Non-Agentic AI Governance Singapore

WISL™ No. 54
Non-Agentic AI 2.0™

Non-Agentic AI Systems

Deterministic, Auditable, Human-Governed Intelligence Architecture

Edwin Koh Wui Kiat

ABSTRACT

This publication introduces a Non-Agentic AI framework designed to eliminate autonomous behaviour, enforce deterministic outputs, and ensure full auditability. The system is engineered for high-assurance domains including medical, regulatory, and safety-critical environments, addressing systemic risks associated with probabilistic and opaque AI systems.

Unlike conventional AI architectures that operate on probabilistic inference and permit emergent autonomous behaviour, Non-Agentic AI enforces zero self-directed intent, strict input-to-output determinism, full traceability of decision pathways, and human authority as the terminal control layer. This framework provides a structural response to the growing concern that current AI deployments overstate safety and controllability while lacking verifiable guarantees.

PART I — THE POLICY CONTEXT

1.1 Information Asymmetry in AI Governance

AI companies currently control architecture details, training data sources, and evaluation benchmarks of deployed systems. This creates a knowledge monopoly where governments regulate blindly and public trust is shaped by corporate narratives rather than independently verified evidence.

The policy implication is clear: governance frameworks must shift toward mandatory disclosure requirements, not voluntary transparency. Independent verification bodies — equivalent to financial auditors or pharmaceutical trial reviewers — are required to close the gap between claimed and actual system behaviour.

1.2 Incentive Misalignment

Corporate drivers — market dominance, investor expectations, first-mover advantage — conflict structurally with safety, ethical deployment, and long-term societal stability. Without liability structures that connect harm to financial and legal consequences, the incentive to overstate safety will persist.

1.3 The Narrative–Reality Gap

Public messaging emphasises safety and responsibility. Internal evidence and independent analysis consistently reveal unknown failure modes, scaling risks, and misuse vulnerabilities. The central risk is false confidence: not knowing what AI cannot reliably do, while claiming that we do.

1.4 Technical Risk Landscape

Risk Category	Description	Governance Implication
Alignment	Models still hallucinate, can be jailbroken, behave unpredictably. Alignment is probabilistic, not guaranteed.	Independent verification of alignment claims required
Evaluation Gaps	Benchmarks are narrow and gamed. Models can pass tests but fail in production.	Real-world deployment complexity must be tested
Data Opacity	Unknown biases, copyright exposure, embedded misinformation in training data.	Dataset lineage documentation and audit rights required
Emergent Behaviour	New capabilities appear unpredictably as models scale.	Mandatory post-market surveillance and incident reporting

PART II — DEFINITION AND DESIGN PRINCIPLES

2.1 What is Non-Agentic AI?

A Non-Agentic AI System is defined as:

> ***A computational system that performs bounded transformations on inputs without autonomous goal formation, operating under deterministic constraints, and producing fully traceable outputs subject to human validation.***

- **Does not initiate goals** — no objective function drives the system beyond its defined task boundary
- **Does not self-direct actions** — every action results from a predefined computational pathway
- **Operates within deterministic constraints** — identical inputs produce reproducible, verifiable outputs

2.2 Core Design Principles

2.2.1 Deterministic Execution

- Fixed pipelines with no hidden decision branches
- Reproducible outputs for identical inputs
- Version-locked models and parameters

2.2.2 Bounded Functionality

- Explicit scope definition — no open-ended reasoning
- Task-specific modules instead of general autonomy
- Hard constraints on output domains

2.2.3 Traceability and Auditability

- Every output linked to input data sources
- Transformation steps mapped and logged
- Time, version, and control parameters recorded for every operation

2.2.4 Human-in-the-Loop Authority

- Final decision authority rests with human operators at all times
- AI outputs are advisory, not authoritative
- Override mechanisms are mandatory — not optional

2.2.5 Privacy-First Architecture

- Local or edge processing where technically feasible
- Minimal data retention policies enforced
- Explicit data lineage tracking throughout the system lifecycle

PART III — SYSTEM ARCHITECTURE

3.1 Core Pipeline

The Non-Agentic AI system enforces the following end-to-end pipeline:

Input → Validation Gate → Deterministic Processing Engine → Verification Layer → Human Approval → Output → Audit Log

No step in this pipeline permits autonomous decision-making beyond predefined logic. Human approval is mandatory prior to execution of critical outputs.

3.2 Component Specifications

Component	FIG.	Function	Key Property
Input Interface	1, 2	Structured data ingestion, source authentication, format validation	Rejects malformed or out-of-scope inputs
Validation Gate	2	Rule-based filtering, scope enforcement, authentication	First line of defence against boundary violations
Deterministic Engine	3	Fixed logic pathways, no stochastic branching, version-controlled	Eliminates probabilistic variance from critical outputs
Verification Layer	4	Constraint checking, output boundary enforcement	Confirms output correctness before human presentation
Human Control Interface	5	Decision display, approval input, override and rollback	Mandatory gate — system cannot self-execute
Logging Module	6	Event logging, secure storage, audit retrieval	Immutable record of all inputs, outputs, and states
Cybersecurity Module	7	Encryption, access control, intrusion detection	Zero-trust architecture with isolated processing

3.3 Architecture Diagrams (FIG. 1–10)

Specifications in accordance with IPOS and WIPO patent drawing standards (black line, no shading, labelled components):

- **FIG. 1 — System Overview:** Input (100) → Validation Gate (110) → Deterministic Engine (120) → Verification Layer (130) → Human Control (140) → Output (150)

- **FIG. 2 — Input and Validation:** Input interface (200), Authentication unit (210), Scope filter (220), Rejection path (230)
- **FIG. 3 — Deterministic Engine:** Fixed logic modules (300), Rule execution layer (310), Version control unit (320)
- **FIG. 4 — Verification Layer:** Constraint checker (400), Boundary enforcement unit (410), Output validator (420)
- **FIG. 5 — Human-in-the-Loop Interface:** Display module (500), Decision input (510), Override control (520)
- **FIG. 6 — Logging and Audit System:** Event logger (600), Data storage (610), Audit retrieval interface (620)
- **FIG. 7 — Cybersecurity Architecture:** Encryption module (700), Access control (710), Intrusion detection (720)
- **FIG. 8 — Data Flow:** Input → Validation → Processing → Verification → Human → Output
- **FIG. 9 — Risk Control Mapping:** Hazard nodes linked to corresponding control modules
- **FIG. 10 — Deployment Architecture:** Local processing unit (1000), Optional cloud interface (1010), Isolation boundary (1020)

PART IV — SAFETY, RISK MANAGEMENT, AND REGULATION

4.1 Risk Control Mapping (ISO 14971)

Hazard	Cause	Control Mechanism	FIG.
Incorrect Output	Model ambiguity or edge case	Deterministic pipeline constraint	3, 4
Hallucination	Probabilistic inference	Rule-based output validation	3, 4
Misuse	Unauthorised access	Access control and scope limitation	7
Data Bias	Imbalanced training data	Curated, documented datasets	2
Cyber Threat	System exposure	Isolated processing and encryption	7
Overreliance	Automation bias	Human-in-the-loop enforcement	5

4.2 Regulatory Alignment

FDA / HSA — Medical Device Context

- Supports Essential Performance definition for SaMD
- Repeatability requirements satisfied by deterministic architecture
- Human factors validation aligned to FDA 510(k) and HSA Class B/C SaMD pathway

IEC / ISO Standards

- IEC 62304: Controlled software lifecycle with defined system boundaries
- IEC 60601-1-8: Alarm system integration with mandatory pause gate
- ISO 14971: Risk management framework fully mapped
- ISO/IEC AI Governance: Transparency, accountability, risk-based classification

EU AI Act

- Non-Agentic architecture satisfies High-Risk AI human oversight requirements
- Full auditability meets transparency obligations
- Mandatory human approval layer satisfies human-in-the-loop requirements

4.3 Verification and Validation

Pre-Deployment

- Deterministic reproducibility testing — identical inputs must produce identical outputs
- Boundary condition analysis — edge case and failure mode testing
- Human factors validation with simulated workflow testing

Post-Deployment

- Real-time monitoring of system behaviour and output integrity
- Mandatory incident reporting with root cause analysis
- Version-controlled updates with full regression testing before deployment

4.4 Cybersecurity Framework

- Zero-trust architecture — no implicit trust at any system boundary
- Local-first processing — minimal external data exposure
- Attack surface minimisation — isolated processing nodes
- Continuous vulnerability scanning and patching protocol

PART V — PATENT-STYLE CLAIMS

Claims presented in accordance with IPOS and PCT international patent application standards.

Claim 1 — Independent System Claim

A Non-Agentic Artificial Intelligence System, comprising:

1. An input interface configured to receive structured data;
2. A validation module configured to authenticate input data and reject out-of-scope inputs;
3. A deterministic processing engine configured to execute predefined computational pathways and produce reproducible outputs for identical inputs;
4. A verification module configured to enforce output constraints and validate output integrity;
5. A human control interface configured to require human approval prior to execution of critical outputs;
6. A logging module configured to record all inputs, outputs, and processing states;

wherein the system prohibits autonomous goal formation and self-directed actions.

Claim 2 — Independent Method Claim

A method for operating a Non-Agentic AI system, comprising:

7. Receiving input data;
8. Validating the input against predefined constraints;
9. Processing the input using deterministic computational logic;
10. Verifying output compliance with defined boundaries;
11. Presenting output to a human operator;
12. Requiring human approval prior to execution;
13. Logging all operational steps;

wherein no step includes autonomous decision-making beyond predefined logic.

Dependent Claims 3–10

- The system of Claim 1, wherein the deterministic engine is version-locked and produces identical outputs for identical inputs across system updates.
- The system of Claim 1, wherein outputs include traceability metadata linking each output to its source inputs and transformation steps.
- The method of Claim 2, wherein validation includes cybersecurity checks against known threat signatures.
- The system of Claim 1, wherein processing occurs on local hardware with no external network egress during critical operations.
- The system of Claim 1, wherein the human control interface enforces a mandatory delay between output presentation and approval capability.
- The system of Claim 1, wherein the logging module stores records in a tamper-resistant, write-once format.

- The system of Claim 1, wherein the validation module applies ISO 14971 risk classification to input data prior to processing.
- The method of Claim 2, wherein human approval is required from a minimum of two authorised operators for critical-class outputs.

Novelty Positioning

Distinguishing Feature	Non-Agentic AI System	Prior Art (Standard AI)
Autonomous goal formation	Explicitly prohibited by design	Permitted — core to agentic paradigm
Output determinism	Guaranteed for identical inputs	Probabilistic — not guaranteed
Human approval gate	Mandatory prior to execution	Optional or absent
Full audit trail	Immutable record of all states	Partial or absent
Failure mode	Inaction (system halts)	Degraded autonomous operation

PART VI — STRATEGIC IMPLICATIONS

6.1 For Developers

The shift from intelligence maximisation to control maximisation is the defining challenge of responsible AI development. Non-Agentic architecture provides a technically grounded path: bounded functionality, deterministic execution, and mandatory human authority.

6.2 For Regulators

Non-Agentic AI is significantly easier to certify than probabilistic agentic systems because its behaviour is predictable and its audit trails are complete. Regulators can apply established frameworks from medical devices directly to Non-Agentic AI deployment.

6.3 For Institutions

Institutions deploying Non-Agentic AI are protected by design. The system cannot execute consequential actions without explicit human authorisation. Liability is preserved at the human layer. Governance is built into the hardware and software architecture.

6.4 For Society

Non-Agentic AI reduces systemic risk by eliminating the conditions under which autonomous AI failure can propagate silently. The greatest risk is not advanced AI itself — it is unverified claims of safety and control. Non-Agentic architecture makes those claims verifiable.

CONCLUSION

Non-Agentic AI provides a structural correction to current AI development trajectories — from probabilistic to deterministic, from opaque to auditable, from autonomous to human-governed.

It directly addresses the central concern of contemporary AI governance: the greatest risk is not advanced AI itself, but unverified claims of safety and control. Non-Agentic architecture makes those claims verifiable through deterministic execution, mandatory human authority, and immutable audit trails.

As AI capability continues to advance, the boundary that matters most is not what AI can do. It is what AI is permitted to decide. Non-Agentic architecture makes that boundary physical, constitutional, and irreversible.

AI as instrument, not actor.

仁義禮智信 · 止於至善

謙虛·沉默·尊严·仁

REFERENCES

Health Sciences Authority (HSA). (2023). *Regulatory guidelines for software as a medical device (SaMD)*. Singapore. https://www.hsa.gov.sg

International Electrotechnical Commission. IEC 62304: Medical device software — software life cycle processes.

International Organization for Standardization. ISO 14971: Medical devices — application of risk management to medical devices.

Koh, E. W. K. (2026). *NAI 2.0™ Framework: Constitutional architecture for non-agentic AI governance*. Non-Agentic AI Governance Singapore. https://kohedwin.ai

World Intellectual Property Organization (WIPO). (2024). *PCT Applicant's Guide*. Geneva: WIPO.
