

Executive Summary for International Standards

(G-OS / PEACE-MAKER Governance Operating System)

1. Purpose and Scope

G-OS / PEACE-MAKER is a **governance operating system** designed to ensure **drift-free, transparent, and institution-compatible AI behavior** across international, national, organizational, and individual layers.

It provides a **model-free, hardware-agnostic governance architecture** that can be applied to any AI system while preserving **human command authority** as the final decision point.

The system is designed to align naturally with major global AI governance frameworks, including:

- **OECD AI Principles**
- **EU AI Act (High-Risk AI Governance)**
- **ISO/IEC 42001 (AI Management Systems)**
- **UN/ITU AI Governance Frameworks**

2. Fundamental Principle (Non-Negotiable)

Human command authority is absolute and cannot be delegated. AI systems may propose, evaluate, and present information, **but only humans make binding decisions.**

This principle prevents escalation, drift, or collapse of governance—even under superhuman AI capability.

3. Unified Causal Line (Core Architecture)

G-OS establishes a **single, immutable causal line**:

Value → Back → Gate → Front → Human

All governance-relevant behavior must be derivable along this line. No bypass, shortcut, or parallel path is permitted.

This structure ensures:

- **Full auditability**
- **Traceability of every output**
- **No internal value drift**
- **Single update pathway**
- **Human-aligned behavior at scale**

4. Externalized Value Layer (Fixed Normative Anchor)

All normative content—laws, ethics, policies, international principles—is externalized into a **Value Layer** that is:

- **Human-defined**
- **Institution-governed**
- **Read-only for AI components**
- **Versioned and auditable**

This prevents internal value drift and enables **scale-free inheritance** across jurisdictions.

5. Layer Roles and Responsibilities

Back Layer — Exploration Without Normativity

Generates proposals, scenarios, and options without embedding values.

Gate Layer — Single Integration Gate

Evaluates all Back proposals against the Value Layer. Ensures **normative consistency**, **single-point evaluation**, and **full logging**.

Front Layer — Stable, Predictable Output

Expresses only Gate-approved outputs in institution-compatible formats.

Human Phase — Final Decision Authority

Humans interpret Front outputs and make binding decisions.

6. Operational Phases (Migration → Parallel → Production)

G-OS defines a **three-phase operational lifecycle**:

1. **Phase 1 — Migration Requirements** Establishes semantic and operational consistency across Production, DR, and Pseudo-G-OS.
2. **Phase 2 — Parallel Operation** Production and Pseudo-G-OS run simultaneously to verify causal and value consistency. A **Front Consistency Score** (e.g., $\geq 90\%$) determines readiness for deployment.
3. **Phase 3 — Production Transition** System becomes operational only after:
 - Zero inconsistencies
 - Verified Value Layer alignment
 - DR readiness
 - Human Phase approval

7. Pseudo-G-OS (Single-Layer Model)

Pseudo-G-OS is a **continuously running mirror environment** used for:

- Parallel consistency verification
- Transparency and auditability
- DR fallback
- Communication-loss operation
- Emergency elevation to Production (Human-approved)

This mechanism is unique globally and provides **unprecedented operational resilience**.

8. Incident Response (Normal & Major Incidents)

Normal Incidents

Follow the causal line in reverse for safe recovery:

Front → Gate → Back

Major Incidents (DR Switching)

- DR1 → DR2 → DR3
- If communication is lost: **Pseudo-G-OS is elevated to Production via Human Phase**
- Full synchronization after recovery

This structure satisfies **ISO 22301 (Business Continuity)** and exceeds typical DR/BCP requirements.

9. International Alignment

G-OS naturally satisfies the core requirements of:

- **OECD**: human-centricity, transparency, accountability
- **EU AI Act**: risk management, logging, human oversight
- **ISO/IEC 42001**: governance processes, operational control
- **UN/ITU**: human rights, sustainability, interoperability

This alignment was not engineered to match these frameworks— it **emerged structurally** from the Value Layer and Gate architecture.

10. Contribution to Global AI Governance

G-OS / PEACE-MAKER provides:

- A **drift-free governance OS**
- A **single causal architecture** applicable across jurisdictions
- A **transparent, auditable, and human-aligned AI governance model**
- A **scalable operational framework** for governments, corporations, and international bodies
- A **model-free, hardware-agnostic blueprint** for long-term AI safety

It is the first system to unify:

- **Externalized normativity**
- **Single update pathway**
- **Parallel causal verification**
- **Human-final governance**
- **Operational DR/BCP integration**

into a single coherent governance OS.