENCE, not SUBSTITUTION. A single three-gate architecture governs AI from protocol ingestion through trial closeout across FDA, ANVISA, and CDSCO.

THE PROCESS IS WORKING

This framework is not asserted — it is being co-developed. A May 2026 working session of 14 practitioners — regulatory-affairs leaders, GCP auditors, ethics-board members, and principal investigators from the US, Brazil, India, and the EU — pressure-tested Version 0.1 across 13 anonymous polls. Engagement verified three open questions through quantitative consensus. That is how a proposal becomes a standard.

Adapter Validation — VERIFIED

Risk-stratified validation proportionate to change magnitude and patient risk, with a mandatory regression floor and a golden evaluation dataset enforced by a CI/CD gate. The CI/CD gate is the mechanism that separates a validation strategy from a validated system.

Continuous Verification — VERIFIED

Risk-classified reverification cadence: CRITICAL — every certificate; HIGH — daily; MODERATE — weekly; LOW — monthly. Working session polls 3, 4, and 5 each independently converged to risk-based approaches — three separate questions, one answer.

Liability — VERIFIED

Allocated by attestation level, PI primary. Liability attaches to people and organizations — sponsor, investigator, participant — never to the AI system. The certificate is upstream evidence; accountability belongs to the person who consumed it.

PCCP Mapping — VERIFIED

PCCP governs when and under what conditions a change may occur. The three gates generate the compliance evidence that a PCCP-governed change was implemented correctly. The certificate differential — $V(n)$ vs. $V(n+1)$ on the same holdout set — is the impact assessment PCCP requires.

EXTERNAL CONVERSATION

The January 2026 FDA-EMA joint Guiding Principles of Good AI Practice document shared directional alignment across the two largest pharmaceutical regulators. Minimum certificate-level specification remains open work — and the institutional will is now on record.

THE DEEPEST QUESTION THAT REMAINS

REGULATORY RULE-ENCODING GOVERNANCE

When jurisdiction-specific regulatory rules are encoded as computable patterns, who certifies that the encoding is accurate? Who bears accountability when an encoding error produces a certificate that is structurally valid but semantically wrong — one that passes all three gates?

No current regulatory framework assigns this responsibility. It is the deepest question this architecture surfaces — named here, not hidden.

THE INVITATION

The standard needs to be co-authored

The FDA-EMA principles call for one. The framework is open for it. The next cohort forms post-DIA. If you work in regulatory affairs, GCP compliance, or data standards — your perspective shapes V2.1. Anonymous polls, named contributions on the published framework, co-authorship on the peer-reviewed deliverable.

Scan to connect — let's talk at DIA, or after.

ACKNOWLEDGMENTS

With thanks to the practitioners who contributed to the May 14, 2026 Co-Development Working Session and subsequent discussion: Gourav Pandey, Principal Manager, R&D GMP Quality; Paul Hanson; Jim Taylor; Thane Carson; Irina Gontschar, MD, PhD; Jessica Stuyvenberg (ARCH Framework, Stuyvenberg Advisory Group). Individual contributions reflect personal expertise, not institutional endorsement. Technical Contributors: Yan Barros — Physics-Informed Neural Network Architecture.
Advisory: Graziela Parente Peduti — Brazil Regulatory Intelligence; Javier Granados II, PharmD — Clinical Research; Michelle Purdom, PhD — NexTrial BioTech Advisor.

Selected References ·
Assoc. of American Cancer Institutes (2018). Clinical Trial Operations Benchmarking Survey. · Shadbolt et al. (2023). Analysis of Rates of Completion, Delays, and Participant Recruitment in Randomized Clinical Trials in Surgery. JAMA Netw Open 6(1):e2250996. · Raissi et al. (2019). Physics-Informed Neural Networks. J. Comput. Phys. 378:686–707. · ICH E6(R3) (2025). Good Clinical Practice. · CFM nº 2.454/2026 (published Feb 2026; effective Aug 2026). · FDA 21 CFR Parts 11, 50, 54, 56, 312, 314. · ANVISA RDC 945/2024; Lei 14.874/2024. · CDSCO NDCTR 2019. · EU AI Act Reg. 2024/1689; CTR 536/2014. · ICON plc (2025). Global Site Survey. · FDA REdI Conference, May 2026.

Steven@NexTrial.ai • NexTrial.ai

Nextrial Github

NexTrial Substack