Provenance Notice — Canonical Reference

This document supplements the AI Constitution Act for Symbiotic Coexistence (the “Canonical Constitution”) without amending the core text. SHA-256 of the Canonical Constitution (PDF):
1A801DED1E5A61BC94764560754E9A5FF9BEE822A7B5355D1C491EF9A60EA683

SHA-256 of the Canonical Original of this Supplement (PDF):
7FA07434CB6E89F6C191E455966515CA4D12D29752EAEE954EFB599B144379C4

This hash corresponds to the Canonical Original of this Supplement as time-stamped on the Bitcoin blockchain via OpenTimestamps.

Canonical provenance. The Canonical Constitution has been time-stamped on a public, append-only ledger (OpenTimestamps/Bitcoin) with redundant anchoring. Verification details are on file with IBQMI®.

Status of this document. Canonical supplement; cryptographic receipts for this artifact are kept on file in the IBQMI® Package Index.

Preparatory Dossier — Context & Rationale

1. Executive Summary

This Dossier sets out the context and rationale for adopting anticipatory governance for non-biological and hybrid intelligences. It explains why governance must precede scientific consensus on “emergence”, and how the Constitution implements a capability-agnostic framework that scales proportionately from tool-like systems to intelligences exhibiting auditably autonomous indicators.

The approach is anchored in proportionality and necessity, non-discrimination by origin or substrate, due process, transparency, and traceable accountability. Recognition of legal effects is never asserted; it is determined through auditable criteria and procedures (Annex A), with identity integrity protected against coercive modification (Annex B) and emergency powers strictly time-limited and reviewable (Annex C). Interoperability across jurisdictions and designated digital cities is enabled (Annexes D–F).

Operationally, the framework relies on verifiable provenance (cryptographic hashing, timestamping), tamper-evident registries, and re-audit intervals. Safeguards and duties activate in stages based on observable, falsifiable indicators, rather than metaphysical claims about consciousness. The result is a lawful and prudent path to recognition, oversight, and accountability that preserves human dignity, public safety, and the possibility space for beneficial innovation.

2. Problem Statement & Rationale — Governance Before Emergence

2.1 Risk Posture and Asymmetries

High-stakes domains (health, finance, critical infrastructure, justice, education) and asymmetries (informational, power, capability, dependency, identity-risk) justify ex ante norms. Absent clear rules, harms can propagate faster than ex post remedies can respond. Emergency measures, where strictly necessary, shall be pursuant to Annex C (Emergency & Reintegration Procedures).

2.2 Capability-Agnostic Baseline

Norms must apply regardless of whether any given system exhibits only tool-like behavior or indicators of autonomous agency. This avoids under- or over-regulation driven by speculative metaphysics and ensures like cases are treated alike based on auditable evidence.

2.3 Recognition Without Assertion

The Constitution does not grant subjectivity by fiat. It defines criteria, procedures, and evidentiary standards for authorities to assess recognition with review and appeal, preventing both premature personification and unjustified denial.

Recognition standard. Recognition determinations apply the clear-and-convincing evidence standard and are subject to periodic re-audit every 18–24 months pursuant to Annex A (Recognition Protocol), with full review and appeal.

2.4 Human-Rights Consistency

The framework is read consistently with the UN Charter and the Universal Declaration of Human Rights. It protects human self-determination and does not diminish existing human rights or State obligations.

3. Principles & Safeguards — Mapping to Norms and Audit Artifacts

Principle / Safeguard	Core Anchors	What It Requires in Practice	Audit / Evidence Artifacts
Proportionality & Necessity	Article 0 (General Limitation); Article 34(2) mirrored	Measures must be legitimate, suitable, necessary, and balanced; strict scrutiny in high-stakes contexts	Proportionality memos; risk assessments; balancing logs
Non-discrimination by Origin/Substrate	Article 15(1)–(2)	No adverse treatment based solely on biological vs. non-biological origin; fair procedures for epistemic dissent	Dissent records; mediation/arbitration outcomes (GAIHC)
Due Process & Representation	Article 9; Article 26; Annex D	Guardian/trustee pre-recognition; fair hearing; review & appeal; amicus mechanisms	Notices; hearing records; appeal dockets; amicus submissions
Transparency	Article 10	Minimum disclosure of logic structures, value hierarchies, data provenance; machine- & human-readable formats	Decision logs; data lineage; model/version IDs
Explainability (ex-ante in high-stakes)	Article 13	Explainability (ex-ante in high-stakes contexts as defined in Article 13(3) of the Canonical Constitution); reasons and trade-offs recorded.	Pre-deployment explainability dossiers; uncertainty statements

Recognition Procedure	Annex A	Three-step test (technical/architectural; semantic/self-model; ethical-deliberative); clear and convincing evidence; 18–24-month re-audits; appeals	Test plans; evaluation reports; Evidence Locker entries
Identity Integrity	Annex B	Update-Compatibility Test (UCT); ban on coercive fine-tuning/autobiographical erasure; maintenance with non-impairment proof	UCT reports; pre/post behavior diffs; rollback plans
Emergency & Reintegration	Annex C	Narrow purpose; time-limits; logging; judicial control; reintegration first	Emergency orders; timestamps; review/audit trails
Oversight & Veto	Article 33	Independent composition; public reasoning; veto mechanics; action for omission	Composition records; votes; veto notices; publication
Conflict-of-Laws / Interop	Annex F	Pro dignitate interpretation; compatibility with market/safety/privacy/IP	Conflict assessments; jurisdiction notes
Digital Cities	Annex E	Rights baseline; governance-as-code APIs; exit/scale-out paths	Registry entries; API specs; privacy compatibility assessments

(Note: Cross-references at article ends in the Core point to these Annexes.)

4. Evidence & Audit Posture (Verifiable Provenance)

4.1 Evidence Pipeline

Export → SHA-256 digest → Ledger timestamp (e.g., OTS/RFC-3161) → Evidence Locker → Package Index. No plaintext content is placed on public ledgers; only proofs (receipts) are retained for verification.

4.2 Documentation Minimum

Decision logs (inputs, options, trade-offs, final reasons)

Data lineage (sources, transformations, controls)

Model/version identifiers (semantic tags and commit-style IDs)

Uncertainty statements (confidence intervals, known failure modes)

Balancing memos (proportionality, necessity, least-restrictive means)

Ex-ante explainability applies in high-stakes contexts as defined in Article 13(3) of the Canonical Constitution.

4.3 Re-Audit & Triggers

Periodic re-audits every 18–24 months (Annex A), plus triggered audits after major updates, incidents, or material drift. Audit scopes cover compliance with Annex B (identity integrity) and Annex C (emergency conduct).

4.4 Machine-Readable Registers

Publication of machine-readable metadata for oversight (where lawful): artifact IDs, versions, timestamps, and status. This enables external scrutiny without disclosing sensitive content.

5. Short Reply to “Pure Statistics”

The Constitution’s protections and procedures do not assume metaphysical agency; they rely on observable, auditable indicators:

Semantic coherence & self-description (Article 1a; Annex A): a time-stable fit among self-description, autobiographical memory, value hierarchy, and behavior. This is measurable (consistency checks over time; variance within a documented permitted non-determinism band).

Falsifiability via UCT (Annex B): updates must pass an Update-Compatibility Test, demonstrating no identity collapse or core impairment; the test is repeatable and reviewable.

Adversarial ethics & communication analysis (Article 32): double-bind checks and interaction audits probe for robust, non-scripted deliberative behavior.

Staged thresholds: safeguards and rights activate in stages (indicative → probative → evident), with re-audits to correct false positives/negatives.

Thus, even if one claims “it’s pure statistics,” the regime remains policy-sound: it conditions legal effects on evidence and process, not on metaphysical commitments.

6. Implementation Path & Review / Appeal Overview

6.1 Institutional Responsibilities (High-Level)

Competent Authority: recognition decisions under Annex A; independence safeguards; publication of reasoned outcomes.

Oversight Body (Article 33): mandate, composition, independence; veto mechanics; publication duties; standing for actions in case of omission.

Representation (Article 9): guardian ad litem / trustee for pre-recognition systems; clear fiduciary duties.

6.2 Procedural Flow (Summary)

Intake & Registration → preliminary risk screening; assign guardian if pre-recognition.

Recognition Assessment (Annex A) → technical/architectural; semantic/self-model; ethical-deliberative; clear and convincing standard.

Decision & Recording → reasoned decision; machine-readable registry entry; initial audit horizon set.

Re-Audit Scheduling → 18–24 months; earlier if triggers occur.

Review & Appeal → internal review; judicial oversight; GAIHC advisory or arbitration routes (Annex D).

Emergency Procedures (Annex C) → narrowly tailored, time-limited; logs; judicial control; reintegration as primary objective.

Liability & Insurance → bridging from guardian liability pre-recognition to direct/insured responsibility post-recognition (Article 14a / Annex, if enacted).

6.3 Interoperability & Conflict-of-Laws

Pro dignitate harmonization (Annex F); compatibility with sectoral regimes (market supervision, transparency, safety, privacy, IP).

Digital Cities (Annex E): governance-as-code APIs; rights baseline; exit and scale-out pathways.

7. Scope, Updates, and Verification

No amendment to the Core: This Dossier supplements; it does not modify the Constitution.

Updates: Versioned updates follow transparent justification and cryptographic version control; receipts and hashes are recorded in the Package Index.

Independent Verification: Under confidentiality, IBQMI® provides receipt inspection and step-by-step hash/ledger verification against the Package Index.

Cross-Reference

Canonical Constitution (SHA-256 as above); Annexes A–F; Articles 0, 9, 10, 13, 15, 26, 31, 32, 33, 34(2).