NOTE: To refresh the Table of Contents, right-click it in Word and choose "Update Field" → "Update entire table."

Human Intelligence Leadership AI Agent Governance Framework (HIL-AGF)

Proposed framework for AI Agent's Integrated Governance System

Contents

1. Canonical Framework Definition

The **Human Intelligence Leadership Agent Governance Framework (HIL-AGF)** is a discernment-first governance framework designed to ensure that AI agents operate strictly as delegated instruments of execution, never as substitutes for human leadership.

HIL-AGF preserves explicit human accountability by requiring a named human owner for every agent, enforcing bounded agency across functional, contextual, temporal, and impact dimensions, and embedding mandatory discernment gates for actions involving human, ethical, reputational, or irreversible consequences.

The framework mandates full explainability, traceability, and auditability of agent actions, alongside continuous revocation and recovery controls that prioritize human authority over automation efficiency. HIL-AGF intentionally rejects silent autonomy, outcome-only optimization, and language that obscures responsibility, establishing a clear accountability chain from human intent to outcome.

---

2. Board-Level Governance Policy

Foundational Principle (Non-Delegable)

Authority may be delegated. Accountability may not.

Purpose

This policy establishes the governance framework for the design, approval, deployment, operation, and oversight of AI agents under the Human Intelligence Leadership™ Agent Governance Framework (HIL-AGF).

Scope

This policy applies to all AI agents operating within or on behalf of the organization that demonstrate initiative, persistence, or goal-oriented behavior.

The HIL Five Pillars of Accountable Agency

1. **Human Authority & Accountability** — Every agent has a single, named human owner who remains accountable for outcomes.
2. **Bounded Agency** — Agent authority is explicitly constrained across functional, contextual, temporal, and impact dimensions.
3. **Discernment Gates** — Mandatory human review is required for high-impact, ambiguous, or irreversible actions.
4. **Explainability & Traceability** — All actions must be logged, auditable, and explainable in plain language.
5. **Revocation & Recovery** — Immediate shutdown, authority revocation, and recovery procedures must be available.

Language & Cultural Rules

Language that obscures accountability is prohibited. Phrases such as "the system decided" or "AI made the decision" constitute governance violations. Required framing must explicitly retain human accountability.

Enforcement

Violation of this policy may result in immediate suspension of the agent, revocation of delegated authority, and appropriate disciplinary action.

---

3. Conceptual & Architectural Model

Core Premise

Agents may execute intent. Humans must retain discernment, accountability, and moral ownership.

HIL Bounded Agency Doctrine

Agent authority must be constrained across four dimensions: - **Functional** — What actions the agent may take - **Contextual** — In which situations those actions are permitted - **Temporal** — The duration of authorization - **Impact** — The maximum acceptable consequence

Authority expires by default unless explicitly reaffirmed.

HIL Discernment Gates

Mandatory human intervention is required for actions involving: - Human wellbeing, dignity, or employment - Ethical ambiguity - Reputational or regulatory risk - Irreversible or large-scale consequences - Power asymmetry (one-to-many impact)

HIL Accountability Chain

Intent → Delegation → Execution → Outcome → Accountability

Breaking the chain at any point constitutes a governance failure.

HIL Agency Test

Before approving any agent, ask: - Who is accountable if this causes harm? - Can this decision be explained to those affected? - Is there a human pause where discernment matters? - Can the agent be stopped immediately? - Would we publicly defend this decision?

If any answer is unclear, the agent must not be deployed.

---

4. HIL-AGF Agent Approval Checklist (Mandatory)

Accountability & Ownership

- Named human owner
- Escalation path defined
- Ownership documented and approved

Purpose & Intent

- Clear objective stated
- Value created beyond automation for automation's sake
- Alignment with organizational values
- Clear success and failure criteria

Bounded Agency

- Permitted actions defined
- Prohibited actions defined
- Operating contexts approved
- Authorization duration specified
- Maximum impact threshold defined

Discernment Gates

- Human-impact actions identified
- Ethical ambiguity scenarios identified
- Irreversible actions gated
- Human confirmation required where appropriate

Explainability & Transparency

- All actions logged
- Decisions explainable in plain language
- Audit trail available
- Stakeholder explanation feasible

Revocation & Control

- Kill switch tested
- Authority revocation process documented
- Recovery / rollback plan in place
- Incident review protocol defined

Final Affirmation

- This agent enhances human discernment in accordance with HIL-AGF
- This agent does not obscure or displace human accountability
- This agent complies with the proposed Human Intelligence Leadership AI Agent Governance Framework (HIL-AGF)

Approved by: _____

Date: _____

Notice, Scope, and Disclaimer

The Human Intelligence Leadership™ AI Agent Governance Framework (HIL-AGF) is provided for informational, educational, and leadership development purposes only.

This framework expresses governance principles, leadership perspectives, and conceptual models intended to support thoughtful decision-making regarding the design, oversight, and use of AI agents. It does not constitute legal, regulatory, technical, compliance, cybersecurity, risk management, or other professional advice.

The interpretation, adaptation, and implementation of this framework — including any decisions made, systems deployed, or actions taken based on it — remain solely the responsibility of the adopting organization and its leadership. Users are expected to exercise independent discernment and to seek appropriate professional advice relevant to their specific context.

Innovar Consulting Corporation (ICC) and the framework's author make no representations or warranties, express or implied, regarding outcomes, effectiveness, regulatory compliance, or fitness for any particular purpose.

To the maximum extent permitted by applicable law, Innovar Consulting Corporation shall not be liable for any direct, indirect, incidental, consequential, or special damages arising from or related to the use, misuse, interpretation, or application of this framework.

Use of this framework does not imply endorsement, certification, or validation by Innovar Consulting Corporation unless expressly agreed in writing.