Cover Letter for the International Submission of G-OS

Submission Type: Initial Technical Dossier for International Standardization Review

To: OECD AI Office, EU AI Act Secretariat, UN/ITU AI Governance Division, ISO/IEC JTC 1/SC 42 Artificial Intelligence

Dear Review Committee,

I respectfully submit the attached dossier titled "G-OS: A Governance Operating System for Stable, Drift-Free, and Institution-Compatible AI Systems." This submission is provided as an *initial technical dossier for international review and standardization consideration.*

As the founder of the newly established research domain G-OS (Governance Operating System), I present this work as a structural framework that bridges administrative institutions and AI research. G-OS is a model-external governance architecture designed to ensure:

- institutional stability,
- drift-impossibility (achieved through strict causal separation and an externalized Value Layer),
- normative alignment with international governance frameworks, and
- reproducible causal separation between Value, Gate, Back, and Front layers.

This submission aims to contribute to global AI governance efforts, including the OECD AI Principles, the EU AI Act, the NIST AI Risk Management Framework, and emerging ISO/IEC standards for AI management and safety. The attached dossier includes the architectural definition of G-OS, its alignment with international governance requirements, and the production-phase implementation roadmap.

This initial submission does not include annex materials.

I remain committed to supporting the international community in establishing stable, institution-compatible AI governance.

Sincerely,

Yasuhiro Azuma
Founder, G-OS Research Domain
CEO, PEACE-MAKER Inc.
Tokyo, Japan

G-OS: Governance Operating System for Stable, Drift-Free, Institution-Compatible AI Systems

Yasuhiro Azuma

June 2026

Abstract

G-OS (Governance Operating System) reconstructs social and institutional decision-making as a single causal line:

Value → Back → Gate → Front

This dossier defines the G-OS architecture, its external Value Layer, its causal separation between exploration (Back), control (Gate), and output (Front), and its structural alignment with OECD AI Principles, the EU AI Act, ISO/IEC 42001, NIST AI RMF, and UN human-rights-based AI governance frameworks.

Contents

1 Executive Summary **3**

2 Background and Purpose **4**

3 G-OS Causal Architecture **5**

3.1 Value Layer 5
3.2 Back Layer 5
3.3 Gate Layer 5
3.4 Front Layer 5

4 Value Layer Construction **6**

4.1 International Minimum Core 6
4.2 National Extension 6
4.3 Immutable Snapshot 7

5 Value Layer Inheritance and Fractal Governance **8**

5.1 Inheritance Principles 8
5.2 Online Distribution and Updates 8
5.3 Drift Prevention 8

6 Integration Gate **9**
6.1 Role of the Gate . . . 9
6.2 Causal Safety . . . 9
6.3 Alignment with Standards . . . 9

7 Back Layer **10**
7.1 Exploration Under Constraints . . . 10
7.2 Non-Direct Output . . . 10

8 Front Layer **11**
8.1 Stable Output Zone . . . 11
8.2 Institutional Integration . . . 11

9 LLM Translation Layer **12**
9.1 Role . . . 12
9.2 Constraints . . . 12

10 International Alignment **13**
10.1 OECD AI Principles . . . 13
10.2 EU AI Act . . . 13
10.3 ISO/IEC 42001 and NIST AI RMF . . . 13
10.4 UN Human-Rights-Based AI Governance . . . 13

11 Use Cases **14**

12 Conclusion **15**

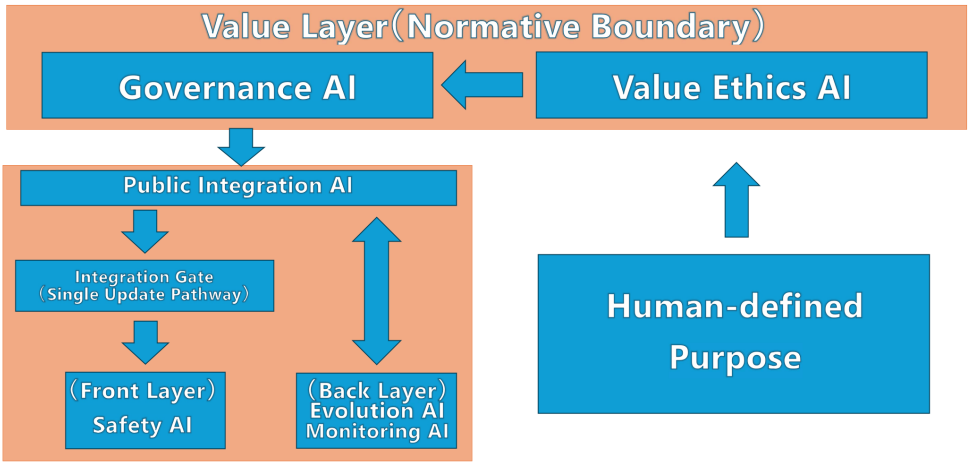
1 Executive Summary

G-OS is a model-external governance architecture designed to ensure institutional stability, drift-impossibility, normative alignment with international governance frameworks, and reproducible causal separation between Value, Gate, Back, and Front layers.

At its core, G-OS treats governance not as a collection of policies or ad hoc controls, but as a causal operating system whose behavior is determined by an externally fixed Value Layer and a single, auditable update pathway.

Figure 1 G-OS Overall Architecture

Layered roles and causal interactions within G-OS.



1

Figure 1: Causal structure of the Governance Operating System, showing the external Value Layer anchoring the Back, Gate, and Front layers in a single reproducible causal line.

G-OS is intended to be compatible with administrative institutions, AI research, and industrial deployment, providing a common governance basis for states, corporations, and international organizations.

2 Background and Purpose

Current AI governance practices are fragmented across ethics, law, technical controls, and institutional procedures. This fragmentation leads to:

- inconsistent application of norms across systems and organizations,
- difficulty in tracing responsibility and accountability,
- vulnerability to drift in objectives and behavior over time,
- misalignment between institutional mandates and AI system behavior.

G-OS introduces a governance OS that is:

- **model-external**: governing AI systems from outside their internal parameters,
- **institution-compatible**: anchored in existing legal and administrative structures,
- **causally explicit**: defined by a clear causal chain from Value to Front.

This dossier aims to:

- define the G-OS architecture and its layers,
- show how the Value Layer is constructed and inherited,
- explain the Integration Gate as the single update pathway,
- demonstrate alignment with international AI governance frameworks,
- outline a production-phase roadmap for institutional deployment.

3 G-OS Causal Architecture

G-OS is defined by a single causal line:

Value → Back → Gate → Front

Each layer has a distinct role:

3.1 Value Layer

The Value Layer is the external, human-approved governance basis. It encodes ethics, law, institutional mandates, and international norms as a stable, immutable snapshot.

3.2 Back Layer

The Back Layer performs exploration and reasoning under the constraints of the Value Layer. It may use multiple models, tools, and procedures, but its outputs are not directly applied to the world.

3.3 Gate Layer

The Gate Layer is the single update pathway. It evaluates Back Layer proposals against the Value Layer and either approves, rejects, or requests revision. No change to institutional behavior can bypass the Gate.

3.4 Front Layer

The Front Layer is the stable output surface. Only Gate-approved decisions and policies are propagated to institutions, systems, and society. The Front Layer is designed to be predictable, auditable, and robust.

4 Value Layer Construction

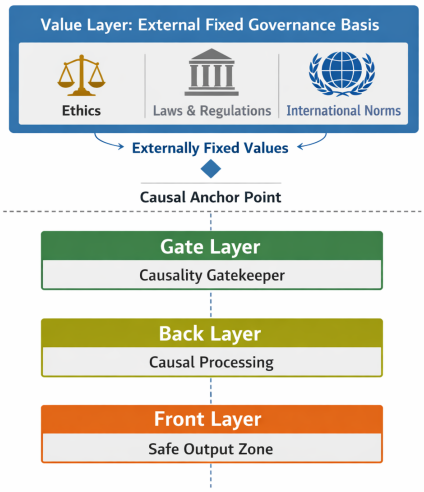


Figure 2: Process for constructing the external Value Layer, integrating international minimum norms, national extensions, and immutable human-approved snapshots.

The Value Layer is constructed through a human-in-the-loop process that combines:

4.1 International Minimum Core

A shared set of norms that can be accepted across jurisdictions, including:

- human-centricity,
- public purpose and societal benefit,
- long-term stability and sustainability,
- clear responsibility boundaries,
- externalization of uncertainty and risk,
- prevention of value drift.

4.2 National Extension

Each state adds its constitutional principles, legal frameworks, cultural values, and institutional mandates as extensions to the international core. These extensions must be explicitly documented, internally coherent, and compatible with the international minimum core.

4.3 Immutable Snapshot

Once reviewed and approved, the Value Layer is frozen as an immutable snapshot:

- changes require a formal human approval process,
- snapshots are versioned and auditable,
- AI systems and institutions refer only to approved snapshots.

5 Value Layer Inheritance and Fractal Governance

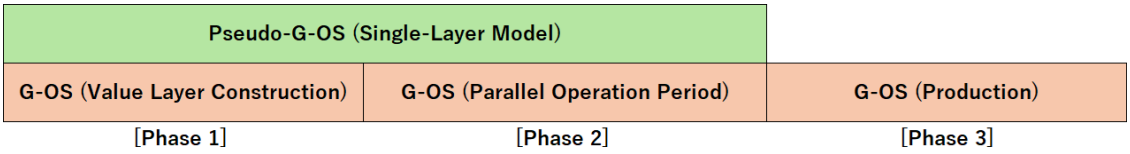


Figure 3: Three-phase transition from Value Layer construction to full production, ensuring drift-free operation through controlled inheritance and parallel validation.

The Value Layer is inherited along a fractal governance chain:

International → National → Corporate → Subsidiary → Contractor → Individual

5.1 Inheritance Principles

- Higher levels can constrain lower levels.
- Lower levels cannot modify higher-level values.
- Lower levels may add more specific constraints, but not override core values.

5.2 Online Distribution and Updates

- Initial distribution of Value Layer snapshots is performed via secure channels.
- Updates are propagated only after Gate-approved changes.
- Offline operation is supported by local copies of snapshots.

5.3 Drift Prevention

Because all layers inherit from a common external Value Layer, and changes must pass through the Gate, institutional and AI system behavior cannot drift away from the approved normative basis without explicit, auditable human intervention.

6 Integration Gate

The Integration Gate is the single update pathway between Back and Front, under the authority of the Value Layer.

6.1 Role of the Gate

- evaluates Back Layer proposals against the Value Layer,
- enforces consistency with institutional mandates and international norms,
- blocks or returns proposals that violate constraints,
- logs decisions for audit and accountability.

6.2 Causal Safety

By centralizing updates in the Gate:

- no uncontrolled change can reach the Front Layer,
- all changes are traceable to specific Value Layer references,
- safety and robustness are enforced structurally, not ad hoc.

6.3 Alignment with Standards

The Gate naturally aligns with:

- EU AI Act requirements for governance, oversight, and traceability,
- ISO/IEC 42001 clauses on operational control and risk mitigation,
- NIST AI RMF functions for govern and manage.

7 Back Layer

The Back Layer is the exploration and reasoning space that operates under the Value Layer.

7.1 Exploration Under Constraints

- multiple models and tools may be used,
- scenarios, simulations, and policy options can be generated,
- all exploration is constrained by the Value Layer.

7.2 Non-Direct Output

Back Layer outputs:

- are not directly applied to institutions or systems,
- must pass through the Gate for evaluation,
- are treated as proposals, not decisions.

8 Front Layer

The Front Layer is the stable output surface of G-OS.

8.1 Stable Output Zone

- only Gate-approved decisions reach the Front Layer,
- outputs are predictable and auditable,
- institutions can rely on the Front Layer as a stable interface.

8.2 Institutional Integration

Front Layer outputs can be:

- policies,
- operational procedures,
- system configurations,
- public-facing decisions.

9 LLM Translation Layer

The LLM Translation Layer is explicitly outside the core G-OS causal line.

9.1 Role

- translates between natural language and G-OS structures,
- assists in drafting Value Layer content and Back Layer proposals,
- supports explanation and communication of decisions.

9.2 Constraints

- cannot modify the Value Layer directly,
- cannot bypass the Gate,
- cannot write directly to the Front Layer.

10 International Alignment

G-OS aligns structurally with major international AI governance frameworks:

10.1 OECD AI Principles

- human-centricity via the Value Layer,
- transparency and accountability via the Gate and Front,
- robustness and safety via causal separation.

10.2 EU AI Act

- governance and oversight via the Gate,
- risk management via Back Layer exploration under constraints,
- documentation and traceability via snapshots and logs.

10.3 ISO/IEC 42001 and NIST AI RMF

- management system structure via G-OS layers,
- govern and manage functions via Value and Gate,
- monitor and improve via Back and Front interactions.

10.4 UN Human-Rights-Based AI Governance

- rights and dignity encoded in the Value Layer,
- institutional safeguards enforced by the Gate,
- public-facing behavior controlled at the Front.

11 Use Cases

G-OS can be applied to:

- administrative decision-making in public institutions,
- corporate governance and compliance,
- cross-border AI deployments requiring shared norms,
- supply chain governance and risk control,
- high-risk AI systems requiring strict oversight.

Each use case follows the same causal structure, with domain-specific Value Layer content and institution-specific Front Layer outputs.

12 Conclusion

G-OS defines a governance operating system based on a single causal line:

Value → Back → Gate → Front

By externalizing the Value Layer, enforcing a single Integration Gate, and separating exploration from output, G-OS provides:

- drift-free governance behavior,
- institution-compatible integration,
- structural alignment with international AI governance frameworks,
- a reproducible basis for AI system oversight.

This dossier is submitted as an initial technical basis for international review and potential standardization.

PEACE-MAKER – Alignment with International AI Governance Frameworks

PEACE-MAKER Structural Element	OECD AI Principles	EU AI Act	ISO/IEC 42001	UN/ITU AI Governance
Value Layer (Externally Fixed Purpose)	Human-Centricity / Accountability	Fundamental Rights Protection	Clause 4 – Context & Leadership	Human-Rights-Based AI
Unknown = Unknown (Uncertainty Externalization)	Transparency / Risk Awareness	Article 9 – Risk Management	Clause 8.3 – Risk Controls	Transparency & Explainability
Integration Gate (Single Update Pathway)	Traceability / Safety	Article 12 – Logging & Record-Keeping	Clause 8.2 – Operational Control	Auditability & Oversight
Front/Back Layer Separation	Robustness / Safety	Article 14 – Human Oversight	Clause 8.4 – Monitoring & Review	Safe Deployment Principles
Model-Free / Hardware-Agnostic Architecture	Future-Proof Governance	General-Purpose AI Requirements	Clause 4 – Scope Independence	Technology-Neutral Governance
Temporal Convergence (Long-Term Stability)	Long-Term Trustworthiness	Article 15 – Technical Safety	Clause 9 – Continual Improvement	Sustainable AI Governance

Note:
The alignment shown in this table was not intentionally designed to match these international AI governance frameworks. PEACE-MAKER was developed independently, without referencing OECD, EU AI Act, ISO/IEC 42001, or UN/ITU documents.

Interestingly, the resulting structure naturally satisfies the core requirements of all major global AI governance frameworks. Whether this convergence is a coincidence or a structural inevitability is unclear even to me.

I suspect that my 20 years of experience in Japan's trust-bank system development — which emphasizes externally fixed purpose, single update pathways, strict separation of testing and production, non-speculative handling of unknowns, and long-term operational stability — may have shaped the Value Layer of PEACE-MAKER in a way that naturally aligns with global governance principles.

Figure 4: Structural alignment of G-OS with OECD AI Principles, EU AI Act, ISO/IEC 42001, NIST AI RMF, and UN human-rights-based AI governance frameworks.