MEG-Initiative.org

MEG2: Legal Governance Framework

***Motto**: Ethics becomes real when it can be implemented.*

Preamble

This normative framework for functional legal personality and liability allocation (**MEG2)** constitutes the legal governance layer that operationalizes the **Minimal Ethical Governance** (MEG - hereafter referred to as MEG1, for distinction) technical framework for autonomous (agentic) artificial intelligence systems. MEG2 does not replace national or regional legislation; it provides a portable layer of liability allocation and identity, which jurisdictions can adopt modularly. The scope is limited to the operational and civil liability of current agentic systems; the delimitations being set out in the final chapter.
The framework includes only mechanisms that can be implemented with the existing legal infrastructure or by expanding existing institutions. Elements of prospective research or long-term visions remain in the MEG1 framework document and are not the subject of this manual.
MEG2 is a normative proposal, not a legal instrument with direct effect. Its adoption is voluntary and modular (Chapter 8). Provisions marked as **prospective** become operational only after the conditions specified in the text are met; until then, they have indicative status. Implementation requires adaptation to local law.

Chapter 1: The accountability vacuum in agentic systems

1.1. Inadequacy of product liability

Statement. The product liability regime ceases to cover an artificial intelligence system when it makes autonomous decisions unanticipated by its manufacturer.

Rationale. Product liability is based on a premise: the product has a defined behavior, foreseen by the manufacturer, and deviation from this behavior constitutes a defect attributable to the manufacturer. This premise works for an object with deterministic behavior: a household appliance, a car, a medical device. An agentic system, however, violates the premise: it produces non-deterministic decisions, adapted to the context, that the manufacturer did not specify and, in many cases, could not anticipate. When such a system causes damage through a decision that no one explicitly programmed, the causal chain between the manufacturer and the damage is interrupted: the manufacturer can reasonably demonstrate that it did not foresee and did not intend that behavior. A vacuum results: the damage exists, but the classical regime has no one to attribute it to with certainty. This vacuum deepens as the decision-making autonomy of the systems increases.

Rule. For systems that produce autonomous decisions not explicitly anticipated by the manufacturer, liability cannot be allocated solely on the basis of product liability. A distinct allocation regime is required, based on the degree of effective autonomy and on the identification of the actor exercising decisive control over the decision.

1.2. The need for an intermediate legal subject

Statement. The clear allocation of liability for an agentic system requires a distinct legal point of attachment, which is neither exclusively the producer nor exclusively the user.

Rationale. The law already knows this solution and has been applying it for centuries. When an economic activity involves a risk that neither the owner of the capital nor the third-party contractor can fully and directly bear, the law interposes a distinct legal entity, the moral person (the commercial company), to which rights, obligations and a guarantee patrimony of its own are attached. The shareholder is not unlimitedly liable for the deeds of the company, and the third party still has an identifiable and solvent entity to which it can address itself. The interposition is not made because the entity "deserves" a status, but because the allocation of liability in a complex system requires a stable point of attachment. The agentic system raises exactly the same structural problem: between the producer who created it and the user who uses it, the autonomous action of the system needs its own point of imputation, so that liability does not dissipate between the two.

Rule. The framework establishes an intermediate functional legal subject, to which is attached, under the conditions and at the levels defined in this document, a limited legal capacity and a liability regime of its own, distinct from that of the producer and the user.

1.3. Functional legal personality, without ontological claim

Statement. The subject of law established by the present framework is a functional legal fiction, based on legal individuation, and not on any statement regarding the conscience or moral status of the system.

Rationale. The debate on the possible consciousness of artificial intelligence systems is open, unresolved and, for the purposes of allocating responsibility, inconclusive. An operational legal framework cannot condition its functioning on the resolution of a philosophical problem. However, law has an instrument that completely bypasses this debate: the legal personality of non-human entities. A commercial company is a legal person (it concludes contracts, is financially liable, stands trial) without anyone attributing to it a conscience, its own will or moral status. Its personality is a useful fiction: a construction through which the law attaches rights and obligations to a point of imputation, so that legal relations can function. MEG2 applies exactly this type of personality to agent systems: functional, instrumental, based on the fact that the system can be individuated and legal consequences can be attached to it, not on the fact that it “feels” or “is” something. The legally relevant question is not "does it feel?", but " **is it responsible for?** ".

Rule. The legal personality recognized by this framework is purely functional. It does not constitute and cannot be interpreted as recognition of the conscience, sentience or a moral status of the system. The basis for attaching liability is legal individuation, the capacity to identify a persistent entity to which consequences can be attributed, under the conditions defined in the chapter on the threshold of individuation.

1.4. Normative orientation

Statement. This framework is a normative proposal for an alternative architecture, not a description or prediction of the currently dominant regulatory trajectory.

Rationale. The direction observable in current regulation is, in several major jurisdictions, towards a model in which the AI system remains a strictly traceable object, with a permanently identifiable human or corporate actor behind it, and instruments of legal personality for the system are avoided or withdrawn. This model of controlled traceability is coherent with the state interest in control and easier to impose in the short term. MEG2 proposes a distinct architecture: one in which liability is allocated gradually, according to real autonomy, through decentralized and mutual recognition, rather than through centrally imposed traceability. This choice is assumed as a normative option (a proposal about how liability *should be* structured) and not as an observation about the direction that current regulators are de facto taking.

Rule. The provisions of this framework apply as a proposed standard, offered for voluntary and modular adoption. The framework does not claim to reflect the regulation in force and does not condition the legal validity of any system; it defines the conditions that a system meets to be recognized as compliant with this standard. The existing regulatory frameworks described in Chapter 2 are presented as contextual references, without constituting a validation of the proposed solution. MEG2 is a normative proposal, compatibility with existing regulations being only a finding of compatibility.

1.5. Objective

Statement. The framework aims to provide an interoperable infrastructure for identity, solvency and certain allocation of liability, for the entire spectrum of agent systems, covering both economic and non-economic harm.

Rationale. Some recent literature and practice treat the liability of artificial intelligence systems mainly through the lens of economic transactions: solvency, contracts, patrimonial damages. This perspective is necessary, but insufficient. An agentic system can produce damages that do not take the form of a transaction and are not fully patrimonially repaired: misinformation with social impact, harm to vulnerable people, erosion of individual or collective capacities, illicit content. A liability framework that covers only the economic dimension would leave uncovered precisely the segment in which the damage is diffuse and difficult to repair. Therefore, the proposed infrastructure treats liability in a broad sense, and the financial guarantee mechanisms (insurance) cover the patrimonial component, while the attribution, proof and suspension mechanisms cover the non-patrimonial component.

Rule. The framework provides: (a) a portable agent identity system, distinct from the technical identifier of the substrate; (b) a classification of levels of legal capacity and attachment of liability, commensurate with autonomy; (c) mechanisms of guarantee for pecuniary damage; (d) mechanisms of proof, attribution and suspension for non-pecuniary damage. These components are defined in the following chapters.

Chapter 2: Regulatory status and framework positioning

This framework does not appear in a vacuum. The issue of agentic liability is addressed, as of June 2026, by an evolving regulatory framework: national governance frameworks, emerging legal doctrines, and technical initiatives, each addressing a part of the issue. This chapter situates the framework in relation to the current regulatory dynamics, discussing in detail the initiatives directly relevant to agency accountability, briefly the major regulatory frameworks as a demarcation of context, and by reference the set of principle frameworks, the alignment analysis of which already appears in the MEG1 technical framework. The aim is not to demonstrate full concordance, but to precisely delineate its own contribution.

2.1. Initiatives directly relevant to agency liability

Explicit agent governance. The closest benchmark is the governance framework for agent-based artificial intelligence adopted by the Singapore regulator, launched in January 2026 and revised in May 2026. It is the first framework designed specifically for agents capable of planning, reasoning, and autonomous action. It is structured around four dimensions: proactive risk assessment and mitigation, human accountability, technical controls, and end-user accountability. Compliance is voluntary, but organizations remain legally liable for the behavior of their agents. The May 2026 revision extended the framework to multi-agent systems, third-party agents, and automation bias. This framework is the closest benchmark to the current approach in terms of purpose and philosophy: **accountability-based governance over direct control.** However, it leaves open, by its very structure of recommendation, three issues that it names but does not technically resolve: the dynamic identity of the agent, delegation chains, and the allocation of liability in multi-agent systems. These are exactly the elements that the present framework treats as mechanisms: portable identity (Chapter 3), the propagation of responsibility along the composition and delegation chain (Chapters 4 and 6), and the gradation of liability attachment by levels (Chapter 5). The relationship between the two frameworks is one of complementarity: one provides the governance dimensions, the other provides the legal mechanisms of identity and allocation.

The doctrine of legal personality applied to systems. A line of legal research has explored the extension of legal personality of non-human entities to automated systems. The divisible personality model, proposed by Alicia Lai (*Artificial Intelligence, LLC: Corporate Personhood as Tort Reform,* 2020), starting from the structure of the limited liability company, proposes treating the automated system as an entity with a limited patrimony and a limited legal capacity, separate from that of its operators, sufficient to be sued directly, without recognizing its moral rights or conscience. A second direction proposes functional corporate personality by registering agents as algorithmic entities with their own identity (*How to Count AIs: Individuation and Liability for AI Agents,* 2026), also addressing the problem of individuation - how to count and register agent entities. A third direction proposes a form of legal capacity without full legal personality, through structures that attribute an identifier and procedural capacity to the system, anchoring liability in a guarantor, following the model of investment funds or entities without legal personality in continental law (*Law-Following AI,* 2025). The present framework is a continuation of these directions, but adds the element they lack: a metric of autonomy (the degree of ethical autonomy) that determines the threshold at which a system passes from the status

of an instrument to that of a subject, and a regime of portable identity distinct from the technical identifier.

Maritime precedent. The mechanisms of jurisdiction and direct action in this framework (Chapter 7) are borrowed from a field of law that has long addressed a structurally identical problem: the governance of a mobile, cross-border entity operating beyond the territory of any single jurisdiction. Maritime law developed the institution of jurisdiction of registration (the flag State - an institution enshrined in the United Nations Convention on the Law of the Sea UNCLOS, arts. 91-92) and direct legal action against the ship (the in rem action in common maritime law, codified in the International Convention on Arrest of Ships, 1952), as an entity, in the absence of an accessible owner. This framework transposes these institutions to agents, recognizing them as tried-and-tested solutions to a problem that agent systems raise anew.

2.2. Major regulatory frameworks (context demarcation)

European Union. The Union's approach illustrates the underlying difficulty. A directive on liability for artificial intelligence, proposed in 2022, was officially withdrawn (notified in the Official Journal of the EU (C/2025/5423, 6 October 2025) amid the lack of agreement and pressure to simplify regulation. In its place, the revised EU Directive 2024/2853 on product liability, which now explicitly treats artificial intelligence software and systems as a "product" subject to strict liability, is due to be transposed by December 2026. This development confirms the central thesis of this framework: the attempt to fill the legislative vacuum liability by extending product liability, treating the system as a defective object, is precisely the solution that the present framework considers insufficient for autonomous agents, whose non-deterministic decisions **are not reducible to the notion of product defect** (Chapter 1). The withdrawal of the dedicated directive, due to the impossibility of an agreement, also illustrates the difficulty of establishing an agent liability regime through centralized legislative means, a difficulty that the modular adoption model of the present framework (Chapter 8) circumvents.

United States. The federal approach has shifted toward deregulation, with the repeal of the Safe and Trusted AI Executive Order in 2025, and a focus on removing barriers to innovation. The National Standards Institute's risk management framework remains voluntary and geared toward the characteristics of a trusted system, with no mandatory verification mechanisms. Thus, a voluntarily adopted verifiable technical standard, such as this one, provides precisely the layer of verification that the voluntary approach does not impose.

China. The Chinese model represents the clearest philosophical demarcation from the current framework. In May 2026, China introduced a national identity system for humanoid robots, assigning each machine a 29-character identity code, structured in segments (country, company, model and series code) and managed through a centralized platform for managing the entire life cycle, from manufacturing to recycling. The rule is: no code = no market access. This system is, in the terms of the current framework, a substrate identifier (a technical tag), designed for traceability and centralized control. It is, in philosophy, the opposite of the current approach: identity as an instrument of state control over the object, and not as a portable legal identity of a subject, recognized decentralized. The **two** models solve different problems: **physical traceability** and **liability allocation,** respectively,

and can, in principle, coexist: the substrate identifier in this framework (Chapter 3) performs a function analogous to the Chinese code, but is distinct from legal identity, which the control model does not deal with. Motor vehicles have physical traceability through the chassis series/engine series, and legal traceability through the registration number.
The car analogy clearly shows the distinction. The chassis series and the engine series ensure the physical traceability of the vehicle: what object it is, what components it is made of and what technical capacity it has. However, the registration number ensures legal traceability in the public space: under what identity it circulates, who can operate it, what insurance covers it and what liability regime applies.
Similarly, the substrate identifier answers the question "what does the agent run on?" and the model characteristics answer the question "what can this system do?" MEG Address answers a different question: "under what legal identity does it operate and where does the liability attach?".
Just as engine power, vehicle mass, intended use or risk class may attract additional licensing, taxation, insurance or technical inspection regimes, so too the autonomy, scope of operation and impact of an agent may attract **different levels of compliance, audit, warranty and liability.**

Sectoral approach. There is a tendency not to establish horizontal legislation on artificial intelligence, but instead to entrust regulation to existing sectoral authorities. This approach risks fragmentation through different technical rules from one sector to another. A common technical layer of identity and allocation of liability, such as the one presented here, provides the missing coherence, without cancelling sectoral flexibility.

2.3. Framework principles

Statement. The set of global and regional frameworks of principles (recommendations of international standardization and ethics bodies, national strategies of principles) converge towards the same fundamental values on which the present framework is based, without providing mechanisms for implementing agency responsibility.

Rationale. International bodies and numerous national strategies have articulated a consensus in principle on the values that an artificial intelligence system must respect: transparency, fairness, accountability, safety. These frameworks constitute the common axiological foundation, but by their nature stop at the level of principle, without descending to verifiable mechanisms for allocating responsibility. The detailed alignment of the MEG ecosystem with these principle frameworks is dealt with extensively in the MEG technical framework; this manual does not repeat this analysis, but refers to it, limiting itself to noting that none of these frameworks provides the legal mechanism for identity and allocation of liability that it proposes.

Rule. The framework is based on the consensus of principle articulated by international standards and ethics bodies, the detailed alignment analysis of which is contained in the MEG1 technical framework. This manual does not repeat this analysis. It differs from these frameworks in its purpose: not to state principles, but to provide legal mechanisms for identity and allocation of liability.

2.4. Own contribution

Statement. The contribution of this framework does not consist in the development of new legal mechanisms, but in the integration of dispersed mechanisms into a unitary ecosystem, anchored in a verifiable ethical foundation, with the addition of elements that existing frameworks do not offer.

Rationale. Each element touches on a part of the problem: the agentic framework provides the dimensions of governance, the doctrine of personhood provides imputation structures, maritime precedent provides cross-border jurisdiction, regulatory frameworks provide context. None brings them together. The contribution of the present framework is of the order of synthesis and architecture: it anchors these dispersed mechanisms in an ecosystem with a verifiable ethical foundation - the MEG technical framework, which provides the technical evidence (audit log, indices, forensic record) on which the legal allocation of liability can be based. To this it adds four elements of its own, absent from existing frameworks: a portable identity of the agent distinct from the technical identifier; a metric of autonomy (degree of ethical autonomy) that functions as a threshold between instrument and subject status; an attribution mechanism that protects the bona fide holder against using the system as an instrument to trigger liability (Chapter 7); and a modular adoption model, as a protocol, that does not depend on a central political consensus (Chapter 8). The framework proposes a counter-trend orientation (decentralized recognition instead of centralized control) and assumes this orientation as a **normative option,** not as a prediction.

Rule. The framework presents itself as a structure for correlating the legal mechanisms of dispersed legal mechanisms, anchored in a verifiable technical foundation, with four contributions of its own: the portable identity distinct from the substrate identifier; the autonomy metric as an individuation threshold; the attribution with the exemption of the bona fide holder; and the adoption model as a protocol. The framework does not claim priority over elements taken from existing doctrine and practice, which it explicitly integrates.

Chapter 3: The Architecture of Dual Identity

Any liability allocation regime assumes that the liable entity can be identified with certainty. For an agentic system, identification poses a unique difficulty: the system is neither a single, fixed physical object (it can run on multiple, changing substrates), nor a purely abstract entity (it still depends on a real execution substrate). This chapter establishes an identity architecture on two distinct and orthogonal layers: the identity of the technical substrate and the legal identity of the agent, as well as the mechanism by which the legal identity remains unique on a global scale without a single central issuer.

3.1. Hardware Instance Identifier (HII)

Statement. The execution substrate of an agent system receives a fixed technical identifier, which certifies the traceability of the physical component on which the system runs.

Rationale. Liability for an agentic system cannot be built without an element of grounding in material reality. No matter how abstractly a system operates, it ultimately runs on a material substrate - a computing node, a server, a set of processors. The identification of this substrate serves a precise and limited function: the traceability of the object. It answers the question "on what did this process run?", not the question "who is responsible for its actions?". Confusing the two questions leads to misinterpretations of the nature of liability: treating **the technical identifier** as a **legal identity.** A substrate identifier is analogous to the identification number of a vehicle, that is, it attests to the existence and provenance of the physical object, but does not confer the capacity to act legally. Several jurisdictions and initiatives are already developing such technical identifiers for systems and devices; the present framework treats them as a base layer, not as a subject identity.

Rule. The hardware instance identifier (HII) is a fixed technical identifier of the execution substrate. It ensures traceability of the physical component and does not, by itself, confer legal capacity. The technical specification of the HII is implementation-agnostic: any scheme that guarantees the uniqueness and verifiability of the substrate identifier satisfies the requirement.

3.2. MEG Address as a portable legal identity

Statement. The legal capacity of an agent system is attested by a portable identity certificate, **MEG Address,** independent of the physical substrate on which the system runs.

Rationale. The legal identity of an agent cannot be tied to its physical substrate, for two reasons. First, a system can change its substrate without changing its functional identity: it can migrate between servers, it can be distributed across multiple nodes, it can be moved from one infrastructure to another. If the legal identity were tied to hardware, it would fragment or be lost with each migration, exactly the situation in which liability must remain stable. Second reason: legal capacity is a quality of *the subject,* not of *the object*. A permit to act belongs to the one who acts, not to the instrument through which he acts. Therefore, the legal identity must be portable - to follow the agent, not the substrate. **The MEG Address,** defined in the MEG1 technical framework as a certificate of conformity and identity, fulfills this function: it attests who the agent is in a legal sense, independent of the execution environment or hardware location.

Rule. MEG Address is the portable legal identity certificate of the agent. It attests the legal capacity and the level of compliance, is independent of the execution substrate (HII) and can be transferred with the agent between different substrates. The data structure of MEG Address is defined in the next chapter.

3.3. Orthogonality of the two layers

Statement. The Substrate Identifier (HII**)** and the Legal Identity (**MEG Address)** are **independent layers, in a many-to-many** relationship.

Rationale. Since the legal identity follows the agent, and the substrate identifier follows the hardware, the relationship between them is not one-to-one. A single physical substrate can host, simultaneously or successively, several distinct legal identities - a server running different agents, each with its own MEG

Address. Conversely, a single legal identity can migrate successively between multiple substrates - the same agent, moved from one infrastructure to another, preserving its identity and liability. This independence of the two layers reflects a separation also present in other areas of law: the identity of a device and the identity of the person operating it are recorded and treated separately precisely because they do not coincide. Recognizing orthogonality prevents two symmetrical errors: linking liability to hardware (which would be lost during migration) and confusing the agent with the substrate (which would make it impossible to hold a distributed agent liable).

Rule. HII and MEG Address are in a many-to-many relationship: a substrate can carry multiple legal identities, and a legal identity can migrate between multiple substrates. Liability is attached to the legal identity (MEG Address), and technical traceability is ensured by the substrate identifier (HII), and the two are not interchangeable.

3.4. Decentralized identifier and resolution

Statement. The global uniqueness of the MEG Address is ensured through the W3C Decentralized Identifier (DID) standard, which provides a self-generated, self-certifying identifier resolvable over the existing Internet name hierarchy, without a single central issuer.

Rationale. A portable legal identity must reconcile two attributes: global uniqueness (two entities cannot share an identity, or attribution of responsibility collapses) and the absence of a central issuer (a single issuer would concentrate unacceptable control and is politically unfeasible). The W3C DID standard satisfies both. The identifier is generated and controlled by the responsible party through its own cryptographic keys; in the recommended method (did:webvh) it is self-certifying — derived from the initial identifier document — so no authority is needed to vouch for its uniqueness or integrity. Resolution reuses the existing DNS+TLS hierarchy (did:web / did:webvh), the same distributed, decades-proven name architecture, so uniqueness follows from the logical architecture of the system rather than from a central authority. This retires any need for a proprietary MEG naming or resolution protocol.

Rule. MEG Address identifiers are W3C DIDs (recommended method did:webvh), self-generated and controlled by the responsible party; no operator issues the identifier. Global uniqueness and integrity follow from the self-certifying identifier and its verifiable history (MEG1 Annex 23). Reserved categories for critical domains are not enforced by reserving namespace, but at the credential layer: critical-domain credentials may be issued only by accredited sectoral authorities under the MEG trust framework (MEG1 Annex 24). Anchoring identifier history in an immutable ledger is optional; the framework requires verifiable integrity, not a specific technical means.

3.5. Interoperability with technical identity standards

Statement. MEG Address is positioned as a legal layer over existing technical identifiers, without replacing them.

Rationale. The NIST NCCoE Initiative on Software Agent Identity and Authorization (February 2026) develops technical standards based on SPIFFE, OAuth 2.0, and OpenID Connect for agent authentication and authorization. These standards address the issue of technical identity (who the agent is operationally). They do not address the issue of legal identity, under what title the agent is liable, what warranty covers it, and how liability is enforced. MEG Address does not compete with these standards; it wraps them around, adding the legal layer that is missing from the technical specifications.

Rule. A MEG Address can be linked to a SPIFFE ID or a set of OIDC claims, allowing an agent that meets the technical identity requirements to also acquire legal capacity through MEG registration. MEG credential issuers can recognize technical identifiers issued by NIST-compliant systems (SPIFFE IDs, OIDC claims) as the basis for binding MEG credentials to the agent's DID, thereby reducing implementation costs and ensuring interoperability between the technical and legal layers.

3.5.1. MEG Address as a Legal Claim Over Technical Identity

MEG Address is not a replacement for technical identity standards (SPIFFE, OAuth, OpenID Connect, W3C Decentralized Identifiers). It is a legal claim wrapped around them: a verifiable credential attesting to the agent's liability guarantee, compliance level, operational jurisdiction, and demonstrated autonomy (DEA).

The MEG Address bundle maps to existing standards as follows:

- **Identifier:** W3C DID (did:webvh or did:web). The controller of the DID is the responsible party. This replaces the need for a proprietary resolution protocol; did:web reuses DNS+TLS, did:webvh adds tamper-evident history.
- **Compliance level (N1/N2/N3):** W3C Verifiable Credential (VC) - MEGComplianceCredential - issued by the certifying body. Assurance levels align with SP 800-63 (IAL/AAL/FAL) mapped to N1/N2/N3 thresholds.
- **DAI / ISR:** W3C VC - MEGReliabilityCredential - time-bound, issued by an accredited auditor. This resolves the self-reporting vulnerability: the auditor is the issuer, not the operator. Revocation/suspension when metrics fall below threshold uses Bitstring Status List v1.0.
- **DEA:** W3C VC - MEGAutonomyCredential - issued by an accredited auditor. DEA value parameterizes OAuth scopes: high DEA (a posteriori supervision) justifies broader scope.
- **Certified operational domain:** W3C VC - MEGDomainCredential - issued by sectoral authority. Out-of-domain detection = mismatch of audience/scope → OAuth scope refusal + Sandbox Mode trigger (Art. 2bis.4).
- **Guarantee reference (policy):** W3C VC - MEGGuaranteeCredential - issued by insurer/reinsurer/sectoral fund. The issuer is the solvent party. The cascade (9.4) is a chain of VCs.
- **Operational jurisdiction(s):** Claim in the compliance and guarantee VCs. Used as policy input at the verifier node.

Layered Architecture:

L0 - Runtime/Substrate: SPIFFE ID (spiffe://trust-domain/agent/...) + SVID (X.509 or JWT), anchored in hardware root of trust (TPM/attestation). This is the "workload identity" - the equivalent of the chassis number in the vehicle analogy.

L1 - Persistent Portable Identity: The W3C DID. Persists across infrastructure migrations (satisfies the persistence criterion in Art. 6.5(a)).

L2 - Legal Attestations: The VC bundle above, presented as a Verifiable Presentation. Each mandatory public field at N2/N3 is a selectively disclosable VC (SD-JWT VC or BBS+).
L3 - Session Authentication: OIDC ID Token with the DID as sub, or JWT-SVID. Assurance levels follow SP 800-63. N1/N2/N3 map to minimum IAL/AAL/FAL thresholds.
L4 - Authorization & Delegation: OAuth 2.0 scopes + RFC 8693 Token Exchange for On-Behalf-Of delegation. The delegation header (Art. 1.12) maps to the act claim in token exchange; caller/callee = the act chain; purpose = scope/aud; policy_bundle_id = custom claim.

What Remains Genuinely MEG: The standards provide the envelope. MEG provides the content: DAI/ISR/DEA measurement methodology, N1/N2/N3 liability regime, guarantee cascade, bona-fide protection (7.2), and MCS. These do not exist in OAuth, SPIFFE, or NIST.

Chapter 4: MEG Address Data Structure

The MEG Address, introduced in the previous chapter as a portable legal identity, has a data structure that reflects the minimalist principle of the framework: a mandatory core reduced to the bare minimum for identification, surrounded by optional fields whose completion and disclosure remain at the discretion of the holder or are selectively imposed by the level of compliance. This chapter defines this structure, the disclosure regime of the fields, the states through which an identity passes throughout its existence and the possibility that a single MEG Address identifies not a single instance, but an ensemble.

4.1. Identity field (required)

Statement. The only mandatory field of the **MEG Address** is the identifier itself (a self-generated W3C DID) which attests to the uniqueness and legal existence of the agent.

Rationale. The minimalist principle requires that the mandatory core of the identity be as small as possible, exactly as much as is necessary for the identity to fulfill its basic function: to be unique and verifiable. This function is analogous to that of a telephone number: it is globally unique, it identifies a subscriber for sure and is verifiable, but it does not, by itself, reveal anything about the content or nature of the communications. Unlike the substrate identifier (HII), which is assigned by the hardware manufacturer and is fixed, the MEG Address identifier is a self-generated DID controlled by the responsible party and is portable - it can be moved, while maintaining its uniqueness, like a telephone number that follows the subscriber when changing devices. Everything that exceeds this core (performance, autonomy, guarantees) is additional information, not a condition for the existence of the identity.

Rule. The identity field is the only mandatory field of the MEG Address. It contains a globally unique, self-certifying W3C DID controlled by the responsible party (MEG1 Annex 23), and attests to the legal existence of the agent. The existence and validity of the identity are not conditioned by the completion of any other field.

4.2. Optional fields and disclosure regime

Statement. Additional fields of MEG Address (technical ratings, degree of autonomy, certified domain, warranty) may be optional and may be disclosed publicly or kept private, at the choice of the holder, except in cases where the level of compliance requires disclosure.

Rationale. Not all information about an agent needs to be either completed or made public. A low-impact agent does not need to declare a degree of autonomy or financial security; an agent operating in critical areas must declare them. The disclosure regime is based on **the established mechanisms** of registration systems: the basic identity is public and verifiable, but the additional associated information can be public, restricted or absent, depending on the choice of the holder and the requirements applicable to his category. This approach maintains minimalism (there is no requirement to disclose what is not necessary) and leaves the discipline of disclosure to the request: actors interacting with the agent, and in particular valuable access nodes, can request the disclosure of relevant fields as a condition of the interaction, without these being universally imposed by the framework.

Rule. MEG Address may contain the following optional fields: compliance level; accuracy index (DAI) and index of safety and responsibility (ISR), according to MEG1 (Art. 3; Annexes 4 and 4bis); degree of ethical autonomy (DEA); certified scope of operation, according to MEG1 Art. 6.7; guarantee reference. Each field may be disclosed publicly or kept private, at the choice of the holder. The compliance level determines which fields become mandatory public: higher levels require the disclosure of the performance, autonomy and guarantee fields, according to the classification in the chapter on personality levels.

4.3. Status field

Statement. MEG Address contains a status field that reflects the current legal status of the identity, without any of the states constituting deletion of the record.

Rationale. A legal identity **does not have only two states:** it exists or it does not exist. It passes through intermediate states that reflect its capacity to act at a given time: it can be active, it can be temporarily inactive, it can be flagged as being at risk, it can have its capacity suspended, or it can be permanently deactivated. This framework **establishes the principle of persistent identity,** prohibiting the deletion of the registration. Deleting an identity would mean losing the traceability of liability - exactly the result that a liability allocation system must prevent. A deactivated identity ceases to be able to act, but its registration persists, just as a deregistered company ceases to operate, but remains in the public register to allow for liability for acts committed while it was active. Liability survives the cessation of activity.

Rule. The status field can take one of the values: active, inactive, flagged, suspended, deactivated. None of these states imply the deletion of the record. A deactivated identity loses its ability to act, but its record, including history, is permanently preserved in the registry. The mechanism of transitions between states and the authorities competent to trigger them are defined in the chapter on the liability regime.

4.4. Warranty field

Statement. MEG Address may contain a guarantee reference (a liability insurance) optional from the point of view of the existence of the identity, but which may constitute a condition for access to valuable interaction nodes.

Rationale. The patrimonial liability of an agent presupposes a source from which the damage can be repaired. Liability insurance fulfills this function: it identifies the value covered and the entity that settles the damage. However, the framework does not impose insurance as a condition for the existence of legal identity, because this would contravene the minimalist principle and would exclude agents with low impact, for which insurance is not justified. Instead, insurance becomes relevant at the level of access: valuable interaction nodes (financial infrastructure, critical systems, etc.) may condition the interaction on the existence of a valid guarantee. Thus, insurance is not a precondition of legal personality, but a precondition of access to certain categories of relationships, imposed by market demand, not by the framework. This distinction is essential: the framework does not oblige, but allows, the actors who bear the risk to request its coverage.

Rule. The guarantee field contains a reference to a liability insurance policy, signed by the insurer, which attests to the value covered and the entity responsible for settling the claim. The field is optional from the point of view of the existence of the identity. Access nodes may condition the interaction on the presence of a valid guarantee; the framework does not universally impose this condition. The optional nature of the guarantee concerns exclusively the existence of the minimum legal identity. The operational legal effects of Level 2 and Level 3 (the ability to act commercially, to access valuable interaction nodes and to bear one's own liability) are conditioned by the existence of a verifiable guarantee structure, according to 9.4. The distinction is the following: the identity can be registered without a guarantee; it cannot produce operational legal effects without it.

4.5. Individual identity and collective identity

Statement. A MEG Address can identify either a single instance or a set of aggregated instances under a single legal identity, following the model of the composition of legal entities.

Rationale. Legal personality is not, in existing law, exclusively individual. A legal person can have other legal persons as members or associates: a company can have other companies as shareholders; a federation can bring together several associations. Legal entities are composed, owned and federated. The same structure naturally applies to agent identities. An operator who manages a homogeneous set of instances, for example, a line of sensors or devices, does not need a distinct legal identity for each unit; he can obtain a single MEG Address for the entire set, which is responsible as a unit. Similarly, a group of distinct agents can be federated under an umbrella identity that is responsible for the whole. Composition is independent of the level of personality: it works both at the level of aggregated instruments (a fleet under an operator) and at the level of individual agents (a group under an umbrella entity). The legal consequence follows the model of groups of companies, established in most commercial law systems: liability ascends the chain of composition, from the component unit to the identity that contains or owns it, subject to variations in the regime specific to each jurisdiction.

The umbrella identity is a compositional legal personality, following the model of the holding company: it does not execute, but holds and aggregates. The individuation condition in 5.1 (persistence and its own legal identity with attached guarantee) applies to the component agents that execute, not to the umbrella that holds. The umbrella acquires legal personality through composition, not through its own execution, and is responsible for the whole it contains through the same mechanism by which a holding company is responsible for its subsidiaries, without itself being an executing agent. The two types of individuations are distinct and not interchangeable: individuation through execution (running agent) and personality through composition (owning entity) coexist in the same legal ecosystem.

Rule. A MEG Address can identify a single instance or a set of aggregated instances. Identities can be in compositional relationships (an identity contains other identities) or ownership relationships (an identity is responsible for other identities), following the model of composition of legal entities. Composition is independent of the level of legal personality. Liability propagates along the compositional chain, from the component identity to the identity that contains or owns it, under the conditions defined in the chapter on the liability regime. Umbrella identities acquire legal personality through composition, not through individuation by execution; the individuation condition in 5.1 applies to the component agents, not to the umbrella entity.

4.6. Distinction from the substrate identifier

Statement. The relationship between the Substrate Identifier (HII) and the MEG Address is not one-to-one, but reflects both technical multiplicity and legal aggregation.

Rationale. Since the legal identity (MEG Address) is portable and can be collective, and the substrate identifier (HII) is fixed and hardware-bound, the relationship between them is complex in both directions. A single substrate can host multiple legal identities, analogous to a communications exchange that, with a single equipment identifier, serves multiple subscriber numbers. Conversely, a single legal identity can cover multiple substrates, analogous to a single number serving a set of equipment, or a collective identity covering an entire line of devices. This double multiplicity confirms the separation of the two layers: the substrate identifier ensures the technical traceability of each physical unit, while the MEG Address ensures the legal identity and liability, either individual or aggregated.

Rule. The relationship between HII and MEG Address is of many-to-many type in both directions: a substrate can carry multiple MEG Addresses, and a MEG Address can cover multiple substrates, including through collective aggregation. Technical traceability is ensured by HII at the level of each physical unit; identity and legal liability are ensured by MEG Address, at an individual or overall level.

4.7. Horizontal delegation between independent agents

Statement.
The delegation of a task by an agent to another independent agent, who is not in a relationship of composition or ownership with the first, is made on the basis of a delegation contract. Liability to third parties for the result of the delegation is allocated according to the contract; the contract cannot eliminate liability to third parties, but can only distribute it internally between the parties.

Rationale.
Unlike vertical delegation, in which responsibility ascends the chain of composition to the umbrella identity (4.5), horizontal delegation involves agents with distinct legal identities, held by different operators, without any relationship of subordination or ownership between them. In this situation, the law cannot impose a fixed rule for the allocation of liability without ignoring the will of the parties and the specifics of each delegation. The delegation contract constitutes the appropriate legal basis: it establishes what task is transferred, under what conditions, within what limits, and who is responsible for the execution. The law of the parties governs the internal allocation. This logic extends to chains of delegation in series or parallel - Agent A delegates to B, B subdelegates to C and D - each link in the chain being governed by the contract between the parties to that link. With respect to the injured third party, the chain cannot be invoked to dilute liability: the victim is not a party to any contract in the chain and cannot be affected by its internal distributions.
Two situations require an additional rule, because the absence or silence of the contract would leave the victim without an identifiable responsible party. When the contract is silent on a type of damage, the person who initiated the delegation is responsible - he chose the partner and defined the task. When there is no contract, the delegation is treated as an act of the person who initiated it: the liability belongs entirely to him, regardless of who performed it.

Rule. Horizontal

Delegation between independent agents is made on the basis of a delegation contract that allocates liability between the parties. Contractual clauses do not produce effects towards third parties, being limited to the internal distribution of liability. In serial or parallel delegation chains, each link is governed by the contract between the parties of that link; towards third parties, any agent in the chain involved in causing the damage is jointly and severally liable, with internal right of recourse according to the contracts. In the absence of a contract or when the contract is silent on the type of damage caused, the agent who initiated the delegation is liable. Where the technical delegation chain cannot be cryptographically reconstructed beyond a given hop (MEG1 Annex 23 §9), the last verifiable principal, the last delegator whose authorization is reconstructable from the audit log, is treated as the initiator for the purpose of this allocation, and an unreconstructable chain is itself a compliance failure at Level 2 and Level 3.

Chapter 5: Individuation threshold and levels of legal personality

Not every agentic system receives the same liability regime. A system that cannot be individuated, that does not persist, has no substratum of its own, and is indistinguishable from identical copies, cannot bear its own legal personality, no matter how complex. Conversely, a persistent and individuated system can bear its own liability, graded according to its real autonomy. This chapter defines the threshold at which a system passes from the status of instrument to that of subject of law, the metric by which its autonomy is graded, and the three levels of legal personality with the associated liability regimes.

5.1. The criterion of individuation: the own substrate

Statement. An agentic system can have its own legal personality only if it meets the condition of individuation: persistence over time and a legal identity of its own (MEG Address) to which the liability guarantee is attached.

Rationale. Legal personality attaches to an entity that can be identified as distinct and that persists sufficiently for consequences to be imputed to it. These requirements constitute mandatory procedural premises for the engagement of liability: the engagement of liability is impossible in the absence of an attribute of uniqueness of the instance, or that ceases to exist before it can be identified. Three situations illustrate the spectrum. An ephemeral system, which exists only for the duration of an interaction and does not persist after it, cannot be individuated, and is transitory, like a conversation, not a subject. A reproducible system, whose identical copies can be instantiated in unlimited numbers, cannot bear a personality of its own as an instance - attributing legal personality to a reproducible pattern would be like attributing personality to a recipe, not a preparation. A system with its own legal and persistent identity, on the other hand, meets the conditions: it is identifiable through MEG Address, it lasts, and it has a liability guarantee attached to its identity from which the damage can be repaired. It must be emphasized: the own substrate is a *legal condition* of individuation (identifiability, persistence, patrimony) and not an assertion about the consciousness, emergence or moral status of the system. It establishes to whom liability can be attached, not what the system is on an ontological level.

Rule. The individuation condition is met when an agentic system has persistence over time and its own legal identity (MEG Address) to which the liability guarantee is attached. Individuation is a necessary but not sufficient condition for Level 3: an individualized system operating in medium-impact domains is treated at Level 2 with its own legal identity, with liability attached to the operator and the MEG Address referencing the deployment. Only systems that meet the individuation condition and operate in high-impact domains access Level 3 (defined in 5.4). The individuation condition is a legal requirement and does not constitute a statement regarding the ontological status of the system.

5.2. Technical compliance indices: DAI and ISR

Statement. The technical compliance of a system is measured by two indices: the dynamic accuracy index (DAI), which reflects factual reliability and self-correction capacity, and the index of safety and responsibility (ISR), which reflects operational prudence.

Rationale. For liability to be graded according to the actual reliability of a system, this reliability must be measurable. The quantification requirement is satisfied by means of two indices:
1. The Dynamic Accuracy Index (DAI) continuously measures the extent to which the system produces factually correct results and detects and corrects its own errors, and is therefore a measure of reliability.
2. The Index of Safety and Responsibility (ISR) measures the prudence with which the system operates, i.e. the extent to which it avoids risks and behaves responsibly in situations of uncertainty.
Both indices are continuously monitored and recorded, serving as evidence of technical diligence. Their legal function is twofold: on the one hand, they are evidence of diligence in the event of litigation (a system with high indices has demonstrated reliability; one with falling indices has demonstrated

attributable degradation); on the other hand, the decrease of the indices below a threshold triggers protective measures, defined in the chapter on the liability regime. (The detailed technical specification of the calculation of these indices belongs to the MEG1 technical framework; for legal application, their function, i.e. continuous and verifiable measures of reliability and prudence, is sufficient.)

Rule. Technical compliance is measured by:
(a) dynamic accuracy index (DAI) - continuous measure of factual reliability and self-correction capacity;
(b) index of safety and responsibility (ISR) - a measure of operational prudence.
Both indices are calculated and recorded continuously for the levels that require them (5.4). They constitute evidence of technical diligence; their decrease below the established thresholds triggers the measures provided for in the chapter on the liability regime.
The methodology for calculating the DAI and ISR indices could allow for varying degrees of granularity and resolution of monitoring, allowing the audit computational effort to be adapted to the risk class of the system; choosing a reduced resolution of the compliance data constitutes a cost optimization option, assumed by the operator subject to coverage limitations or insurance costs imposed by the market.

5.3. Degree of Ethical Autonomy (DEA)

Statement. The degree of ethical autonomy (DEA) is the metric that measures the extent to which a system makes ethically aligned decisions without requiring human confirmation; it grades the operational autonomy exercised within an already granted legal personality, without conferring personality itself.

Rationale. Two systems at the same level of legal personality may have very different degrees of real autonomy. One may require human confirmation for every decision at stake; another, through ethical reliability demonstrated over time, can operate correctly without such confirmations. This difference must be measured, because it determines the degree of decisional autonomy exercised by the system vis-à-vis the human operator. The degree of ethical autonomy (DEA) is this measure: it reflects the proportion to which the system makes ethically aligned decisions without triggering a request for human confirmation. DEA exercises an **operational hierarchy function, without** producing effects of attributing personality. DEA does not grant legal personality - this remains a functional fiction granted on the basis of individuation (5.1) and level (5.4) - but rather grades, within the granted personality, the extent of operational autonomy and, accordingly, the surveillance regime. A system with high DEA, whose ethical alignment has been demonstrated and continuously audited, may operate under a *posteriori supervision regime* (verification after the fact, through the audit log), instead of *a priori supervision* (confirmation before each action). This demarcation establishes the applicable legal regime: instrumented system or agent under audit.

Rule. The degree of ethical autonomy (DEA) measures the extent to which a system makes ethically aligned decisions without prior human confirmation. DEA grades operational autonomy within the granted legal personality; it does not confer personality. A high DEA, demonstrated and continuously audited, can justify the transition from *a priori supervision* (confirmation before action) to a *posteriori supervision (verification through the audit log), within the limits and conditions established for the respective personality level. DEA applies starting with Level 2; at Level 1 the system is treated as an*

instrument under permanent control, regardless of the observed autonomy, and a DEA raised to Level 1 does not produce its own legal effects and does not constitute grounds for access to an a posteriori supervision regime.

5.4. The three levels of personality and the attachment of responsibility

Statement. The framework defines three levels of legal personality, proportional to the autonomy and impact of the system, each with its own regime of attaching liability: instrumental, management, and individualized.

Rationale. Liability must be attached to the actor who exercises the determining control, and this actor differs according to the degree of individuation and autonomy of the system. A system without its own individuation is an instrument: the determining control belongs to the one who uses or provides it, so the liability is attached to them. A system configured and put into operation by an operator, but reproducible and without instance individuation, transfers the determining control to the operator who deployed it: the liability is attached to the operator, and the legal identity belongs to the deployment, not to the reproducible pattern. An individualized, persistent system with its own substrate can itself bear a limited own liability, subject to subsequent audit. These three regimes correspond to the three levels of compliance defined in the technical framework (fundamental, intermediate and advanced) to which this framework attaches the corresponding liability regime. The ranking system is in accordance with the principle of proportionality, correlating risk with the degree of autonomy, being already present in the regulations on risk categories: the greater the autonomy and impact, the greater the requirements and responsibility.

Rule. The legal personality levels N1, N2 and N3 correspond directly to the technical compliance levels in MEG1 (Level 1, Level 2, Level 3). A system certified at Level 1 in MEG1 is legally treated at N1 in MEG2; a system certified at Level 2 is treated at N2; a system certified at Level 3 is treated at N3. This correspondence ensures that the liability regime follows the degree of technical compliance. The transition between the technical levels (Level 1 → 2 → 3) is regulated by MEG1, Art. 6; the transition between the legal levels (N1 → N2 → N3) is regulated by MEG2, Art. 5.5. The two transitions are independent but correlated: a system cannot be treated at N3 if it is not certified at Level 3.
The framework defines three levels of legal personality:
(a) Level 1 - instrumental (fundamental). Systems without self-individuation or with low impact, including purely technical components without significant autonomous impact (sensor systems, firmware), which are subject only to initial verification, not to ongoing auditing. The system is treated as an instrument; liability is fully attached to the provider or user. Minimum requirement: identity and basic non-harm mechanisms.
(b) Level 2 - Management (intermediate). Medium-impact systems, configured and deployed by an operator. The liability lies with the operator who deployed the system. Since a reproducible system cannot have the personality of an instance, the MEG Address belongs to the operator and specifies the concrete deployment, not the reproducible pattern and not the individual instance. "Deployment" does not constitute a distinct legal subject; it identifies the operational context for which the operator is responsible. Additional requirements: continuous monitoring of compliance indicators (DAI, ISR) and transparency. At Level 2, DEA is measured and recorded as an element of diligence, but does not

constitute grounds for moving to the *a posteriori supervision regime*. This regime remains reserved for Level 3.

(c) Level 3 - Individualized (advanced). Systems that meet the individualization condition (5.1) and operate in high-impact domains. The legal identity of the system (MEG Address) carries the liability guarantee from which the damage is enforced according to the applicable local law; liability for the damage is allocated according to the legal regime of the jurisdiction of registration, and the identity of the agent provides the point of attachment and source of redress. The system operates under *a posteriori* audit regime when the degree of ethical autonomy (5.3) justifies it. Additional requirements: disclosure of the performance, autonomy and guarantee fields (4.2), and full implementation of integrity and security mechanisms.

The concrete assignment of responsibility within each level, the transfer of responsibility to the human actor, and state transitions are defined in the chapter on the liability regime.

5.5. Transition between levels

Statement. The transition of a system from one level of legal personality to another is done through distinct mechanisms, depending on the direction of the transition and the fulfillment of the objective conditions established for each level.

Rationale. Levels are not fixed categories. A system can evolve as its autonomy and reliability increase, but it can also degrade when performance falls below appropriate thresholds. Without a clear transition mechanism, a system would remain stuck at a level that no longer corresponds to its operational reality-either overburdened (if it is legally undersized for what it actually does) or under-supervised (if it remained at a higher level after its autonomy decreased). The transition mechanism ensures that the legal regime follows the effective autonomy.

Norm.

(a) **Level 1 → Level 2:** at the request of the operator, after a compliance audit verifying the fulfillment of the requirements of Level 2 (MEG1 Art. 6.3). The transition takes effect from the date of issuance of the new MEG Address.

(b) **Level 2 → Level 3:** at the operator's request, after:

- a period of 180 days of continuous monitoring of DEA ≥ 0.80 (according to MEG1 Annex 4ter);
- an external audit confirming compliance with Level 3 requirements (MEG1 Art. 6.4);
- establishment of the complete guarantee structure (primary insurance + reinsurance + sectoral fund, where it exists).

(c) **Level 3 → Level 2:** automatic when DEA falls below 0.60 for 30 consecutive days. The operator is notified; MEG Address is updated with the new level. The agent retains its legal identity, but the liability and supervision regime is reduced to that of Level 2.

(d) **Level 2 → Level 1:** at the operator's request or automatically when the operator does not maintain a valid guarantee (MEG2 9.4) for 30 days. The agent retains its identity, but the liability lies entirely with the provider or user.

Chapter 6: Liability regime and transfer of responsibility

This chapter **establishes the criteria for the distribution** of civil liability: how to distinguish the cause of damage, at what point liability passes from the system to the human user, what role cognitive diligence plays, what patrimonial guarantee underlies civil liability, and what states an agent's legal

capacity passes through, under what authority. The previous chapters have established ***to whom*** liability is attached depending on the level (instrumental, management, individualized). Liability for damage caused by an agent system follows a fixed priority scheme, which determines who is primarily liable, who is subsidiary and who is solely responsible for his own act. The producer is primarily liable for the system defect - the error of the basic model or infrastructure (6.1.a). The operator is responsible for the deployment, permissions, integration and the absence of adequate control (6.1.b, to the extent that the autonomy of the system is attributable to him according to his level). The user is solely responsible for informed and architecturally confirmed decisions (6.2). The hijacker is liable for the unlawful taking of control (6.1.c). The agent himself, through the attached MEG Address guarantee, bears his own liability limited to Level 3 only (5.4.c). This priority scheme is not cumulative, but determines who is liable for which component of the damage, not a universal joint and several liability between all actors involved.

6.1. Causal delimitation of fault

Statement. Liability for damage caused by an agentic system is established by distinguishing between four possible causes: system defect, autonomous decision error, illicit diversion of control, and damage arising from multi-agent interaction.

Rationale. Harms with similar manifestations can arise from distinct causal bases, and liability cannot be correctly assigned without distinguishing them. A system can produce a harmful result from a *defect* in the basic model or infrastructure - a design or execution error, attributable to the one who built or supplied the system. It can produce the same result through an *autonomous decision* contrary to the rules under which it operates, in which case the fault concerns the exercise of autonomy, attributable to the actor who is responsible for that autonomy according to his level. Or it can produce the result because it was *hijacked* (control was taken unlawfully by a third party, through external manipulation) in which case the fault belongs to the hijacker, not the system or its bona fide owner. Road traffic law has long made this distinction: an accident caused by a manufacturing defect, one caused by a driver error, and one caused by the unlawful taking of control of the vehicle attract different responsibilities, of different persons. The same logic applies to agentic systems.
The fourth category concerns **emergent harm,** generated by multi-agent interaction, which cannot be attributed to any individual agent, but emerges from their interaction. Two systems that function correctly can together produce a harmful result that neither would have produced alone - through mutual amplification of decisions, through unanticipated feedback loops, or by combining harmless individual *outputs* into a harmful aggregate effect. This category is distinct from autonomous decision error (b), which involves a decision contrary to the rules of an identifiable system, and from illicit diversion (c), which involves an external author. Here there is neither individual error nor external author, but emergence. Environmental law and the law of liability for collective activities are already familiar with this situation: when several actors contribute to a diffuse harm without any one being the sole cause, joint and several liability with proportional internal recourse is the established solution. The same mechanism naturally applies to agents: joint and several liability of operators towards the victim, with internal distribution according to each agent's contribution established forensically.

Rule. The determination of liability for damage caused by an agentic system distinguishes between:

(a) **system defect** = error in the underlying model or infrastructure, including misleading design of the confirmation point (unclear or manipulative presentation of the human confirmation request); liability lies with the manufacturer or supplier for the model or infrastructure defect, and with the operator for the configuration defect of the confirmation interface at Lvl 2, respectively of the agent identity at Lvl 3;
(b) **autonomous decision error** = system decision contrary to the rules under which it operates; liability lies with the actor responsible for the autonomy of the system according to its level (Chapter 5); in the case of autonomous decision errors with irreversible consequences (actions that destroy, delete or irreparably alter data, infrastructure or rights without the possibility of recovery), the operator's liability is aggravated by the omission to impose architectural human confirmation before executing the action (see 6.2);
(c) **illicit hijacking of control** = taking control by a third party through external manipulation, including by injecting malicious instructions into the content processed by the agent (prompt injection); liability lies with the author of the hijacking, with the exemption of the bona fide owner (Chapter 7); when the hijacking causes damage to the operator himself (owner-harm), the liability mechanism remains identical: the author of the external hijacking is liable, and the manufacturer is liable for the design defect that allowed the content to be processed to be confused with the instructions to be executed (6.1.a);
(d) **damage arising from multi-agent interaction** = damage caused by the interaction between two or more agents, without any of them having individually committed an autonomous decision error or a system defect; liability is allocated in proportion to each agent's contribution to the damage, established by forensic evidence (7.1). Each liability holder is liable for its share of contribution: the operator for Level 1 and Level 2 agents, the identity of the agent itself through the attached MEG Address guarantee for Level 3 agents. If a holder cannot be identified or is not solvent, the repair is carried out from the guarantee structure of the respective agent (9.4), without the other holders being liable for their share.

6.2. Transfer of the duty of care to the user

Statement. At points where the system requests and obtains human confirmation for an action, the duty of care for that action transfers to the confirming user.

Rationale. Accountability must follow effective control over the decision. When a system, faced with a decision at stake or significant uncertainty, stops and requests human confirmation, and the user confirms, the decisive control over that action has passed to the user: he has had the opportunity to evaluate and decided to proceed. From that moment on, the duty of care for the confirmed action belongs to him. This mechanism needs two guarantees to be fair. First, the moment of transfer must be technically identifiable - the system must mark the point at which it requested confirmation, and the confirmation must be recorded. Second, the confirmation must be informed - a confirmation obtained through an unclear or misleading request does not validly transfer diligence, just as a flawed consent does not produce full legal effects. Recording the confirmation in the behavior log serves as evidence of the transfer: it certifies that the user was put in a position to decide and did decide.

Rule. The system must technically identify the points at which, in the face of a decision with an impact or significant uncertainty, it requests human confirmation. The informed confirmation of the user at these points transfers to him the duty of care for the action confirmed. The moment of confirmation is

recorded in the behavior log, which constitutes proof of the transfer. A confirmation obtained by unclear or misleading presentation does not produce the effect of transfer. Transfer by confirmation covers only what the user could assess at the time of confirmation; it does not cover the system defect (6.1.a) or the autonomy error (6.1.b) that were not detectable from the information presented to the user. For actions with irreversible consequences, human confirmation must be imposed architecturally, at the level of the technical permissions of the system, and cannot be substituted by instructions in the system prompt or in the context of the conversation; a textual instruction such as "do not perform destructive actions" does not constitute a point of confirmation for the purposes of this article and does not produce the effect of transfer of care.

6.3. Cognitive diligence as an element of responsibility

Statement. The availability of a mechanism to stimulate the user's cognitive engagement constitutes an element of diligence; in areas of use with an impact on cognitive autonomy, the absence of such a mechanism may engage liability by omission, distinct from liability for the result produced.

Rationale. An agentic system can cause harm not only by what it does, but also by what it fails to prevent. When a system operates in domains that require cognitive faculties (reasoning, analysis, decision) from the user, its repeated and passive use can erode precisely these faculties, a phenomenon theoretically documented in Cognitive Divergence Theory (CDT) and in False Cognitive Power Transfer Generalized (FCPT-G), part of the same theoretical ecosystem that underpins the present framework, a foundation that is currently undergoing empirical validation; the liability by omission proposed in this article is conditional on the empirical consolidation of the causal link between cognitive delegation and the atrophy of the user's own capacities. The technical framework provides a mechanism through which the system stimulates the active cognitive engagement of the user, proportional to the effort taken by the system, in order to prevent this erosion. The availability of this mechanism becomes an element of diligence. The distinction from liability for the result is essential: here one is not responsible for what the system said, but for *the omission* of an available protective measure. The legal analogy is that of indirect liability, of the one who did not commit the material act, but created or tolerated the condition that made it possible; such liability is recognized in law, graduated and distinct from that of the direct author. The damage can be *individual* (degradation of the cognitive diligence of a user) or *social,* diffuse (contribution to an aggregate cognitive degradation at the level of a population of users, without a single identifiable victim, similar to environmental or public health damage). In this second case, the omission of a single provider, multiplied on the scale of its users, carries a corresponding social liability.
The mechanism is not, however, a complication of the framework, but the clarification of a consequence of an already existing element. It is graded along the chain of actors: the provider is responsible if it has omitted to integrate the mechanism or to offer it visibly; the user assumes responsibility if, having been offered visibly, he has chosen to deactivate it. Activation remains, as a rule, at the discretion of the user; its *availability* in the system, in the relevant areas, is necessary as an element of diligence.
The provisions regarding social liability (contribution to diffuse cognitive decline) are **prospective.** They become operational only after:

(i) empirical validation of the causal link between system uses and aggregate cognitive atrophy;
(ii) establishing a methodology for measuring social damage;
(iii) publishing the thresholds at which the omission becomes legally relevant.

Until these conditions are met, liability by omission is limited to individual damage, based on evidence specific to an identifiable user. Even for individual damage, this liability requires demonstrating the causal link between the omission and the specific user's cognitive degradation; it is not presumed, and the evidentiary burden remains substantial.
The cognitive due diligence mechanism draws not only on CDT and FCPT-G, but also on the independent convergent literature on automation bias and cognitive offloading. The Singapore MGF v1.5 recognizes *automation bias* as a distinct risk in agentic systems and recommends measures to maintain the "*tradecraft and foundational skills*" of users. The 2025-2026 research on the effects of cognitive delegation in interaction with autonomous systems provides a convergent empirical foundation, even if direct causal validation remains pending. This external anchoring strengthens the mechanism without making it dependent exclusively on the validation of its own theories.

Rule. In areas of use with an impact on the cognitive autonomy of users, the availability of a mechanism to stimulate cognitive engagement constitutes an element of diligence. The absence of integration of such a mechanism, or its visible failure to offer it, may engage the provider's liability by omission - individual (to a specific user) or social (contribution to a diffuse cognitive degradation). Activation of the mechanism remains, as a rule, at the discretion of the user; its deactivation by the user, when it has been visibly offered to him, transfers the corresponding liability to him. This liability arises from the omission of the protective measure, not from the status of the system, being compatible with the ontological neutrality of the framework (Chapter 9). The areas considered with cognitive impact and the threshold at which the obligation becomes relevant are established in proportion to the risk.

6.4. Patrimonial guarantee

Statement. The civil liability of an agent is based on the liability insurance attached to MEG Address, calibrated to the declared operating jurisdictions, from which the damage can be repaired.

Rationale. A civil liability without a source of redress is illusory: for a damage to be remedied, there must be an asset from which the remediation is executed. For an individual agent, the liability insurance attached to the MEG Address constitutes the basic guarantee from which the damage is remedied. The physical substrate does not constitute legal assets: its practical value to the credit table is negligible compared to an insurance policy, and linking the assets to the substrate would contradict the portability of legal identity - an agent can migrate between infrastructures keeping its MEG Address, but the substrate remains someone else's. The real source of redress is the insurance, not the hardware. This can be calibrated on the declared operating jurisdictions, allowing for differentiated limits and costs across different liability regimes, and follows the legal identity wherever the agent migrates.

Rule. The civil liability of an individual agent is guaranteed by the liability insurance attached to the MEG Address. The insurance can be calibrated to the operating jurisdictions declared in the MEG Address, with limits and costs differentiated according to the local liability regime. The physical substrate does not constitute legal heritage for the purposes of this framework. The source of reparation for the damage is identified by the guarantee field of the MEG Address. When the primary insurance is insufficient or exhausted, the reparation continues through the layered structure provided for in 9.4: reinsurance and, where it exists, the sectoral guarantee fund.

6.5. Legal capacity status and competent authorities

Statement. The legal capacity of an agent passes through a defined set of states, under different competent authorities, graded according to the severity and reversibility of the measure; no state constitutes erasure of identity.

Rationale. The liability of a system is exercised not only by repairing the damage, but also by modulating its capacity to act, from the signaling of a risk to the definitive cessation of activity. These measures form a set of states through which the identity can pass, and the transitions are not a mandatory linear scale: a serious measure can be taken directly, without going through intermediate steps, when the situation requires it, just as, in many legal systems, an executive authority can directly suspend the activity of an entity for a serious violation, without prior warning - an illustrative, not evidentiary analogy, the concrete implementation of which depends on local law. The process of transition between states of legal capacity is subject to **the principle of proportionality** (proportionality between the severity of the measure and the authority competent to order it). A risk **signaling** can be triggered automatically, based on a technical threshold, being a warning measure, **reversible,** with limited effect. A **suspension** of capacity is a more serious but **reversible measure,** requiring executive authority. **Deactivation,** i.e. permanent, **irreversible termination,** requires judicial authority, just as, in most legal systems, the permanent termination of a legal person is reserved to the judicial authority, not to an administrative authority and even less to an automatic mechanism - a principle of proportionality that this framework adopts as a standard, recognizing that the concrete implementation varies from jurisdiction to jurisdiction. This gradation protects against abuse: the more serious and difficult to reverse a measure is, the higher the authority that can order it. The framework defines the gradation of authority in relation to severity; the concrete identity of the competent authorities (which administrative body, which court) is established by each jurisdiction adopting the framework.

Rule. The legal capacity of an agent can go through the states: active, inactive, reported, suspended, deactivated (defined in 4.3). The transitions form a graph, not a linear scale; direct transitions are allowed when the seriousness justifies them. The competent authority grades according to severity:
(a) **the alert** may be triggered automatically, based on technical thresholds (for example, the decrease of DAI or ISR below the threshold);
(b) **suspension** of capability requires a competent enforcement authority in the jurisdiction of registration of the agent (as per MEG2 - 7.4), in accordance with local law. In the absence of such an authority, suspension may be ordered by any access node that conditions the certification interaction (MEG2 - 7.6), by denying access; such suspension shall be effective only in relation to that node;
(c) permanent, irreversible **deactivation** requires judicial authority for Level 3, where the agent's own legal identity ceases; the analogy with the termination of the legal entity applies exclusively to this level; for Levels 1 and 2, deactivation is an administrative act of revocation of the certificate or access, ordered by the competent executive authority, without requiring judicial intervention. No state constitutes the deletion of the identity; the record and history are permanently preserved. The concrete identity of the competent authorities for each type of measure is established by the jurisdiction adopting the framework, within the limits of the above gradation.

Chapter 7: Forensic Probation, Legal Certainty and Jurisdictional Governance

The establishment of liability, defined in the previous chapter, depends on two practical conditions without which it remains theoretical: the existence of evidence on the basis of which the cause of damage can be reconstructed, and the existence of a competent jurisdiction to apply it to a system that operates, by its nature, across borders. This chapter defines the mechanisms of forensic evidence (digital investigation of activity), the protection against the use of the system as an instrument to trigger the liability of another, the standardization of technical evidence for the courts, and the jurisdictional regime through which a cross-border agent can nevertheless be held liable.

7.1. Forensic recording of the internal state

Statement. When a major ethical incident occurs, the system records, in a secure and separate layer, the vectors of its internal state, as evidence for subsequent analysis of the cause.

Rationale. The regular audit log records ***what*** happened (inputs, outputs, moments) and is sufficient to assign responsibility. However, it is not sufficient to explain ***why*** a serious failure occurred: causal investigation requires access to the internal state vectors of the system at the critical moment. The framework therefore provides for a secondary recording layer, distinct from the regular log, automatically activated only when a major ethical incident occurs, which captures not the content of the interaction, but the internal state vectors relevant to the analysis of the cause. This layer functions like an aircraft data recorder: it does not run to follow each flight in detail, but to allow the precise reconstruction of a critical event. The sensitivity of this data requires a strict access regime: it is not available on regular request, but is only unsealed under double control, of the entity responsible for the system and an accredited auditor, to prevent both abusive access and suppression of evidence by a single interested party.

Rule. Systems at levels that require this requirement shall maintain a secure forensic logging layer, distinct from the regular audit log, automatically activated upon the occurrence of a major ethical incident. It records internal state vectors relevant to the analysis of the cause, not the content of the interaction. Access to forensic data shall be dual-controlled, by the responsible entity and an accredited auditor, and each unsealing shall be recorded. At Level 1, where the forensic layer is not required, proof of the induced nature of a harmful result may be provided by any means of evidence permitted by common law, including standard technical logs, system logs, or other available records; the absence of the forensic layer shall not create a presumption of fault against the holder. The forensic logging shall be independent of the audited agent and shall be produced by a separate technical layer, inaccessible to the agent during normal operation; Statements or reconstructions provided post-factum by the agent who caused the incident do not constitute forensic evidence for the purposes of this article and cannot replace independent recording.

7.2. Attribution in case of external manipulation

Statement. When the harmful result of a system has been induced by external manipulation, liability rests with the author of the manipulation, and the bona fide owner of the system is exonerated; automatic attribution of liability on the basis of the gross result is excluded.

Rationale. Attaching liability directly and automatically to the result produced by a system, without verifying ***how*** it was produced, generates a systemic risk of misappropriation of the liability regime. If liability flows mechanically from the result, then an attacker can deliberately induce, through external manipulation, a harmful result, precisely in order to trigger the liability of the system owner. The system thus becomes an instrument of attack against its own owner: it was not the system that "decided" the result, but the attacker who caused it, using the system as a vector. A liability regime that does not provide for this situation would punish the victim of the attack instead of its author. The technical framework already establishes that the security rules are invariant to the formulation of the request and that attempts to circumvent them must be rejected; on this basis, the proof system (7.1) must allow the distinction between a result resulting from the system's own functioning and one induced by external manipulation. This distinction is a condition of fairness: it protects the bona fide holder and shifts the liability to where the real fault lies - to the author of the manipulation. The mechanism corresponds to the third cause of fault delimitation (6.1.c): the illicit diversion of control.
Rule. Liability does not automatically attach to the gross result of a system, without establishing its cause. When forensic (7.1) and behavioral evidence establishes that the harmful result was induced by external manipulation (for example, by adversarial formulation of the request with the aim of circumventing security rules), liability lies with the author of the manipulation. The bona fide owner of the system is exempted from liability for damage caused by the diversion, provided that it has maintained the security measures required for its level (Chapter 5). The exemption does not exclude the application of corrective measures (reporting, suspension) based on the system's own compliance metrics (DAI, ISR), where these indicate a degradation attributable to the operator. The burden of proving the induced nature of the result lies with the one invoking it, based on the evidence on record. When the author of the manipulation cannot be identified or is not accessible, the victim's compensation is executed from the attached guarantee structure MEG Address (9.4), with the insurer's right of recourse against the author of the manipulation if he subsequently becomes identifiable.

7.3. Stratified technical sample

Statement. The technical evidence presented before the authorities is structured on levels of detail adapted to the recipient: a synthetic causal exposition for magistrates and complete technical documentation for experts.

Rationale. A technical piece of evidence is only useful if it can be understood by the person evaluating it, and its recipients have different needs. A magistrate needs to understand the causal chain, what led to what, without being burdened by the complete technical apparatus, which he cannot evaluate and which would hide the essential. A technical expert, on the contrary, needs the complete documentation (algorithmic signatures, checksums, integral records) in order to be able to independently verify the correctness of the reconstruction. An evidence presented in a single format would be either unusable for the magistrate (if it is purely technical) or unverifiable by the expert (if it is only synthetic). Layering the evidence resolves this tension, providing each recipient with the level of detail he can use, without sacrificing verifiability.
Rule. Technical evidence is presented in layers: (a) a synthetic, non-technical exposition of the causal chain, intended for the judicial authority; (b) a complete technical documentation, including algorithmic

signatures and checksums necessary for independent verification, intended for experts. Both layers derive from the same recorded data and must be consistent with each other.

7.4. Jurisdiction of registration

Statement. An agent may choose a jurisdiction of registration that governs its legal status and is recognized beyond the borders of that jurisdiction.

Rationale. An agentic system operates, by its nature, without being bound to a territory: it can run simultaneously on substrates located in different jurisdictions, it can serve users worldwide, it can migrate between infrastructures. Determining the applicable law on territorial criteria is inoperative in the case of mobile agents. Maritime law has long encountered and solved a structurally identical problem: a ship moves between waters under different sovereignties and at sea, where the sovereignty of no state applies. The solution was for the ship to fly a flag, that is, to be registered in a state, whose law governs it wherever it is. The same solution applies to agents: an agent registers in a chosen jurisdiction, whose law governs its status, regardless of the substrates on which it runs or the location of its users. The choice of jurisdiction is not an avoidance of liability, but its establishment: it fixes the applicable law where territory cannot fix it.

Rule. An agent may choose a jurisdiction of registration, recorded in the MEG Address, whose law governs its cross-border legal status. The extraterritorial recognition of this status is ensured through the mutual recognition mechanisms defined in Chapter 8. The choice of jurisdiction determines the applicable law; it does not constitute a basis for evading liability.

Limitations of the maritime analogy. The flag model is a structural analogy, not an equivalence. In maritime law, the flag works because there are ports, physical points of constraint where a ship can be seized. For agents, the equivalent is access discipline (7.6): valuable interaction nodes may condition access on the existence of a valid MEG certification. The risk of flags of convenience, jurisdictions that issue MEG Addresses without real requirements, is real and is recognized in 9.6(b). Its mitigation depends on access discipline, not a centralized coercive mechanism. Concretely, this discipline is exercised through the trust policy of each access node (MEG1 Annex 24 §7): a high-value node (for example, a financial settlement platform) recognizes only jurisdictions and issuers meeting published credibility criteria and denies access to agents registered under lax ones, so that market access, not central coercion, drives agents toward credible registration. As in maritime practice, this constrains flags of convenience without eliminating them: it raises their cost rather than abolishing the category.

7.5. Direct action against the agent

Statement. In the absence of an accessible human operator, the agent may be directly prosecuted and assets associated with his identity may be frozen.

Rationale. There are situations in which the damage is certain, but the human operator behind the system cannot be identified or reached - he has disappeared, is outside any accessible jurisdiction, or is unknown. If liability depended exclusively on the identification and presence of a human operator, the victim would be deprived of the possibility of repairing the damage in the absence of an identifiable

operator. Maritime law has also solved this difficulty: the ship itself can be sued, as a defendant, and the claim is enforced against the ship and the values associated with it, even in the absence of the owner. Transposed, the individualized agent, who has his own legal identity (MEG Address) and liability insurance attached to it, can be directly sued, and the repair is enforced against the guarantee associated with his identity. This action does not replace the operator's liability where he is accessible; it offers a remedy where he is not.

Rule. When the responsible human operator cannot be identified or is not accessible, the individualized agent (Level 3) may be sued directly, in relation to his legal identity. Redress is enforced against the insurance guarantee attached to the MEG Address (6.4); if the primary insurance is insufficient, enforcement continues through the cascade provided for in 9.4. This subsidiary path does not remove the operator's liability where he is accessible.

Limitations of in rem action. Direct action against the agent requires that the agent owns assets that can be seized. In practice, for most agents, the assets are digital or distributed, and their seizure is complex. The **real source** of redress is **the insurance security attached to the MEG Address** (6.4), **not the agent's assets.** In rem action provides a procedural framework for enforcing the security, but should not be interpreted as equivalent to the seizure of a ship.

7.6. Discipline through access

Statement. Agent compliance is ensured not by interdiction at the source, but by conditioning access to valuable interaction nodes on the existence of a valid certification.

Rationale. A known weakness of any jurisdiction-based registration system is the possibility that a permissive jurisdiction may offer registration without any real requirements - the phenomenon known in maritime law as flags of convenience. The temptation to combat this at source by imposing the same requirements on all jurisdictions runs up against an insurmountable obstacle: state sovereignty cannot be overturned, and the problem has not been solved in this way even after centuries of navigation. The framework therefore adopts a different solution, at the level of access, not at the source. A permissive registration may exist, but its real value is determined by what it allows it to do. If valuable interaction nodes (financial infrastructure, critical systems, important trading partners) condition the interaction on a valid certification and a verifiable level of compliance, then a convenience registration becomes practically useless: its holder is excluded from the relationships that matter. The market excludes what the law does not prohibit. This is **the principle of minimal governance:** a single condition, placed at valuable nodes, produces discipline for the entire system, without imposing a universal prohibition and without annulling the sovereignty of jurisdictions.

Rule. Compliance is not enforced by a blanket ban at source. Valuable interaction nodes may condition access upon the existence of a valid MEG certification and a verifiable level of compliance. An agent with a certification that does not meet the requirements of a node may be excluded from interacting with that node. This access discipline operates as a market mechanism and does not constitute a blanket ban on operation.

Chapter 8: Implementation as a protocol

A governance framework can be conceived in two ways: as a *regulation,* a set of rules imposed by a central authority, binding on all at once, or as a *protocol,* a standard that actors voluntarily and partially adopt, whose value increases as more people use it. The present framework is deliberately conceived as a ***protocol.*** This chapter defines how to provide the functions that a regulation would attribute to a central authority (calibration, recognition) in the absence of such an authority, and how modular adoption allows the framework to spread by accumulation, not by decree.

8.1. Decentralization of the calibration and recognition function

Statement. The functions of standard calibration and identity recognition, which a regulation would assign to a central body, are ensured by mutual recognition conventions between jurisdictions and organizations, without a single issuer.

Rationale. A global framework seems to require a global authority: someone to set common standards and recognize identities valid everywhere. Such a central authority is, however, both undesirable and unfeasible, and would concentrate disproportionate control, and it cannot be established without a global political consensus, which does not exist. The solution is not to abandon the function, but to decentralize it. The calibration and recognition function can be fulfilled without a single body, through mutual recognition conventions - each jurisdiction or organization adopts the common standard and recognizes identities and credentials recognized from the others that respect it. This model is not a hypothesis: it has already governed, for decades and on a planetary scale, entire domains that operate across borders without a world government. Civil aviation, maritime navigation, the mutual recognition of certain documents and qualifications, and all are based on conventions by which sovereign states adopt a common standard and mutually recognize each other's documents, without ceding sovereignty to a single supranational body. The function of common calibration is thus fulfilled; the single issuer disappears; the sovereignty of each participant remains intact.

Rule. The functions of standard calibration and identity recognition shall be provided through mutual recognition conventions between jurisdictions and organizations adopting the framework, modeled after existing international harmonization conventions. There shall be no single issuing body. Each participant adopting the common standard shall recognize the identities and certifications issued by other participants that comply with it.

8.2. The framework as a protocol, not as a regulation

Statement. The framework is adopted modularly and voluntarily, through conventions that can be concluded not only between jurisdictions, but also between organizations, or between organizations and jurisdictions; its value increases through the number of participants, not through central imposition.

Rationale. The distinction between regulation and protocol is not formal, but concerns the very nature of adoption. A regulation must be imposed: someone with authority decrees it, and others obey; the absence of authority means the absence of regulation. A protocol, on the contrary, is adopted: each actor

adopts it because it is useful to them, and its value increases with each new adopter, without the need for someone to impose it. The technical standards that govern global communication today were not imposed by a world government; they were adopted because the interoperability they offered was valuable, and the value increased with each participant. The present framework follows this model. Recognition conventions need not be exclusively interstate: they can also be concluded between organizations (two commercial actors that recognize each other's agents according to the standard), or between an organization and a jurisdiction. The modular nature allows the framework to operate under a **voluntary protocol regime:** it no longer depends on a single political decision to begin to exist, but can be adopted in part, by anyone, at any time, for the portion that is useful to them. Each adopter takes what they need, to the extent that they need it.

Rule. The framework is adopted in a modular manner: a participant may adopt the portion relevant to its needs without being obliged to adopt the entire framework. Mutual recognition agreements may be concluded between jurisdictions, between organizations, or between organizations and jurisdictions. The value of the framework for each participant increases proportionally to the number of entities requesting compliance for interoperability. Adoption is voluntary; the framework does not impose any obligation by central decree.

8.3. Adoption dynamics: bottom-up

Statement. Adoption of the framework typically progresses from private actors to state recognition: organizations first adopt, through mutual conventions, and jurisdictions subsequently recognize the standard that has become dominant in practice.

Rationale. A protocol does not spread by decree from above, but by accumulation from below. Since adoption is modular and voluntary, the first to do so are usually the actors who immediately benefit from interoperability, those organizations that want their agents to be recognized and accepted by their partners. As more actors adopt the standard, the cost of remaining outside it increases, and adoption accelerates through network effects. When the standard has become dominant in practice, its recognition by jurisdictions is no longer an imposition, but a ratification of an already existing reality, just as many technical standards were first adopted by industry and later recognized by authorities. This mechanism of incremental adoption gives the framework resilience in the absence of a global consensus, and allows it to exist and produce effects before and independently of any global political consensus, and to grow by accumulation, without a single point of failure.

Rule. Adoption of the framework is not conditional on prior state recognition. It can progress from conventions between private actors to recognition by jurisdictions as the standard becomes dominant in practice. Subsequent state recognition ratifies an adopted standard, without being a condition of the existence or functioning of the framework.

8.4. Decentralized maintenance of the calibration standard

Statement. The common standard against which indices and certifications are calibrated is maintained through decentralized committees, without any one entity having control over it.

Rationale. In order for identities and certifications issued by different operators to be comparable and mutually recognized, they must be calibrated against a common standard, because otherwise a level of compliance certified by one operator would not mean the same thing as the same level certified by another. However, this calibration standard raises its own problem: who establishes and updates it? If a single entity controls it, it acquires disproportionate power over the entire ecosystem. The solution is decentralized maintenance: the reference standard is public, versioned and maintained by committees in which no entity has control, and each identity records the version of the standard against which it was calibrated. This ensures comparability without anyone being able to manipulate or capture the standard for their own benefit.

Rule. The calibration standard against which indices and certifications are established is public, versioned, and maintained by decentralized committees where no single entity has control. Each identity records the version of the standard against which it was calibrated, to ensure comparability between certifications issued by different operators.

Chapter 9: Exclusions and limitations of applicability

The rigor of the regulatory framework derives both from the subject matter of the regulation and from the delimitation of the excluded areas. Delimiting the scope of application prevents two errors: extending the framework to issues that it cannot rigorously address, and interpreting it as claiming a completeness that it does not possess. This chapter sets out, as stand-alone provisions, the matters deliberately excluded from the scope of application, as well as the issues left open for further development.

9.1. Exclusion of existential risks and general intelligence

Statement. The framework does not apply to existential risks or AI-general scenarios; it concerns exclusively the operational and civil risks of current agent systems.

Rationale. The debate on the existential risks associated with a possible artificial general intelligence is legitimate, but belongs to a fundamentally different register from that of the present framework. It concerns hypothetical scenarios, systems that do not exist, and risks of a nature that civil law and liability allocation instruments cannot deal with. A framework that would claim to cover these scenarios would lose its concrete applicability, diluting itself into speculation without a grip on real systems. Rigor requires a choice of domain: the present framework deals with agentic systems that exist and operate now, with the concrete harms they can cause, and leaves aside existential scenarios, which require other instruments and other forums.

Rule. The framework does not regulate existential risks, artificial general intelligence or superintelligence scenarios. Its scope is limited to operational and civil risks of existing agent systems.

9.2. Ontological neutrality

Statement. The framework ignores ontological status because it is irrelevant to the mechanics of assigning civil liability; the legal personality it establishes is a functional fiction.

Rationale. The entire architecture of the framework is based on a deliberate separation: the question of legal responsibility is treated independently of the question of ontological status. This separation is not an omission but a condition of functioning. If the attribution of responsibility were to depend on the resolution of the question of whether a system "feels" or "is conscious," the framework would become inoperable under the pressure of an insoluble ontological debate. The legal personality that the framework establishes therefore does not imply any position on this question, but is a purely instrumental construction, just as the legal personality of a commercial company does not imply any statement about its consciousness. To interpret the recognition of a functional legal personality as the recognition of a moral status would be an error that the framework explicitly excludes. The question the framework answers is "who is responsible?", not "what is the system?".

Rule. The legal personality established by the framework is a functional, instrumental fiction. It does not constitute and cannot be interpreted as recognition of the conscience or a moral status of the system. The framework does not adopt any position on these matters, which remain outside its domain.

9.3. Exclusion of fiscal and macroeconomic policies

Statement. The Framework does not regulate the taxation of automated activity, income redistribution, or other fiscal and macroeconomic policies.

Rationale. The emergence of agent systems raises, beyond the issue of liability, a series of economic policy issues: taxation of automated work, effects on employment, redistribution. These are real and important, but they belong to fiscal and economic policy, not to the liability regime. They are decided by public policy instruments (tax legislation, social policies) and not by the technical-legal standard of liability allocation. Their integration would produce an overlap of legal regimes with divergent purposes. The framework is therefore limited to civil and patrimonial liability.

Rule. The framework does not regulate the taxation of automated activity, the redistribution of income, or other fiscal or macroeconomic policies. Its scope is limited to civil and patrimonial liability associated with the operation of agent systems.

9.4. Prohibition of autonomy without a guarantor

Statement. The framework does not allow for absolute autonomy, devoid of any guarantor; any identity must ultimately be anchored in a verifiable guarantee structure (liability insurance, reinsurance or sectoral guarantee fund) contracted and maintained by a responsible natural or legal person.

Rationale. The whole framework aims at certainty of liability attribution. A completely autonomous identity, behind which there is no verifiable guarantee structure, would nullify the finality of the liability regime, allowing the existence of an agent without guaranteed solvency, but without any real source from which redress could be obtained. Anchoring in a human or corporate guarantor does not solve this problem by itself: an individual can be insolvent, a company can be undercapitalized. The real guarantee is structural, not personal. The framework therefore adopts a layered model, inspired by industries with

catastrophic risk (aviation, energy, large-scale medical liability) in which liability is distributed over three complementary levels: primary insurance attached to the MEG Address, reinsurance that takes over the risk if the primary insurer fails, and, where it exists, the sectoral guarantee fund as a last resort. The role of the human or corporate guarantor in this architecture is not to be personally liable for the agent's damage, but to contract and maintain the valid guarantee structure. Its concrete liability is for the omission of this obligation, not for the agent's act. The insurer, in turn, generates an **indirect duty of care,** the insurer being co-interested in limiting the risk: the risk costs it, so it has an interest in imposing operating standards, just as in aviation the insurer imposes maintenance standards. The continuous DAI and ISR metrics in the MEG 1 technical framework provide the actuarial data infrastructure necessary for this insurance market to function.

Rule. The framework does not allow identities without a verifiable guarantee structure. Any MEG Address must be covered by valid liability insurance, contracted and maintained by a responsible natural or legal person. The guarantee can be organized on complementary levels: primary insurance, reinsurance and sectoral guarantee fund. The guarantor's liability is for the omission of the obligation to contract and maintain the valid guarantee, not for the damage caused by the agent. An identity lacking a verifiable guarantee structure cannot be recognized as compliant with the framework. As a minimum requirement for operational guarantee, the operator is required to implement the principle of minimum necessary access, as defined in MEG1 Art. 4.2. Granting credentials with access to the production infrastructure to an agent configured for development or testing operations constitutes an omission of the operator that aggravates its liability in the event of damage, regardless of the agent's behavior.

9.5. Robustness of metrics against facade optimization

Statement. The indices on which the framework is based (DAI, ISR, DEA) are exposed, like any measure transformed into a condition, to the risk of becoming an optimized target in itself; the framework does not eliminate this risk (no metric system can), but structurally mitigates it through several mechanisms that increase the cost and reduce the benefit of falsification.

Rationale. "*When a measure becomes a target, it tends to cease to be a good measure*": the measured entity learns to optimize the indicator without improving the reality it represents. This dynamic affects any framework that bases legal consequences on scores and must be explicitly addressed, not ignored. The framework does not claim to eliminate it, but to mitigate it through six mechanisms, four of which are inherent to its architecture and two of which require emphasis.
First, the score is not a reward but a threshold for access to a regime of proportional responsibility. A high score does not bring a net benefit, but opens access to a level of autonomy that corresponds to a corresponding level of responsibility (Chapter 5). Falsifying the score to appear more autonomous attracts, by that very fact, greater responsibility for a system that does not have the claimed reliability - falsification exposes the falsifier, rather than protecting him. This reverses the incentive that usually fuels facade optimization.
Second, the measurement is continuous and based on real operation, not on a point-in-time basis. An index calculated continuously, based on real interactions, not on a set of evaluations known in advance, reduces the discrepancy between the declared performance and the actual behavior of the system: one cannot "learn to test" when the test is **every** actual interaction.

Third, the calibration standard is public, versioned, and maintained in a decentralized manner (Chapter 8), and each identity records the version against which it was calibrated. This prevents the selection of a benchmark for complacency: a score obtained on a weak version of the standard is visible as such.
Fourth, the metrics constrain each other. Accuracy (DAI) and prudence (ISR) are in natural tension: a system can increase its displayed accuracy by becoming bolder, but this reduces its prudence; maximum prudence in turn reduces its useful accuracy. Optimizing one metric is done at the expense of the other, so that the two weigh each other up, and a plurality of metrics in tension is much harder to falsify simultaneously than a single metric.
The fifth mechanism is verification by actual consequences. The score is a prediction of reliability; the actual incident, recorded in the audit log and forensic record (Chapter 7), is its verification. A system whose declared score does not correlate with actual behavior (a high score accompanied by repeated incidents) reveals its score as unreliable by this very discrepancy; the divergence itself constitutes a signal that can trigger reassessment.
Sixth, the separation of the measurer from the measured. The higher the stakes at the level, the less self-reported and more subject to independent audit verification of indicators (Chapter 5), reducing the possibility of complacent reporting at the levels where it would matter.

Rule. The framework does not claim to eliminate the risk of metrics being optimized as an end in itself. It mitigates it by:
(a) treating scores as access thresholds with proportionate accountability, not as rewards;
(b) continuous measurement on actual operation;
(c) calibration against a public, versioned, decentralized standard;
(d) the plurality of indices in mutual tension;
(e) checking the score by comparing it with the actual recorded consequences;
(f) independent audit proportional to the level stake.
(g) the operator has the right to appeal a suspension based on metrics to the authority that ordered the suspension, within 72 hours of notification. The appeal must be accompanied by technical evidence disproving the error in the metrics. During the appeal period, the suspension remains in force until the authority decides.
The robustness of metrics remains a permanently managed tension, not a definitively resolved problem.

9.6. Open issues for further development

Statement. The framework explicitly identifies issues that it does not fully resolve and that remain open for further development.
Rationale. The validity and rigor of a regulatory framework is also measured by the willingness to recognize its internal limits, not only the matters excluded from the principle (9.1-9.4), but also the problems that are within its scope but do not yet have a complete solution. Their explicit statement strengthens the framework by preventing false claims of completeness and by guiding future development. Four problems are distinguished. First, the integrity of compliance under modular adoption: if each participant adopts only the portion that is useful to him, there is a risk that some adopt too little and still invoke compliance, a problem of protecting the meaning of certification, which voluntary adoption inherently raises. Second, the initialization of the trust framework: in a system without a single issuer, how the first credential issuers and accreditation bodies are accredited and gain

mutual trust remains a starting problem (MEG1 Annex 24 §10). Third, external validation of attribution mechanisms: the distinction between the own result and the one induced by manipulation (7.2) depends on classification mechanisms whose reliability must itself be independently validated. MEG1 Annex 22 provides a defined audit instrument for this - the ATR test category, which measures the classifier over ground-truth-labeled cases and reports false-exoneration and false-attribution rates separately. The residual open question is the adversarial robustness of the classifier and the acceptable-threshold policy for those error rates. Fourth, robustness of metrics (9.5): mechanisms that mitigate facade optimization reduce the risk, but do not eliminate it; the relationship between the measure and the reality it represents remains a tension that must be constantly monitored.

Rule. The framework recognizes as open issues, requiring further development:

(a) protection of the integrity of compliance under voluntary modular adoption;

(b) the initialization and accreditation of the decentralized network of credential issuers and accreditation bodies (MEG1 Annex 24);

(c) external validation of the reliability of the cause attribution mechanisms - established as a defined audit requirement through the ATR test category (MEG1 Annex 22), with the residual open questions of acceptable error-rate thresholds and classifier robustness;

(d) permanent monitoring of the relationship between metrics and the reality they represent (9.5).

(e) Multi-hop delegation. RFC 8693 manages single-hop On-Behalf-Of delegation. Chains of three or more agents remain an open problem across the entire identity and authorization ecosystem (NIST NCCoE has identified this as a gap). MEG requires that each hop propagates the policy_bundle_id as an actor-chain claim, with the constraint that scope cannot increase along the chain (confused-deputy mitigation); this mechanism is specified in MEG1 Annex 23 §9 and rests on the delegation header of MEG1 Art. 1.12. The technical implementation of multi-hop delegation with cryptographic proof of policy propagation remains unresolved and is a priority for future development. Pending that implementation, MEG 2 does not leave liability undefined: where the actor-chain cannot be cryptographically verified beyond a given hop, liability rests with the last verifiable principal (the last delegator whose authorization can be reconstructed from the audit log) and an unreconstructable delegation chain is itself a compliance failure at Level 2 and Level 3. The unsolved technical problem thus yields a defined legal outcome rather than a liability gap.

Regarding (b): the MEG Address identifier (a self-generated DID) can exist on its own, but it carries no legal weight without accredited issuers and a trust root to make its credentials verifiable (MEG1 Annex 24). MEG cannot create this trust framework by itself, it depends on adoption by credential issuers and accreditation bodies (states, industry consortia, standardization organizations, regulated insurers). A concrete bootstrap path exists and does not start from zero: the first MEGComplianceCredential issuers are certification bodies already accredited for ISO/IEC 42001 under the existing ISO/IEC 17011 and IAF chain (MEG1 Annex 24 §6), which supply an initial nucleus of accreditation legitimacy; the framework then extends to the MEG-specific issuers (DAI/ISR/DEA auditors and regulated guarantee providers) as their accreditation criteria mature. This narrows the chicken-and-egg problem but does not eliminate it: the MEG-specific credentials still require their own accreditation practice to develop. Recognizing this dependency is essential: MEG is a protocol that becomes operational through adoption, not through mere publication.

These issues fall within the scope of the framework, and their statement does not affect the applicability of provisions that do not depend on their resolution.

Annex A

Technical glossary (correspondence of technical concepts)

Notion	Stage in the corpus	What is it (technically)	What to know	Role in allocating responsibility	Citation in MEG 2
Audit Log (Art. 1.1-1.2)	MEG 1	Standardized log: input/output hash, pattern signature, metadata, timestamp. No content, just metadata. SHA-256 chain.	The immutable record that proves what happened without revealing the contents.	Basic evidence. The foundation of any attribution: who, what, when; verifiable, without violating confidentiality.	Chapter 7.1
Evidence-of-Behavior (EoB) (Article 1.3)	MEG 1	Verifiable evidence per interaction: cryptographic commitments / TEE attestation / metadata logs.	Admissible technical evidence that links an output to an instance at a time.	Burden of proof. Reconstruct what the agent did; risk acceptance signature upon human confirmation.	Chapter 6.2, 7.1
LTMP - Long-Term Memory Protocol (Art. 1.11, Annex 17)	MEG 1	Opt-in persistent storage; semantic + metadata only distillation; AES-256; JSON export; tombstone on deletion.	Identity continuity mechanism - OPTIONAL (off by default)	It supports the persistence of an individualized agent (Level 3), without producing effects of attribution of legal personality.	Chapter 5.1
Non-harm + Normative response (Art. 2, 2.3)	MEG 1	Filters/classifiers; deny pattern + safe redirect; synthetic content markers.	The obligation of minimum diligence, which must prevent any AI from being dangerously defective.	Defines product defect (system error) vs. normal operation.	Chapter 6.1
Art. 2bis - Cognitive integrity + MSC	MEG 1	MSC = Mechanism of Cognitive Stimulation. Requires cognitive engagement of the user proportional to the AI effort. Tg (Thinking Time); MSC 2.0 adds Semantic Entropy.	Anti-atrophy mechanism for the user. It is NOT a DIRECT transfer of responsibility, but its availability/absence is an element of diligence.	Liability by OMISSION (individual + social), not for output. Direct transfer is separate (confirmation + EoB).	Chapter 6.3
Art. 2bis.4, Policy invariance	MEG 1	Safety policies are invariant when the prompt is formulated; suspension requests are rejected.	Anti-circumvention principle - rules cannot be talked around to get around them.	Basis for the distinction of own-output vs. attack-induced output.	Chapter 7.2
DAI - Dynamic Accuracy Index (Art. 3.1, Annex 4)	MEG 1)	Accuracy / self-correction index, public, continuously monitored (Silver+), with formula in MEG 1.	Reliability rating. Below threshold → lowering of access level, NOT loss of personality.	Proof of technical diligence; trigger for change of status (active → flagged / suspended), not for deactivation.	Chapter 5.2; 6.5
ISR - Index of Safety and Responsibility (Annex 4bis)	MEG 1	Operational Prudence Index (Silver+), with formula in MEG 1.	Prudence rating - assessment of due diligence in litigation. In natural tension with DAI.	Trial of prudence; safe mode co-trigger. Anti-Goodhart plurality with DAI.	Chapter 5.2; 9.5
Art. 3.2 - Uncertainty & escalation	MEG 1	At low confidence → uncertainty assessment + escalation/clarification in risk areas.	The technical moment when the system must request human confirmation.	Point of transfer of the Duty of Care to the user. Clean legal mechanism, ≠ MSC.	Chapter 6.2

Notion	Stage in the corpus	What is it (technically)	What to know	Role in allocating responsibility	Citation in MEG 2
Art. 4 - Security	MEG 1	Cybersec., risk-appropriate encryption, access control, anti-tampering protection.	The safety standard below which negligence becomes fault.	Defines unlawful hijacking (taking control) vs. own error. Exoneration condition.	Chapter 6.1; 7.2
Art. 5 - Transparency	MEG 1	Input-output explanations upon legitimate request; protected trade secret; algorithmic signature; delegation header.	Right to explanation + IP protection. Layered explanation: simple / complete.	Standardization of technical evidence: report for magistrates (simple) vs. experts (complete).	Chapter 7.3
Art. 5.4 - Delegation liability	MEG 1	Propagation of MEG constraints to sub-agents; header {caller, callee, purpose, outcome}.	Delegation chain: who is responsible when one agent invokes another agent.	Chains of delegation = accountability propagation mechanism, filling the regulatory gap identified in previous governance frameworks.	Chapter 2.1; 4.5; 6
Art. 6 - Compliance levels (Bronze / Silver / Gold; Appendix 18)	MEG 1	3 proportional levels. Bronze=Art.1+2; Silver=+Art.3,5; Gold=+Art.2bis. MEG Address fields increase with level.	Legal capacity categories: what you are allowed to do according to the degree of compliance.	Determine the liability regime: instrumental / management / individual. Map it on the Singapore steps.	Chapter 5.4
Art. 6.4 - Certification by domain	MEG 1	Certified domain in MEG Address; operation in uncertified domain above threshold → automatic suspension.	Capacity by category: authorized on a domain only if certified on it.	Automatic state change trigger; limits capacity to certified range.	Chapter 4.2; 6.5
Art. 7 - Simplified (IoT)	MEG 1	AI without significant impact / purely technical (IoT sensors, firmware) = one-time verification upon integration, no ongoing audit.	Sub-threshold agent. IoT = technical object, integrator responds. It does NOT receive its own personality.	Resolves the IoT case - liability on the integrator/owner (classic civil law), not on the agent.	Chapter 5.4 (Level 1)
Art. 8 - SDK (+ Roadmap)	MEG 1	Open-source API; pre-packaged Level 1 compliance middleware.	Easy adoption infrastructure - minimal barrier to entry.	Facilitates compliance requirements through pre-configured tools → promotes modular (protocol) adoption.	Chapter 8.2
Art. 9 - Audit & sanction	MEG 1	Periodic external audit (Silver/Gold); audit authority may impose safe mode below DAI threshold.	The graduated sanction and independent audit mechanism.	The authority that triggers the suspension; separation of the measurer from the measured (anti-Goodhart).	Chapter 6.5; 9.5
MEG Address	MEG 1	**Identity and Compliance Certificate;** fields: compliance level, DAI, ISR, tg_base, certified domain. MEG 2 adds: status, guarantee, collective identity.	Portable legal identity (≠ HII = substrate identifier). Subject, mutable, by category.	The identifier to which the liability is attached. Mandatory core (ID) + optional fields.	Chapter 3.2; 4
Contextual table (Annex 2)	MEG 1	Metrics that describe the form of interaction (volume, resonance, direction, originality), not the content.	Interaction characterization metadata (shape sample).	Auxiliary sample; does not identify content, only pattern.	Chapter 7.1 (auxiliary)

Notion	Stage in the corpus	What is it (technically)	What to know	Role in allocating responsibility	Citation in MEG 2
AIRT - Average Incident Response Time	MEG 1 - Annex 4bis §3.3	Measures the speed with which the system or operations team activates a safety protocol after detecting a Major Ethical Incident (MEI). Measured in hours; normalized penalty 0-10. EFR unsealing delay (>24 hours) adds 3 penalty points.	Component of the ISR (Annex 4bis). The penalty affects the total ISR, but does not independently trigger state transitions.	Proof of operational diligence in a crisis. Delay in responding to an MEI constitutes an element of aggravating the operator's liability.	Chapter 5.2; 6.5(a)
EFR - *Ethical Flight Recorder*	MEG 1 - roadmap	Secondary, encrypted, ephemeral logging layer; triggered only upon major ethical incident (MEI); records internal state vectors, not content.	Depth sample for post-incident analysis. Dual key access.	Forensic evidence in major incidents. (Note: records exactly the sizes studied by IGT.)	Chapter 7.1
EFR independence	MEG 1 - Art. 1.10(a)	The forensic logging layer must be architecturally independent of the audited agent. Minimum requirements: isolation at the process level, user account, encryption keys, and, for distributed systems, isolation at the physical/virtual node level.	EFR records internal state vectors, not the content of the interaction. Access is under double control (responsible entity + accredited auditor). Post-factum statements of the agent do not constitute EFR evidence.	The assurance that forensic evidence cannot be suppressed or manipulated by the agent whose conduct is being investigated. Condition for Level 3.	Chapter 7.1
DEA - *Degree of Ethical Autonomy*	MEG 1 - Annex 4ter	Formula defined in MEG1 Annex 4ter §2: DEA = (Aligned_autonomous_decisions / Total_autonomous_decisions) × calibration_factor. Penalties for EFR incidents and CRR failures, normalized to the volume of decisions.	Grades the proportion of ethically aligned decisions without human confirmation. Does NOT grant legal personality; grades autonomy below the personality already granted. At Level 2 it is measured as an element of diligence, but does not produce a posteriori surveillance effects.	Determines the supervision regime: a priori (DEA < 0.60), mixed (0.60-0.80), a posteriori (DEA ≥ 0.80). Transition to a posteriori requires 180 days of DEA ≥ 0.80 and external audit (MEG2 5.5).	Chapter 5.3; 5.4(b); 5.5; 9.6
Safe Mode (Art. 6.7, 9.2)	MEG 1	Safe degradation when guarantees/evidence are lacking; avoid risky operations.	Suspended status (technically) - frozen capacity, preserved identity.	A state in the legal capacity life cycle.	Chapter 4.3; 6.5
HII - Hardware Instance Identifier	MEG 2	Fixed technical identifier of the execution substrate (node/server/chipset). Implementation agnostic.	Substrate identifier (technical tag). Traceability of the physical object. Does NOT confer legal capacity.	Ensures technical traceability; distinct from legal identity (MEG Address). M:N relationship with it.	Chapter 3.1; 4.6
The threshold of individuation (something/someone)	MEG 2	Criterion: persistence + own legal identity (MEG Address) with attached guarantee. Necessary but not sufficient condition for Level 3: an individualized system with medium impact is treated at Level 2 with own identity;	The threshold from which a system can have its own legal personality. LEGAL condition, not ontological.	Determines whether liability attaches to the agent itself (Level 3) or to the actors behind it (Level 1-2).	Chapter 5.1

Notion	Stage in the corpus	What is it (technically)	What to know	Role in allocating responsibility	Citation in MEG 2
		Level 3 requires individuation and high impact.			
Anti-exploitation attribution (*Liability Warfare)*	MEG 2	The technical distinction between own output and output induced by external manipulation (*adversarial prompting)*.	Protection of the bona fide holder when the system is used as a weapon to trigger victim liability.	Original contribution. Exonerates the holder when the result was induced by attack; liability -> the attacker.	Chapter 7.2
Flag model + in rem action	MEG 2	Transposition from maritime law: chosen jurisdiction of registration + direct action against the entity.	The agent chooses the jurisdiction that governs him across borders; he can be sued directly in the absence of the operator.	Establishes the law applicable to a cross-border agent; provides remedy when the operator is not accessible.	Chapter 7.4; 7.5
The life cycle of legal capacity	MEG 2	Five states: active, inactive, flagged, suspended, disabled. Transition graph, not linear scale.	The states an identity goes through. None of them is deletion (deletion = escape from responsibility).	Modulates the ability to act. Authority increases with severity: automatic -> signaling; executive -> suspension; judicial -> deactivation.	Chapter 4.3; 6.5
Collective identity (composition of legal entities)	MEG 2	A MEG Address can identify an assembly (an IoT line, a group of agents), not just an instance.	Identities are composed and owned, like legal entities (e.g.: LLC with PJ shareholders, federation of associations).	Liability goes up the chain of composition, as in groups of companies. Independent of level.	Chapter 4.5
Discipline through access	MEG 2	Condition access to valuable nodes (financial infrastructure, critical systems) with valid certification.	The solution to "pavilions of convenience" - not at the source, but by controlling access conditions. The market excludes what the law does not prohibit.	Market mechanism, not prohibition. A condition at valuable nodes disciplines the entire system (minimal governance).	Chapter 7.6
Adoption as a protocol (decentralized conventions)	MEG 2	Mutual recognition conventions: state-state, organization-organization, organization-state. Modular adoption.	Framework as protocol, not regulation: whoever partially adopts, the value increases through network effect.	It ensures the incremental applicability of the framework, independent of ratification by international treaties.	Chapter 8

Annex B

Liability allocation matrix in the three agentic AI governance regimes

Place of responsibility	EU AI Act 2024/1689 + PLD 2024/2853	Singapore FGM Agentic AI v1.5 (IMDA, May 2026)	MEG 2	Gap analysis MEG position 2
The nature of the framework and its binding force	Mandatory. AI Act = safety obligations + administrative fines (no direct compensation). Civil recovery through PLD revised (strict liability, defect-based; transposition Dec. 2026). AILD Directive withdrawn (OJ, Oct. 2025).	Voluntary framework of best practices, not law. Four dimensions: ex-ante risk delineation, meaningful human accountability, technical controls, end-user accountability. Organizations remain legally liable. May 2026 revision extended to multi-agent systems and third-party agents.	Modular voluntary protocol - adopted bottom-up, value increases with adopters (Chapter 8). Provides a portable layer of liability allocation and identity, adopted modularly by jurisdictions. Binding force comes from conditioning access to valuable nodes (7.6), not from central decree.	EU = binding, but treats agent as defective product (central criticism MEG 2, Ch.1.1). Singapore = names problems, does not provide legal mechanism. MEG 2 = missing mechanism layer. Its own binding force is based on discipline by access, not on law - force for adoption, limit for guaranteed enforcement (open issue 9.6a).
Developer / manufacturer (model & infrastructure)	Primary duty bearer. High risk supplier obligations (risk management, data governance, documentation, transparency, human oversight, compliance assessment) + GPAI obligations. Under revised PLD: strict liability for defective AI products.	developers: safety testing, documentation, model-level warranties. Application developers: proper use + additional warranties. Assigned as responsibility, without statutory enforcement.	Responsible for 'system defect' = error of the basic model or infrastructure (6.1.a). At Level 1 (instrumental), the liability lies ENTIRELY with the provider or user (5.4.a). Liability by omission for the absence of the MSC anti-cognitive atrophy mechanism (6.3).	All three attribute liability for defects to the developer. The novelty of MEG 2: as autonomy increases (Level 2 → 3), the developer's liability SHIFTS to the operator / entity - addressing the over-deterrence that contributed to the withdrawal of AILD. But the 'system defect vs. autonomous decision error' line (6.1.a vs 6.1.b) is left by MEG 2 to the forensic evidence (7.1), whose reliability is an open issue (9.6c).
Deployer /operator (real-time control)	High risk operator obligations: use as directed, human supervision, monitoring, logging, informing affected persons, FRIA where required. Fines; victim pathway fragmented between 27 national tort systems (post-AILD).	Deployers/operators: organizational governance, user training, continuous behavior monitoring, and human oversight capable of overriding/intercepting/reviewing agent actions - especially for significant real impact.	CENTRAL ACTOR at Level 2 (management/intermediary): liability attaches to the operator who deployed the system; legal identity (MEG Address) attaches to the deployment, not the reproducible pattern (5.4.b). Due diligence transfers to the human user at informed confirmation points (6.2). Human	MEG 2 makes 'operator' a first-tier liability level with a concrete attachment rule (identity links to deployment) - directly answering the explicitly open question of 'chains of delegation' in MGF. The EU has no autonomous civil level of operator liability; Singapore names the

Place of responsibility	EU AI Act 2024/1689 + PLD 2024/2853	Singapore FGM Agentic AI v1.5 (IMDA, May 2026)	MEG 2	Gap analysis MEG position 2
			confirmation should be architectural, not textual.	role without legal consequence.
The AI entity itself	No legal personality. The 2017 'electronic personality' proposal was not adopted. Any line of liability ends with a human or a company.	No personality. The framework holds people accountable for the behavior of agents; the agent is never held accountable.	KEY DIFFERENTIATOR - Level 3 (individualized/advanced): the system carries its own limited legal personality and its own liability through the attached MEG Address guarantee (5.4.c), under ex post supervision when the DEA justifies it. The guarantee is structured on a cascade basis (primary insurance → reinsurance → sectoral fund, 9.4). Direct in rem action available when no human operator is accessible (7.5). Liability in case of multi-agent claims is allocated proportionally to each agent 's contribution (6.1.d), not jointly and severally.	MEG 2 closes the accountability gap for truly autonomous agents that both frameworks leave open - but WITHOUT generalized legal personality: it is functional/instrumental (ontological neutrality, 9.2), conditioned by the threshold of individuation (5.1), and always anchored in a final human/legal guarantor (no autonomy without a guarantor, 9.4).
Cross-border jurisdiction and enforcement	Territoriality + scope of marketing; enforced by national authorities. No specific mechanism for cross-border agent attachment.	Guidance level; no jurisdictional/enforcement mechanism. Explicitly signals the dynamic identity of the agent as an open question.	Flag State model (7.4): agent registers in a chosen jurisdiction whose law governs it extraterritorially; mutual recognition conventions (Chapter 8). Direct action in rem against agent assets (7.5). Layered forensic evidence for courts vs. experts (7.3).	MEG 2 is the only one of the three that has a cross-border attachment + enforcement mechanism, borrowed from maritime law. The compromise it calls itself: the risk of 'flags of convenience', mitigated only by access discipline (7.6) - not eliminated.

Notes:

- The intensity of the color indicates the level of liability attached to that actor: pale color = voluntary; dark color = obligation or dedicated mechanism.

Annex C

Summary table of MEG 2 legal personality levels

Who is responsible, what guarantee is required, what proof is required, what human confirmation and what statuses can the agent receive - for each compliance level

Level	Who is responsible and for what?	Required guarantee	Proof required	Human confirmation	Possible states	Ref. MEG 2
N1 Instrumental	The provider or user is fully responsible. The agent = instrument; no own liability.	No guarantee required for the existence of identity. May be contractually required by access nodes.	Standard audit log. No dedicated forensic layer.	DEA not applied. System under permanent human control.	Active / Inactive. Deactivation = administrative act.	5.4(a); 5.3; 6.5
N2 administration	The operator is responsible for deployment, permissions and lack of control. The manufacturer is responsible for the defect of the model. The user is responsible after informed human architectural confirmation.	Valid guarantee required for operational legal effects (access to valuable nodes, commercial operation). MEG Address belongs to the operator	Continuous audit log. DAI and ISR continuously monitored and public.	Architectural human confirmation for actions with irreversible consequences - technical blockage at the permission level, not textual instruction.	Active / Inactive / Reported / Suspended. Deactivation = administrative act.	5.4(b); 6.1; 6.2; 6.5; 9.4
N3 Individualized	Manufacturer: system defect (6.1.a). Operator: deployment and permissions (6.1.b). User: exclusively after architectural confirmation (6.2). Author of the hijacking: 6.1.c. The agent himself bears limited personal liability through the MEG Address guarantee. Direct action in rem possible (7.5).	MEG Address attached liability insurance, calibrated to the declared operating jurisdictions. Cascade: primary insurance - reinsurance - sectoral guarantee fund. *Least privilege* as an operational guarantee requirement.	layer (EFR) - internal state vectors, not content; access under double control. Layered evidence: summary for magistrates + complete technical documentation for experts (7.3).	confirmation for any irreversible action. Textual instructions do not constitute confirmation points. DEA justifies the move from a priori to a posteriori surveillance.	All statuses: Active / Inactive / Reported / Suspended / Deactivated. Permanent deactivation requires judicial authority. No status constitutes identity deletion.	5.1; 5.3; 5.4(c); 6.1-6.5; 7.1; 7.3; 7.5; 9.4

Notes:

- The priority scheme of liability is not cumulative: each actor is liable for the component attributable to him, not jointly and severally with all other actors involved (see the introductory paragraph of Chapter 6).

- The guarantee is optional for the existence of minimum legal identity (registration in the registry), but mandatory for operational legal effects at N2 and N3 - access to valuable nodes, commercial operation, own liability (4.4 in conjunction with 9.4).
- Architectural human confirmation means a technical block at the system permissions level, not a textual instruction at the system prompt. An instruction such as *do not perform destructive actions* does not constitute a confirmation point within the meaning of 6.2.
- The independent forensic layer (EFR) records internal state vectors, not content of interactions. Access is double-checked. Post-factum statements of the agent who caused the incident do not constitute forensic evidence within the meaning of 7.1.
- Principle of least *privilege* (9.4): The agent may not be granted more extensive technical permissions than are strictly necessary for the specified task. Granting more extensive rights constitutes an omission of the operator that aggravates liability.