MEG-Initiative.org

MEG 2: Legal Governance Framework
Executive Technical-Legal Brief

Adrian Stan | ORCID 0009-0003-1457-5155 | meg-initiative.org

The Problem: The Accountability Gap

The next major AI governance failure may not begin with a hallucination.
It may begin with an action.
An AI agent sends a communication. Changes a customer record. Approves a workflow step. Modifies code. Blocks a transaction. Requests a payment. Negotiates with another agent.
The system may have performed exactly as designed. Yet the organization may still be unable to answer the most fundamental question:
Who authorized the machine to take that action?
This is the **accountability gap**.
Traditional product liability remains the right instrument for defective AI products - and the revised EU Product Liability Directive (2024/2853) reflects that. But product liability is built for a static, predictable product. It does not fully cover a genuinely autonomous decision that no one explicitly programmed or anticipated. The harm exists, but the legal system has no clear attachment point.
The result: AI agents are practically uninsurable at scale. Enterprise adoption in high-stakes domains slows. And when something goes wrong, everyone points at everyone else.
MEG 2 fills that gap. It does not replace product liability. It adds what product liability cannot provide: a portable legal identity for the agent itself, a graduated liability regime, and a verifiable chain of accountability.

The Solution: A Portable Liability Layer

MEG 2 is a **voluntary, modular protocol** - adopted like a technical standard, not imposed by regulation. It provides a portable legal identity and liability layer for AI systems, from simple classifiers to fully autonomous agents.
Its practical force comes from the market:
- **Insurers** may require it before writing a policy.
- **High-value access nodes** (financial platforms, critical infrastructure) may condition access on it.
- **Contracts** may incorporate it by reference.
- **Regulators** may adopt it as a recognized conformity standard.

This is the same discipline-through-adoption logic that underlies voluntary conformity frameworks - extended with a verifiable layer and a functioning reference implementation.

MEG 2 and the Singapore Model AI Governance Framework for Agentic AI v1.5 are complementary, not competing. The MGF provides the governance dimensions (what good agentic governance requires). MEG 2 provides the legal mechanisms beneath them (how liability attaches, how identity persists, how cross-border enforcement works). It answers the three questions the MGF explicitly leaves open: dynamic agent identity, delegation chains, and multi-agent liability.
MEG 2 depends on MEG 1 as its technical evidence layer - the Audit Log, EoB, DAI, ISR, DEA, EFR, delegation headers and MEG Address. MEG 2 does not replace that layer. It assigns **legal effect** to the evidence it produces.

Key Mechanisms

Mechanism	What it does
MEG Address	A portable legal identity that follows the agent wherever it runs - independent of the server, persistent across infrastructure migrations. Like a ship's flag registration: the agent carries its legal identity with it.
Liability Container	A functional legal construct that holds obligations, procedural standing, and a dedicated guarantee. It makes no claim about the agent's consciousness or moral status - it simply answers: *who is responsible for what this agent does?*
Earned Autonomy (DEA + DAI + ISR)	Liability is graduated, not binary. As demonstrated autonomy increases, liability moves from developer → operator → the agent's own guarantee. This addresses the over-deterrence that contributed to the EU AI Liability Directive's withdrawal.
Architectural Human Confirmation	A human gate that cannot be bypassed - not a textual instruction the agent can ignore. The distinction is critical: "do not delete production data" is a textual instruction; a technical permission block is architectural confirmation. Only the latter constitutes a valid transfer of diligence.
Ethical Flight Recorder (EFR)	Captures internal state vectors (**not conversation content**) at the moment of a major incident. Access requires dual authorization. It transforms post-incident liability disputes from credibility contests into objective technical audits.
Flag-State Jurisdiction + Action in Rem	The agent declares an operational jurisdiction through its MEG Address, enabling cross-border enforcement. Direct action against the agent's insurance is available when the human operator is unreachable.

Who Responds for What

MEG 2's liability schema is **not cumulative**. Each actor responds for the component attributable to them - not jointly and severally for all harm caused by the system.

Actor	Responds for	Condition
Developer / Producer	System defect - errors of the base model or infrastructure	Primary; liability recedes as autonomy rises (Level 2 → 3)
Operator	Deployment, permissions, integration, lack of control	Primary for operational decisions; aggravated by omission of architectural safeguards
User	Confirmed decisions only	Exclusively after informed architectural confirmation - not mere textual acknowledgement
Attacker / Hijacker	Unlawful takeover of control (including prompt injection)	Full liability; operator of good faith exonerated
Agent (MEG Address)	Own limited liability at Level 3 only	Backed by liability insurance; cascades to reinsurance and sectoral guarantee fund

A 90-Day Path to Defensible AI

MEG is designed for incremental, practical adoption. You do not need to implement everything at once.

Phase	What to do
Days 1–30: Discover	Identify which AI systems in your organization can act, not just output. Map their identities, owners, tools, and permissions.
Days 31–60: Bound	Assign MEG Addresses to material systems. Define their compliance level (N1/N2/N3). Attach the guarantee (liability insurance).
Days 61–90: Prove	Implement the Audit Log and EFR. Test the architectural confirmation mechanism. Verify against the Reference Prompts Registry.

The objective is not to eliminate autonomy. It is to **make autonomy governable**.

The Board-Level Test

Every board should be able to answer these questions:

1. Which AI systems in our organization can take actions, not only produce outputs?
2. Does every material agent have a verifiable legal identity (MEG Address)?
3. Are prohibited actions technically blocked, or merely described in policy?
4. Can we reconstruct any material action from identity to outcome?
5. Can we prove who authorized the machine to act?

If you cannot answer these questions, your governance has not yet caught up with your deployment.

Economic and Regulatory Impact

For	What MEG 2 offers
Insurers	An actuarial anchor - a persistent legal identity with continuous behavioral metrics (DAI, ISR). Munich Re/HSB's March 2026 launch of a commercial AI liability product confirms the market is moving in this direction.
Developers	As demonstrated autonomy increases, the victim gains an additional, solvent, directly actionable debtor (the agent's own guarantee) - without losing existing actions against the developer for defects or against the operator for negligent deployment. The agent's own guarantee absorbs operational liability at Level 3, protecting the parent corporation from systemic exposure.
Operators	A first-class liability tier with a concrete attachment rule - identity binds to the deployment, not the reproducible template. This directly answers the delegation-chain problem that existing frameworks leave open.
Regulators	A bottom-up protocol that works across borders without requiring global political consensus. The retreat of comprehensive mandates (EU AI Liability Directive withdrawn; Colorado's risk-based AI Act repealed) reinforces the case for a voluntary, verifiable protocol that does not depend on legislative consensus.

Current Status and Adoption Path

MEG 2 is ready for modular adoption.

- The technical specification (**MEG v5.0**) is published on Zenodo (DOI: 10.5281/zenodo.21280679) and at meg-initiative.org.
- The legal governance framework (MEG 2) is published under CC BY 4.0 (DOI: 10.5281/zenodo.21280675), together with Annex A (Technical Concepts Glossary), Annex B (Liability Comparison Matrix across EU, Singapore and MEG 2) and Annex C (Level Summary Table).
- A working reference implementation generates and verifies MEG Addresses as W3C did:web identifiers at registry.meg-initiative.org.

We propose MEG 2 be considered as an **implementation-layer companion standard** to the Singapore MGF v1.5, translating governance principles into verifiable identity, graduated liability, and cross-border enforceability.

One Last Thought

The AI governance debate has spent years asking whether artificial intelligence is trustworthy.
Agentic AI forces a more concrete question:

Has your organization created a trustworthy system for delegating machine authority?

The decisive distinction of the agentic era will not be between organizations that use autonomous AI and organizations that do not. It will be between those that can **defend the authority they delegated** and those that discover that authority only after something has gone wrong.

The enterprise may automate the action. It cannot automate away its accountability.

MEG 2 is the infrastructure that makes that accountability verifiable.