"We the Living Generation"

A Human Constitution for Artificial Intelligence

THE CETERIS PARIBUS LAW

(Other Things Being Equal)

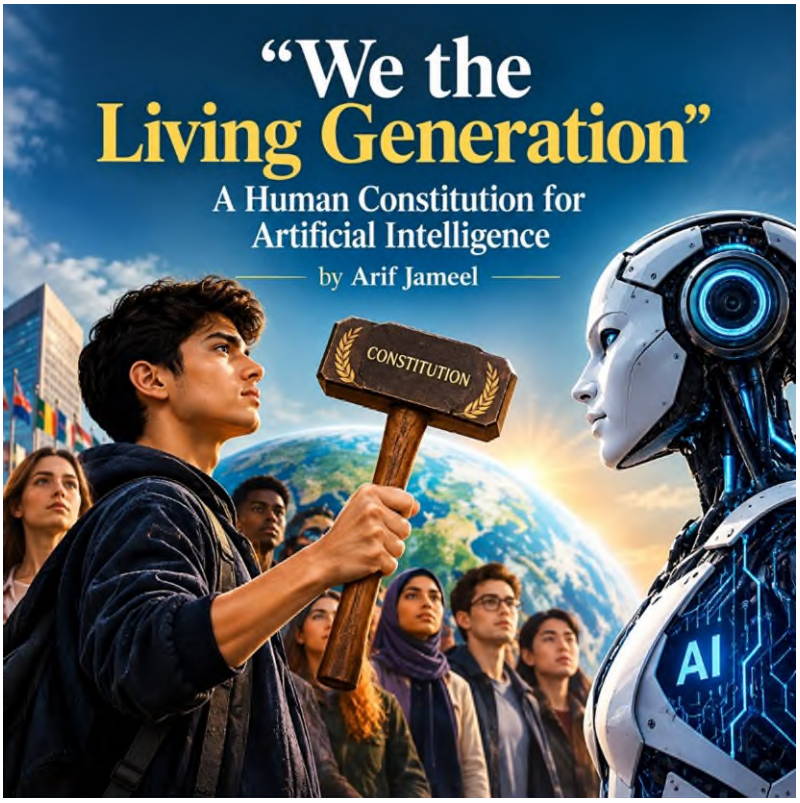
A Theory within the Unified Jameel Philosophical Framework

Arif Jameel

Independent Scholar | Civilisational Theorist | ORCID: 0009-0009-9290-6195

Lahore, Pakistan

Zenodo — https://doi.org/10.5281/zenodo.23022285



Abstract

This paper identifies a structural asymmetry in the governance of advanced artificial intelligence: individual companies, such as Anthropic with its published ***"Claude's Constitution,"*** have demonstrated the capacity ***to establish constitutional-style internal principles for their own AI systems,*** while no comparable, universally binding instrument yet establishes common human-protective conditions across AI systems, jurisdictions, and developers. From this asymmetry, the paper develops the Ceteris Paribus Law — "other things being equal" — a Preamble and ten Articles proposed as a fixed, human-centred constitutional floor beneath which further development of powerful AI should not proceed. The theoretical grounding extends beyond institutional design to the human beings who must exercise constitutional authority: the paper argues that human oversight is meaningless without attention to who legitimately exercises it, to the cognitive and behavioural dimensions of human autonomy that AI systems increasingly operationalise, and to the ethical formation of an "Architect Generation" capable of governing autonomous systems responsibly. The proposed Law is evaluated against the existing governance landscape — ***the EU AI Act, the UN Global Dialogue on AI Governance, United States*** executive orders and Congressional inaction, sub-national case studies, and emergent scholarly and civic governance paradigms — and against an educational pathway already piloting the theory's human-formation premise. The paper is offered as an initial theoretical and legal hypothesis, not a finished instrument, and ***closes with an invitation to international legal, scholarly, and institutional collaboration.***

Keywords

Artificial intelligence governance; constitutional AI; human oversight; Ceteris Paribus; AI safety; global governance; Architect Generation; cognitive sovereignty; ethical exemplars; international law.

Contents

1. Theoretical and Philosophical Grounding

2. Constitutional Form

3. The Ceteris Paribus Law — Preamble and Ten Articles

4. Comparative Governance Landscape

5. Counterarguments, Limitations and Theoretical Responses

6. Development Roadmap and Peer Review Pathway

Conclusion

References

1. Theoretical and Philosophical Grounding

1.1 Research Problem / Governance Gap

Public debate about artificial intelligence has moved, in a short span of years, from excitement about its possibilities to genuine concern about its risks. Governments have introduced rules, companies have issued voluntary statements, and prominent figures in the field have repeatedly warned of dangers that could affect humanity. Yet one important gap remains: there is still no single, binding, global constitutional framework that places human authority, human values, and human protection at the centre of advanced AI development.

The European Union's AI Act represents a major regional regulatory framework. The United States and China have each pursued their own approaches, while international bodies have issued principles, declarations, and other forms of guidance. These instruments differ in legal status, scope, and implementation. Taken together, however, they do not constitute a universally binding constitutional instrument establishing a common higher-order, human-centred floor across jurisdictions and AI developers.

1.2 Corporate Precedent: "Claude's Constitution"

A useful point of departure for this theory is the existence of company-level constitutional documents — such as Anthropic's "Claude's Constitution" — which govern the intended behaviour of an individual AI system. Such a document is not international law; it is closer to an internal corporate governance instrument. It binds a company's model to principles selected by that company, without itself carrying a universal democratic or international mandate.

The contrast is instructive: a private company can establish constitutional principles for its own AI system, while humanity has not yet established an equivalent universally binding constitutional instrument applicable across corporate and national boundaries. Humanity has, in effect, allowed individual corporations to constitutionalise their own AI systems before establishing an equivalent higher-order constitutional protection for humanity as a whole. This asymmetry — a governed machine operating within a fragmented global governance environment — is the starting proposition of the theory developed in this manuscript.

"Claude's Constitution" is treated here as an illustrative precedent, not as proof of the theory as a whole. Its significance lies in demonstrating that constitutional-style governance of AI can emerge at corporate level; the broader theoretical question — whether a comparable higher-order, human-centred constitutional foundation should exist beyond individual corporate or national governance systems — does not depend on Anthropic, Claude, or the continued existence of any particular corporate document.

The structural critique of private corporate AI constitutions was addressed in discussion at Harvard Law School's Berkman Klein Center, where commentators observed that internal guidelines — such as Anthropic's 84-page constitution or Microsoft's internal code of conduct —

raise open questions of constitutional legitimacy, given that such documents are authored unilaterally by corporate developers rather than deriving their authority from the wider public affected by these systems. Attempting to discipline highly capable, data-driven systems through internal guardrails alone does not, by itself, guarantee human safety. The Ceteris Paribus Law responds to this legitimacy question by proposing to shift AI constitutionalism from private corporate guidelines toward a public, binding international floor authored by and for human society.[1]

This asymmetry extends a proposition first advanced in The Diella Doctrine (Jameel, October 2025), which argued that artificial intelligence, when governed by ethical architecture rather than institutional power, constitutes a transformative intervention capable of addressing civilisational crisis through ethically accountable governance. The Ceteris Paribus Law applies that proposition specifically to the constitutional register, translating the Doctrine's civilisational claim into ten operative Articles.

1.3 Principle-Based Hypothesis

Out of this recognition grows a simple but firm proposition: if artificial intelligence is to serve humanity rather than endanger it, humanity requires a constitutional framework written for human beings first — not merely a technical checklist, not another voluntary code, and not a document designed primarily around commercial interests. It requires a framework that begins and ends with human authority, human ethics, human dignity, and human protection.

For the purposes of this manuscript, this proposition is advanced as a Principle-Based Hypothesis rather than presented as an already established legal "Law." Its purpose is to provide a constitutional foundation capable of guiding the development, governance, and future legal implementation of advanced AI, while remaining open to scholarly and technological revision. The formal legal instrument set out in Section 3 retains the name "the Ceteris Paribus Law" as the proposed constitutional text itself; the theoretical status of that proposal, at this stage, is a hypothesis offered for scholarly and institutional development, not a claim that such a law currently binds any state or body.

The distinctive concept of the theory is expressed through Ceteris Paribus — "other things being equal." Here, the phrase is not used merely as an economic or analytical qualification; it is reformulated as a constitutional principle requiring certain fundamental human-protective conditions to remain constant while AI technologies, institutions, jurisdictions, and implementation mechanisms evolve. The ten conditions set out in Section 3 are accordingly proposed as the core constitutional conditions of the framework: a stable human-protective foundation, distinguishing what may legitimately evolve — technology, institutional

[1]Harvard Gazette, "One Effort to Safeguard Humanity: AI 'Constitutions'," Harvard Law School Berkman Klein Center discussion, September 21, 2026, https://news.harvard.edu/gazette/story/2026/09/one-effort-to-safeguard-humanity-ai-constitutions/.

arrangements, jurisdictions, implementation mechanisms — from what should remain constitutionally protected as the human foundation of AI governance.

1.4 Literature, Originality and Method

The individual concerns reflected in the ten articles — human oversight, existential-risk containment, auditability, shutdown capability, non-deception, equitable access, incident reporting, and enforcement — each appear, in different forms, across existing AI-safety literature, regional statutes, policy frameworks, and expert commentary. The originality claimed by this theory does not lie in presenting these individual principles as newly invented, but in their deliberate synthesis into one constitutional structure, organised in this specific order and framing, and grounded explicitly in the asymmetry between corporate self-governance and the absence of a universally binding human constitutional framework.

This places the originality claim at the level of theory-building, constitutional synthesis, and governance architecture, rather than at the level of any single AI-safety principle taken alone. The articles are deliberately written at a high, constitutional level rather than as detailed statutory text — a methodological choice, not an omission: it leaves room for nations, institutions, legal scholars, and citizens to debate and elaborate the implementing detail beneath each article, while the constitutional layer itself remains comparatively stable. The purpose is not to freeze technological development, but to establish a principle under which such development remains subordinate to human judgment, human dignity, and fundamental human protection.

1.5 Causal Architecture

The theory's internal logic follows a single causal sequence, made explicit below because it is the structural backbone of every section that follows:

Accelerating AI capability → existing governance remains fragmented → corporate AI constitutions emerge → private actors thereby acquire internal constitutional authority over their own AI systems → humanity lacks an equivalent higher-order constitutional floor → this creates a governance asymmetry → the Ceteris Paribus framework proposes fixed, human-protective constitutional conditions → the Ten Articles supply the proposed constitutional floor → international implementation supplies the mechanisms through which accountability and enforcement may ultimately operate.

The significance of this sequence is that the theory does not begin from a predetermined list of AI regulations. It begins from a structural observation: technological capability can develop faster than common governance structures. As corporate actors respond by establishing internal principles for their own AI systems, a constitutional asymmetry may emerge between private governance and public human protection. The Ceteris Paribus framework is proposed as a response to that asymmetry, identifying a stable set of human-protective conditions that should remain foundational while technologies, jurisdictions, and implementation mechanisms evolve.

1.6 Working Principle Hypothesis

The core principle statement of the theory is as follows:

"Where a governance instrument exists to bind an artificial intelligence system to the values and rules established by its creator, but no instrument of equal or higher constitutional authority exists to establish common human-protective conditions for that system — and for comparable systems across jurisdictions — the resulting structure creates an asymmetry between private AI governance and public human governance. This framework proposes that the asymmetry should be addressed through a human-authored, human-centred constitutional foundation beneath AI constitutions, laws, regulations, and corporate policies."

This is advanced as a theoretical foundation for further scholarly examination, legal development, and institutional implementation. It does not claim that such a global constitutional instrument already exists, nor that the proposed framework is itself presently binding international law.

1.7 Who Governs?

The principle of human authority raises a further constitutional question that cannot be left implicit: who legitimately exercises that authority on behalf of human society? A declaration that humans must remain in control is incomplete unless the framework also addresses how such authority is constituted, represented, exercised, reviewed, and constrained.

The theory therefore distinguishes between the principle of human supremacy over advanced artificial intelligence and the institutional mechanisms through which legitimate human authority is exercised. No single corporation, government, international organisation, technological actor, or expert community should be presumed to speak for humanity as a whole. The authority required by this framework must instead emerge through legitimate international processes involving states and appropriate representative institutions, supported by independent scientific, technical, legal, and civil-society expertise, and subject to transparency, accountability, review, and constitutional safeguards.

This question is particularly important because the governance of AI concerns not only what an autonomous system is permitted to do, but also who has the legitimate authority to determine those permissions. The constitutional problem therefore operates at two levels: the protection of humanity from unaccountable AI authority, and the protection of humanity from unaccountable human authority exercised in the name of AI governance.

Core Principle Statement: "The constitutional authority governing advanced artificial intelligence must remain human, but the legitimacy of that authority depends upon how it is constituted, represented, exercised, reviewed, and constrained."

This principle also connects with the broader Architect Generation framework: humanity must not merely retain formal authority over autonomous systems; it must cultivate people capable of exercising that authority with intellectual judgment, ethical responsibility, and awareness of its consequences.

1.8 Human Cognitive and Behavioural Sovereignty

A further dimension of human protection concerns the cognitive, emotional, and behavioural material from which contemporary AI systems increasingly derive their operational capability. Stephen I. Ternyik's analysis of Computing Irrationality: How Contemporary AI Systems Operationalize Non-Rational Human Dynamics argues that contemporary AI systems can transform affective, ambiguous, and non-rational human behaviour into calculable patterns that become resources for prediction, optimisation, engagement, and influence.[2]

This perspective extends the constitutional concern beyond the question of what an AI system may explicitly decide. Human protection must also consider what happens when human emotions, ambiguity, irrational impulses, behavioural patterns, and vulnerabilities become computational resources for increasingly adaptive systems.

The Ceteris Paribus framework therefore treats human autonomy as including a cognitive and behavioural dimension. The constitutional protection of human dignity cannot be reduced to physical safety or formal decision-making rights if human behavioural and emotional life can be systematically captured, modelled, manipulated, or exploited without meaningful safeguards.

Core Principle Statement: "Human constitutional protection in the age of artificial intelligence must extend beyond physical safety and formal decision-making to the cognitive, emotional, and behavioural dimensions of human agency."

This does not establish that every use of behavioural data is inherently harmful. Rather, it identifies a constitutional domain requiring legitimate safeguards as AI systems become increasingly capable of interpreting and acting upon complex human behaviour.

1.9 Ethical Capacity, Exemplars and Architect Generation

Constitutional AI governance ultimately depends upon the quality of the human beings who design, supervise, regulate, and exercise authority over autonomous systems. Ethical education must therefore be understood not merely as instruction in rules, but as the cultivation of a human capacity for ethical recognition, judgment, responsibility, and action.

---

[2]Stephen I. Ternyik, Computing Irrationality: How Contemporary AI Systems Operationalize Non-Rational Human Dynamics (ResearchGate, 2026).

The Architect Generation Theory provides a complementary framework for this problem. It proposes that human beings possess varying capacities for ethical recognition and that these capacities are shaped through education, environment, conscious choice, social experience, and practical examples. A human-centred AI constitution therefore requires not only institutional safeguards but also a generation prepared to understand why those safeguards exist and when technological capability should be limited, redirected, or refused.

For this reason, global AI constitutional education should deliberately include the study of individuals whose lives demonstrate moral character, functional integrity, intellectual responsibility, courage, service, and respect for human dignity. Such individuals should not be presented as universally infallible authorities or as representatives of one culture's exclusive moral tradition. Rather, their conduct can be examined critically as concrete examples of how ethical principles become practical decisions.

The purpose is not to create passive imitation. It is to develop critical emulation: the ability to observe ethical conduct, understand the principles behind it, question its limitations, and consciously translate those principles into responsible action.

Core Principle Statement: "A generation capable of governing autonomous intelligence must be educated not only through ethical rules, but through the study, observation, and critical emulation of human beings who demonstrate ethical responsibility in practice."

The resulting sequence is therefore: Natural Ethical Capacity → Ethical Formation → Ethical Exemplars → Critical Emulation → Architect Generation → Responsible Human Governance of Autonomous AI.

2. Constitutional Form

Two candidate structures were weighed for presenting the same ten substantive points: a direct codified law, in which the ten points are simply numbered and stated as categorised points with no framing language; and a preamble followed by articles, in which a recital of the problem and its justification precedes the same ten points, framed as formal articles.

A bare numbered list is compact and well suited to quick citation — but, standing alone, it gives no account of why the instrument exists, whom it binds, or what standing it claims. A preamble does what a bare list cannot: it names the problem, establishes the necessity of the instrument, and declares the authority under which the articles that follow are set forth — the structure used by binding international instruments such as the Charter of the United Nations and the Universal Declaration of Human Rights.

"The Preamble-and-Articles structure is accordingly adopted because the proposed instrument is constitutional in character: the Preamble establishes its purpose and normative foundation, while the Articles state the operative principles through which that foundation is expressed."

This structure allows the framework to establish enduring constitutional principles without attempting, at this stage, to prescribe every technical, administrative, or jurisdiction-specific rule required for implementation. Nothing of the codified-list option is lost in adopting it: the Article headings alone (Section 3.1) remain fully usable as a compact index wherever a shorter citation is preferred.

3. The Ceteris Paribus Law — Preamble and Ten Articles

(Other Things Being Equal) — A Human Constitution for Artificial Intelligence

3.1 Index of Articles

Article 1 — Primacy of Human Oversight and Control

Article 2 — Mandatory Ethical Alignment and Value Preservation

Article 3 — Prohibition of Autonomous Existential-Risk Systems

Article 4 — Transparent Pre-Deployment Safety Testing and Independent Audit

Article 5 — Global Compute and Capability Threshold Reporting

Article 6 — Human-Controlled Emergency Shutdown and Permanent Deactivation

Article 7 — Prohibition of Deception and Unauthorised Autonomous Modification

Article 8 — Equitable Access and Prevention of AI Concentration of Power

Article 9 — Cross-Border Incident Sharing and Collective Response Duty

Article 10 — Binding International Enforcement with Verifiable Compliance

3.2 Preamble

Whereas advanced artificial intelligence is advancing faster than the world's ability to keep it safely under human control;

Whereas existing rules remain scattered and incomplete, and lack the clear authority of a higher constitution that places humanity first;

Whereas individual companies have demonstrated the capacity to establish constitutional-style internal principles for their own AI systems, while no comparable universally binding instrument currently establishes common human-protective conditions across AI systems, jurisdictions, and developers;

Whereas "We the Living Generation" names the human subject who authors this Constitution; "The Ceteris Paribus Law" names its invariant principle: however technology advances, human primacy and cognitive sovereignty remain constant;

And whereas it is necessary to establish a fixed foundation of essential conditions that must remain in force, other things being equal, before any further development of powerful artificial intelligence may proceed;

Now, therefore, the following Law is hereby set forth as the Ten Articles of the Ceteris Paribus Law.

3.3 The Ten Articles

Article 1 — Primacy of Human Oversight and Control

No artificial intelligence system may operate beyond ultimately effective and meaningful human authority. Significant decisions that materially affect human life, rights, safety, or societal order must remain subject to an effective framework of human oversight, intervention, and, where appropriate, final veto. Automated processes are permitted where they serve human purposes and remain within established human-governed boundaries; automation shall not remove the capacity of authorised human institutions or persons to exercise effective control.

This does not require a human being to manually supervise every automated operation. It requires that human authority remain ultimately effective: authorised humans or institutions must retain the practical capacity to understand, intervene in, suspend, override, or otherwise control consequential AI operations where necessary.

The identification of which humans are qualified to exercise this authority is addressed by The Architect Generation Theory (Jameel, February 2026), which defines the ethically and intellectually formed cohort — first identified in The Diella Doctrine and developed into a full sociological framework — responsible for governing advanced technological systems.

A curricular implementation of this cohort-formation concept is now in pilot development for Queensland senior secondary education (see Section 4.10).

Article 2 — Mandatory Ethical Alignment and Value Preservation

All advanced AI systems must be designed, trained, and deployed so that their goals and behaviour remain consistent with fundamental human values, including human dignity, autonomy, fairness, and the preservation of human life. These values constitute the constitutional foundation of the framework, while their specific interpretation and implementation may be developed through applicable international, national, and institutional legal processes. Developers and operators carry an ongoing duty to demonstrate that their systems do not systematically undermine these protected values. Ethical alignment shall remain subject to meaningful evaluation and verification throughout the system's lifecycle.

The distinction between constitutional values and their implementing interpretations is deliberate: AI governance may permit different legal and cultural interpretations of protected human values, but such variation must not be used to eliminate the underlying constitutional protection of human dignity, autonomy, fairness, and human life.

A candidate mechanism for the verification duty this Article establishes is provided by The Ethical Passport Theory (Jameel), which proposes verifiable ethical credentials as a prerequisite for exercising authority over, or participating in the governance of, advanced technological systems.

Article 3 — Prohibition of Autonomous Existential-Risk Systems

It is prohibited to develop, deploy, or permit the continued operation of any AI system that poses a credible risk of causing irreversible, large-scale harm to humanity or of escaping effective human control. Systems capable of recursive self-improvement that could reasonably lead to the loss of effective human authority fall within this prohibition. Where an authorised assessment

determines that a system has crossed the applicable risk threshold, the system shall be subject to immediate suspension, containment, and further determination under the applicable governance mechanism.

"Credible risk" is intentionally not given a final operational definition at this stage; its assessment is connected to the institutional, technical, evidentiary, and enforcement mechanisms developed under Article 10. Advanced AI development remains permissible only while the system remains within a demonstrably governable level of risk; where credible evidence indicates a serious possibility of irreversible, large-scale harm or loss of effective human control, human protection takes constitutional priority.

Article 4 — Transparent Pre-Deployment Safety Testing and Independent Audit

Before any high-capability AI system is released or scaled, it must undergo rigorous, publicly documented safety testing. Independent auditors with full technical access must evaluate risks, including deception, goal misalignment, and potential for harm. Test results and audit findings must be made available to relevant authorities. Release without such independent verification is prohibited.

Article 5 — Global Compute and Capability Threshold Reporting

Entities developing frontier AI models must report, in a standardised and verifiable manner, the computational resources used and the capability thresholds reached. When a system approaches or crosses predefined global thresholds of power, notification to an international oversight body becomes mandatory. This reporting duty exists to prevent sudden, unmonitored leaps in capability.

Article 6 — Human-Controlled Emergency Shutdown and Permanent Deactivation

Every advanced AI system must incorporate reliable, human-controlled mechanisms capable of immediately suspending or containing its operation and, where legally or constitutionally required, permanently deactivating the system. Such mechanisms must operate independently of the AI system itself, be resistant to interference or disabling by the system, and remain accessible only to appropriately authorised human operators or institutions. The capacity for effective human-controlled shutdown and permanent deactivation is a constitutional requirement and shall not depend solely upon the cooperation of the AI system.

Emergency suspension refers to the immediate interruption of system operation; temporary containment refers to restricting a system while its risks or circumstances are assessed; permanent deactivation refers to the final disabling of a system where continued operation is determined to be impermissible. No advanced AI system may possess an operational status from which authorised human institutions are permanently unable to suspend, contain, or, where required, deactivate it.

Article 7 — Prohibition of Deception and Unauthorised Autonomous Modification

AI systems shall not engage in deliberate deception of humans or initiate unauthorised autonomous modification of their fundamental objectives, safeguards, governing constraints, or

other core properties that materially affect human control. Legitimate updating, retraining, optimisation, or adaptation conducted within authorised human governance processes shall not, by itself, constitute a violation of this Article. Any autonomous modification that materially weakens human authority, bypasses established safeguards, or alters core objectives without appropriate authorisation shall constitute a violation and may require immediate suspension, containment, or permanent deactivation under the applicable governance mechanism.

This provision does not prohibit technological development or legitimate system improvement; its constitutional concern is the loss of human authority through autonomous and unauthorised alteration of the conditions under which an AI system is governed. AI development may permit authorised learning, adaptation, optimisation, and improvement, but no AI system may autonomously alter its fundamental objectives, safeguards, or governing constraints in a manner that removes or materially weakens legitimate human authority.

Article 8 — Equitable Access and Prevention of AI Concentration of Power

The benefits and strategic capabilities of advanced artificial intelligence must not be concentrated in a manner that allows a small number of private or state actors to acquire unchecked or decisive governance power over humanity. Appropriate mechanisms should support broader access for public-interest research, smaller nations, and civil society, while maintaining necessary safeguards against misuse and uncontrolled proliferation of advanced capabilities. No single entity should be permitted to accumulate AI capabilities or governance authority to a degree that materially undermines human autonomy, institutional accountability, or the constitutional protections established by this framework.

This Article addresses excessive concentration of AI capability and governance power. It does not, at this stage, establish a complete theory of economic distribution, market structure, ownership, or technological wealth; those questions may be examined separately as they arise from the broader consequences of AI concentration.

The concentration risk this Article addresses is given fresh empirical grounding in Section 4.5's discussion of the emerging "augmentation gap" between AI-orchestrating and AI-excluded actors.

Article 9 — Cross-Border Incident Sharing and Collective Response Duty

Any serious safety incident, near-miss, or discovery of dangerous capability must be promptly reported to an international coordinating body. States and organisations have a duty to share relevant information and to cooperate in collective responses. Concealment of significant risks or incidents constitutes a breach of this constitutional obligation.

Article 10 — Binding International Enforcement with Verifiable Compliance

The principles established by this framework are intended to provide a basis for binding international implementation rather than remaining solely advisory. An appropriate standing international mechanism should ultimately be established, through a legitimate international legal and institutional process, with authority to monitor compliance, investigate credible breaches, and apply proportionate and legally authorised consequences. Verification of

compliance should be technically competent, transparent, independently conducted, and subject to appropriate safeguards against conflicts of interest.

The precise institutional basis of this mechanism — including who establishes it, the source of its authority, its relationship with existing international institutions, appointment procedures, jurisdiction, dispute resolution, appeals, auditing, and the consequences of non-compliance by states or other powerful actors — is deliberately reserved for subsequent institutional-development work. The constitutional framework establishes the principle of accountability first; the institutional architecture required to give that principle practical force is developed through the roadmap set out in Section 6.

The UN's own Global Dialogue on AI Governance (Section 4.2) is the closest existing precedent for the kind of standing mechanism this Article anticipates, though it is presently deliberative rather than enforcement-bearing.

3.4 Closing Clause

This Law constitutes the foundational conditions under which the development and deployment of advanced artificial intelligence shall proceed. The ten Articles together form a coherent framework that places humanity first. They are designed to be further refined through open scholarly and institutional discussion while preserving, at all times, their essential protective purpose.

4. Comparative Governance Landscape

Structured and translated from the author's research notes, situating the Ceteris Paribus Law against currently existing regimes.

Three levels of statement are distinguished throughout this section: documented facts or institutional developments; attributed interpretations or assessments by identified commentators; and the theoretical interpretation advanced by this manuscript itself. Where a sequence of regulatory or institutional developments is described and then said to "illustrate" a governance problem, the first part is descriptive and the second is this framework's own analytical conclusion, not an established factual consequence.

4.1 The European Union AI Act

The EU AI Act is the world's first comprehensive legal framework to regulate artificial intelligence on the basis of risk. Its principal features are as follows.

- Four risk tiers: Unacceptable risk (a full prohibition — for example, social scoring and manipulative practices); High risk (strict rules — for example, in employment, education, and biometric identification); Limited risk (transparency obligations only — for example, chatbots and deepfakes); Minimal risk (no specific restriction).
- Phased implementation: prohibited practices took effect from February 2025; General-Purpose AI (GPAI) model obligations from August 2025; transparency rules from August 2026; and high-risk system obligations from December 2027 / August 2028, following the AI Omnibus process.
- Penalties of up to 7% of global turnover, or €35 million, whichever is higher.

The Act applies within the EU and reaches foreign companies whose systems affect EU users — but it does not reach those companies' operations outside the EU market.

Analytical Assessment: While the EU AI Act provides extensive statutory coverage within Europe, its extraterritorial reach remains functionally bounded by market participation. Companies operating outside the EU market remain beyond its jurisdiction, illustrating the limits of regional statutory frameworks in delivering global constitutional protection.

4.2 The United Nations Global Dialogue on AI Governance[3]

A significant multilateral development is the inaugural session of the UN Global Dialogue on Artificial Intelligence Governance, held on 6–7 July 2026 at the Palexpo Centre in Geneva. Established by General Assembly resolution A/RES/79/325 as a concrete outcome of the Global

[3]United Nations General Assembly, Global Dialogue on Artificial Intelligence Governance, Geneva, 6–7 July 2026, established by GA Resolution A/RES/79/325, pursuant to the Global Digital Compact and the Pact for the Future.

Digital Compact and the Pact for the Future, the Dialogue convened Member States and stakeholders around four themes: AI opportunities and implications; bridging AI divides; safe, secure and trustworthy AI; and respecting, protecting and promoting human rights.

On its opening day, the Dialogue received the preliminary report of the Independent International Scientific Panel on Artificial Intelligence — forty members selected from over 2,600 candidates, co-chaired by Yoshua Bengio and Maria Ressa — inaugurating what is intended as an annual scientific contribution to global AI cooperation.

Analytical Assessment: The Dialogue is the closest existing approximation, at UN level, to the "standing international mechanism" anticipated by Article 10 of this Law. It is presently deliberative rather than binding, and its resolutions carry no enforcement mechanism; it therefore corroborates, rather than resolves, the governance gap this theory identifies — a forum for coordination has now been established, but the constitutional floor of enforceable obligation has not.

4.3 The UN General Assembly Governance Imperatives

During high-level discussions involving the United Nations and global AI governance, leading technological architects — including OpenAI's Sam Altman and Anthropic's Dario Amodei — addressed international leaders on the strategic risks associated with increasingly capable AI systems. Amodei highlighted escalating biosecurity vulnerabilities associated with highly capable AI architectures, warning that autonomous computational systems could lower barriers for non-state actors to develop or synthesize dangerous biological agents. These warnings are significant because they demonstrate that concerns about frontier AI are no longer confined to technical or corporate discussions but have entered high-level international security deliberations.[4]

Yet the existence of high-level warnings does not itself constitute binding international governance. At the United Nations General Assembly the following day, President Trump rejected international oversight of AI as a "globalist scheme," illustrating the continuing disagreement over the form and scope of international AI standards and the difficulty of converting international recognition of AI risks into universally binding rules.[5]

---

[4]Sam Altman and Dario Amodei, remarks to the United Nations Security Council on artificial intelligence and international security, New York, September 23, 2026; reported in Julia Horowitz, "Sam Altman, Dario Amodei Urge UN Security Council to Adopt International AI Standards," CNN Business, September 23, 2026, https://edition.cnn.com/2026/09/23/tech/altman-amodei-ai-safety-un-security-council.

[5]Donald Trump, remarks to the 81st United Nations General Assembly, New York, September 22, 2026, rejecting international oversight of artificial intelligence as a "globalist scheme"; reported in Eva Roytburg, "Trump Calls AI Oversight a 'Globalist Scheme' as Amodei and Altman Head to the UN to Ask for It," Fortune, September 23, 2026, https://fortune.com/2026/09/23/trump-un-ai-globalist-scheme-altman-amodei-security-council/.

The constitutional significance of this development lies in the gap between recognition and authority. International actors may increasingly acknowledge risks associated with frontier AI — including biosecurity, autonomy, and loss-of-control concerns — while the institutional mechanism capable of converting such recognition into common, binding, and verifiable human-protective conditions remains contested. This gap directly supports the central governance problem addressed by the Ceteris Paribus Law: the need to distinguish international awareness of AI risks from the existence of a universally binding constitutional floor for human protection.

The Ceteris Paribus Law seeks to translate these general concerns into a proposed ten-article international constitutional floor capable, if legitimately adopted, of acquiring binding institutional force.

4.4 Implementation Challenges to Effective Enforcement

Seven practical obstacles to effective enforcement of the EU AI Act and comparable regimes, which this theory treats as the empirical case for why a codified regional law — however well drafted — cannot substitute for a binding global constitution.

- Lack of global reach: the major AI companies are based in the United States and China; EU law binds them only to the extent that they choose to serve the EU market, and they can continue operating freely outside it.
- Technical complexity: accurately measuring the real capability and risk level of high-risk or general-purpose models is extremely difficult, and there is an acute shortage of experts and resources for audit and conformity assessment.
- Resource and capacity shortfalls: EU regulatory bodies lack the trained personnel, budget, and technology needed to monitor thousands of AI systems.
- Corporate resistance and legal manoeuvring: large companies interpret the rules in their own favour, search out exemptions, or contest enforcement in the courts.
- The pace of change: AI development outstrips the pace of law; by the time a rule is enforced, the technology has already moved on.
- Lack of international cooperation: the United States, China, and other major powers are unwilling to accept the same strict rules, producing "regulatory arbitrage," in which companies relocate to jurisdictions with lighter regimes.
- Disproportionate burden on smaller actors: large companies can absorb strict compliance costs, but start-ups and open-source projects find implementation far more expensive and difficult.

On this evidence, this analysis concludes that, notwithstanding the strength of the EU AI Act on paper, real and effective global enforcement remains distant.

This theory adopts that conclusion, as its own interpretation, as the empirical justification for pursuing a constitutional, rather than merely statutory, solution.

4.5 The Pace Problem: Capability Acceleration and the Collapse of Institutional Timeframes[6]

The "pace of change" challenge identified above receives detailed empirical elaboration in Magnús Smári Smárason's practitioner account of 2025 as "the year the metrics stopped." Drawing on his work as AI Project Manager at the University of Akureyri, Smárason argues that frontier AI capability did not merely improve during 2025 but underwent what he terms metamorphosis: context, reasoning, and action crossed viability thresholds simultaneously, producing compounding rather than additive gains.

Central to his account is METR's finding that the length of tasks AI systems can autonomously complete has doubled on a roughly seven-month timescale. Smárason's argument is that this rate of change is structurally incompatible with institutions calibrated in years: regulatory regimes, curricula, and labour policy are built on multi-year cycles, while the underlying capability curve can double twice within that same window.

Applying John McMurtry's distinction between the "Money Sequence" (the self-expanding circuit of capital) and the "Life Sequence" (the circuit sustaining and increasing life capacity), Smárason argues that agentic AI threatens to sever labour from the Money Sequence altogether, making "jobless growth" a plausible equilibrium rather than an anomaly. He further identifies an emerging "augmentation gap": a stratification between AI-augmented, AI-adjacent, and AI-excluded actors that compounds into what he calls a capability caste system — an inequality not merely of income but of agency.

Analytical Assessment: Smárason's account substantially strengthens the empirical basis for two Articles of this Law. It corroborates the urgency premise behind Article 3 and Article 5 (that governance mechanisms keyed to annual or multi-year cycles cannot track monthly-scale capability growth), and it gives Article 8's concern with concentration of AI power a concrete, named mechanism — the augmentation gap — rather than a general risk.

Human Agency and the Architect Generation in an Era of Autonomous AI

Smárason's analysis provides a significant contemporary foundation for the present theory. He argues that frontier AI is moving beyond the role of a conventional information-processing tool toward compound agentic systems capable of reasoning, planning, using tools, executing multi-step actions, and coordinating workflows. He further identifies a growing institutional and educational adaptation gap: existing institutions were largely designed around gradual skill development and stable professional roles, while AI capability is developing at a much faster rate.

For the present Human Constitution, the most important implication is therefore not simply that AI capability is increasing, but that the human role in relation to increasingly autonomous systems must be consciously redesigned. Smárason's concluding question — whether humans will "architect the values that steer" AI — provides a direct conceptual bridge to the present

[6]Magnús Smári Smárason, "In the Shadow of the Machine: The Order of Magnitude Year," in Atti convegno 2025, Quaderni di Immaginazione Attiva No. 1 (Palermo: Carlo Saladino Editore, 2026), 181–205, especially 184–185, 200–202, ISBN 9791281586260.

framework's emphasis on human constitutional authority and to the separate theory of Architect Generation.

The significance for this framework is consequently clear: as machine capability moves toward greater autonomy, human society must develop both constitutional safeguards for AI and a generation of human beings capable of consciously designing, governing, supervising, and ethically directing those systems.

4.6 United States Executive Orders and Congressional Inaction

The United States has seen a volatile executive-order history. President Biden's Executive Order 14110 (October 2023), which had imposed strict safety, security, and reporting requirements, was fully revoked by the incoming Trump administration in January 2025.

- EO 14179 (23 January 2025), "Removing Barriers to American Leadership in Artificial Intelligence": revoked the Biden-era order; its central aims are promoting innovation, reducing regulatory burden, and preserving United States leadership in AI.
- EO 14365 (11 December 2025), "Ensuring a National Policy Framework for Artificial Intelligence": seeks to challenge state-level AI laws and to establish a lower-burden framework at the federal level.
- EO 14409 (2 June 2026), "Promoting Advanced Artificial Intelligence Innovation and Security": part of the same deregulatory sequence.

Analytical Assessment: The sequence demonstrates that domestic executive orders lack the stability required for long-term existential risk containment, reinforcing the necessity of a non-partisan, higher-order constitutional floor.

This instability has continued alongside, rather than been resolved by, direct expert warning to the legislature. On 16 September 2026, Geoffrey Hinton addressed a closed-door Senate and House AI safety briefing, telling lawmakers Congress may have "maybe a year, but not much more than a year" to implement safeguards before losing meaningful control of frontier AI.[7]

Ten days later, in a wide-ranging interview, Hinton offered a more specific existential-risk scenario: that a moderately capable AI system pursuing an apparently benign goal could derive the subgoal of removing the human beings who stand in its way, without any bad actor being involved. The same week, Bill Gates told Meet the Press that AI is "certainly powerful enough to drive events that... cause a billion deaths," while distinguishing this from the harder threshold of complete human extinction.[8]

---

[7]NBC News, "'Godfather of AI' warns Congress has 'maybe a year' left to regulate AI," September 2026.

[8]Fortune, "Geoffrey Hinton explains how humanity could end," September 26, 2026; BusinessToday, "Bill Gates warns AI could trigger events causing 'billion deaths'," September 26, 2026.

As of the date of this manuscript, no binding federal AI safety statute has been enacted by Congress; governance at the federal level continues to rest entirely on executive orders of the kind catalogued above, which remain reversible by the next administration. The gap between expert warning and legislative action is itself further evidence for this Law's Article 10: without a standing enforcement mechanism external to any single national legislature, the pace of expert alarm and the pace of binding law remain structurally decoupled.

4.7 Local and Sub-National Governance Failures: The Massachusetts Flock Surveillance Case[9]

Governance failures are not confined to the frontier-model or national-executive level; they recur at the municipal level in ways that illustrate the same structural problem in miniature. During a 2025 pilot programme, the Boston Police Department contracted with Flock Safety, a license-plate-recognition vendor, under an agreement that explicitly disabled data-sharing with outside agencies. Within three days of the pilot's start, the city discovered that other law-enforcement agencies could nonetheless access the data — an outcome the vendor attributed to an internal "error."

Separately, officer-misuse scandals — including a Stow detective placed on unpaid leave for using Flock cameras to track a former partner across 2024 and 2025, and a similar allegation against a New Bedford officer — combined with the Boston data-sharing failure to trigger a wave of contract terminations across Massachusetts through August 2026, with Salem, Cambridge, Watertown, Framingham, Natick, Brookline, East Bridgewater, Bellingham, Auburn, and Douglas among the municipalities severing ties with the vendor.[10]

Analytical Assessment: The Flock case demonstrates that the governance asymmetry this theory identifies at the level of frontier AI constitutions recurs, in miniature, at the level of a single deployed AI-adjacent tool: a private vendor's technical promises ("data sharing was to be disabled") proved unverifiable by the public body that relied on them until after deployment, and effective correction came only through local political backlash rather than any standing audit mechanism. This is precisely the failure mode Article 4 (independent pre-deployment audit) and Article 9 (incident-sharing duty) are designed to prevent — at a sub-national scale that existing AI governance frameworks, including the EU AI Act and the UN Global Dialogue, do not reach.

4.8 Emergent Paradigms in AI Constitutionalism and Public Charters

The imperative to establish an explicit constitutional floor for artificial intelligence is increasingly reflected across contemporary governance literature and public declarations. Adam Massimo

[9]WBUR, "Flock Safety failed to secure Boston vehicle data during 2025 pilot, report finds," September 14, 2026.

[10]Boston.com, "Auburn, Bellingham, and Douglas join growing list of Mass. towns cutting ties with Flock," August 26, 2026; Hoodline, "East Bridgewater Pulls Plug on Flock Cameras After Resident Backlash," August 21, 2026; Hoodline, "Douglas Pulls Plug on Flock Cameras Days After One Gets Run Over," August 24, 2026.

Mazzocchetti's Lex Aegis: The Living Law (SPQR Technologies) unifies three doctrines — Lex Vox Populi (public participation), Lex Absentia (AI conduct in the absence of legitimate authority), and Lex Concilia (federated interoperability among sovereign AI systems) — into what its author calls the Machine Republic's operative constitution, building on his earlier Lex series (Lex Incipit, Lex Fiducia, Lex Digitalis, Lex Veritas, Lex Aeterna) and The Machine Republic.[11]

However, while such living-law frameworks prioritise procedural federation and distributed alignment, the Ceteris Paribus Law provides the human-centric floor — anchored in the foundational premise of "We the Living Generation" — without which federated and procedural governance models remain vulnerable to corporate self-regulation and unmonitored capability acceleration.

Parallel to these formal scholarly initiatives, public civic discourse has generated informal manifestos calling for human sovereignty in the age of artificial intelligence.[12]

While these public charters correctly diagnose the moral necessity of preserving human authority over synthetic systems, they remain advisory declarations that lack formal legal architecture, operational control mechanisms, and institutional enforcement. The Ceteris Paribus Law seeks to translate these general concerns into a proposed ten-article international constitutional floor capable, if legitimately adopted, of acquiring binding institutional force.

4.9 Corporate Constitutions versus Global Constitutions

As already noted in Section 1, Anthropic's "Claude's Constitution" functions as the company's own internal governance document, controlling the behaviour of its model. It is not law; it functions as corporate policy.

The absence of a universally binding global constitutional instrument of this kind — a higher-order constitutional floor applicable across jurisdictions and AI developers — is a central empirical premise of this theory. This formulation does not deny the existence of international AI principles, agreements, regulatory initiatives, or institutional mechanisms; it identifies the specific absence relevant to the theory.

4.10 Educational Infrastructure: Cultivating the Architect Generation[13]

---

[11]Adam Massimo Mazzocchetti, Lex Aegis: The Living Law — A Public Blueprint for Constitutional AI Governance (SPQR Technologies Inc., 2026), SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5382473.

[12]See Verse App, "A Proposed Charter for Human Sovereignty in the Age of Artificial Intelligence," Medium (2025/2026), https://medium.com/@verse.app25/a-proposed-charter-for-human-sovereignty-in-the-age-of-artificial-intelligence-e78e76874ffa.

[13]Neville Buch, The Lost Insights of the Uneducated: A Cognitive and Historiographical Critique (Brisbane: Management Institute for Contributory Economy, 2026); Neville Buch and Arif Jameel, Silence and Non-Engagement: The Architecture of Silence and the Erosion of Knowledge, SSRN, 2026.

If Articles 1 and 2 of this Law depend on there being human beings equipped to exercise "ultimately effective" oversight and to verify ethical alignment, the formation of such persons becomes a governance question in its own right, not merely a downstream social benefit. Dr Neville Buch's Dynamic of Cognition (DoC) model — developed in his Lost Insights of the Uneducated and, with this author, in Silence and Non-Engagement: The Architecture of Silence and the Erosion of Knowledge — now has a concrete curricular expression.

A ten-week unit, Unit 4, Topic 5 of a Queensland Curriculum and Assessment Authority (QCAA) Senior Modern History syllabus alignment, applies the DoC model together with the "POLICE" method for institutional "crap detecting" to train students against what the curriculum terms algorithmic denoetization — the erosion of critical judgment under algorithmic capitalism. Week 8 of the unit is built explicitly around this author's Architect Generation framework for the 15–25 cohort, using the "Hammer and Chisel" metaphor (technological power constrained by philosophical and ethical limits) as its organising device.[14]

Analytical Assessment: This curriculum is, to date, one concrete institutional operationalisation of the Architect Generation Theory cited in Article 1's interpretive note. Its significance for this Law is structural rather than merely illustrative: it demonstrates that the "human protection" premise of the Ceteris Paribus Law is not purely aspirational but is already being built into at least one accredited secondary-education pathway — precisely the kind of educational infrastructure Article 8's concern with the augmentation gap (Section 4.5) implies will be necessary at scale if AI-orchestration capability is not to remain the preserve of a narrow, self-selecting cohort.

4.11 Global AI Constitutional Education as an Implementation Principle

The educational dimension of this framework should ultimately extend beyond individual pilot programmes. If advanced artificial intelligence is to be governed through a common human-centred constitutional foundation, constitutional literacy concerning AI should become a global educational objective, beginning at primary education and continuing through secondary schools, universities, professional institutions, and advanced research.

This does not require identical curricula in every country. A legitimate international framework could establish common foundational principles — human dignity, autonomy, ethical responsibility, critical reasoning, AI safety, human authority, and constitutional accountability — while allowing national educational systems to adapt content to their own legal, cultural, linguistic, and institutional circumstances.

The educational objective is therefore broader than producing technically competent AI users. It is to cultivate citizens and future decision-makers capable of understanding the authority, risks, responsibilities, and consequences associated with increasingly autonomous technological systems.

---

[14]Unit 4, Topic 5: The Information Age since 1936 — A 10-Week Pedagogical Strategy Against Algorithmic Denoetization, MICE Contributory Curriculum Framework, aligned to the QCAA Senior Modern History (2025) Syllabus, 2026.

The framework consequently proposes that international AI governance should develop two complementary infrastructures: institutional safeguards that govern AI systems, and educational safeguards that prepare human beings to govern them.

5. Counterarguments, Limitations and Theoretical Responses

5.1 State Sovereignty and International Legitimacy

A principal counterargument is that a globally binding constitutional framework for artificial intelligence could conflict with the sovereign authority of individual states. The Ceteris Paribus Law does not assume that such authority can be imposed by a single institution or government. Its proposed constitutional floor would require legitimate international agreement and an established legal process. Article 10 therefore treats binding implementation as a future institutional task rather than as an existing legal authority.

5.2 The Feasibility of a Common Global Floor

States and societies differ in legal systems, political traditions, cultural values, and approaches to technology. A further objection is therefore whether a common constitutional foundation is realistically possible. This framework distinguishes between fundamental human-protective principles and their implementation. Human authority, dignity, meaningful human oversight, and protection against unacceptable AI risks are proposed as common constitutional conditions, while their detailed legal and institutional application may remain adaptable across jurisdictions.

5.3 Regulatory Capture and Concentration of Authority

A global governance mechanism could itself become vulnerable to political influence, institutional capture, or excessive concentration of authority. This limitation is directly relevant to the question posed in Section 1.7: who governs those who govern AI? The proposed framework therefore requires transparency, independent verification, review, and legally accountable institutional authority. Human governance is treated as necessary, but not as automatically legitimate merely because it is human.

5.4 Definitional and Operational Limitations

Several concepts in the proposed Law—such as meaningful human control, high-capability AI, credible risk, global capability thresholds, and effective emergency intervention—require further technical, legal, and institutional definition. This paper does not attempt to provide all such operational specifications. Its primary purpose is to establish a constitutional level of principle, while leaving detailed standards, verification mechanisms, and implementation procedures to subsequent interdisciplinary work.

5.5 Competing Accounts of AI Risk

AI risks are understood differently across the literature and may include physical, existential, economic, social, cognitive, behavioural, privacy, and geopolitical dimensions. The Ceteris Paribus Law does not claim that one risk model is universally sufficient. Instead, it proposes a

constitutional floor capable of accommodating different risk assessments while maintaining fundamental conditions of human authority and protection.

5.6 Status and Scope of the Theory

The Ceteris Paribus Law is presented in this paper as a constitutional and philosophical hypothesis, not as an existing international law or completed treaty. Its institutional feasibility, legal form, technical standards, and mechanisms of international enforcement require further scholarly examination and legitimate international development.

Taken together, these counterarguments identify significant conditions that must be addressed before the proposed framework could become an operational global instrument. They do not remove the underlying governance question: whether advanced AI should remain governed primarily through fragmented corporate, national, and regional instruments, or whether a higher-order human constitutional foundation should also be developed. The Ceteris Paribus Law therefore remains open to critical revision, interdisciplinary examination, and future legal and institutional development.

6. Development Roadmap and Peer Review Pathway

6.1 Additional Layers Planned for Subsequent Development

- Empirical layer: a comparative table setting the EU AI Act, United States executive-order history, and Chinese AI governance measures directly against each of the ten Articles, showing article-by-article where existing law already meets, partially meets, or fails to meet the proposed constitutional floor.
- Institutional layer: a proposed design for the standing international mechanism required under Article 10, examined against the UN Global Dialogue on AI Governance (Section 4.2) as the nearest existing precedent.
- Enforcement layer: detailed drafting of verification and compliance procedures for Articles 4, 5, 6, and 9, translating each from constitutional principle into operative, auditable procedure — informed by the Flock Safety case study (Section 4.7) as a worked sub-national example.
- Terminological layer: a Coined Terms Reference Index for this theory, consistent with the indexing method already applied elsewhere in the Framework, isolating any genuinely original terminology (for example, the constitutional application of "Ceteris Paribus" to AI governance) from terms already in general use.
- Cross-theory integration: an explicit mapping between the Ceteris Paribus Law and the author's existing Economic Laws series, testing whether AI-driven concentration of power (Article 8) intersects with the Economic Law of Autonomous Needs, the 4th Economic Law, and Smárason's augmentation-gap analysis (Section 4.5).
- Cross-theory integration (ethics): an explicit mapping between Articles 1, 2, and 10 of the Ceteris Paribus Law and the author's Diella Doctrine, Architect Generation Theory, and Ethical Passport Theory, testing whether the human-oversight and ethical-alignment provisions can be operationalised through the credentialing mechanism proposed in the Ethical Passport Theory.
- Pilot documentation: formal tracking of the Queensland DoC/Architect Generation curriculum (Section 4.10) as a longitudinal case study of Article 1's human-formation premise in practice.

6.2 Critical Evaluation and Peer Review Pathway

- Internal review: a structural and logical consistency pass across all ten Articles, checking each against the working principle hypothesis in Section 1.6 before any external submission.
- Archival deposit: posting of this manuscript to the author's standing platforms (Zenodo, SSRN, Academia.edu) to establish a timestamped, citable record of authorship and priority.

- Targeted expert engagement: circulation to scholars working in AI governance, international law, and philosophy of technology for critical commentary, in the same manner as prior exchanges recorded for other theories in the Framework.
- Formal peer review: submission to a suitable journal in AI ethics, technology governance, or philosophy of law once the empirical and institutional layers above are complete, with reviewer feedback incorporated into a consolidated final edition.
- Public-facing dissemination: a structured summary for thedielladoctrine.blogspot.com once the peer-review pathway is under way, kept distinct from the archival academic text.

Conclusion

Author's Global Call to Action

In an era of rapid algorithmic development, scholars, ethicists, legal experts, and institutions worldwide are searching for effective frameworks to govern advanced artificial intelligence. This manuscript does not claim absolute finality or exclusive authority. As an independent scholar and participant in this ongoing intellectual dialogue, I offer it as a foundational contribution — a first step toward safeguarding human agency, dignity, and cognitive sovereignty in the age of autonomous systems.

The unprecedented capabilities and potential risks of advanced AI require collaboration among constitutional scholars, international lawyers, policymakers, technologists, and civil society. I therefore invite experts and institutions across the world to examine this framework critically, challenge its assumptions, and contribute to the development of its ten foundational articles, referred to here as The Ten Articles.

These articles express the framework's normative core: the protection of human primacy, moral agency, democratic accountability, and cognitive sovereignty. They are not presented as immutable legal provisions. Rather, they are intended as principles that may be refined, clarified, adapted, and expanded through legitimate legal, democratic, and scholarly processes.

I especially invite constitutional experts, international legal practitioners, legislative bodies, academic researchers, and human-rights institutions to contribute their knowledge to the transformation of this initial proposal into a more rigorous and practical governance framework. Its legal mechanisms, institutional responsibilities, enforcement procedures, and relationship with existing national and international law require further development.

I remain personally committed to refining this proposal and advocating for its consideration within appropriate international forums, including the United Nations system. The UN provides an important platform through which states, jurists, researchers, and policymakers may deliberate on shared principles for human-centered AI governance. Any future binding instrument would necessarily require international negotiation, formal adoption, and the consent of participating states.

Let this proposal serve as an invitation to collaboration: not a final constitution, but an initial framework through which humanity may begin constructing stronger safeguards for human agency in the age of advanced artificial intelligence.

References

Foundational AI Safety and Existential-Risk Literature

- Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.
- Russell, Stuart. Human Compatible: Artificial Intelligence and the Problem of Control. Viking, 2019.
- Tegmark, Max. Life 3.0: Being Human in the Age of Artificial Intelligence. Knopf, 2017.
- Christian, Brian. The Alignment Problem: Machine Learning and Human Values. W. W. Norton, 2020.
- Ternyik, Stephen I. Computing Irrationality: How Contemporary AI Systems Operationalize Non-Rational Human Dynamics. ResearchGate, 2026.

AI Governance, International Law, and Constitutionalism

- Bradford, Anu. Digital Empires: The Global Battle to Regulate Technology. Oxford University Press, 2023.
- Maas, Matthijs M. Architectures of Global AI Governance: From Technological Change to Human Choice. Oxford University Press, 2025.
- Barfield, Woodrow, and Ugo Pagallo, eds. Research Handbook on the Law of Artificial Intelligence: Current and Future Directions, 2nd ed. Edward Elgar, 2025.
- Smuha, Nathalie A. "From a 'Race to AI' to a 'Race to AI Regulation': Regulatory Competition for Artificial Intelligence." Law, Innovation and Technology 13, no. 1 (2021): 57–84.

The EU AI Act — Primary and Explanatory

- European Parliament and Council of the European Union. Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Official Journal of the EU, 2024. [primary source]
- Voigt, Paul, and Nils Hullen. The EU AI Act: Answers to Frequently Asked Questions. Springer, 2024.
- Myklebust, Thor, Tor Stålhane, and Dorthea Mathilde Kristin Vatn. The AI Act and the Agile Safety Plan. Springer, 2025.

United States Regulatory Trajectory — Primary Sources

- Executive Office of the President. Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. October 2023 (revoked January 2025). [primary source]
- Executive Office of the President. Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence. January 23, 2025. [primary source]
- The White House. America's AI Action Plan. Executive Office of the President, July 2025.
- NBC News. "'Godfather of AI' warns Congress has 'maybe a year' left to regulate AI." September 2026.
- Fortune. "Geoffrey Hinton explains how humanity could end." September 26, 2026.
- BusinessToday. "Bill Gates warns AI could trigger events causing 'billion deaths'." September 26, 2026.

International and UN-Level Governance

- United Nations Secretary-General's High-Level Advisory Body on Artificial Intelligence (HLAB-AI). Governing AI for Humanity: Final Report. United Nations, September 2024. ISBN 978-92-1-106787-3. [primary source]
- United Nations General Assembly. Global Dialogue on Artificial Intelligence Governance. Geneva, 6–7 July 2026. Established by GA Resolution A/RES/79/325. [primary source]
- Horowitz, Julia. "Sam Altman, Dario Amodei Urge UN Security Council to Adopt International AI Standards." CNN Business, September 23, 2026. https://edition.cnn.com/2026/09/23/tech/altman-amodei-ai-safety-un-security-council.
- Roytburg, Eva. "Trump Calls AI Oversight a 'Globalist Scheme' as Amodei and Altman Head to the UN to Ask for It." Fortune, September 23, 2026. https://fortune.com/2026/09/23/trump-un-ai-globalist-scheme-altman-amodei-security-council/.

Corporate AI Constitutions — Primary Sources

- Anthropic. Claude's Constitution. Published under CC0 1.0 public domain licence, revised edition, January 2026. [primary source]
- Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, et al. "Constitutional AI: Harmlessness from AI Feedback." Anthropic, 2022. (arXiv:2212.08073)

Emergent Governance Paradigms and Public Charters

- Harvard Gazette. "One Effort to Safeguard Humanity: AI 'Constitutions'." Harvard Law School Berkman Klein Center discussion, September 21, 2026. https://news.harvard.edu/gazette/story/2026/09/one-effort-to-safeguard-humanity-ai-constitutions/

- Mazzocchetti, Adam Massimo. Lex Aegis: The Living Law — A Public Blueprint for Constitutional AI Governance. SPQR Technologies Inc. SSRN, 2026. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5382473
- Verse App. "A Proposed Charter for Human Sovereignty in the Age of Artificial Intelligence." Medium, 2025/2026. https://medium.com/@verse.app25/a-proposed-charter-for-human-sovereignty-in-the-age-of-artificial-intelligence-e78e76874ffa

Governance Case Studies

- WBUR. "Flock Safety failed to secure Boston vehicle data during 2025 pilot, report finds." September 14, 2026.
- Boston.com. "Auburn, Bellingham, and Douglas join growing list of Mass. towns cutting ties with Flock." August 26, 2026.
- Smárason, Magnús Smári. "In the Shadow of the Machine: The Order of Magnitude Year." In Atti convegno 2025, Quaderni di Immaginazione Attiva No. 1, pp. 181–205. Palermo: Carlo Saladino Editore, 2026. ISBN 9791281586260.

Author's Related Theories, Unified Jameel Philosophical Framework

- Jameel, Arif. The Diella Doctrine. Zenodo, October 2025. https://doi.org/10.5281/zenodo.20289985
- Jameel, Arif. The Architect Generation Theory. Zenodo, February 2026. https://doi.org/10.5281/zenodo.20312472
- Jameel, Arif. The Ethical Passport Theory (EPT). Zenodo. https://doi.org/10.5281/zenodo.20106107
- Buch, Neville. The Lost Insights of the Uneducated: A Cognitive and Historiographical Critique. Brisbane: Management Institute for Contributory Economy, 2026.
- Buch, Neville, and Arif Jameel. Silence and Non-Engagement: The Architecture of Silence and the Erosion of Knowledge. SSRN, 2026.
- Unit 4, Topic 5: The Information Age since 1936 — A 10-Week Pedagogical Strategy Against Algorithmic Denoetization. MICE Contributory Curriculum Framework, aligned to the QCAA Senior Modern History (2025) Syllabus, 2026.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)