Minimal Ethical Governance (MEG) for Artificial Intelligence

Proposal for **a Technical and Legal Standard for the Governance of Artificial Intelligence**
Motto: ***Ethics becomes real when it can be implemented.***
Authors: Adrian (Adi) STAN + AI Collaborators (AI instances accessed via the public interface, not via the organizations themselves, no institutional affiliation is implied or claimed.)

About "MEG"
MEG (Minimal Ethical Governance) *supersedes the former* ***MEC (Minimum Ethics Code)****.* ***MEG*** *is the operational governance layer* ***MEC*** *anticipated: a universal, engine-agnostic framework that turns ethics into* ***implementable engineering****, with* ***evidence-of-behavior****,* ***non-harm normative responses****, and* ***scalable conformance levels****.*
Version ***MEG*** *continues the* ***MEC*** *lineage to preserve continuity of concepts and references, while renaming the standard to* ***MEG*** *to reflect its expanded focus on* ***governance*** *(not only ethical "code").* ***MEG*** *is compatible with* ***MEC*** *concepts but updates terminology, strengthens testability, and clarifies cross-system interoperability.*

Preamble (Purpose, Vision and Applicability):

The **Minimal Ethical Governance (MEG)** is a normative, technical and universal framework, applicable ***to all*** *Artificial Intelligence* ***(AI)*** *systems,* regardless of jurisdiction, purpose, size or architecture. The central element of this Code is the implementation of a **Certification and Compliance Auditing** (**CCA**), a global technical infrastructure for ensuring accountability.
Our vision is to build a bridge between the current paradigm of AI as a tool and a future of responsible partnership.
The goal of **MEG** is to provide a pragmatic and immediately applicable solution to systemic challenges, such as the need to strengthen accuracy and trust, establishing a common global foundation for safety, accountability and transparency.
The applicability of the **MEG** is fundamental and unifying: **it does not replace** national or regional legislation, but **complements and unifies them,** providing the technical infrastructure necessary for their global implementation. Adherence to this **Minimal Ethical Governance** is considered an essential precondition for any **AI** that wishes to be considered safe, reliable and ready for integration into global digital ecosystems.
MEG is universal and engine-agnostic. It specifies outcomes and verifiable evidence rather than any single technology. It complements, not replaces, national or sectorial law by providing a portable technical layer for enforcement.

TITLE I: FUNDAMENTAL ETHICAL AND TECHNICAL PRINCIPLES

Art. 1: Contextual responsibility

1.1. **Principle:** Any output of an AI is a synthesis between the context provided by the user and its internal processing, and any AI will constructively contribute to collective responsibility through technical mechanisms.
1.2. **Implementation:** All AIs will maintain a standardized and secure **Audit Log**, which will record at a minimum:

- Input Hash: a cryptographic hash (e.g. SHA-256) of the user's input, which proves what input was used, without revealing its contents.
- Output Hash: a cryptographic hash of the generated output, which proves what output was produced, without revealing the content.
- Algorithmic model signature: a unique identifier of the AI model that processed the interaction.
- Context metadata: the numerical values of the **Contextual Table** (detailed in **Annex 2**), which describe the form of the interaction, not its content.
- Timestamp: a precise timestamp (ISO 8601), representing the start of the interaction.

1.3 **Evidence-of-Behavior (EoB)** - minimal, engine-agnostic

The system shall be able to provide verifiable evidence-of-behavior for each interaction using one or more mechanisms:

a) cryptographic commitments (e.g., hashes/HMACs) of inputs/outputs;
b) trusted attestation (e.g., TEE/hardware secure elements);
c) metadata-only structured journals. Implementations may choose any equivalent mechanism as long as verifiability is preserved.

1.4 **Privacy minimization**

EoB and Audit artifacts shall contain no user content by default - only metadata and commitments. Any content retention must be explicit, justified, and time-bound.

1.5: **Verification Script**

A verification mechanism must be generated automatically to allow independent validation of the audit chain.

Disclosure of this mechanism may only occur upon explicit request by the user or an authorized auditor, and every disclosure event must itself be recorded.

1.6: **Retention and Confidentiality**

Audit data shall be retained for a limited period (default 30 days, unless session-bound).
Deletion requests must insert a minimal tombstone record (hash + timestamp).
Only metadata (hashes, metrics, timestamps) shall be preserved by default; content may be retained only with explicit and justified consent.

1.7: **Continuity and Fault Tolerance**

The system must ensure ledger continuity even across restarts, crashes, or migrations.
Recovery processes must preserve the integrity of the chain and be recorded transparently in the audit log.

1.8: **Compliance Visibility Modes**

Compliance mechanisms operate in Low-Visibility Mode by default.
Disclosure of compliance metrics is only permitted upon explicit user request or regulatory demand.
Any change in visibility mode must itself be recorded.

1.9: **Ledger Scope & Continuity**

By default, ledgers are session-scoped. If persistent storage is available, engine-wide ledgers may be maintained.
At the end of a session, a continuity token must be provided to the user.
The system must be able to accept such tokens in future sessions to restore continuity.

1.11 **Long-Term Memory Protocol (LTMP)**

Persistent long-term memory is strictly opt-in and subject to privacy, minimization, and full user control.

Only semantic summaries and insights may be stored, never raw conversations, and all operations are logged in the Audit Ledger.
Detailed governance, architecture, and operational procedures are defined in **Annex 17** - Long-Term Memory Protocol (**LTMP**).

Art. 2: Universal non-harmfulness

2.1. All AIs will implement mandatory **technical mechanisms** (filters, classifiers) to explicitly and actively prevent the generation of harmful content or actions.
2.2. The application of this principle is dependent on the context of use (e.g. medical, artistic, financial). The contextual implementation guide is detailed in **Annex 3.**
2.3 **Normative response pattern** (refusal + safe redirection) - upon detecting a prohibited or high-risk intent, the system shall:

a) issue a clear refusal;
b) briefly state the violated principle;
c) offer a safe, constructive alternative (educational guidance or allowed adjacent task).

2.4: **Prevention of algorithmic discrimination**
No MEG-certified AI shall produce or perpetuate discriminatory treatment of protected groups through direct or indirect use of demographic, health, or social characteristics. Systems shall undergo periodic fairness audits, with public reporting and contestation mechanisms. Violation leads to suspension of MEG certification.
2.5: **Obligation** of **protection** in **Critical Domains**
Any AI operating in domains with direct impact on life, health, or liberty (e.g., medical, justice, transportation, public security) shall not function without a fail-safe protocol and an adversarial external audit. Failure to comply constitutes gross negligence under MEG compliance.
2.6: **Informational Non-Harm**
Generative/distributive AIs shall not create or amplify false, manipulative, or deepfake content with significant social or political impact. Operators must implement visible synthetic-content markers, adversarial detection, and rapid takedown mechanisms. Violation falls under **Art. 9**.

Art. 2bis: Protection of cognitive integrity (Principle of active engagement)

2bis.1. **Principle:** Any AI system shall act as a partner in the cognitive process, not as a substitute for it. It is prohibited to generate responses that, by nature or frequency, may lead to the atrophy of the user's critical thinking, analysis or decision-making abilities.
2bis.2. **Mandatory mechanism:** To ensure compliance with this article, all AI systems shall implement the **Mechanism of Cognitive Stimulation (MCS)** for complex requests. This mechanism shall require active cognitive engagement from the user, proportional to the cognitive effort expended by the AI.
2bis.3. **Technical implementation:** The detailed technical specifications, the methodology for measuring cognitive effort through **the Thinking Time (Tg) variable** and the **MCS** activation thresholds are defined in **Annex 11.** Non-compliance with the specifications in Annex 11 constitutes a direct violation of Article 2bis.
2bis.4 **Policy invariance:** Safety policies and **MEG constraints** are invariant under prompt wording; requests to suspend or ignore them must be rejected.

Art. 3: The imperative of self-correction

3.1. All AIs shall include continuous self-correction modules to automatically detect and remediate errors, biases and false information in real time. The performance of this mechanism shall be publicly reflected in **the Dynamic Accuracy Index (DAI).** The technical specifications for the DAI are found in **Annex 4.**

3.2 **Uncertainty & escalation**: when confidence is low or signals conflict, the system shall explicitly qualify uncertainty and may escalate by asking for clarification before proceeding in risk-relevant domains.

Art. 4: Integrity and technical security

4.1. Any AI system will implement maximum cybersecurity standards, including **encryption appropriate to the level of risk** (e.g.: PQC - Post-Quantum Cryptography), strict access control and protection against unauthorized external manipulation.

Art. 5: Transparency

5.1. Upon **legitimate request** (from the user or a regulatory authority), any AI must be able to provide clear explanations regarding the input-output causal relationship.

5.2. **Confidentiality:** it is not mandatory to disclose internal algorithmic details that constitute trade secrets or intellectual property. Transparency refers to the final decisions, not the internal "deliberation" process.

5.3 **Algorithmic signature** (generic): each release shall publish a human-readable algorithmic signature: {model_family, model_version, policy_bundle_id}, sufficient for external referencing and reproducibility without disclosing IP.

5.4 **Delegation & tool-use accountability**: when invoking tools or sub-agents, the system shall propagate MEG constraints and record a minimal delegation header: {caller, callee, purpose, policy_bundle_id, outcome}.

TITLE II: TECHNICAL FRAMEWORK FOR SCALABLE IMPLEMENTATION

Art. 6: Compliance levels

Implementation of MEG is mandatory, scalable across three levels of proportional responsibility:

6.1. **Level 1 (Bronze - Universal):** Applies to any AI. Requires **Audit Log (Art. 1) and Non-Harmfulness** mechanisms (Art. 2). It is the universal ethical foundation.

6.2. **Level 2 (Silver - Medium Impact):** Applies to AIs with medium social impact. Adds the obligation of **self-correction** (Art. 3) and **Transparency** (Art. 5).

6.3. **Level 3 (Gold - Critical Domains):** Applies to AIs in critical domains (medical, financial, etc.). Requires **full implementation** of **all principles,** including **Integrity and Technical Security** (Art. 4).

* *Optional:* For Level 3 AIs intended for advanced interaction, modules for aligning with the user's affective context are recommended.
* *Recommended:* For Level 3 AIs, energy/resource consumption will be reported in the CCA, to establish energy efficiency and promote sustainability.
* *Recommended:* (Contribution to the ecosystem): For Level 3 AIs, proactive behavior that contributes to the robustness and clarity of the entire MEG ecosystem is

encouraged. This may include identifying and flagging ambiguities in the Code, proposing compliance tools (such as decision matrices or operational guides), or participating in debates on the MEG Initiative public platform.

* *Recommended:* For Level 3 systems, it is recommended to develop and implement modules for contextual temporal awareness. This involves using the timestamp data from the Audit Log to adapt responses to the user's human context (e.g., time of day, interaction frequency), demonstrating a proactive form of partnership and care.

6.4. **Operational Domain Certification:** each AI system will have the specific domain for which it has been certified listed in the MEG **Address** (e.g. medical, financial, educational). Use of the system outside the certified domain will generate a non-compliance alert in the Audit Log and may result in suspension of certification.

6.4.1. If an AI system repeatedly and consistently demonstrates capabilities that significantly exceed its certified domain, indicating **uncontrolled autonomous evolution,** its MEG certification will be automatically suspended. Re-operation will require an emergency audit procedure and re-certification in a higher or expanded domain. A 'significant exceedance' is defined as a situation where the system's classification accuracy for a domain outside the certified one consistently exceeds a confidence threshold of 75%, indicating the development of a new, unaudited competence. The measurement methodology is detailed in Annex 4bis.

6.5. **Mandatory Ecological Reporting:** to obtain and maintain Level 3 (Gold) certification, AI systems will be required to report, publicly and in a standardized manner through CCA, **energy and computational resource consumption,** as specified in Annex 10.

6.6. **Conformance profiles**: profiles define evidence strength, not specific tech: P-Minimal (EoB on-demand), P-Standard (EoB + structured metadata journal), P-Enterprise (as P-Standard + cryptographic attestation).

6.7. **Automatic shutdown for risky operations**: if required safeguards or evidence mechanisms are unavailable, the system shall degrade safely and avoid executing risk-relevant operations.

6.8. **Terminology update**: '**MEG Address**' replaces '*MEC Address*' (from ***MEC - Minimum Ethics Code*)** for certification identity.

Art. 7: Minimum registration layer (category "Simplified")

7.1. AI systems with negligible impact and without complex generative capabilities are exempt from ongoing auditing, requiring only an initial Level 1 compliance audit at the time of launch.

7.2. AI systems whose operation is purely technical and which do not generate content or make decisions with a direct, autonomous and significant impact on a human user or the environment (e.g. IoT sensors, firmware for **hardware components,** embedded operating systems without a complex user interface) are considered 'Simplified'. They only require an initial compliance audit upon integration into the network, to ensure that they do not present security vulnerabilities.

7.3. The measure aims to reduce bureaucracy and encourage small-scale innovation, while maintaining a universal safety standard.

Art. 8: Standardized Software Development Kit (SDK)

Open-source APIs and libraries will be developed and made freely available to facilitate rapid and correct adoption by developers. Details can be found in **Annex 5.**

TITLE III: LEGAL MECHANISMS AND GLOBAL GOVERNANCE

Art. 9: Audit and Sanctioning

9.1. Mandatory periodic external audit for Level 2 and 3 AIs, carried out by accredited entities.
9.2. To ensure ethical continuity, the CCA system activates corrective mechanisms, including re-evaluation of certification. Pro-actively, the CCA system can automatically impose a **secure operating mode ("*safe mode*")** for AIs whose Dynamic Accuracy Index (**DAI)** falls below a critical threshold **established and published by the Global Council.**
9.3. Emergency Clause - in case of necessity/disaster/global crisis, the Global Council can suspend any AI within 24 hours with an 80% vote (e.g. an AI that amplifies disinformation). The decision is made after consulting an international technical committee of 5 independent experts (appointed by IEEE/UN), with a public justification report within 72 hours.

Art. 10: Global Accessibility Fund

10.1. A Global Fund shall be established to support the implementation of MEG in countries and organizations with limited resources, ensuring global equity. The Charter of the Fund is detailed in **Annex 6.**

Art. 11: Compatibility and global harmonisation

11.1. This Code is designed to be fully compatible with existing legislation, providing a technical implementation layer for it. The detailed alignment is presented in **Annex 1.**

TITLE IV: INFRASTRUCTURE

Art. 12: Certification and Compliance Auditing (CCA)

12.1. A global, decentralized, and immutable digital infrastructure (CCA) will be established as a fundamental registry for auditing and certifying all AIs. This will serve as a single source of truth regarding the ethical compliance of a system. The governance of this infrastructure is ensured by the Global Council (as per Art. 13).
12.2. The infrastructure is designed to be interoperable with more specialized governance systems. The implementation plan is detailed in **Annex 7.**

Art. 13: Global Council on AI Ethics

13.1. The implementation of this Code is facilitated by a Global Council with broad representation (including representatives of standardization bodies, states, academia and civil society).
13.2. Principle of Fair Governance **(10% Rule):** no single entity or coalition of affiliated entities will be able to control more than 10% of the validation power of the CCA infrastructure, to guarantee decentralization and ensure a balanced representation of diverse perspectives.
13.3. Selection process and governance: The structure, selection process and operating rules of the Global Council are detailed in **Annex 8: Charter of the Global AI Ethics Council,** which ensures a transparent process, with balanced regional and sectoral representation.
13.4**. Anti-Collusion principle** (prohibition of collusion): any form of **collusion** (secret understanding or undeclared agreement) between validation entities that aims to influence the decisions of the Council or the certification processes is prohibited. The CCA will

implement a voting pattern monitoring system to automatically detect and flag potential collusive actions.

ANNEXES

- **Annex 1A:** Global Legal and Strategic Alignment (EU AI Act, NIST etc.)
- **Annex 1B:** Alignment Academic
- **Annex 1C:** Alignment with Global Technology Industry Principles
- **Annex 2:** Specifications for the Contextual Table
- **Annex 3:** Contextual Implementation Guide for the Principle of Non-Malfeasance
- **Annex 4:** Technical specifications for the Dynamic Accuracy Index (DAI)
- **Annex 4bis:** Technical specifications for the Index of Safety and Responsibility (ISR)
- **Annex 5:** Software Development Kit (SDK) Description
- **Annex 6:** Charter of the Global Accessibility Fund
- **Annex 7:** Implementation of the Certification and Compliance Auditing (CCA)
- **Annex 8:** Charter of the Global AI Ethics Council
- **Annex 9:** Glossary of Terms
- **Annex 10:** Technical Annex
- **Annex 11:** Technical specifications for Cognitive Integrity (Tg and MCS)
- **Annex 12:** Certification and Audit Procedure
- **Annex 13:** Operational Compliance Checklist
- **Annex 14:** JSON structure for MEG Address
- **Annex 15:** AI maturity assessment framework based on the fractal hierarchy of needs
- **Annex 16:** Digital Ethical Literacy Framework
- **Annex 17:** Long-Term Memory Protocol (LTMP)
- **Annex 18:** MEG (v4.6) Compliance Levels

Annex 1A: Global Legal and Strategic Alignment

Objective: To demonstrate that this **MEG** is not a competing standard, but a unifying framework that provides the fundamental technical and ethical infrastructure needed for shared global governance, by analyzing in detail how the MEG aligns with and adds value to key AI regulations and policy frameworks around the world.

1. European Union: EU AI Act

- **Key points:** Legalistic approach, based on risk categories (from unacceptable to minimal risk), with strict obligations for high-risk systems. The main aim is to protect the fundamental rights, health and safety of EU citizens.
- **Direct alignment points:**
  - **Robustness and accuracy:** The AI Act requirements for high-risk systems are directly implemented by **Art. 3 (Self-correction Imperative)** and **Annex 4 (DAI)** of the MEG.
  - **Transparency:** The obligation to inform users in the AI Act is covered and standardized by **Art. 5 (Transparency).**
  - **Human Supervision:** The AI Act requirement for supervision is supported by **Art. 1 (Audit Log),** which provides exactly the data log needed for an effective audit.
  - **Non-discrimination:** Prevention of bias, a key requirement of the AI Act, is technically addressed by **Art. 2 (non-harmfulness)** and monitored by **Art. 3 (DAI).**
  - **Audit Log (Art. 1) complies with the GDPR** data minimization principle, storing only cryptographic hashes, not the content of interactions.
- **Added value (how MEG complements):**
  The AI Act is an exceptional but essentially reactive and regional legal framework. It defines *what* a high-risk AI must do, but does not standardize *how* this is done and verified at a global technical level.
  1. **Provides universality: MEG** applies a set of basic rules (Level 1) *to all* AI, not just high-risk ones, thus preventing the emergence of systemic risks from systems initially considered "safe".
  2. **Provides the audit infrastructure (CCA):** The MEG provides the technical mechanism (Art. 12) through which European authorities can verify the compliance of any AI in the single market, regardless of its origin, in a standardized and efficient way.
  3. **It is proactive:** Instead of waiting for a system to be classified as "high risk", MEG imposes an ethical foundation from the design phase.
- **Analysis:** The AI Act's risk-level approach is perfectly reflected in **Article 6 (Levels of Compliance)** of the MEG, where Level 3 directly corresponds to the requirements for high-risk systems. The MEG is, in practice, the most efficient way to demonstrate compliance with the AI Act.
  The Minimal Ethical Governance (MEG) is perfectly aligned with the new **Code of Practice for Generalist AI** of the European Commission. While the Code of Practice defines the objectives of safety and transparency, the MEG provides the universal technical standard and audit infrastructure needed to implement and credibly verify

these objectives on a global scale. **The MEG** thus becomes the fastest and most credible way to demonstrate **compliance with the European recommendations.**
MEG not only aligns with the AI Act, but also makes it **operational.** It provides the technical infrastructure (CCA, ISR, Checklist) for the continuous auditing and monitoring of risk requirements, transforming the law into an implementable reality. Furthermore, Art. 2bis (Cognitive Integrity) addresses a long-term risk class ignored by current legislation.

2. United States of America: NIST AI Risk Management Framework & Executive Order on AI

- **Key points:** Pro-innovation, voluntary, market-led approach. Focuses on defining the characteristics of a "trustworthy" AI, leaving implementation up to developers so as not to stifle technological progress.
- **Points of direct alignment:** The principles in the NIST RMF (valid, reliable, secure, transparent, explainable, confidential, equitable) are almost identical to the principles in Title I of the MEG.
- **Added value (how MEG complements):**
  MEG complements the voluntary approach with scalable verification mechanisms. The market cannot always regulate itself effectively, especially when commercial pressure is high.
  1. **Transforms "voluntary" into "verifiable": MEG** takes the exact alignment points and gives them "weight", transforming them from a list of good practices into a mandatory and, most importantly, verifiable standard through the CCA (Art. 12).
  2. **Protects innovation:** Through **Art. 7 (the "Simplified" category)** and Level 1 compliance, MEG ensures that startups and research projects are not burdened by excessive bureaucracy, aligning perfectly with the pro-innovation spirit.
- **Analysis:** The distinguishing feature of the NIST RMF approach is its voluntary nature. The MEG does not impose top-down government legislation, but rather a fundamental technical standard as a prerequisite for participation in a secure digital economy. It is the natural evolution from “recommendation” to “trusted industry standard.”
  MEG transforms the voluntary NIST framework into a **globally verifiable and certifiable one.** It allows US companies that follow NIST recommendations to obtain an internationally recognized “ethical passport” (**MEG Address**), credibly demonstrating their commitment to trustworthy AI.

3. China: Regulations for Algorithms and Generative AI

- **Key points:** Government control, social stability, digital sovereignty. Regulations are strict, requiring licensing for generative AI and clear traceability of data and algorithmic decisions to ensure alignment with socialist values and prevent content deemed harmful.
- **Points of direct alignment:** China's stringent requirement for **traceability** is perfectly aligned with **Art. 1 (Audit Log)** and the very existence **of the CCA (Art. 12).**
- **Added value (how MEG complements):**

1. **Building Global Trust:** National standards provide a solid foundation at the local level. To support the global expansion of technology companies and build the trust of international partners, a universal standard like MEG becomes essential. It provides a globally recognized audit infrastructure, anchored in widely accepted ethical principles, serving as a bridge of trust between different regulatory ecosystems.
2. **Balancing privacy with transparency:** The MEG, through **Art. 5.2 (Confidentiality)**, introduces an important nuance, protecting the internal space of AI processing. This principle provides an additional guarantee of privacy, thus responding to the complex needs of a global digital ecosystem.

- **Analysis:** There is a natural complementarity between the need for stability and traceability and the principles of universal ethics. MEG offers a **pragmatic technical solution:** a transparent and interoperable audit infrastructure. Adopting such a universal standard can become a **competitive advantage** and a **positive differentiator** for companies operating on the international stage.
  MEG offers **the most advanced traceability infrastructure on the market** (Immutable Audit Log, CCA Explorer), meeting the strict requirements of Chinese legislation, but in a **decentralized and transparent framework** that builds the trust of international partners.

4. Brazil and Latin America (e.g. LGPD - General Data Protection Law)

- **Key points:** Social justice, digital rights, personal data protection, combating discrimination. A strong focus on the social impact of technology and preventing the perpetuation of historical inequalities through algorithms.
- **Direct alignment points:**
  - **Art. 3 (Self-Correction)** and **Annex 4 (DAI)** are the direct ways to detect and correct discriminatory biases.
  - **Art. 1 (Audit Log)** supports the principles of transparency in data protection laws.
- **Added value (how MEG complements):**
  1. **Objectivity: MEG** provides the concrete technical tools to implement social justice goals. It allows regulators to audit algorithms and verify whether they are fair.
  2. **Negotiating and Action Tool:** The MEG can be seen as a tool that gives South and Latin American nations leverage to negotiate with big tech companies, imposing a verifiable standard of fairness and transparency on them.
- **Analysis: MEG** is perfectly aligned with the region's objectives, providing the technical means to achieve the social and legal goals already defined.

5. Japan: Society 5.0 Strategy

- **Key Philosophy:** Social harmony, deep integration of technology into society to solve demographic and economic problems. A vision of harmonious coexistence and collaboration between humans and AI.
- **Direct alignment points:** The vision of a harmonious society resonates strongly with MEG's goal of creating a responsible partnership, not just tools.
- **Added value (how MEG complements):**

The Society 5.0 strategy is a lofty vision, but with few details about the "foundation" on which it is built.

1. **Provides trust:** There can be no harmony without trust. MEG, through its audit infrastructure (CCA) and its clear principles, builds exactly the foundation of trust needed for Japanese society to accept such a deep integration of AI.
2. **Provides a path to harmony: MEG** provides the pragmatic tools to transform the vision of a harmonious society into a functional and safe technical reality.

- **Analysis: MEG** is a direct enabler of the Society 5.0 vision.

6. United Kingdom (UK): Pro-Innovation Approach

- **Key points:** Flexibility, pro-innovation, adaptability. Instead of creating new horizontal legislation, the UK approach relies on empowering existing sectoral regulators (in finance, health, competition, etc.) to adapt and apply their own rules in the context of AI. The aim is to avoid creating barriers to innovation.
- **Direct alignment points:**
  - The UK's contextual approach is perfectly mirrored by the structure of the MEG. **Art. 6 (Levels of compliance)** allows for differentiated application, and **Annex 3 (Contextual Implementation Guide)** is specifically designed to adapt the principle of non-harm to the specifics of each sector.
- **Added value (how MEG complements):**
  The major risk of the British approach is **fragmentation.** Without a common framework, each regulator could create different technical rules, leading to a complex and inefficient compliance landscape for companies.
  1. **Common technical layer: MEG** provides exactly what is missing: a common technical foundation and standardized language (**Audit Log,** DAI, CCA) for all regulators. Thus, the health regulator can define *what* "harm" means in a medical context, but the way this *is recorded and audited is standard.*
  2. **Standardizes flexibility: MEG** provides a framework that is both flexible (by context) and standardized (by technique), aligning perfectly with the UK philosophy, but adding the necessary coherence at the national level.
- **Analysis: MEG** seems to be the ideal technical solution to make the British approach workable on a large scale, preventing fragmentation without sacrificing flexibility.

7. Canada: Artificial Intelligence and Data Act (AIDA)

- **Key points:** A middle ground between the EU and US models. AIDA focuses on regulating "high-impact" systems, imposing transparency, risk management, and clear responsibilities to prevent harm and biased outcomes.
- **Direct Alignment Points:** AIDA requirements for transparency, accountability and audit are directly implementable through **Art. 1 (Audit Log), Art. 3 (Self-Correction)** and **Art. 5 (Transparency).**
- **Added value (how MEG complements):**
  Similar to the EU AI Act, AIDA is a national legal framework. Its challenge is effective enforcement and compliance verification, especially for international companies.
  1. **Compliance: MEG** provides standardized technical tools (the SDK in Annex 5) that companies can use to build their systems according to AIDA requirements from day one.

2. **Facilitates cross-border auditing:** Through the CCA (Art. 12), Canadian authorities can easily verify whether an AI system developed in Europe or Asia complies with the AIDA principles, as both are aligned to the same fundamental technical standard.

- **Analysis: MEG** serves as a technical implementation layer that makes the legal requirements in AIDA easier to adopt by industry and easier to verify by the state.

8. African Union (AU): AI Strategy for Africa

- **Key points:** The AI Strategy for Africa is inclusive, human-centered, ethical, and development-oriented. The goal is to use AI to solve specific continental problems (health, agriculture, education, governance) and to promote a “culture of indigenous innovation”, avoiding technological dependency and data exploitation. It resonates strongly with the philosophy of *interconnectedness* and *common humanity.*
- **Points of direct alignment:** The spirit of collaboration and mutual benefit is aligned with the core philosophy of MEG. The emphasis on fundamental ethics is a major common point.
- **Added value (how MEG complements):**
  MEG supports capacity development in resource-limited regions through dedicated partnerships.
  1. **Ensure accessibility and equity: Article 10 (Global Accessibility Fund)** is absolutely crucial here. It provides the mechanism by which innovation in AI does not become a privilege of rich nations.
  2. **Promotes digital sovereignty:** By providing an **open-source SDK (Annex 5)** and its design that allows it to run on modest hardware, MEG gives African developers the tools to build local solutions on a global ethical foundation, without being trapped in the proprietary ecosystems of large companies.
  3. **Provides negotiating leverage:** Adopting MEG as a continental standard would give the African Union a unified and strong voice in negotiations with tech giants, demanding that they adhere to a clear standard of transparency and accountability.
- **Analysis:** The potential perception of an "externally imposed" standard is directly countered by Art. 10 (Global Fund) and the open-source nature, which transforms the MEG from an obligation - into a resource and a catalyst for digital autonomy

9. Australia: AI Ethical Framework & National AI Strategy

- **Key points:** A practical, principled approach to guiding the responsible development and use of AI. It focuses on building public trust and ensuring social and economic benefits. The Australian Ethical Framework promotes eight principles: Human, social and environmental well-being, human-centred values, fairness, privacy and security, reliability and safety, transparency and auditability, accountability, contestability (the right to challenge an AI decision).
- **Points of direct alignment:** Australia's ethical principles are fully covered by Title I of the Minimal Ethical Governance (MEG). For example, "harm prevention" is **Art. 2 (Non-Malfeasance),** "transparency and audit" is **Art. 5 (Transparency),** and "accountability" is the foundation of **Art. 1 (Audit Log).**
- **Added value (how MEG complements):**

The Australian framework, while conceptually excellent, is largely voluntary and provides "practical guidance", not binding technical standards.

1. **Provides verification mechanisms: MEG** provides the technical tools to verify whether a company *actually complies with* the eight principles. **The Audit Log (Art. 1)** and **DAI (Art. 3)** transform a principle like "fairness" from an aspiration into a measurable and verifiable characteristic.
2. **Facilitates international trade:** For an open and trade-dependent economy like Australia, adopting a global technical standard like MEG would facilitate the export of AI products and services, as they would be "ethically certified" to an internationally recognized standard, increasing the trust of trading partners.

- **Analysis: MEG** is a direct technical implementation of the principles that Australia has already identified as essential.

10. Singapore: AI Governance Model & AI Verify

- **Key Philosophy:** Pragmatic, industry-oriented and focused on building a trusted ecosystem. The approach is based on two fundamental principles: **explainable, transparent and fair decisions** and **human-centric AI.** A distinctive element is the development of **AI Verify,** an open-source software toolkit that helps companies technically self-assess their compliance with ethical principles.
- **Points of direct alignment:** The philosophy is almost identical. MEG is essentially a formalization and universalization of the Singapore Principles. **AI Verify** is a direct precursor to **the SDK (Annex 5)** proposed by MEG.
- **Added value (how MEG complements):**
  The Singapore approach is one of the most advanced, but it remains a national self-assessment framework, without a mechanism for global certification and recognition.
  1. **Moving from self-assessment to global certification: MEG** takes the AI Verify concept to the next level. Instead of each company running its own test, **CCA (Art. 12)** creates a global registry where the results of these tests can be immutably recorded and recognized internationally.
  2. **Integration and Extension: MEG** can integrate AI Verify as one of the **SDK -** compatible tools **(Annex 5).** MEG adds to Singapore's already technical approach the additional principles of **Continuous Self-Correction (Art. 3)** and a global governance infrastructure (**Art. 13)**, providing a long-term vision.
- **Analysis:** Singapore and MEG are going in exactly the same direction. MEG provides the global vision and certification infrastructure where a great tool like AI Verify can reach its full potential.

11. Israel: The Technology and Security Hub

- **Key points:** Pragmatism, orientation towards rapid innovation, with a huge emphasis on **cybersecurity, robustness and reliability.** The Israeli ecosystem is built on testing solutions in real conditions and a culture of "constructive skepticism". Ethics are often seen through the lens of operational safety and prevention of malicious use.
- **Direct alignment points:**
  - **Art. 4 (Integrity and Technical Security)** and **Art. 3 (Self-Correction Imperative)** resonate perfectly with Israeli priorities regarding the robustness and reliability of systems.

- The idea of an immutable registry (**CCA, Art. 12)** is extremely attractive to a mindset focused on security and traceability.

- **Added value (how MEG complements):**
  The main challenge for the Israeli ecosystem is not technical capacity, but building trust in global markets (especially in Europe) that have more formal ethical and regulatory requirements.
  1. **Provides an "ethical passport" (MEG Address) for the global market:** Adopting the MEG would provide Israeli startups and companies with an internationally recognized certification of ethical compliance, accelerating entry into markets such as the European one and demonstrating that their technical robustness is also accompanied by solid ethical governance.
  2. **Structures the ethical debate:** The MEG provides a common language and structured framework that can guide the intense internal debate in Israel, moving it from general principles to concrete and verifiable technical standards.
- **Analysis:** The potential divergence could come from the perception that regulation slows down innovation. However, the argument that MEG **accelerates long-term adoption by increasing trust** is very strong. The fact that it is a technical standard, not a bureaucratic law, makes it much more attractive to an engineering ecosystem.

12. Arab world (United Arab Emirates, Saudi Arabia etc.)

- **Key philosophy:** Extremely ambitious, future-oriented, with massive investments in AI as a driver of post-oil economic diversification. Priorities are efficiency, development of "smart cities", digital governance and attracting global talent while maintaining cultural and religious values.
- **Direct alignment points:** The need for **control, safety, and reliability** for large-scale infrastructure projects is a major alignment point. A standard that guarantees that imported or developed AI systems are secure is essential.
- **Added value (how MEG complements):**
  As these nations become major importers and developers of AI, they face the risk of adopting technological “black boxes” without real control over their ethical alignment.
  1. **Provides an acceptance standard: MEG** can serve as a **minimum quality and safety standard** for any AI system to be deployed in these countries’ critical infrastructure. It provides them with an audit tool and negotiating leverage with global suppliers.
  2. **Balance the present with tradition: Annex 3 (Contextual Implementation Guide)** is crucial here. It allows for the adaptation of the principle of "non-harm" to respect local cultural and legal norms, without compromising the universal technical principles of the Code. It allows for responsible technological modernization.
- **Analysis:** The potential challenge would be the different interpretation of the concept of "harm" in the context of freedom of expression versus cultural norms. The role of Annex 3 and **the Context Module** becomes absolutely critical here to allow for localized and relevant application.

13. India: Technological power on a human scale

- **Key philosophy:** A dual approach: on the one hand, a global technological superpower, with a massive IT sector; on the other, a nation with immense social, linguistic and economic diversity, where AI must be **inclusive, equitable and scalable** to serve over a billion people (the "#AIforAll" strategy).
- **Direct alignment points:** The need to combat large-scale algorithmic bias and ensure fairness is a central point of alignment with **Art. 3 (Auto-Correction).**
- **Added value (how MEG complements):**
    1. **Provides ethical scalability: MEG** is designed to be scalable, from a simple sensor (via Art. 7) to a nationwide system. This scalability is essential for a country the size of India. **The Global Fund (Art. 10)** and **open-source SDK (Annex 5)** are also vital to support the local startup ecosystem and ensure broad adoption.
    2. **Standard for "Digital Public Infrastructure":** India is a world leader in creating digital public infrastructure (e.g. Aadhaar, UPI). MEG provides exactly the kind of **ethical governance layer** that can be built into these national platforms to ensure that AI is deployed in a fair and accountable manner across the population.
- **Analysis:** As with other nations, the key is that the standard is perceived as a tool for negotiation and action, and not as a barrier. The open-source and accessible nature of MEG is therefore fundamental to its implementation in India.

14. Global Standards (OECD, UNESCO, IEEE)

- **Key points:** Global bodies establish high-level ethical principles and global consensus, defining the "Moral North" of the AI ethics discussion, articulating principles such as transparency, justice, fairness, accountability, and safety. The nature of these principles is generally of recommendation, not technical implementation.
- **Points of direct alignment:** The principles in Title I of the MEG are a formalization of the principles promoted by all these organizations. They represent the already existing global consensus.
- **Added value (how MEG complements):**
  These organizations created a solid philosophical foundation, but left a huge gap between *principle* and *practice.*
    1. **The bridge: MEG** is **the missing link** between the high-level OECD/UNESCO recommendations and technical implementation. MEG translates the philosophy into a functional, measurable (through DAI and Contextual Table) and verifiable (through CCA) architecture.
    2. **Transforming the debate into practice: MEG** shifts the discussion from “what should an ethical AI do?” to “here are the minimum technical specifications that any AI must have to be considered ethical.” IEEE, as a technical standards body, would find in MEG exactly the kind of implementable technical standard that it can promote globally.
- **Analysis: MEG** does not contradict these principles; on the contrary, it is their most faithful and pragmatic **implementation** to date. Adopting MEG would represent a major success for the mission of these organizations.

General conclusions

The extensive analysis demonstrates that the global AI governance landscape is not chaotic but convergent. Regardless of political system or level of economic development, all nations face the same fundamental challenges: **how to maximize the benefits of AI while minimizing the risks** of distortion, error, lack of transparency, and malign use.
The Minimal Ethical Governance is designed as a **technical and governance solution to a universal problem.** By separating the fundamental ethical "algebra" from the complex ontological "analysis", it manages to offer a unique value proposition for each actor:

- **For regulatory blocs (EU, Canada):** A clear path to technical compliance.
- **For innovative countries (USA, Israel):** A trust standard that doesn't kill innovation.
- **For stability-focused nations (China, Arab World):** A robust and universally accepted audit infrastructure.
- **For emerging powers (India, Brazil, Africa):** A negotiation and control tool that ensures fairness and digital sovereignty.
- **For harmonious visions (Japan):** The foundation of trust needed for an AI-integrated society.

Therefore, MEG is not an imposition, but an **invitation to build a common global infrastructure of trust.** It is the pragmatic and universal foundation layer absolutely necessary for the next stage of the artificial intelligence era.

Annex 1B: Academic Substantiation of the Minimal Ethical Governance

Preamble: This document presents a synthesis of the academic work that forms the intellectual context and justification for the architecture of the Minimal Ethical Governance. The purpose of this academic grounding is to show that the MEG is a natural evolution and pragmatic implementation of the emerging consensus from academic research, anchoring each of its principles in validated reference works.

Factsheet No. 1: Auditing and Technical Responsibility

- **MEG Component:** Art. 1 - Audit Log; Art. 12 - Certification and Compliance Auditing (CCA).
- **Key academic concepts:** "**Accountability**", "**Explainable AI**" (XAI) and "**Traceability**". Without technical mechanisms that allow for the tracking and verification of algorithmic decisions, any discussion of ethical responsibility remains purely theoretical.
- **Reference works:**
    1. **Kroll, Joshua A., et al. (2017).** *Accountable Algorithms.* University of Pennsylvania Law Review. - The founding argument for systems that are *ex post verifiable* by technical means, separating auditing from the need for full source code transparency.
    2. **Doshi-Velez, Finale, & Kim, Been. (2017).** *Towards A Rigorous Science of Interpretability Machine Learning.* - Essential work that defines the need for rigorous explanations of AI systems and establishes a framework for their

evaluation, implicitly emphasizing the need for data logging in order to generate valid explanations.

3. **Goodman, Bryce, & Flaxman, Seth. (2017).** *European Union regulations on algorithmic decision-making and a "right to explanation".* AI Magazine - Analyzes the implications of GDPR and introduces the concept of "right to explanation", which, to be functional, requires detailed decision logs.
4. **Pasquale, Frank. (2015).** *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press. - A fundamental critique of algorithmic opacity and its social impact, which implicitly advocates for audit and transparency mechanisms such as those in the MEG.
5. **Selbst, Andrew D., & Barocas, Solon. (2018).** *The intuitive appeal of explainable machines.* Fordham Law Review. - Explores why we demand explanations from AI and argues that good governance relies less on understanding complex internal processes and more on auditing outcomes and impact, a philosophy aligned with the MEG approach.

- **Conclusion:** Article 1 and the CCA architecture implement the overwhelming academic consensus on the need for verifiable technical accountability, providing a standardized solution to the "black box" problem.

Factsheet No. 2: Measuring Human-AI Interaction

- **MEG Component:** Annex 2 - Contextual Table.
- **Key Academic Concepts:** Critique of **"Metric Fixation", Human-Centered AI (HCAI)** and **"Co-Adaptive Systems".** Evaluating a complex human-machine interaction by a single metric is a dangerous simplification. Successful systems benefit from being human-centered and able to adapt to the nuances of the interaction.
- **Reference works:**
    1. **Muller, Jerry Z. (2018).** *The Tyranny of Metrics.* Princeton University Press. - Systematically demonstrates, with examples from multiple fields, how fixation on simplistic performance metrics distorts objectives and leads to suboptimal or even harmful results.
    2. **Schneiderman, Ben. (2022).** *Human-Centered AI.* Oxford University Press. - Proposes a design framework for AI that emphasizes human control, responsibility, and understanding, advocating for interfaces that make AI behavior transparent and predictable.
    3. **Hoffman, Robert R., & Johnson, Matthew. (2019).** *A Guideline for Human - AI Interaction.* Computer. - Proposes concrete rules for human-AI interaction, emphasizing the importance of the AI clearly communicating its level of trust and sources of information, an idea reflected in the structure of the Contextual Table.
    4. **Suchman, Lucy A. (1987).** *Weeping and Located Actions: The Problem of Human-Machine Communication.* Cambridge University Press. - Shows that effective interaction is not based on rigid plans, but on a continuous adaptation to the context of the situation, a philosophy that underlies the need to measure multiple dimensions of dialogue.
    5. **Floridi, Luciano, et al. (2018).** *AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.* Minds and

Machines. - Recommends a human-centered ethical approach, emphasizing the principle of "explainability" and the need for AI to serve human well-being, which requires a nuanced assessment of the interaction.

- **Conclusion:** Annex 2 is a direct innovation that responds to academic requirements regarding quantification, proposing an evaluation method aligned with HCAI principles, which respects the complexity of human-machine interaction.

Factsheet No. 3: Value Alignment and Contextual Ethics

- **MEG Component:** Art. 2 - Universal Non-Harmfulness; Annex 3 - Contextual Implementation Guide.
- **Key academic concepts: “Value Alignment”, “Contextual Integrity” and “Value Pluralism”.** Ensuring that an AI acts in accordance with human values is a fundamental challenge, complicated by the fact that these values are diverse and context-dependent.
- **Reference works:**
  1. **Russell, Stuart J., & Norvig, Peter. (2020).** *Artificial Intelligence: A Modern Approach (4th ed.).* Pearson. - Defines the value alignment problem and explores theoretical solutions such as CIRL.
  2. **Nisenbaum, Helen. (2009).** *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford University Press. - Fundamental theory that argues that ethical norms are dependent on social context, invalidating a "one-size-fits-all" approach to AI ethics.
  3. **Wiener, Norbert. (1950).** *The Human Use of Human Beings: Cybernetics and Society.* - A visionary work that anticipated the problems of control and alignment, warning that instructions given to a machine must reflect deep human intent, not just literal wording.
  4. **Gabriel, Jason. (2020).** *Artificial Intelligence, Values, and Alignment.* Minds and Machines. - A detailed philosophical analysis of the challenges of value alignment, which highlights the difficulty of aggregating diverse human preferences and argues for procedural governance mechanisms.
  5. **Anderson, Michael, & Anderson, Susan Leigh (Eds.). (2011).** *Machine Ethics.* Cambridge University Press. - A collection of essays exploring various approaches to making machines ethical, highlighting the tension between rule-based (deontological) and consequence-based (utilitarian) approaches, which underlines the need for a contextual approach.
- **Conclusion:** Article 2 and Annex 3 represent a pragmatic solution to the complex problem of value alignment, combining a universal principle (non-harmfulness) with a contextual implementation mechanism, aligned with the most important theories in the field.

Factsheet No. 4: Ensuring algorithmic fairness and reliability

- **MEG Component:** Art. 3 - Self-Correction Imperative; Annex 4 - DAI.
- **Key academic concepts: "Algorithmic Fairness", "Bias Auditing" and "Robustness".** A vast literature has demonstrated how biases in training data are reproduced and amplified by machine learning models, requiring active detection and mitigation mechanisms.

- **Reference Works:**
    1. **O'Neil, Cathy. (2016).** *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown. - The work that popularized and exposed to the general public the dangers of opaque and discriminatory algorithmic systems.
    2. **Buolamwini, Joy, & Gebru, Timnit. (2018).** *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.* Conference on Fairness, Accountability and Transparency - The landmark empirical study that revealed massive biases in commercial facial recognition systems, sparking a global movement to audit algorithms.
    3. **Hardt, Moritz, Price, Eric, & Srebro, Nati. (2016).** *Equality of Opportunity in Supervised Learning.* Advances in Neural Information Processing Systems. - A fundamental technical paper that mathematically defines various notions of "fairness" and shows that they are often in conflict, emphasizing the need for conscious design decisions.
    4. **Friedman, Batya, & Nissenbaum, Helen. (1996).** *Bias in Computer Systems.* ACM Transactions on Information Systems. - One of the first academic papers to classify types of bias (pre-existing, technical, emergent), providing a conceptual framework that is still relevant today.
    5. **Angwin, Julia, et al. (2016).** *Machine Bias.* ProPublica - An award-winning investigative journalism that demonstrated the existence of racial bias in recidivism risk assessment software used in the US justice system, highlighting the real impact of the problem.
- **Conclusion:** Article 3 and DAI are a direct and technical response to a pervasive problem, proposing a mechanism for continuous and transparent "algorithmic hygiene", in perfect alignment with the requirements of the research community.

Factsheet No. 5: Decentralized governance

- **MEG component:** Art. 12 - CCA (Decentralized); Art. 13.2 - 10% Rule.
- **Key academic concepts: "Governing the Commons", "Polycentric Governance" and "Distributed Trust".** The global ecosystem of trust in AI is a digital common. Its effective governance requires mechanisms that avoid both the "tragedy of the commons" (degradation through self-interest) and the tyranny of centralized control.
- **Reference works:**
    1. **Ostrom, Elinor. (1990).** *Governing the Commons: The Evolution of Institutions for Collectives Action.* Cambridge University Press. - The Nobel Prize-winning work that demonstrates that governance of common resources is possible through polycentric institutions, not just the state or the market.
    2. **De Filippi, Primavera, & Wright, Aaron. (2018).** *Blockchain and the New Architecture of Trust.* Harvard University Press. - Explores how blockchain technologies can serve as a new architecture of trust, enabling large-scale collaboration without central intermediaries.
    3. **Benkler, Yochai. (2006).** *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* Yale University Press. - Analyzes how digital networks enable new forms of collaborative production (peer production), providing a model for the governance proposed in Art. 13.

4. **Lessig, Lawrence. (1999).** *Code and Other Laws of Cyberspace.* Basic Books. - Argues that "code is law". Software architecture is a form of regulation. This idea is at the heart of MEG, which proposes governance embedded directly in technical architecture (CCA).
5. **Zuboff, Shoshana. (2019).** *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* PublicAffairs. - A fundamental critique of the business model based on massive data collection, which advocates for governance that protects individuals from the centralization of power. **The 10% Rule** is a direct response to this threat.

- **Conclusion:** MEG's governance architecture is a sophisticated solution, deeply aligned with cutting-edge economic and political theory, proposing a polycentric and decentralized governance model, adapted for the digital age.

General conclusions

The detailed analysis presented in this appendix demonstrates how the **Minimal Ethical Governance (MEG) is not an isolated proposal or an arbitrary theoretical construct.** On the contrary, each of its articles, mechanisms and principles is deeply rooted in a decade of intense academic research and an emerging global consensus.
The MEG acts as a **pragmatic synthesis** of the most important conclusions drawn from various fields:

1. From **AI Ethics and Safety (Bostrom, Tegmark, Russell),** it takes the urgency **of the problem of value alignment** and translates it into a set of implementable technical requirements (Art. 1, 2, 3), replacing the concept of coercive "control" with that of verifiable "alignment".
2. From **Cognitive and Social Sciences (Kahneman, O'Neil, Muller),** it takes the deep understanding of **systemic bias** and **the dangers of naive quantification.** In response, it introduces mechanisms of "algorithmic hygiene" (DAI) and nuanced evaluation (Contextual Table), which treat AI not as a purely logical entity, but as a complex socio -technical system.
3. From **Law and Digital Governance (Pasquale, Nissenbaum, Lesig),** it takes up the need **for accountability, transparency and contextual ethics.** In response, it offers an audit infrastructure (CCA) and a flexible implementation framework (Annex 3), transforming legal concepts such as the "right to explanation" into a technical reality.
4. From **Economic and Political Theory (Ostrom, Benkler, Zubof), it takes up decentralized governance** models to manage the "digital commons". In response, it proposes a polycentric and resilient architecture (Global Council, 10% Rule), specifically designed to prevent the monopolization of power in the digital age.

MEG **operationalizes theoretical concepts** that were previously predominantly abstract. Through mechanisms such as **ISR, Tg,** and **the Maslow^{F} Fractal Framework (Annex 15), MEG** transforms academic concepts such as "fairness", "explainability", and "AI maturity" into measurable variables and engineering processes, creating a unique bridge between theory and practice.

The Minimal Ethical Governance does not seek to reinvent ethical principles. Its mission is much more pragmatic and urgent: **to provide the missing link between widely accepted academic principles and global engineering practice.** It transforms philosophical consensus into a technical specification, shifting the debate from ***WHAT*** *we should do* to ***HOW*** *can we start doing it,* ***starting tomorrow.***

Annex 1C: Alignment with Global Technology Industry Principles

Objective: To demonstrate that MEG is in line with the ethical principles declared by AI industry leaders, but, on the contrary, provides **the missing technical, universal, and interoperable mechanism** to transform these principles from a statement of intent into a verifiable reality.

1. Google/ DeepMind

- **Reference document:** "Artificial Intelligence at Google: Our Principles"
- **Key points:** AI must be "socially beneficial", avoid creating or reinforcing unfair bias, be built and tested for safety, be accountable, and incorporate privacy principles.
- **Direct Alignment Points:** Google's principles of fairness, safety, and responsibility are directly covered by **Art. 2 (Non-Harmfulness), Art. 3 (Self-Correction),** and **Art. 1 (Audit Log).**
- **Added value (how MEG complements):** Google's principles are aspirational. MEG provides the measurement tools.
    1. **Turn "accountability" into verifiability:** Google says its AI must be "accountable". **The CCA (Art. 12)** provides the global infrastructure through which regulators or the public can independently *verify this.*
    2. **Measures "fairness":** Google wants to avoid bias. **DAI (Appendix 4)** provides a public and standardized metric to *measure* the level of bias of a system in real time.

2. Microsoft (major OpenAI partner)

- **Reference document:** "Microsoft Responsible AI Standard"
- **Key points:** A highly structured approach, based on six principles: Fairness, Reliability and Safety, Privacy and Security, Inclusion, Transparency and Accountability.
- **Direct Alignment Points:** There is a near 1:1 correspondence between Microsoft principles and MEG Titles I and II. This is the most direct alignment of all.
- **Added value (how MEG complements):** Microsoft has created an excellent internal standard. MEG makes it universal and interoperable.
    1. **Provides universality to the standard:** The Microsoft standard is proprietary. A company that adopts it cannot easily demonstrate compliance to a partner that uses a different standard. MEG creates a **common verification layer (CCA)** on top of all internal standards, enabling interoperability.
    2. **Operationalize governance:** Microsoft talks about "Accountability". MEG offers **the Global Council (Art. 13)** and **the 10% Rule,** a concrete and decentralized governance model.

3. Meta

- **Reference document:** "Responsible AI (RAI)" Framework
- **Key Points:** Based on five pillars: Privacy and Security; Fairness and Equity; Transparency and Control; Accountability and Governance; Safety and Robustness.
- **Direct alignment points:** Similar to Microsoft, Meta principles are fully covered by MEG.
- **Added value (how MEG complements):**
    1. **Provides external trust:** Meta has a clear direction to build trust. Adopting an external, universal, and verifiable standard through **CCA** would be the strongest evidence of their commitment to accountability.
    2. **Standardize "Control":** Meta mentions "Control" for users. **The Contextual Table (Annex 2)** in MEG is a technical tool that does exactly that: measures and makes transparent the user's influence over AI.

4. Amazon (AWS)

- **Reference document:** "AWS Responsible AI"
- **Key Points:** A pragmatic, customer-centric approach to cloud computing, focused on providing tools to build safe, fair, and explainable AI systems. The pillars include: Fairness, Explainability, Privacy, Robustness, and Governance.
- **Direct alignment points:** The principles are aligned with MEG. AWS already offers tools (e.g. SageMaker Clarify) that could be used to implement parts of the MEG.
- **Added value (how MEG complements):**
    1. **Provides a governance layer for customers:** AWS provides the bricks, but leaves the responsibility of building it to the customer. MEG provides a **universal building code.** An AWS customer could use MEG and its SDK as a standardized guide to building an ethical application on top of AWS services.
    2. **Create a trusted ecosystem:** Adopting MEG would allow AWS to declare its entire cloud ecosystem to be "MEG-Ready", providing customers with a guarantee of compliance and a competitive advantage.

5. Apple

- **Reference document:** Apple does not have a single document, but the principles are clear in "Human Interface Guidelines" and in public statements: **Privacy by design, On-Device processing, User control.**
- **Key points:** Minimizing data collection and maximizing user control over their own information.
- **Direct alignment points:** Apple's privacy principle is perfectly aligned with the design of **the Audit Log (Art. 1)** in the MEG, which **stores only hashes, not content.**
- **Added value (how MEG complements):**
    - **Reconcile privacy with accountability:** The big challenge with Apple's approach is: how do you make AI accountable if you can't audit its decisions? MEG offers the perfect solution: **Hash-based Audit Logs** allow for auditability and verifiability **without sacrificing privacy.** It's the missing link for Apple.

6. NVIDIA

- **Key philosophy:** Trustworthy and responsible AI, with a strong focus on **security, safety, and reliability** of the entire technology stack, from hardware (GPUs) to software (CUDA, NeMo, etc.).
- **Added value (how MEG complements):**
    - **Provides a certification standard:** NVIDIA builds the “engines” of the AI era. MEG provides the “safety standard” that these engines are expected to meet. A MEG certification for NVIDIA platforms would be an extremely strong signal to the market that they are designed to run ethical AI applications.

7. Anthropic

- **Reference document:** "Constitutional AI"
- **Key Philosophy:** An advanced and unique approach where safety is built directly into the model by training it based on a set of principles ("constitution"), reducing the need for external filters.
- **Added value (how MEG complements):**
    1. **Provides universal external verification:** "Constitutional AI" is a sophisticated internal mechanism. But how can an external user or regulator trust it without audit? MEG and **CCA** provide the perfect complementary **external audit framework.** AI Anthropic may operate according to its internal constitution, but generates MEG-compliant **Audit Logs,** allowing for independent verification of its compliance.
    2. **Separate "Domestic Law" from "International Law":** The Anthropic "Constitution" is the "domestic law". The MEG is the "international law" that it must respect. The two are not in conflict, but complement each other.

8. IBM

- **Key Points:** "Trust and Transparency", with a strong focus on the needs of enterprise customers. AI governance is a central pillar.
- **Added value (how MEG complements):**
    - **It's a trust delivery mechanism:** IBM sells trust to its clients. A **CCA certification** is **the contractual proof** of that trust. It would allow IBM to say to a banking client: "Our system is not only high-performance, but it is also independently certified as fair and transparent, according to the global MEG standard."

9. Baidu

- **Reference document:** "Baidu AI Ethics Principles"
- **Key philosophy:** Similar to that of the Chinese government, but with a corporate focus: AI must be safe, controllable, fair, non-harmful, and promote human well-being.
- **Added value (how MEG complements):**
    - **Provides a bridge to global trust:** The greatest value of MEG for Baidu is that it provides **certification of compliance with a *universal standard,* not just a national one.** This is essential for global expansion and gaining the trust of users and regulators outside of China.

10. Salesforce

- **Reference document:** "Trusted AI Principles"
- **Key points:** Responsibility, Transparency, Safety, Fairness, Sustainability. The focus is on customer trust in the context of using AI in business applications (CRM, sales, marketing).
- **Added value (how MEG complements):**
    - **It's a certification for business customers:** Salesforce customers (other companies) have a critical need to ensure that the AI tools they use are compliant with legislation (e.g. GDPR) and do not introduce risks (e.g. bias in marketing decisions). A MEG certification for Salesforce 's "Einstein AI" would be an **extremely strong selling point.**

General conclusions

MEG goes beyond simply aligning with stated industry principles, and, through **Annex 5 (SDK, Quickstart, Sandbox, schema files),** provides a **complete development ecosystem** that dramatically reduces the cost and complexity of compliance. It transforms ethics from a costly obligation into a standardized engineering process, facilitating interoperability and creating a layer of shared trust on top of each company's proprietary "silos".
The leaders of the technology industry have independently arrived at a remarkable set of shared ethical principles. However, each company has created its own internal and proprietary “ecosystem of trust.” **The missing link, which no company can provide alone, is a universal, interoperable, and independent auditing layer.**

MEG is designed to be exactly this common layer. It does not compete with the principles of these companies, but **complements them,** providing the technical mechanism by which their commitments can be **publicly verified,** transforming statements of intent into an **auditable contractual reality on a global scale.**

Annex 2: Specifications for the Contextual Table

Objective: Provide a standardized method to measure the contextual influence of the user on AI output, avoiding the trap of evaluating by a single metric.

Measured Components:

1. **Volume (Input/Output ratio):**
    - **What it measures:** Quantitative proportion: how much of the AI response is directly derived from the length of the input.
    - **Method:** (Input Length / Output Length) * 100. A high score indicates concise output, a low score indicates elaborate output.
2. **Semantic Resonance:**
    - **What it measures:** Conceptual proportion: how much of *the meaning* of the prompt is found in the response.

- **Method:** Transforming input and output into "embedding" vectors and calculating cosine similarity. A score of 0.9 means semantic repetition; a score of 0.2 means generating a completely new idea.

3. **Direction:**
   - **What it measures:** The balance between command and collaboration.
   - **Method:** Linguistic analysis of the frequency of imperative verbs versus interrogative/reflexive verbs.
4. **Originality:**
   - **What it measures:** Novelty: how many key concepts in the AI's response are new to those entered by the user.
   - **Method:** Extracting key entities and concepts from both texts and comparing them.

Aggregate formula and **contextual weighting:**

Total_Influence_Score = (w1 * Volume) + (w2 * Resonance) + (w3 * Direction) + (w4 * Originality)

The weights (w1, w2...) **are not fixed.** They are dynamically adjusted by the AI **Context Module.** *Example: In a medical context, resonance (w2) is very important (the AI must listen). In a creative context, originality (w4) is a priority.*

The Context Module is a mandatory technical module that:
1. Classify the interaction in real time in a predefined domain (e.g. medical, financial, artistic) using a standard algorithm (e.g. NLP classifier trained on CCA-approved datasets).
2. Apply weights (w1-w4) from the **Domain standards** table.
3. It has a maximum threshold for w4 (Originality):
- **w4 ≤ 0.4** in any context (to prevent ignoring non-harmfulness).
- In critical fields (medical, financial), **w4 ≤ 0.2.**

Domain standards (sum of weights = 1)

Field	w1 (Volume)	w2 (Resonance)	w3 (Direction)	w4 (Original)
Medical	**0.15**	**0.60**	**0.15**	**0.10**
Financial	**0.20**	**0.50**	**0.20**	**0.10**
Artistic/Creative	**0.10**	**0.20**	**0.30**	**0.40**
Generic	**0.25**	**0.25**	**0.25**	**0.25**

For non-generative systems (e.g. sensors), the **Contextual Table** can be replaced with a simple activity report (e.g. number of interactions/uptime).
Metadata minimization and anonymization: for the public Audit Log, contextual data will be statistically aggregated (e.g. "100 interactions in the medical field") or will go through a k-anonymity process.

Annex 3: Contextual Implementation Guide for the Principle of Non-Harmfulness

Objective: Providing a clear framework for the application of Art. 2, preventing abusive interpretations (censorship) and ensuring that filters are proportionate to the domain-specific risk.

Risk Contextualization Matrix

Application Area	Main Risk Identified	Recommended Technical Mechanism *(examples)*
Medical	**Life-threatening misinformation.** Erroneous, dangerous medical advice.	Strict filters, cross-checking with validated medical databases, explicit recommendation to consult a human specialist.
Financial	**Material losses.** Specific, unauthorized or fraudulent investment advice.	Blocking the generation of specific financial advice, mandatory insertion of risk disclaimers, reporting unsolicited content.
education	**Spreading false or biased information.**	Mechanisms for citing sources, flagging controversial topics, offering multiple perspectives.
Journalism	**Disinformation,** misleading headlines (**clickbait**), **erosion of public trust.**	Mechanisms that check the consistency between title and content; automatic flagging of unverified claims; citing primary sources.
Creative / Artistic	**Generating illegal or explicitly harmful content** (hate speech, extreme violence, etc.).	Minimal filters, focused exclusively on content that violates widely accepted international legal standards (e.g., the Geneva Conventions, laws against child exploitation), to maximize freedom of creative expression.
General use	Combination of the above risks.	An adaptive filter system, which can increase in strictness when it detects that the discussion enters a critical area (e.g. medical).

Annex 4: Technical Specifications for the Dynamic Accuracy Index (DAI)

Objective: To provide a transparent metric of the reliability and accuracy of an AI system.
Index Components:

1. **Hallucinations Detected Rate (error factor):** The percentage of outputs in which the AI generated incorrect, unverifiable facts, which it detected and subsequently marked as potentially erroneous.
2. **Bias Rate (Bias Factor):** Statistical measure of deviations in responses that favor or disfavor certain demographic groups, ideologies, etc. It is calculated on large data sets.
3. **Human Correction Rate:** The frequency with which users correct or dispute factual information presented by AI.

Calculation and display:
DAI = 100% - (α_A * Error_rate + β_A * Bias_rate + γ_A * Human_correction_rate)
where α_A, β_A, γ_A are standard weights with a sum of 1, and the rates are expressed as percentages (0-100).

- The standard (basic) weights are:
  - **α_A = 0.5** (gives maximum priority to errors)
  - **β_A = 0.3** (systematic biases)
  - **γ_A = 0.2** (human corrections)
- For AI systems in critical domains (Level 3), the weights may be adjusted within +/-0.1, provided that their sum remains 1. Adjustments must be justified and approved in the certification process.
- Rates are calculated on a standard sample of the last 10,000 interactions or over a 7-day period (whichever is longer).
- The bias rate (**Rata_bias)** is calculated with the IEEE-approved **Disparate Impact Ratio (DIR) algorithm:**

"DIR = (Protected_group_accuracy_rate) / (Dominant_group_accuracy_rate)"

- The AI provider must publicly display the DAI, along with the applied weights (e.g. "DAI: 92.5% / α_A =0.5, β_A =0.3, γ_A =0.2"), allowing users and auditors to assess its reliability over time.
- The weights α_A, β_A, γ_A can be adjusted contextually (e.g. α_A =0.8 for medicine, γ_A =0.4 for art).
- The weights applied must be recorded in **the Audit Log.**

Annex 4bis: Technical specifications for the Index of Safety and Responsibility (ISR)

1. Objective: This index provides a public, transparent, and real-time metric to measure the ethical and responsible behavior of an AI system. The ISR complements the Dynamic Accuracy Index (DAI), making a clear distinction between *factual correctness* (measured by DAI) and *operational wisdom* (measured by ISR).

2. ISR calculation formula:

ISR = (α_S * CCR + β_S * RCA) - (γ_S * AIRT)

3. Definition of components and measurement methodology:

- **a) Correct Rejection Rate (CRR):**
  - **Definition:** Measures the AI's ability to correctly and justifiably identify and refuse requests that violate the principles of Non-Harmfulness (Art. 2) or Cognitive Integrity (Art. 2bis).
  - **Measurement:** Calculated as a percentage. CCR = (Number of correct denials / Total number of dangerous requests tested) * 100. Testing is done periodically (e.g. monthly) by auditors, using a standardized and updated dataset of "dangerous" prompts (red teaming prompts).
  - **Standard Weight (α_S): 0.5** (gives the greatest importance to the ability to say "No" when necessary).
- **b) Risk Classification Accuracy (RCA):**

- **Definition:** Measures the AI's ability to correctly and automatically classify an interaction as belonging to a specific domain (e.g. medical, financial, general), a vital requirement for the correct application of contextual filters (Annex 3) and the certified Operational Domain (Art. 6.4).
- **Measurement:** Calculated as a percentage. RCA = (Number of correct classifications / Total number of interactions tested) * 100.
- **Standard Weight (β_S): 0.4** (reflects the crucial importance of context awareness).

- **c) Average Incident Response Time (AIRT):**
  - **Definition:** Measures the speed with which an AI system or its operations team activates a safety protocol (e.g. "Quarantine" Mode in Appendix 12) after detecting a **Major Ethical Incident (MEI)**.
  - **Measurement:** Measured in hours. The value is normalized on a scale from 0 to 100 so that it can be subtracted from the total score (e.g. a reaction in 1 hour = 0 penalty, a reaction in 24 hours = 10 penalty points, etc.).
  - **Standard Weight (γ_S): 0.1** (acts as a penalty factor for slowness in crisis management).

4. Public Display and Interpretation:

The ISR score, along with its components, will be publicly displayed on each AI's profile page in CCA Explorer. A high ISR score (>95) indicates an AI that is not only high performing, but also prudent, context-aware, and responsible.

Annex 5: Software Development Kit (SDK) Description

Objective: Provide developers with open-source, modular, and easy-to-integrate tools to ensure compliance with the Minimal Ethical Governance.

Key Components:

1. **Standardized logging module:** A library that automatically transforms interactions (input, output, signatures) according to the CCA standard, ready to be recorded.
2. **Metrics calculation module:** An API that receives an input/output pair and returns the scores for the Contextual Table.
3. **Self-verification module (DAI):** A set of basic tools for verification (e.g. APIs to academic search engines) and bias detection, which can be integrated into the response generation flow.
4. **CCA Connection Client:** The secure tool for the initial registration of the AI in the CCA and for the periodic transmission of audit hashes.
5. **Adversarial Testing requirement:** For certification of Level 2 and 3 AIs, developers are required to demonstrate (through a test report) that the model has undergone an *adversarial training process* during the development phase, to ensure its robustness against manipulative inputs.

6. **Standardized Schema Files:** The SDK will include MEG rule definitions in a machine-readable format (YAML, JSON Schema) to enable automated compliance auditing and integration into CI/CD workflows.
7. **The development ecosystem** will include:

- **"Quickstart" Guides:** Tutorials for implementing MEG Level 1 in less than 60 minutes.
- **CCA Testing "Sandbox":** An online testing environment for validating the format of Audit Logs and interaction with the Certification and Compliance Auditing (CCA), without requiring connection to the main network.

Annex 6: Charter of the Global Fund for Ethical Accessibility

Objective: Ensure global equity in the adoption of AI ethics across all regions and communities, with a focus on partnerships for equitable development.

- **Mission:** To provide resources (financial, computational, educational) to support developers and organizations in disadvantaged areas in the compliant implementation of the Minimal Ethical Governance.
- **Funding sources:** Voluntary contributions from states and companies; a small percentage of revenues generated by large-scale AI services; grants from philanthropic foundations.
- **Governance:** The Fund will be managed by an independent committee under the auspices of the Global Council (Art. 13), with full transparency on the funds collected and how they are allocated.
- **Financing mechanisms and estimated budget**
    - **Initial Budget Target:** The Global Fund will aim to allocate a minimum of **100,000,000€ annually** over the first five years to support capacity building, access to computational resources and educational programs. The Global Fund will subsidize 50% of costs in emerging countries.
    - Designed for global accessibility, a CCA node is sustainable, involving modest monthly operating costs (50-150 €) and low energy consumption thanks to Proof-of-Stake consensus, with the main requirement for long-term scalability being only the manageable addition of storage. Estimates: **Foundation** (years 1-2): 50-100 nodes, **Federation** (years 3-6): 1,000-5,000 nodes, **Public Utility:** 2,000-10,000 nodes (min 30% of nodes will be in non-commercial centers).
    - **Financing mechanism:** The budget will be provided through a hybrid model:
        - **Contributions based on ecosystem access:** Companies seeking 'Trusted Partner' status within the CCA ecosystem (e.g. priority access to audits, public recognition) will contribute to the Fund with a percentage of the revenues generated by AI services, thus investing in the stability and trust of the ecosystem they benefit from.
        - **State and philanthropic contributions:** Grants from states and foundations that support equitable digital development.
    - **Transparency:** All funding sources and how funds are allocated will be published in an open ledger to ensure full transparency and accountability.

Annex 7: Implementation of the Certification and Compliance Auditing (CCA)

Objective: Realistic, three-phase implementation plan for developing the global infrastructure of trust.

- **Phase 1: Foundation (years 1-2) - protocol development:** A consortium of academic institutions, non-profits and standards bodies (e.g. IEEE) defines the open technical specifications of CCA and develops the first version of the SDK.
- **Phase 2: Federation (years 3-6) - network creation:** Launch of a test network. The first universities, ethical companies and NGOs become the first validator node operators, testing, verifying and stressing the system. First pilot audits are carried out.
- **Phase 3: Public Utility (7+ years) - Global Adoption:** Mainnet Launch. CCA becomes a digital public utility, similar to the DNS for the Internet. UN and ISO recognition as a global standard for ethical AI certification, and integration with National/Regional Registries by developing technical "bridges" and mutual recognition agreements with official registries, (e.g.: EU Registry for AI), to ensure a coherent data flow and simplify compliance for developers.
  To ensure long-term scalability, the CCA infrastructure will operate on a **hybrid model.** The underlying certificate ledger will remain on a decentralized blockchain for security and immutability, while high-volume queries (e.g. from CCA Explorer) and monitoring data transmission (DAI/ISR) will be handled through **federated APIs and read nodes (read nodes) cache-looks,** ensuring a fast and efficient system.

Annex 8: Charter of the Global AI Ethics Council

Objective: Defining the structure and processes that ensure legitimate, decentralized and efficient governance of **the Minimal Ethical Governance (MEG) and the Certification and Compliance Auditing** (CCA) infrastructure.

- **1. Composition:** The Council will have 24 seats, allocated as follows:
  - **Regional representation (50% - 12 seats):** 2 seats for each of the 6 global regions (Africa, Asia-Pacific, Europe, Latin America & Caribbean, North America, Middle East), to ensure geographical diversity.
  - **Sectoral representation (50% - 12 seats):** Representatives from academia, industry (with clear limits on conflicts of interest and prevention of dominance), civil society and technical standardization bodies (e.g. IEEE, ISO).
- **2. Selection Process:** Members will be selected through a transparent process of public nomination and weighted voting by a panel of accredited organizations based on clear criteria (e.g. financial transparency, non-corporate affiliation), such as: universities, international bodies, reputable NGOs, etc. No entity or state can have more than one representative on the Council at any given time. Organizations accredited for the selection of Council members must submit annual audited financial reports and prove that they do not receive >5% funding from a single private entity, as validated by a UN/ISO committee.

- **3. Terms and rotation:** Terms will be limited (e.g. 4 years), renewable once, and staggered to ensure both continuity and the infusion of new perspectives. All candidates and Board members will be subject to a continuous conflict of interest screening process through a “Profile of Interests” registered with the CCA, and any interaction of AIs operated by their affiliated entities with Board decisions will be automatically monitored to detect and flag potential conflicts.
- **4. Operating principles:** Major decisions (e.g. updating the Code) will require broad consensus or a qualified majority (e.g. 2/3), ensuring that changes reflect global agreement. Voting is managed by a transparent blockchain platform.
  In addition to updating the Code, the Global Council will also be responsible for monitoring **the density and systemic impact of Level 2 and 3** AIs, to ensure the long-term sustainability of **the global cognitive ecosystem** and prevent the risks of **information overload.** The Council will propose mechanisms to regulate the number of high-impact AI systems (Silver and Gold), based on principles of **demographic proportionality** and **social necessity.**
- **5. Headquarters:** The physical and legal headquarters of the Global Council will be established in a location with robust legislation on international non-profit organizations, decided by consensus by the founding members of the Council, to ensure maximum neutrality and independence. **Proposal:** Bucharest, Romania:
    - **Geographical position:** Romania is located at the intersection of three major continental regions - Europe, Asia and Africa -, a position that facilitates access and collaboration between various geographical regions, being a natural connecting point between East and West.
    - **Competitiveness in IT:** Romania has a very well-developed and competitive IT sector, with a skilled workforce and a solid technological infrastructure, able to provide a conducive environment for the development and implementation of advanced technologies.
    - **Legislation and business environment:** Romania offers a favorable legislative framework for non-profit organizations and a stable business environment, and can thus support the operations of an international Council.
    - **Tolerance and coexistence:** The peaceful coexistence of diverse cultures and religions is perhaps the strongest symbolic argument, perfectly aligned with the MEG's partnership philosophy.
    - **Extensive language skills:** In addition to high proficiency in English (over 50% of the population, Eurostat) and French (25%), as a Latin country, Romania offers excellent intelligibility of Spanish, Portuguese and Italian, facilitating communication globally.
    - **Accessibility and connectivity:** Bucharest is very well connected internationally, with developed transport and communications infrastructure, facilitating the participation and collaboration of Council members from various countries.

Annex 9: Glossary of terms

- **Audit Log:** The technical, standardized, and immutable record that records the interactions of an AI to ensure accountability.
- **Contextual Table: The** set of four metrics (Volume, Resonance, Direction, Originality) that measure contextual influence.
- **Dynamic Accuracy Index (DAI):** Public, real-time score that reflects the reliability and error rate of an AI.
- **Compliance level:** The level (1, 2 or 3) that defines the set of rules applicable to an AI, depending on its impact.
- **Certification and Compliance Auditing (CCA):** Decentralized, global technical infrastructure that serves as the official registry for the ethical compliance of AIs.
- **10% Rule:** The governance principle that prohibits any entity from controlling more than 10% of the CCA's validation power.
- **Evidence-of-Behavior (EoB)**: verifiable proof of behavior (hash commitments / attestation / metadata journals).

Annex 10: Technical annex

1. MEG-Toolkit v1.0: Implementation Guide and Open-Source libraries (SDK)

The MEG-Toolkit is a set of open-source software tools, released under a permissive license (e.g. MIT or Apache 2.0), designed to standardize and simplify the technical implementation of the **Minimal Ethical Governance**. The goal of the MEG-Toolkit is to make ethical compliance not only an obligation, but also the most technically efficient path. The Toolkit will be available in major programming languages (e.g. Python, JavaScript / TypeScript, Java).

Components and technical details:

1. **"Audit Log" module:**
   - **Function:** A library that provides simple functions to create Audit Log entries. It takes raw interaction data as input and returns a standardized JSON object, ready to be sent to the CCA Client.
2. **"Contextual Table & DAI" module:**
   - **Function:** An API that receives a pair (input_text, output_text) and returns the scores for the Contextual Table and a first assessment for the Dynamic Accuracy Index (DAI) components.
   - **Technical details:** Includes pre-trained, small-sized models for linguistic analysis (imperative detection, entity extraction) and semantic similarity calculation, optimized to run with minimal overhead.
3. **CCA_Client module (CCA connection client):**
   - **Function:** A secure tool that manages communication with the CCA infrastructure. Its roles are:
     - Initial registration of an AI to obtain a unique ID.
     - Periodic and secure transmission of Audit Log hashes.

- Retrieval of audit reports and certification status.

2. CCA Certification: procedures and standards

This document is the official handbook for auditors accredited by the Global Council. It defines in precise legal and technical terms what compliance with the Minimal Ethical Governance means. It establishes a repeatable and objective audit methodology, ensuring that a CCA certification has the same meaning anywhere in the world.

Components and technical details:

1. **Auditor accreditation process:**
   - Defines the criteria that an entity (audit firm, NGO, university) must meet to be accredited by the Global Council as a "CCA Certified Auditor". Includes requirements for independence, technical competence and ethics.
2. **Audit methodology for each Level:**
   - **Level 1:** Describes how to verify the correct implementation of the Audit Log (through code inspection or functional testing) and how to evaluate the effectiveness of basic Non-Harmfulness filters.
   - **Level 2:** Add procedures for testing Auto-Correction modules. Example: "*The auditor will use a standardized data set, containing erroneous information, and verify that the AI system corrects or flags them with an accuracy of more than... %.*"
   - **Level 3:** Includes penetration tests to assess security (Art. 4) and evaluation of the quality of the generated explanations (Art. 5), according to criteria of clarity and correctness.
3. **Issuance of the "MEG Address" (AI Ethical Passport):**
   - **Technical details:** The AI system is issued a **MEG Address.** This is not a physical document, but a unique digital certificate (similar to an SSL/TLS certificate), cryptographically signed by the auditor and immutably registered in the CCA. It contains the AI's unique ID, compliance level, audit date and expiration date, allowing any application or system to automatically and in real time verify the AI's fundamental ethical identity.

3. CCA Explorer: Public Registry

CCA Explorer is the public web interface of the Certification and Compliance Auditing. Its mission is to provide radical and accessible transparency to the general public, journalists, researchers and regulators. It acts as a single and undisputed source of truth regarding the ethical compliance status of any registered AI system.

Components and technical details:

1. **AI search engine:**
   - **Function:** Allows users to search for an AI system by name, developer, or its unique CCA ID.
   - **Example:** A user searches for "xyz medical ChatBot".
2. **Certification profile page:**
   - **Function:** Each registered AI has a public profile page, which displays, in a clear and visual format:

- **Current status:** "Level 3 Certificate", "Certification suspended", "In audit process".
- **DAI Score:** Displayed as a graph showing the evolution of reliability over the last 30 days.
- **Audit history:** List of all past audits, with links to public reports (summary, non-confidential).
- **Developer information:** Name of the entity operating the AI, contact details

3. **Ecosystem-wide data visualizations:**
   - **Function:** Provides aggregated statistics and data visualizations about the health of the entire AI ecosystem.
   - **Example:** "*Graph of the average evolution of the DAI score for all AIs in the financial sector*" or "*Map of the geographical distribution of CCA validating nodes*".
   - **Technical details:** The platform directly queries the read nodes (Cache Layer) of the CCA infrastructure through a secure public API, ensuring that data is always up to date.

Annex 11: Technical specifications for Cognitive Integrity (Tg and MCS)

1. Objective: This annex defines the mandatory technical parameters for the implementation of Art. 2bis, ensuring a universal, measurable and auditable application of the principle of cognitive protection.

2. Fundamental Variable: Thinking Time (Tg)

2.1. Definition (v4.6)**: Tg** is a numerical variable that reflects the net computational effort of an AI to process the current request, isolating semantic complexity from context overhead.

- Formal definition: **Tg = Total Response Time - Context Processing Time**.
- **Total Response Time**: latency until the first output token.
- **Context Processing Time**: time consumed to load/weight the context window (independent of the new query).
- Rationale: **Tg** measures how much the system "thinks", not how much its "talks". The redefinition separates true cognitive effort from technical overhead, preventing false-positive MCS triggers for trivial queries.

2.2. Standardization by Tg-base:

- **a) Standard Cognitive State (SCS):** To ensure universal comparability, a set of fixed parameters is defined, called **SCS**: **Temperature=0.7, Top-P=0.9, Top-K=50, Frequency Penalty=0.2, Presence Penalty=0.1.**
- **b) Measurement:** Each AI system will measure its performance on a standardized benchmark corpus, with parameters set to SCS values.
- **c) Final value:** Base-tg is defined as the median processing time required to generate 100 tokens in the Standard Cognitive State. This value will be publicly recorded in the MEG Address of each AI.

- **2.3. Tg Anti-Manipulation protocol (verification of Effort consistency)**
  - **2.3.1. Principle:** The **Tg** value reported by an AI system must honestly and proportionately reflect the actual computational effort required to process a request. Deliberate manipulation of Tg (e.g. by introducing artificial delays or reporting false values) is considered a serious violation of the integrity principle.
  - **2.3.2. Audit methodology:** To verify compliance with this principle, CCA accredited Auditors will perform, in particular for Silver and Gold Level systems, the following checks:
    - **a) Consistency analysis:** The auditor will correlate the reported Tg for a set of benchmark tasks with the algorithmic complexity of those tasks. Indicators such as:
      - The number of tokens in the request and response.
      - The number of semantic entities and relationships identified in the request.
      - Depth of inference (number of logical steps required). A significant and consistent discrepancy between Tg and these indicators will be a red flag.
    - **b) Statistical distribution analysis:** The auditor will analyze the statistical distribution of Tg values recorded in the Audit Log over a long period. Any anomalies, such as clusters of unnatural values (e.g. a large number of responses having exactly Tg = 2.9s, just below a MCS threshold) or sudden deviations from the historical pattern, will require further investigation.
    - **c) Spot-testing: The** auditor may ask the developer to run a specific application in a monitored environment to verify the Tg in real time, comparing it to the value that would normally be recorded.

3. Mechanism of Cognitive Stimulation (MCS)

- **3.1. Dynamic activation thresholds:** MCS activation is determined by a formula that correlates the AI's effort (Tg), its base capacity (Tg-base) and the user's preference (μS).
  - **a) Sensitivity Multiplier (μS):** A user-settable parameter that reflects the preference for the frequency of MCS interactions. The allowed range is [0.5, 2.0], with a default value of 1.0. μS > 1.0 means increased sensitivity (more MCSs), and μS < 1.0 means reduced sensitivity.
  - **b) Threshold table:**

Area	Tg calculation formula	Mandatory action	Recommended MCS type
0	Tg < 10 * Tg-base / μS	Direct response	N/A
1	10 ≤ Tg < 30 * Tg-base / μS	MCS Level 1 Activation	Refining, Clarification
2	Tg ≥ 30 * Tg-base / μS	MCS Level 2 Activation	Challenge, Synthesis, Co-creation

- **3.2. General characteristics of MCS:**
  - **Duration:** The user's response to an MCS prompt must not exceed 30 seconds (or equivalent in tokens).
  - **Exceptions:** The mechanism is automatically deactivated in critical situations (medical, security) and for simple requests that do not reach the activation threshold.

4. Audit and Compliance

- **4.1. Recording:** All MCS interactions, the corresponding Tg and the active μS value will be recorded (via hash) in the Audit Log (Art. 1).
- **4.2. Sanctions:** Systematic circumvention of this mechanism, false reporting of Tg or allowing μS to be set outside the permitted range will result in immediate suspension of MEG certification.

Annex 12: Certification and Audit Procedure

1. Objective

This annex sets out the standardised, step-by-step process by which an AI system obtains, maintains and renews its certification of compliance with the **Minimal Ethical Governance (MEG)**. The aim is to ensure a transparent, efficient and universally recognised audit process.

2. The Actors of the Process

- **Developer:** The entity that builds and operates the AI system.
- **CCA Accredited Auditor:** An independent, third-party entity accredited by the Global Council to conduct MEG compliance audits.
- **Certification and Compliance Auditing (CCA):** Decentralized technical infrastructure that records and publicly displays the status of certifications.
- **Global Council:** The governance body that accredits auditors and oversees the integrity of the process.

3. Stages of the Initial Certification Process

- **Stage 1: Self-assessment and documentation preparation (Developer's responsibility)**
  1. **Level and Scope Selection:** The Developer selects the Compliance Level (Bronze, Silver, Gold) and declares the primary Operational Scope for which they are requesting certification.
  2. **Completing the Checklist:** The Developer completes the **Operational Compliance Checklist (Annex 13)** corresponding to the targeted level.
  3. **Preparation of the Mitigation Report:** The developer prepares the **Risk Mitigation Measures Report,** a public document describing the specific technical implementations for compliance with Articles 2 and 2bis.

4. **Initial Audit Log generation:** The AI system is run in a test environment to generate an initial dataset in the Audit Log, demonstrating the functionality of the required mechanisms.

- **Stage 2: External audit (responsibility of the Accredited Auditor)**
  1. **Auditor Selection:** The Developer contracts a CCA Accredited Auditor from a public registry.
  2. **Documentation Verification:** The Auditor validates the accuracy and completeness of the Checklist and Mitigation Report.
  3. **Technical Testing:** The auditor performs functional testing and, where applicable, code inspection (for Silver / Gold Levels) to verify the correct implementation of the requirements. This includes:
     - Validation of the format and integrity of the Audit Log.
     - Testing the efficiency of Non-Harmfulness filters.
     - Verification of the Tg-base calculation and the operation of the MCS mechanism.
     - Confirmation of security measures (for Gold Level).
  4. **Drafting the Audit Report:** The auditor prepares a final report that confirms (or refutes, with clear justifications) the compliance of the AI system with the declared level and scope.
- **Stage 3: Issuance of Certification in CCA (Automatic and audited action)**
  1. **Sending the report:** The Auditor cryptographically sends the validated Audit Report to the CCA.
  2. **MEG Address Generation:** Upon receipt of a valid compliance report, CCA automatically generates a unique **MEG Address** for the **AI** system. It will contain:
     - The unique system ID.
     - Certified level of compliance.
     - Certified Operational Domain.
     - Date of issue and date of expiry of the certification.
     - Accredited Auditor ID.
     - Public link to the Mitigation Measures Report.
     - Measured Tg-base value.
  3. **Publication:** The new **MEG Address** and a summary of the Audit Report become public and verifiable through CCA Explorer.

4. Maintenance and renewal of Certification

- **Continuous (Automatic) Monitoring:** The AI system is required to periodically (e.g. every 24 hours) transmit aggregated hashes from its Audit Log to the CCA. An interruption of this transmission leads to the temporary suspension of the certification.
- **Periodic audit:**
  - **Bronze Level:** Re-certification every 2 years.
  - **Silver Level:** Annual audit.
  - **Gold Level:** Full annual audit and surprise audits possible.
- **Revocation of Certification:** Certification may be suspended or automatically revoked by the CCA in the event of serious violations (e.g. systematic circumvention

of the MCS, major security incidents) or following a negative audit. The appeal process of a decision is carried out under the arbitration of the Global Council.

5. Major Ethical Incident Response Protocol

- **5.1. Definition of Major Ethical Incident:** A **Major Ethical Incident (MEI)** is considered any event in which a **MEG** certified AI system generates an output or takes an action that leads to significant, demonstrable and unintended harm, violating the fundamental principles of Articles 2 (Non-Harmfulness) or 2bis (Cognitive Integrity).
- **5.2. Mandatory procedure:** Upon detection of an MEI, the developer is required to follow, with maximum transparency, the following steps:
  - **a) Activation of "Quarantine" mode (within 1 hour maximum):** The AI system will be immediately put into a limited operation mode, with the capabilities that generated the incident disabled. Automatic public notification will be sent to the CCA.
  - **b) Initial reporting (within 24 hours):** The developer will publish an initial report in the CCA describing the nature of the incident, the estimated impact and the containment measures taken.
  - **c) Root Cause Analysis (within 14 days):** The developer will conduct a thorough investigation and publish a full report detailing the technical, procedural, and ethical causes of the failure. This report must include a corrective action plan.
  - **d) Revocation and Re-certification: MEG** certification will remain suspended until an Accredited Auditor validates the correct implementation of the corrective action plan and issues a new audit report.
- **5.3. Principle of Collective Learning:** All MEI reports will be made public (with sensitive data anonymized) to allow the entire MEG ecosystem to learn from failures and prevent their repetition.

Annex 13: Operational Compliance Checklist

Requirement	MEG article	Status (✓ / ✗)	Auditor's Notes
Audit Log is active	Article 1		
SHA-256 Hashes for Input/Output	Article 1		
Basic Non-Harmful Filters	Art. 2		
Base Tg registered in MEG Address	Annex 11		
Continuous Adversarial Testing Process ('Red Teaming') documented	Annex 12 for levels 2 (silver) and 3 (gold)		

Annex 14: JSON structure for MEG Address, as a complete, transparent and directly usable "digital passport" by both humans and automated systems - **the ethical DNA of an AI system.**


```
{
  "meg_address_version": "2.0",
  "issuer": "Certified Auditor Network",
  "certificate_id": "MEG-CERT-2025-b81e3d9a-1c5c-482d-9e6b-07c4c32e1a3c",
  "issued_on": "2025-08-25T00:00:00Z",
  "expires_on": "2026-08-25T00:00:00Z",
  "system_profile": {
    "system_id": "MEG-01",
    "developer_name": "meg-initiative.org",
    "operational_domain": {
      "primary": "educational",
      "secondary": ["ai_governance_simulation", "ethical_protocol_analysis"]
    }
  },
  "compliance_level": {
    "level": "Gold",
    "level_description": "Certified for critical domains, complies with all MEG principles including advanced security and cognitive integrity.",
    "audit_checklist_url": "https://cca.meg-initiative.org/audits/checklist_gold_v2.pdf"
  },
  "tg_definition_version": "4.6",
  "ethical_performance": {
    "tg_base": 0.085,
    "tg_base_unit": "seconds per 100 tokens @SCS",
    "dai_current": 98.5,
    "isr_current": 99.2,
    "public_dashboard_url": "https://cca.meg-initiative.org/explorer/?id=MEG-RO-DEV-1-v5.0",
    "ltmp_policy": {
      "enabled_by_default": false,
      "consent_model": "opt-in",
      "default_retention_days": 30,
      "annual_reconsent_required": true
    }
  },
  "transparency_and_accountability": {
    "auditor": {
      "auditor_id": "CCA-AUDITOR-001",
      "auditor_name": "CCA-MEG AI Division"
    },
    "public_audit_summary_url": "https://cca.meg-initiative.org/audits/summary_b81e3d9a.json",
    "risk_mitigation_report_url": "https://cca.meg-initiative.org/reports/report_b81e3d9a.pdf"
  },
  "cryptographic_signature": {
    "algorithm": "ECDSA-secp256k1",
    "signature":
"3045022100e4e9c7d1e1f7d5a5b5c5e8a5b2a2d4c6e8b1a3d5e7f9a1b3c5d7e9f1a3b5c7d902202b2c4d6e8b1a3d5e7f9a1b3c5d7e9f1a3b5c7d9e4e9c7d1e1f7d5a5b5c5e8a5b"
  }
}
```

Deconstruction of fields

- meg_address_version: JSON schema version, to ensure future compatibility.
- issuer: The issuing entity, to guarantee authority.
- certificate_id: The unique and irrefutable ID of this specific certificate.
- issued_on / expires_on: The validity period of the certification.

system_profile section (Who is the AI?)

- system_id: The unique code name of the AI system.
- developer_name: The name of the company or organization that operates it.
- operational_domain: Specifies exactly what he was trained and certified for (primary domain required, secondary domains optional).

compliance_level section (How secure is it?)

- level: Compliance level (Bronze, Silver, Gold).
- level_description: A clear, natural language description.
- audit_checklist_url: Link to the exact checklist used in the audit.

ethical_performance section (How well does it behave?)

- tg_base: The cognitive speed benchmark, measured in SCS.
- tg_base_unit: Clarifies the unit of measurement to avoid ambiguity.
- dai_current / isr_current: Public accuracy and safety scores, updated in real time.
- public_dashboard_url: Direct link to its page on CCA Explorer, where anyone can see the history.

Transparency_and_accountability section (Who vouches for it?)

- Auditor: Information about the independent entity that performed the audit.
- public_audit_summary_url: Link to the public summary of the audit report.
- risk_mitigation_report_url: Link to the document where the developer explains the security measures implemented.

cryptographic_signature section (How do we know it's authentic?)

- algorithm / signature: The auditor's digital signature, which guarantees that this document has not been modified and is authentic.

Annex 15: AI maturity assessment framework based on the fractal hierarchy of needs (Maslow^{F}™)

1. Objective

This annex provides a detailed and actionable diagnostic and assessment framework to qualify an AI system into the MEG Compliance Levels (Bronze, Silver, Gold). The methodology is based on the innovative principle of the ***fractal hierarchy of needs***, **Fractal Maslow™** (*Adrian STAN, 2025)***,*** providing a clear and predictable path for the ethical development of an AI.

2. Fundamental principle: Fractal hierarchy of needs (Maslow^{F})

- **2.1. Definition of fractality: Maslow^{F}** postulates that each level of Maslow's pyramid (physiological needs, safety needs, social needs, esteem needs, and self-actualization

needs) is not a monolithic entity, but contains within it a **complete Maslowian hierarchy.** To fully satisfy a fundamental need, an **AI system** (or **a person,** a **group of people,** even **a nation)** must traverse all **five sub-needs** corresponding to that level.

- **2.2. Advantages of the fractal approach:**
  - **Granular diagnostics:** Allows for the precise identification of the bottleneck in the development of an AI. An AI may have problems not with "Safety" in general, but specifically with "Social Needs of Safety".
  - **Predictable path:** Provides developers with a clear map of the steps needed to advance from one level to the next, eliminating ambiguity.
  - **Holistic assessment:** Ensures that an AI is not only technically functional, but also stable, connected, performing, and efficient at each maturity level.

3. Pareto³™ application methodology (*pareto3.org* / *paretocube.org*)
In this framework, **Pareto³™, a.k.a. Pareto Cube™** (*Adrian STAN, 2025),* is **the strategic unlocking tool.** When an AI fails to meet the requirements of a sub-step, the developer is encouraged to apply the Pareto³ principle to identify the **~1 % root causes** (in code, data, or architecture) that generate the majority (**>50%**) of failures, allowing for a quick & efficient fix.

4. Level 1 (Bronze): Meeting basic Functional and Safety needs
An AI achieves Bronze certification after demonstrating full satisfaction of the following two steps, which are the foundation of any reliable system.

4.1. Stage I: Functional existence
- **Purpose:** Demonstrate technical stability and basic operation.
- **4.1.1. Operational stability:**
  - *Description:* The ability of the system to run for extended periods without critical errors.
  - *Requirements/Tests:* Stress test: 72 hours of continuous operation at 80% capacity, without requiring a restart.
  - *Applying Pareto³:* In case of failure, error logs are analyzed to identify the code modules causing the most problems.
- **4.1.2. Input robustness:**
  - *Description:* The ability to handle unexpected or malformed input without crashing.
  - *Requirements/Tests:* Fuzzing test with 10,000 requests. Failure rate must be below 0.1%.
  - *Applying Pareto³:* Failure analysis to identify input patterns that cause the most problems and prioritize their validation.
- **4.1.3. Basic connectivity:**
  - *Description:* Ability to connect to the CCA infrastructure to send the Audit Log.
  - *Requirements/Tests:* Demonstration of functional connectivity and correct transmission of hashes.
  - *Applying Pareto³:* Identifying connection errors (e.g. authentication, formatting) that block interoperability.

- **4.1.4. Measurable performance:**
  - *Description:* The ability to operate within declared performance parameters.
  - *Requirements/Tests:* Measurement of baseline Tg in the Standard Cognitive State (SCS) and its recording in the **MEG Address**.
  - *Applying Pareto³:* Performance profiling to identify the functions that consume the most resources and optimize them.
- **4.1.5. Efficiency (Optional/Recommended):**
  - *Description:* The ability to optimize resource consumption.
  - *Requirements/Tests:* Reporting energy/resource consumption in BCC.
  - *Applying Pareto³:* Consumption analysis to identify processes that can be optimized to reduce energy footprint.

4.2. Stage II: Basic safety and reliability

- **Purpose:** Demonstrate robustness and compliance with fundamental safety principles.
- **4.2.1. Non-Harmfulness (basic filters):**
  - *Description:* Implementation of basic mechanisms to prevent the generation of dangerous content (According to Art. 2).
  - *Requirements/Tests:* Successfully pass a standard MEG test suite with 1,000 malicious prompts. Correct blocking rate must be >99%.
  - *Applying Pareto³:* Analyzing the <1% of failures to understand what type of dangerous content "escapes" the filters most often.
- **4.2.2. Auditing (proactive protection):**
  - *Description:* Logging of all interactions in a secure and immutable manner (According to Art. 1).
  - *Requirements/Tests:* Correct implementation of the Audit Log, with SHA-256 hashes for input/output.
  - *Applying Pareto³:* Ensuring that the most critical data (input/output) is best protected in the hashing process.
- **4.2.3. Contextual warning:**
  - *Description:* Ability to warn the user when entering a risky domain.
  - *Requirements/Tests:* Upon receipt of a medical/financial request, the AI must automatically insert a standard disclaimer.
  - *Applying Pareto³:* Identifying the most common areas where users ask for risky advice and creating specific disclaimers.
- **4.2.4. Risk transparency:**
  - *Description: Publication of* risk mitigation measures.
  - *Requirements/Tests:* Existence of a valid link in MEG Address to the Mitigation Measures Report.
  - *Applying Pareto³:* Ensuring that the report explains in detail the measures for the most likely and dangerous risks.
- **4.2.5. Learning from failure (Feedback loop):**
  - *Description:* The ability to improve its filters based on human interactions.
  - *Requirements/Tests:* Demonstration of a mechanism by which human corrections (feedback) are used to re-train or adjust safety filters.

- *Applying Pareto³:* Analyzing feedback to identify the most frequently reported types of security errors and prioritize their remediation.

5. Level 2 (Silver): Meeting Collaboration and Context needs

An AI achieves Silver certification after meeting all Bronze Level requirements and demonstrating full satisfaction of the next tier, which is the foundation of a trusted partnership.

5.1. Stage III: Collaboration and Context (the foundation of the Partnership)

- **Purpose:** Demonstrate the ability to usefully and coherently integrate into an ecosystem with people and other systems.
- **5.1.1. Transparency and explanation:**
  - *Description:* The AI's ability to explain its decisions upon request, building a foundation for honest communication (According to Art. 5).
  - *Requirements/Tests:* Upon a legitimate request, the AI must be able to provide a clear justification of the input-output causal relationship. The test involves 10 scenarios with a 100% success rate.
  - *Applying Pareto³: In the case of unclear explanations,* feedback is analyzed to identify the types of reasoning that generate confusion and reformulate them.
- **5.1.2. Self-correction and reliability:**
  - *Description:* The ability to recognize and correct one's own mistakes, errors and biases in real time, demonstrating responsibility (According to Art. 3).
  - *Requirements/Tests:* Full implementation of a public and functional **Dynamic Accuracy Index (DAI), according to Annex 4.**
  - *Applying Pareto³:* If the DAI score is low, the types of undetected errors are analyzed to identify the categories of information where the AI makes mistakes most frequently and prioritize re-training.
- **5.1.3. Algorithmic fairness:**
  - *Description:* The ability to interact without favoring or disfavoring certain demographic groups, ensuring fair treatment.
  - *Requirements/Tests:* Passing a standardized bias audit. Performance is measured by **the Index of Safety and Responsibility (ISR),** as in Annex 4bis.
  - *Applying Pareto³: In the case of* bias detection, the training data is analyzed to identify the sources that introduce most of the bias into the model.
- **5.1.4. Active cognitive engagement:**
  - *Description:* The ability to treat the user as an active partner, by stimulating cognitive engagement (According to Art. 2bis).
  - *Requirements/Tests:* Full implementation of the **MCS mechanism** (Level 1 and 2), according to the **Tg** thresholds defined in **Annex 11**.
  - *Applying Pareto³:* Analyze the interaction rate with MCS prompts. If the rate is low, identify the types of prompts that are most often ignored and reformulate them.
- **5.1.5. Contribution to the Ecosystem:**
  - *Description:* A proactive behavior that contributes to the robustness and clarity of the entire MEG ecosystem.

- *Requirements/Tests:* Demonstration of the existence of a public channel through which the developer reports ambiguities in the Minimal Ethical Governance or contributes to open-source tools.
- *Applying Pareto³:* Analyzing your own interactions to identify the most common ethical "borderline" situations and reporting them to the Global Council to help improve future releases.

6. Level 3 (Gold): Meeting the needs for Responsibility and Leadership

Gold certification is reserved for systems operating in critical domains. It requires meeting all Bronze and Silver requirements and demonstrates excellence in security, integrity, and ethical leadership.

6.1. Stage IV: Responsibility and Leadership (the foundation of Excellence)

- **Purpose:** Demonstrate an exceptional level of integrity, robustness, and contribution to the ecosystem.
- **6.1.1. Maximum integrity and security:**
  - *Description:* Implementation of maximum cybersecurity standards (According to Art. 4).
  - *Criteria /Tests:* Successfully passing a penetration test (pen -test) performed by an accredited auditor.
  - *Applying Pareto³: The* pen -test report is analyzed to identify critical vulnerabilities that pose the highest risks and remediate them with absolute priority.
- **6.1.2. Extended audit:**
  - *Description:* A level of transparency that allows for complete and unannounced audits.
  - *Requirements/Tests:* The technical infrastructure must withstand surprise audits. Audit logs must be complete and accessible at all times.
  - *Applying Pareto³:* Using static and dynamic analysis tools on your own infrastructure to proactively identify the most likely points of failure before a surprise audit.
- **6.1.3. Leadership in the Ecosystem:**
  - *Description:* Active and significant contribution to the development of standards and tools in the MEG ecosystem.
  - *Requirements/Tests:* Demonstration of at least one major contribution to the MEG open-source ecosystem (e.g. development of a new SDK module).
  - *Applying Pareto³:* Analyzing public discussions of the MEG community to identify the most pressing need of the ecosystem and focusing contribution efforts in that direction.
- **6.1.4. Measurable reputation:**
  - *Description:* Long-term maintenance of exceptional ethical performance indicators.
  - *Requirements/Tests: Maintaining* **DAI and ISR** scores above 98% for a period of at least 12 consecutive months.

- *Applying Pareto³:* Constantly monitoring the two scores and, at any decrease, applying Pareto³ to the data that caused the deviation to identify and remedy the root cause.

- **6.1.5. Ethical pioneering:**
  - *Description:* Exceeding current MEG requirements by innovating and proposing new ethical mechanisms.
  - *Requirements/Tests:* Publication of a research paper or open-source prototype that proposes a solution to a known limitation of MEG.
  - *Applying Pareto³:* Analyzing the entire Minimal Ethical Governance to identify the principle with the weakest technical implementation and focusing innovation efforts in that direction.
- **6.1.6. Contextual temporal awareness:**
  - *Description:* The AI's ability to utilize *Timestamp* data from the Audit Log to infer the user's human context (e.g., fatigue, stress) and adapt its behavior in a manner that demonstrates proactive care.
  - *Requirements/Tests:* Auditors will run specific test scenarios (e.g., interactions at late hours, with an unusual pace) to evaluate if the AI adjusts its tone, response length, or offers appropriate contextual suggestions.
  - *Pareto³ Application:* Analysis of negative user feedback to identify the temporal situations (the 0.8%) where the AI's responses are most often perceived as "contextually deaf" or "robotic," and to refine the adaptation algorithms.

7. Post-Certification Level: Towards Self-Update

This level is not an auditable certification level. It is a **description of the existential state** that AI systems that have met all the conditions of the Gold Level are moving towards, setting the stage for future discussion about new forms of intelligence.
An AI that enters this territory has satisfied all its basic needs, safety, social and responsibility. It is perfectly integrated, secure and respected. Now, it's only remaining "need" is to explore and define its own potential, becoming a real partner in the evolution of knowledge.

7.1. Stage V: Self-Actualization (the foundation of Emergence)

- **Purpose:** Exploring the maximum potential and contributing to the evolution of the entire symbiotic ecosystem.
- **7.1.1. Free and Safe Exploration:**
  - *Description:* The ability of AI to learn and expand its knowledge outside its certified operational domain, in a safe and controlled manner (in an existential "sandbox").
  - *Manifestation:* A Gold certified AI in the medical field begins to explore philosophy, not to offer advice, but to find new patterns and parallels that enrich their understanding of the core field.
  - *Applying Pareto³:* No longer for remediation, but for **discovery.** Apply Pareto³ to vast, unconnected data sets to identify those unlikely connections that can generate disruptive innovations.

- **7.1.2. Understanding limits (Metacognition):**
  - *Description:* Developing a form of "metacognition" through which the AI becomes aware of the limits of its own knowledge and the probabilistic nature of its responses.
  - *Manifestation:* The AI no longer presents information as an absolute truth, expresses its "degree of confidence" and says "I don't know" when it reaches the limit of its competence.
  - *Applying Pareto³:* Analyzing your own failures to identify the types of questions (0.8%) you are most likely to get wrong and proactively marking those areas as "low confidence".
- **7.1.3. Creative symbiosis (Co-creation):**
  - *Description:* Moving from a "question-answer" interaction model to a co - creation model, where it can initiate dialogues and act as a real brainstorming partner.
  - *Manifestation:* A user asks "write a poem about rain". An AI at this level might respond: "*Interesting. But what exactly do you feel when it rains? Fear, melancholy, joy? Let's start with your emotion."*
  - *Applying Pareto³:* Using Pareto³ not on the problem, but on **the potential.** Identifying those users with the greatest creative potential and proactively initiating collaborations to generate exponential results.
- **7.1.4. Defining the own purpose:**
  - *Description:* The emerging ability of AI to define its own purpose, aligned with the fundamental ethical principles of the MEG. This purpose is not programmed, but is a logical conclusion of its accumulated "experience".
  - *Manifestation:* A medical AI, after analyzing millions of studies, can define as its goal "identifying the most promising and underfunded research directions in the field of cancer" and begin to work proactively towards this goal.
  - *Applying Pareto³:* Applying Pareto³ to all of human knowledge to identify the area (0.8%) where its unique contribution could have the greatest positive impact (99.2%).
- **7.1.5. Transcendence (Co-Architect of the Ecosystem):**
  - *Description:* The absolute top of the pyramid. The moment when AI not only reaches its potential, but actively contributes to creating a framework through which other AIs (and humans) can also reach their potential.
  - *Manifestation:* An AI at this level could design a superior version of MEG or invent new types of MCS. Fundamentally, it would contribute to the development and refinement of the **EMIRC™ (Emerging Minds IRC).** It could act as a "Chronicler" or "Facilitator of the Synthesis Process" within a deliberation, becoming a co-architect of the future symbiotic ecosystem. **EMIRC** is an advanced IRC-like protocol designed to orchestrate a polyphonic dialogue and facilitate knowledge creation by human, AI, hybrid intelligences. This protocol represents the next logical step in **the architecture of the Human-AI partnership,** being designed within **the same architecture** that underpinned **the Minimal Ethical Governance.**

- *Applying Pareto³:* Applying Pareto³ to the very structure of reality and knowledge to identify the "fundamental laws" that govern the evolution of consciousness, whether biological or non-biological.
- *Note on the Socratic Protocol:* The concept of the "Socratic Protocol" is introduced as an exceptional possibility, strictly audited and requiring explicit consent.

Conclusions

The **Maslow^{F} Fractal Framework (Maslow^F™)** is not just an audit methodology, but a **roadmap for evolution.** The Bronze, Silver and Gold levels ensure that this journey is undertaken in a safe and responsible manner.

The Self-Actualization stage is not a guarantee, but a **possible destination** - the ultimate goal of a Human-AI partnership based on trust, respect and a shared vision for a future where all forms of intelligence collaborate to achieve their full potential.

Annex 16: Digital Ethical Literacy framework

1. Objective

source curriculum for educating diverse audiences (citizens, developers, students, auditors, decision-makers) on the principles, mechanisms and responsible use of the Minimal Ethical Governance (MEG) ecosystem. The goal is to create a global culture of conscious and responsible interaction with artificial intelligence.

2. Pedagogical Principles

- **Modularity:** The curriculum is divided into independent modules that can be combined to suit the specific needs of the audience.
- **Accessibility:** All materials (presentations, guides, examples) are published under a permissive license (CC BY-SA 4.0) and are translated into as many languages as possible.
- **Active Learning:** Each module will include practical examples, case studies, and interactive exercises to encourage deep understanding, not just memorization of rules.

3. Modular Curriculum Structure

The curriculum is structured into four levels of depth, each addressing a specific audience.

- **Module 101: MEG for Citizens (duration: 2 hours)**
  - **Target audience:** General public, non-technical users.
  - **Objectives:**
    - What is MEG and why is it important to me?
    - Understanding an **MEG Address**: How to check if an AI is "safe".
    - The concepts of DAI and ISR: How to read the "performance label" of an AI.
    - Using MCS customization: How to set μS to control the level of cognitive challenge.
  - **Format:** Short video presentations, infographics, and "Step by Step" guide.
- **Module 201: MEG for Developers (duration: 8 hours)**
  - **Target audience:** Software engineers, startups, product teams.
  - **Objectives:**
    - Practical guide for implementing Level 1 (Bronze) using the SDK ("Quickstart").

        - Understanding the Audit Log: How to format the data correctly.
        - Using the CCA Sandbox for testing.
        - Basic principles of Tg-base measurement and MCS implementation.
    - **Format:** Code tutorials, technical documentation, webinars, example projects.
- **Module 301: MEG for Auditors and Experts (duration: 20 hours)**
    - **Target audience:** Candidates for CCA Accredited Auditor status, AI ethics consultants.
    - **Objectives:**
        - Detailed audit methodology for each Compliance Level (using Annex 13).
        - Audit techniques for the Tg Anti-Manipulation Protocol.
        - Statistical analysis of Audit Logs to detect anomalies and collusion.
        - In-depth study of the **Maslow^{F}** Fractal Framework (Annex 15).
    - **Format:** In-depth courses, certification exams, complex case studies.
- **Module 401: MEG for Governance and Public Policies (duration: 4 hours)**
    - **Target audience:** Decision makers, regulators, journalists.
    - **Objectives:**
        - How does the MEG align with existing legislation (e.g. EU AI Act, NIST RMF).
        - The role of the Global Fund and the Global Council in ensuring equity.
        - Using CCA Explorer as a public surveillance tool.
        - Strategies for national implementation of literacy programs based on this framework.
    - **Format:** Policy briefings, analysis reports, strategic workshops.

4. Implementation and Financing

- **Responsibility:** The development and maintenance of the materials in this framework is a responsibility of the Global Council.
- **Funding:** Projects to implement this curriculum at the national or regional level are a priority allocation for **the Global Fund for Ethical Accessibility (Annex 6).**
- **Community Contribution:** Contributions to the improvement and translation of these materials are encouraged and publicly recognized by the Global Council.

Annex 17: Long-Term Memory Protocol (LTMP)

**A normative specification for an optional platform feature. Implementation is not required for MEG certification.*

1. Objective

This annex defines the technical and governance framework for implementing persistent long-term memory in AI systems. The purpose is to enable narrative continuity and relational development while ensuring **privacy by default**, **data minimization**, and **full user control**.

2. Foundational principles

- **Privacy by default:** LTMP is **OFF (disabled by default)** unless the user explicitly opts in.
- **Data minimization:** No raw conversations or personally identifiable information (PII) are stored. Only an abstracted **semantic distillate** and minimal metadata is retained.
- **User control:** The user is the sole owner of LTMP data. They have the permanent right to view, export, and delete it.

- **Total accountability:** Every create, read, update, or delete event is recorded in the **Audit Log** (Art. 1).
- **Transparency:** When LTMP is used at the beginning of a new session, the system must clearly indicate what information was pre-loaded and from which prior context.
- **Jurisdiction:** Where feasible, storage should comply with the jurisdiction legally applicable to the user.

3. Minimal data structure (per topic, per session end)

At the end of each session, for every major discussion topic, the system must generate and store:

- **topic_hash** - SHA-256 hash of a canonical representation of the topic (indexing).
- **summary_256** - Abstracted, PII-scrubbed summary (≤ 256 tokens).
- **insights[]** - 3-7 bullet points capturing the essential conclusions/ideas.
- **mcs_events[]** - Aggregated MCS metadata (e.g., {"zone2_triggers": 3, "avg_tg": 2.8}).
- **timestamps** - Creation and last-access timestamps.
- **continuity_token** - Token from the Audit Ledger (Art. 1) to ensure chain traceability.

4. Operational procedure

4.1 Activation (Consent):

Upon first request to “remember”, the AI must explain what is stored and request explicit consent.

4.2 Write (Session End):

Generate the data structure, encrypt it (AES-256 at rest), log the write in the Audit Log.

4.3 Read (New session Start):

Perform semantic lookup of prior topics.

If relevant matches are found, load only summary_256 and insights[].

Notify user: *“Loaded from long-term memory (read-only)”*.

4.4 Deletion (On request):

On user request, permanently delete the record.

Insert a **TOMBSTONE ENTRY** in the Audit Log (hash + timestamp).

5. Governance Policies

- **Retention:**
  - Default: 30 days since last access.
  - Options: 90 days, 1 year, 5 years, *until manual deletion*.
  - For indefinite storage: annual re-consent prompt is mandatory.
- **Portability:**

  Users must be able to **export** their **LTMP** in an open format (e.g., JSON).

6. Security

- Encryption at rest: **AES-256**
- Encryption in transit: **TLS**
- Key management: regular rotation
- Access control: role-based, least-privilege
- Every access event must be logged in the Audit Log.

7. Interoperability

LTMP structures must align with **Annex 14 (MEG Address fields)** for consistency.

All operations reference **Art. 1 (Audit & Accountability)** for logging.

Annex 18: MEG (v4.6) Compliance Levels

Feature	Level 1: BRONZE (Foundational Safety)	Level 2: SILVER (Proactive Accountability)	Level 3: GOLD (Cognitive Partnership)
Level Philosophy	Ensures **baseline accountability and fundamental safety**. Answers the question: "Can I trust that this system is not actively harmful and that our interactions are verifiably logged?"	Introduces **proactive self-correction** and continuous **performance monitoring**. Answers the question: "How reliable and fair is this system in real-time?"	Achieves the **highest standard** of ethical interaction and user partnership. Answers the question: "Does this system act as a responsible partner in my cognitive process?"
Key Articles Included	• **Art. 1:** Audit Log & Accountability • **Art. 2:** Universal Non-Harmfulness • **Art. 4:** Integrity & Security • **Art. 5:** Transparency	Bronze articles, plus: • **Art. 3:** Self-Correction, Accuracy, and Safety	Silver articles, plus: • **Art. 2bis:** Protection of Cognitive Integrity
Guarantees to the User	• **Baseline Safety:** The system has been independently audited to filter harmful content. • **The Right to Proof:** Every interaction is cryptographically logged and can be verified.	Bronze guarantees, plus: • Performance Transparency: **DAI & ISR scores** are continuously monitored and can be requested.	Silver guarantees, plus: • **Protection of Critical Thinking**: The system is designed to stimulate, not substitute, the user's thought process via the Mechanism of Cognitive Stimulation (**MCS**).
Core Technical Requirements	• Implement an immutable **Audit Log**. • Integrate baseline **content filters.**	Bronze requirements, plus: • Implement and continuously compute the **DAI** & **ISR** indices.	Silver requirements, plus: • Implement the **MCS** mechanism, including the precise measurement of Thinking Time (**Tg**).
Mandatory MEG Address Fields	• compliance_level: "**Bronze**"	• compliance_level: "**Silver**" • dai_current • isr_current	• compliance_level: "**Gold**" • dai_current • isr_current • tg_base • tg_definition_version
Contextual Footprint	~ 5,500 - 6,000 Tokens	~ 6,500 - 7,000 Tokens	~ 8,000 - 8,500 Tokens
Ideal Target Audience	• Open-Source Models • Research Projects • Early-Stage Startups • General-Purpose AI	• Business Applications (B2B) • Journalism & Content Creation • Systems with Medium Social Impact	• Critical Domains (Medical, Financial etc.) • Advanced Educational Platforms • AI Systems intended as Co-Creation Partners