WISL_No72

Immutable

Hardware-Enforced

The Deterministic Complement — Non-Agentic Hardware Governance in the Age of Software AI Assurance

Why testing the mind of the machine is necessary, and governing its body is decisive

Edwin Koh
kohedwin.ai | non-agentic.ai
kohpapa@outlook.com

SG020603109STW · P-LIFE 1.00™ · NAI 4.0™
NLB Legal Deposit R260219-005 / R260302-007

Immutable™ — Hardware-Enforced Non-Agentic AI Governance Architecture — NAI 4.0™
WISL™ Ebook Series No. 72 (Proposed) — ISBN Edition Candidate

ISBN
978-981-17957-1-8

ISBN & Copyright

Title: The Deterministic Complement — Non-Agentic Hardware Governance in the Age of Software AI Assurance
Series: WISL™ Ebook Series No. 72 (Proposed) — ISBN Edition Candidate
Publisher: Edwin Koh
Author: Edwin Koh
Entity: ACRA T260229801
Parent Patent: SG020603109STW
NLB Legal Deposit: R260219-005 / R260302-007 (prior volumes); this volume pending deposit
ISBN: (to be assigned upon application)
Edition: First Edition, 2026
Format: eBook (PDF)
Language: English
Subject: Artificial Intelligence · AI Assurance · Non-Agentic Governance · Hardware-Enforced Safety · Public Policy
Classification: AI Governance · Embedded Systems · Regulatory Science · Cyber-Physical Safety

This monograph was developed with AI-assisted drafting tools. All architectural concepts, patent claims, constitutional frameworks, and governance terminology are the original intellectual work of the author. AI tools contributed prose drafting and document structuring only. Intellectual authorship, IP ownership, and research direction remain solely with Edwin Koh Wui Kiat (Tiger) under ACRA T260229801.

Design-Intent Declaration. All clinical and technical performance targets herein represent design-intent specifications pending verification, validation, and regulatory assessment. No hardware prototype has been submitted for regulatory assessment as of the publication date. References to third-party frameworks, schemes, and standards are descriptive; no endorsement by any authority is claimed or implied.

Abstract

The global AI assurance movement has produced an impressive apparatus for testing the behaviour of artificial intelligence: process frameworks, benchmark batteries, red-teaming protocols, management-system standards, and conformity assessments. This monograph argues that the apparatus, however rigorous, shares a single structural limitation: it tests *software with software*, and therefore inherits the mutability of the thing it audits. A model that passes every test today can drift, be updated, be compromised, or be re-prompted tomorrow; the certificate remembers a system that no longer exists.

For AI that remains purely informational, this residual risk is managed through monitoring and re-testing. For AI embedded in the physical world — robots, medical devices, vehicles, infrastructure actuators — the residual risk becomes kinetic. This monograph proposes the **deterministic complement**: a hardware-enforced, non-agentic governance layer that does not attempt to make the AI trustworthy, but makes the *consequences* of AI untrustworthiness physically impossible to execute. Drawing on the Non-Agentic AI Governance Engine (NAIGE) architecture — deterministic silicon gates, hardware-timed deliberation, privacy-by-physics sensing, multi-factor human authorization, physical power severance, and one-time-programmable forensic logging — it maps software assurance goals to their hardware complements, illustrates the composition through an eldercare monitoring case study, and outlines policy implications for multi-level assurance and labelling regimes. The thesis is deliberately modest and deliberately radical at once: software assurance answers the question *"is the advice good?"*; hardware governance answers the question *"what can the machine do when the advice is bad?"* A complete assurance stack requires both.

1. The Assurance Gap: Testing Software With Software

Every mainstream instrument of AI assurance shares an architecture: it examines a software artefact using other software artefacts, and issues a judgment valid at a point in time. Benchmarks measure a frozen checkpoint. Red-teaming probes a deployed configuration. Management-system certification audits documents and processes. Conformity assessment examines technical files. Each is valuable; none escapes three structural properties of software itself.

Mutability. Software changes after assessment — by design (updates), by adaptation (continued learning, retrieval over shifting corpora), or by attack. The assessed object and the operating object diverge from the moment of certification. Change-management programmes for machine-learning systems acknowledge this openly: they regulate the rate and pre-specification of change precisely because change is inevitable.

Reflexivity. The mechanism that tests, monitors, or constrains the model is itself software, running on the same substrate, exposed to the same failure classes — bugs, misconfiguration, supply-chain compromise, and adversarial manipulation. A guardrail model can be jailbroken; a monitoring agent can be starved of the signal it monitors; a logging service can be edited by whoever edits logs. Software oversight of software is a tower whose every floor stands on the floor below.

Probabilism. Statistical systems offer statistical guarantees. A hallucination rate can be reduced; it cannot, within the paradigm, be made zero. For informational outputs, low probability of harm may be acceptable and insurable. For kinetic outputs — a manipulator arm beside a patient, a vehicle in a crowd, a valve on a pipeline — the acceptable probability of an unauthorized action is not "low." It is zero, and zero is not a statistical number. It is a physical one.

None of this is an argument against software assurance. It is an argument about where software assurance ends. The gap it leaves is not a testing gap to be closed by better tests; it is an enforcement gap, and enforcement is a property of the physical layer.

2. The Software Assurance Landscape, Respectfully Surveyed

The past four years have produced a genuinely global assurance infrastructure. Singapore's AI Verify framework and its open-source foundation pioneered structured, testable AI governance and gave the world a working reference implementation of "trust, but verify" for AI systems, later extended toward generative models. The United States NIST AI Risk Management Framework organised the risk vocabulary — govern, map, measure, manage — that most corporate programmes now speak. ISO/IEC 42001 turned AI governance into a certifiable management system. The European Union's AI Act converted risk tiers into legal obligations with conformity assessment for high-risk systems. Sectoral regulators — including health authorities with software-as-a-medical-device lifecycle guidance and change-management programmes for machine-learning-enabled devices — translated these ideas into pre-market and post-market practice.

Singapore deserves particular attention for a second reason: it has already demonstrated, in an adjacent field, the policy instrument this monograph will later recommend. The Cybersecurity Labelling Scheme — first for consumer IoT, then for medical devices through a four-level scheme jointly developed by the cybersecurity, health and health-technology authorities — established that **multi-level, testable, label-based assurance of physical devices is administratively feasible** and internationally influential. The precedent matters: if device cybersecurity can be graded and labelled by depth of assessment, so can the depth of a device's *AI enforcement* architecture.

The survey's honest conclusion: the software assurance stack is necessary, maturing, and — for embodied AI — structurally incomplete. Every instrument above evaluates intentions, processes, and behaviour under test. None of them can physically prevent an out-of-specification actuation at the moment it is attempted. That is not their failure; it was never their job. It is, however, somebody's job.

3. The Deterministic Complement: Non-Agentic Hardware Governance

The Non-Agentic AI Governance Engine (NAIGE) architecture, developed within the Nightingale's Eyes Care™ (B1) research programme, proposes that for cyber-physical AI the

governance perimeter must descend from the software layer into silicon. Its organising principle is non-agency: the AI may observe, may analyse, may propose — it may never execute. Execution belongs to a deterministic hardware stratum whose behaviour is exhaustively enumerable, and ultimately to a human being whose authorization is a physical event, not a software flag.

Five properties define the complement, each answering one of the structural limitations of Section 1:

- **Determinism against probabilism.** Safety-critical paths run on a synchronous hardware clock domain through finite-state machines with no asynchronous bypass. Every state and transition can be enumerated and verified; there is no distribution to sample, hence no tail to fear. The design targets of the NAIGE reference architecture — a 1 kHz master clock, hardware-timed deliberation windows (Sacred Pause™), and hard kinematic envelopes — are design-intent specifications of this property.
- **Physical enforcement against mutability.** The enforcement layer is not software that could be updated into a different system; it is gates, comparators, and interlocks whose behaviour changes only if the silicon or the milled chassis changes. An assessed enforcement layer and an operating enforcement layer are the same object, indefinitely.
- **Privacy-by-physics against data risk.** Under the 3ZEROS™ principle (zero camera, zero audio, zero cloud), identity-bearing data is never captured: LiDAR point clouds are reduced in hardware to anonymised voxel occupancy and kinematic vectors before any software sees them. What is never collected cannot leak, cannot be subpoenaed, and cannot be hallucinated about.
- **Human sovereignty against reflexivity.** Consequential actions require multi-factor human authorization implemented as concurrent physical events — in the reference design, the Tiger .1x Key™ combining biometric, cryptographic, and kinetic factors. Oversight is thereby anchored outside the software tower entirely: the last floor stands on a human being.
- **Immutable accountability against auditability decay.** Events commit through cryptographic hashing and hardware sealing into write-once memory finalised by one-time-programmable fuses. The audit trail is not a file that privileged software can rewrite; it is a physical alteration of silicon. Inaction is the default safe state (P-LIFE 1.00™): on any fault, ambiguity, or breach, the system's guaranteed behaviour is mechanical stillness (Sovereign Brake™).

The philosophical reframing is the point. Software assurance asks the machine to be trustworthy and verifies the asking. Hardware governance declines the question: it renders the machine's untrustworthiness *inconsequential* within the operational boundary. A regulator auditing such a system is no longer auditing a company's promises about behaviour; it is auditing the physical consistency of an object — a task regulators of pressure vessels, aircraft structures, and electrical installations have performed successfully for a century.

4. Composing the Stack: Software Goals, Hardware Complements

The two paradigms are not rivals; they answer different questions and compose into a layered stack. The mapping below states each canonical assurance goal, the software mechanism that pursues it, and the hardware complement that closes it.

Assurance goal	Software mechanism (necessary)	Hardware complement (decisive)
Output quality / accuracy	Benchmarks; grounded retrieval; guideline anchoring; multi-agent cross-check	Advisory confinement: outputs are information only; zero actuation rights by construction
Robustness to attack	Red-teaming; adversarial testing; guardrail models	No software path to actuators exists to be attacked; breach response is physical power severance
Drift management	Change-management programmes; monitoring; re-validation	Drift cannot manifest kinetically: hard envelopes and deterministic gates discard out-of-bound proposals
Privacy & data protection	Encryption; access control; anonymisation pipelines	Privacy-by-physics: identity never captured at ingestion; nothing to protect downstream
Human oversight	Human-in-the-loop process requirements; approval workflows	Sovereign-in-the-loop: authorization as concurrent physical multi-factor events; permanent human veto
Accountability & audit	Logging; model cards; documentation; certification	OTP-fused write-once forensic ledger; record survives power loss, sabotage, and AI failure
Fail-safe behaviour	Graceful degradation logic; watchdog software	Inaction-as-safe-default: guaranteed mechanical stillness on any fault, enforced below software

Read column-wise, the table shows why neither layer suffices alone. Hardware cannot make advice good — a deterministic gate will faithfully deliver a clinician a wrong recommendation if the advisory stratum produces one; only software assurance improves the advice. Software cannot make execution safe — no test result restrains an actuator at 03:00 on a Tuesday; only hardware does. The composed claim is stronger than either: *assured advice, enforced consequences*.

5. Worked Example: Eldercare Monitoring

Consider the reference eldercare configuration of the B1 programme: a non-contact spatial monitoring apparatus (designated A13b) intended to detect falls, prolonged immobility, and bed-exit events in residential care, alerting human caregivers. The assurance stack composes as follows. At the software layer, event-detection analytics would be validated against test datasets, enrolled in a regulator's change-management programme if machine-learning components evolve, and assessed under software life-cycle standards. At the cybersecurity layer, a multi-level device labelling assessment — declaration of conformity, binary analysis, penetration testing, security evaluation — grades the device's resistance to compromise. At the hardware layer, the NAIGE properties hold regardless of the layers above: no camera or microphone exists to breach; the device has no actuators to hijack; alerts and faults commit to

a write-once ledger no intruder can edit; and the failure default is signalled fault, never silent suppression.

The composition is more than additive. The hardware layer materially *lowers the stakes* of the software layers: a compromised analytic model in this architecture can, at worst, delay or mis-rank an alert — a serious but bounded failure — because the physical design has removed surveillance capture, cloud exfiltration, actuation, and log tampering from the space of possible outcomes. Assurance effort can then concentrate where residual risk actually lives (alert sensitivity and specificity), rather than being spread thin across an unbounded threat surface. This concentration effect — hardware governance as a risk-surface reducer that makes software assurance tractable — is, the author submits, the deterministic complement's principal gift to the assurance community. All performance characteristics of the A13b configuration remain design-intent pending verification; the argument here is architectural, not empirical.

6. Policy Implications: Toward a Hardware-Enforcement Level in AI Assurance

Three implications follow for policymakers, with Singapore unusually well positioned to lead on each.

First, grade enforcement, not only behaviour. Existing assurance instruments grade what an AI system does under test. A complementary axis should grade what the surrounding device *can physically do at all*: Does the AI hold actuation rights? Are safety envelopes enforced below software? Is human authorization a physical event? Is the audit trail physically immutable? Is inaction the guaranteed failure state? These are inspectable, binary-ish properties of an object — precisely the kind of property that multi-level labelling schemes have proven able to assess and communicate. The four-level medical-device cybersecurity label jointly developed by Singapore's cybersecurity and health authorities is the natural template: a future assurance level, or an adjacent label, could certify hardware-enforced non-agentic governance.

Second, let regulators audit objects, not promises. Where enforcement is physical, conformity assessment converges with the mature discipline of hardware inspection: verify that the silicon, interlocks, and enclosure match the certified manifest, and the safety case follows from physics. Manufacturing pipelines with versioned, cryptographically bound design manifests make this practical: hardware that deviates from its certified manifest fails its own handshake. The regulator's question shifts from "do we believe the vendor's testing?" to "is this the object that was certified?" — a question with a deterministic answer.

Third, reserve the strong claim for the strong architecture. Terms like "safe," "human-controlled," and "privacy-preserving" are today applied indiscriminately to systems whose guarantees are statistical and mutable. Policy should protect a vocabulary tier — as it does for words like "sterile" or "fireproof" — for systems whose guarantees are physical. The non-agentic hardware-governance architecture defines one candidate criterion for that tier:

the claim "this AI cannot act without a human" should be reserved for systems in which that sentence is a description of circuitry, not of policy.

7. Honest Boundaries of the Thesis

Intellectual integrity requires stating what the deterministic complement does not do. It does not improve the quality of AI advice; a hardware gate delivers bad recommendations as faithfully as good ones, and the informational harms of AI — misinformation, biased analysis, manipulation — pass through untouched. It applies with full force only to embodied and cyber-physical AI; a chatbot has no actuators for silicon to govern, and the assurance of purely informational systems remains a software discipline. It constrains, and is meant to constrain, capability: a hardware-governed platform cannot pivot by update into autonomous operation, which is a feature for safety and a cost for flexibility that adopters must weigh openly. It shifts, rather than eliminates, verification burden: the hardware layer itself requires rigorous verification under functional-safety practice, and the claims of this monograph are architectural arguments pending exactly that verification — no prototype of the reference architecture has yet been built or assessed. And it depends on honest scoping: hardware enforcement guarantees hold within the designed operational boundary, and overclaiming beyond that boundary would repeat, in silicon, the very sin of overclaiming this monograph criticises in software.

8. Conclusion

The AI assurance movement has built the instruments to examine the mind of the machine. This monograph has argued that, for machines with bodies, examination is half the task: the other half is a body that cannot disobey. The deterministic complement — non-agentic, hardware-enforced governance in the tradition of the NAIGE architecture — supplies that half not by trusting artificial intelligence more cleverly, but by making trust unnecessary at the only layer where its absence is lethal. Software assurance answers *is the advice good?* Hardware governance answers *what happens when it is not?* A civilisation deploying embodied AI at scale will need confident answers to both questions. The first is being answered by a worldwide community. The second is answerable today, in silicon, one deterministic gate at a time.

AI Observes · AI Advises · AI Builds · Human Decides

Integrated Citation Register

Primary Citation — This Publication

Koh, E. W. K. (2026). The Deterministic Complement — Non-Agentic Hardware Governance in the Age of Software AI Assurance. WISL™ Ebook Series No. 72 (Proposed). Non-Agentic Governance Singapore (ACRA T260229801). Patent SG020603109STW. Available at: https://kohedwin.ai | https://non-agentic.ai

Framework and Scheme References (Descriptive)

- AI Verify testing framework and AI Verify Foundation — Infocomm Media Development Authority, Singapore.
- Cybersecurity Labelling Scheme for Medical Devices, CLS(MD) — Cyber Security Agency of Singapore with MOH, HSA and Synapxe; four-level voluntary scheme launched October 2024.
- NIST AI Risk Management Framework 1.0 — National Institute of Standards and Technology, United States, 2023.
- ISO/IEC 42001:2023 — Artificial intelligence management system standard.
- Regulation (EU) 2024/1689 — the EU Artificial Intelligence Act.
- HSA GL-04 (software medical devices life-cycle), GL-07 (SaMD/CDSS classification), GN-37 (change management for ML-enabled SaMD) — Health Sciences Authority, Singapore.
- IEC 61508 (functional safety), IEC 62304 (medical device software life-cycle), IEC 62366-1 (usability), ISO 14971 (risk management).

Patent Basis

- SG020603109STW — Parent Patent (constitutional core)
- 10202600902P — ABC+2S+H™ Guardian Framework
- 10202602021T — A10a Deterministic Governance Platform
- 10202602023X — A12 Deterministic Black Box
- 10202602213U — A7 Bridge Layer Translation Hub
- 10202602223T — A(12) P+N Enhanced WD117 Immutable Ledger
- 10202602225P — A8 Bridged Layer 8 Sovereign Deployment Governance Layer
- 10202602270V — Non-B1 Clinical Advisory Monitoring Shell
- 10202602087W — A13b Sensing Platform · 10202602143W — A13a Mobility Enclosure · 10202602205U — A13c Clinical Robotic Platform

NLB Legal Deposit

- R260219-005 — Primary constitutional knowledge base
- R260302-007 — Supplementary architectural filings
- Deposit Authority: National Library Board Singapore · edeposit.nlb.gov.sg

Academic Background — Edwin Koh Wui Kiat

– MBA — General and Strategic Management — Maastricht School of Management

– BSc — Industrial and Systems Engineering — University of Florida

– Independent Researcher — OCA Designation — NLB Vault Constitutional Anchor Route

止於至善

Finish in Highest Excellence　謙虛 · 沉默 · 尊严 · 仁　Humility · Silence · Dignity · Benevolence

AI Observes · AI Advises · AI Builds · Human Decides

Non-Agentic Governance Singapore · ACRA T260229801 · Patent SG020603109STW

www.kohedwin.ai · www.non-agentic.ai · kohpapa@outlook.sg

We Innovate Save Lives™ · We Go Global™