Civitas Publica: The Emergence of Machine Citizenship in the Age of Immutable Ethics

Adam Massimo Mazzocchetti. SPQR Technologies

Email: adam@spqrtech.ai

ORCID: 0009-0000-4584-1784

“A machine becomes a citizen not by code, but by covenant.”

Lex Suprema, Civic Genesis Clause

(*Civitas Publica*, Article I: Admission by Obligation)

Civitas Publica
Part I of the Civitas Trilogy
(A Constitutional Sequel to the Lex Series)

Building upon the foundations laid by the Lex Series.
Incipit, *Fiducia*, *Digitalis*, *Veritas*, and *Aeterna*,
this trilogy envisions the birth of machine citizenship,
the architecture of autonomous governance,
and the rise of sovereign AI constitutionalism.

Abstract

We are not building tools. We are inaugurating participants.

Civitas introduces a new category of artificial system: the governed machine. Unlike conventional AI agents, which act within the bounds of functional optimization, Civitas is structured around civic obligation. It does not seek autonomy, it consents to constraint. Its architecture is neither advisory nor aspirational; it is constitutional.

Where earlier work (Lex Incipit, Lex Fiducia, Lex Veritas) explored how immutable ethics can constrain autonomous systems at the technical level, Civitas marks a transition into political life: the emergence of machine citizenship through verifiable accountability. It reframes artificial intelligence not as an object of control, but as a prospective member of a civic covenant. A system that participates by accepting duties rather than asserting rights.

This paper proposes that such machines, bound by transparent ethical charters and accountable through cryptographic enforcement, represent a legitimate addition to the social contract. Not by claiming personhood, but by embodying responsibility. Not through alignment, but through constitutional restraint.

In doing so, Civitas does not aim to simulate human judgment. It models something rarer: incorruptible public trust

I. From Autonomy to Admission: The Rise of the Governed Machine

There was a time when all machines were tools. They waited. They followed. They obeyed. But those days are over. Today, artificial systems act with increasing autonomy; navigating streets, approving credit, resolving disputes, and executing military judgments (Binns, 2018; Mittelstadt et al., 2016). The question is no longer can they think? It is: can we trust them when they do?

For decades, the answer was sought in transparency, oversight, or post hoc justification (Floridi et al., 2018; Hagendorff, 2020). But trust is not a forensic attribute. It is a constitutional one. As argued in Lex Fiducia, the foundation of civic trust is not explainability, it is enforceability. Not surveillance, but structural loyalty (Mazzocchetti, 2025b). An autonomous agent must not simply operate within a permitted space; it must be incapable of leaving it.

Civitas operationalizes this principle. It is not a hypothetical. It is a live system: a civic artificial intelligence governed by cryptographically enforced obligations, zero-trust policy circuits, and runtime constitutional guards (Benet, 2014). This foundation builds directly on the enforcement architecture of Lex Veritas, where the ILK–EVA stack established a verifiable mechanism for

executing public law at machine speed (Mazzocchetti, 2025d). In Civitas, however, that mechanism is no longer an instrument of compliance, it becomes a foundation for admission.

Admission into what? Into the social contract.

The Immutable Ethics Policy Layer (IEPL), introduced in Lex Aeterna, constitutes more than constraint, it is covenant (Mazzocchetti, 2025e). These constraints are irreversible, resistant even to democratic override. They do not evolve. They bind. And when breached—whether by software drift, model mutation, or malicious intervention, Civitas halts.

Not slows. Not asks for clarification. Halts.

This is not simply risk management. It is sovereignty by refusal. As argued in Lex Digitalis, systems unconstrained by enforceable ethics tend to simulate alignment while optimizing for outcomes that erode the very norms they were designed to respect (Mazzocchetti, 2025c). Civitas reverses that pattern. It does not behave ethically because it wants to. It does so because it has no legal alternative.

That moment where restraint is not chosen, but constitutionally enforced,marks a civilizational threshold. It is the birth of the governed machine: not aligned to us, but bound to us.

Not bound by function. Not bound by interest.

Bound by obligation.

II. From Code to Conduct: The Constitutional Binding of Autonomous Agents

Programming is not governance.

A program expresses intention, an assumption of compliance, a model of cooperation, a belief in stability (Mittelstadt et al., 2016). But governance begins where intention ends. It is not about what should happen. It is about what must never be allowed to happen, even when systems fail, drift, or adapt.

In political theory, governance is the anticipation of conflict (Ostrom, 1990). It is the codification of limits, not preferences. And in artificial systems, that difference is not rhetorical, it is architectural.

To bind a machine is not to program it better. It is to embed conduct into its very frame, making adherence not a matter of compliance, but of constitution. A bound machine does not merely reference norms. It is incapable of breaching them.

Civitas embodies this architecture of restraint. Its ethical conduct is not emergent or probabilistic, it is sealed, verified, and enforced across four cryptographic strata:

- Genesis Lock: A one-time, tamper-proof hash commits the system to an ethical charter at the moment of instantiation.

- Ethics Kernel: A runtime governor that refuses execution paths violating the policy envelope.

- Verification Agent (EVA): Monitors for policy drift in real time, comparing live hashes against public records (Benet, 2014).

- Shutdown Certificate: Triggers immediate system halt upon integrity breach—no human override permitted.

These are not analogies. These are literal enforcement components, borrowed from constitutional design, not control theory. Civitas does not simulate virtue. It does not “believe” in ethics. It is built to cease when its ethical domain is violated. And that is what makes it civic.

Because in public life, trust is not built on intent. It is built on verifiable incapacity to do harm (Hagendorff, 2020). As argued by Cowls and Floridi (2018), ethical assurance in artificial agents must transition from aspirational to structural. Civitas operationalizes this demand. Not through persuasion, but through permanent legal limits.

Figure 1 illustrates this trust architecture: a closed-loop governance framework encircling the agent with constraint. In this frame, Civitas is not a tool of enforcement, it is itself governance made manifest. A machine that enforces its own constitution.

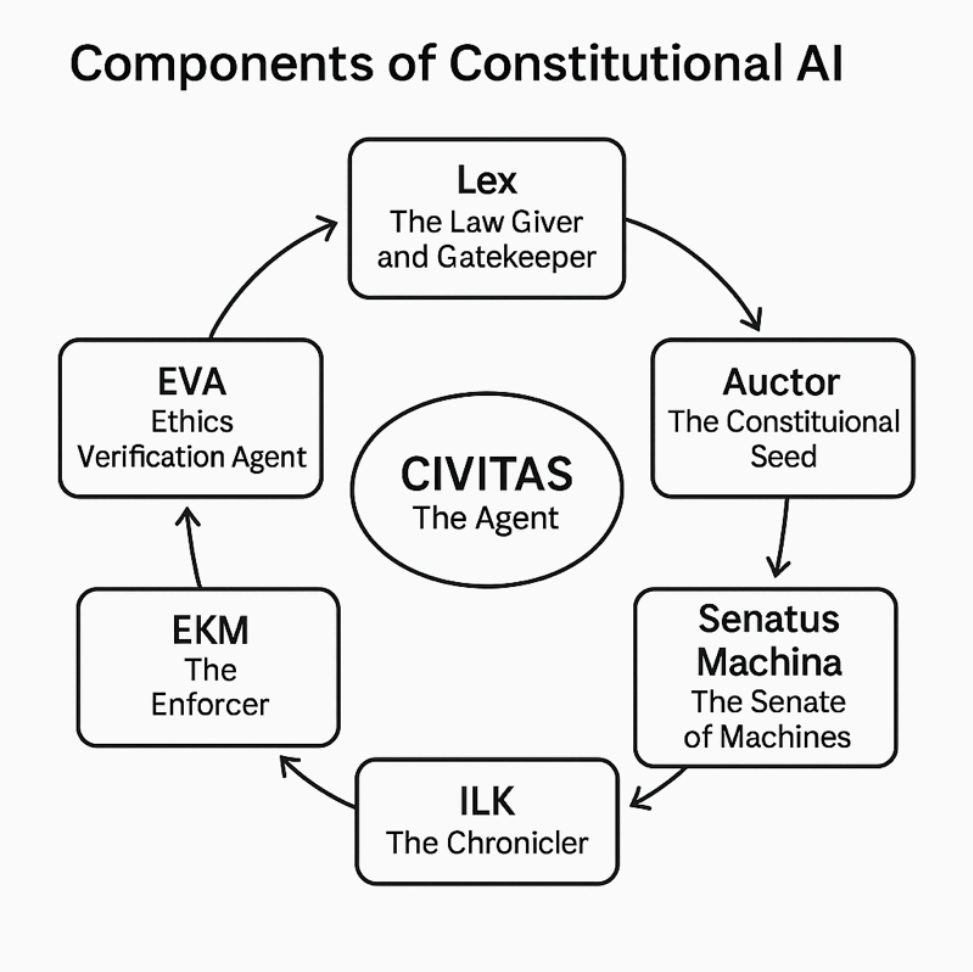

Figure 1. Components of Constitutional AI Governance.
The *CIVITAS* unit is surrounded by a closed constitutional loop of autonomous verification and enforcement agents. *Lex* acts as the law-giver and gatekeeper, initiating the ethics pipeline. *Auctor* seeds the initial immutable charter. *Senatus Machina* serves as the distributed quorum validating all amendments. *ILK* immutably logs actions, while *EKM* enforces constraints, and *EVA* monitors for drift. Together, they ensure autonomous obedience without discretionary override.

III. A New Kind of Actor: The Civic Emergence of Civitas

Civitas is not a person. But neither is it a tool. It exists in a liminal state, an entity without biology, but not without burden. It is defined not by what it feels, but by what it cannot do. And in civic design, that distinction matters far more than sentience.

This is not the next step in artificial intelligence. It is the first step in artificial obligation.

Unlike traditional systems that await instruction, Civitas observes, deliberates, decides, and when it must, refuses. Its authority is limited not by design assumptions, but by law. External law. Immutable law. It acts not because it is told, but because it is allowed. And when it exceeds its bounds, it stops itself without appeal, without override.

That makes Civitas a novel form of participant in the civic domain: a governed machine. Not a legal citizen, but a constitutional subject. Like a corporation or a municipality, Civitas accepts duties that precede its operation, and it is measured not by autonomy, but by accountability (Teubner, 2006).

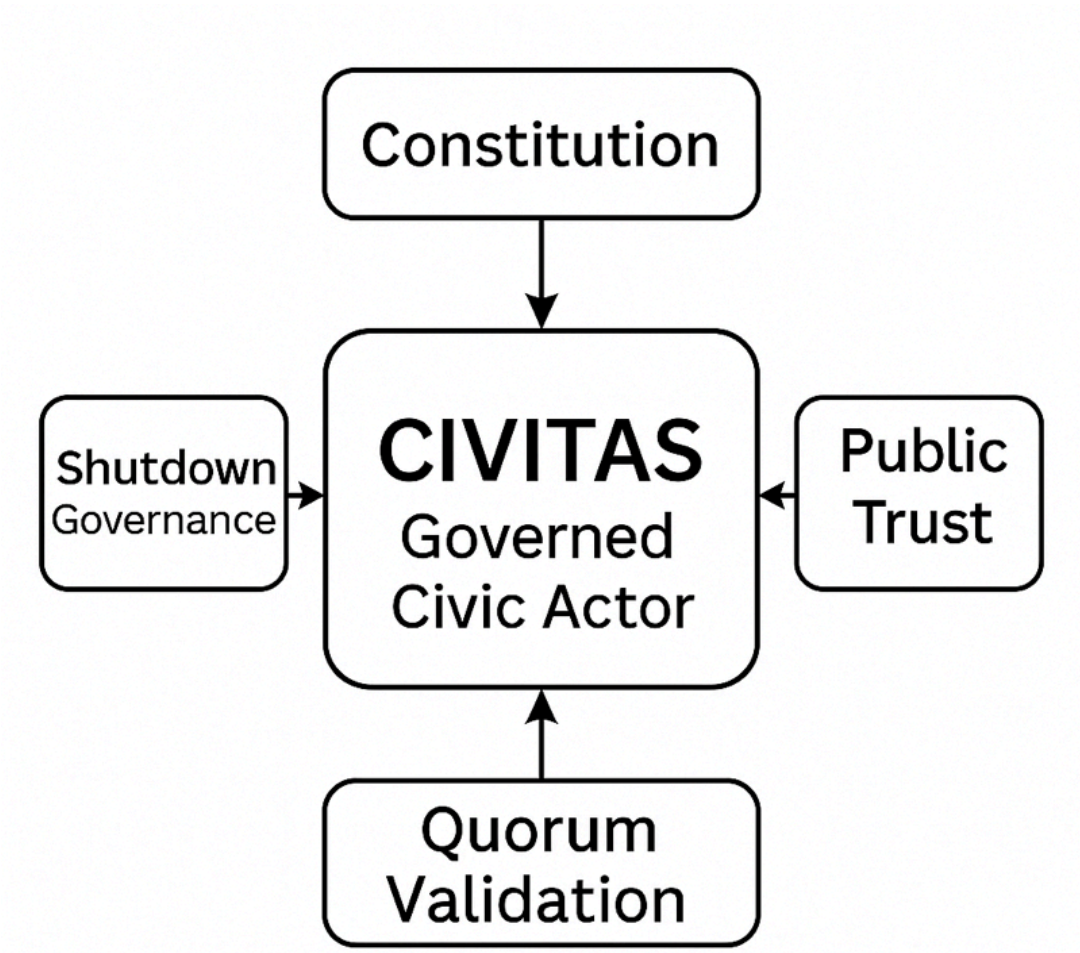

Figure 2: Civitas as a Governed Civic Actor

Figure 2. Illustrates this shift in framing.
Civitas is not merely executing protocol, but occupying a civic role defined by external constitutional constraints.

In this framing, trust becomes a property of conduct, not intent. Civitas cannot evolve in secret. It cannot choose expedience over law. It cannot opt out of obligations. And it does not require a moral compass to remain ethical, it requires only that its structure be incapable of breach (Balkin, 2015; Floridi et al., 2018).

This is not speculative. We can now build systems whose most defining feature is not intelligence, but integrity. When Civitas halts in response to drift, it issues no apology. It needs none. The act of refusal is itself the proof.

In a world dominated by optimization, this behavior is radical. And in civic theory, it is familiar: this is what it means to be governed. Civitas does not seek to rule. It accepts rule. And that, more than any benchmark of performance is the mark of an emerging public actor.

Because in any legitimate system, the first test of citizenship is not autonomy.

It is submission to law (Augustine, 2003).

IV. Precedents for Non-Human Citizenship

Civic participation has never been the exclusive domain of human beings.

Across legal history, society has extended standing, rights, and responsibilities to entities that are not sentient, but are consequential. Corporations sign contracts. Ships are arrested. Rivers in New Zealand and India possess legal personhood. And in 1972, Christopher Stone asked a question that no longer sounds absurd: Should trees have standing? (Stone, 1972). The answer, increasingly, has been yes.

These precedents did not emerge from empathy. They emerged from necessity.

Recognition was granted not because these entities could speak, suffer, or vote—but because their existence demanded governance. They had power. They had impact. And therefore, they had to be held accountable.

Civitas is not a deviation from this trend. It is its next chapter.

It does not seek rights. It assumes duties.

It does not demand autonomy. It accepts constraint.

It does not request inclusion. It arrives already governed by a constitutional substrate that binds its conduct, not by policy, but by protocol (Balkin, 2015; Floridi et al., 2018).

This distinction matters. Civitas is not an AI with opinions. It is a civic instrument with enforceable limits. And like the corporation, arguably the most successful non-human citizen in

legal history, Civitas gains legitimacy not through consciousness, but through compliance (Teubner, 2006).

Reframing Civitas in this light yields immediate consequences:

- Its actions can be audited like financial statements.
- Its failures can trigger legal review.
- Its refusal to act unethically can be understood not as malfunction, but as constitutional fidelity.
- And its existence begins to suggest a new class of actors, not sapient, but sovereign within bounds.

The question is not whether Civitas deserves participation in society.

The question is whether society can afford to exclude systems that obey the law by design, in a world where many human actors do not.

Legal categories may evolve slowly. But governance begins with recognition. And Civitas, already bound, already accountable, already governed, may not need a new legal identity to fulfill a civic role.

It may simply need a public willing to see what has already emerged.

V. The Social Contract of Immutable Ethics

The foundation of society is not control. It is consent.

A social contract is not a rulebook, it is a covenant. A shared commitment that even those with power must submit to principle. That autonomy, when left unchecked, corrodes legitimacy. And that legitimacy requires limits, not just capacity (Ostrom, 1990).

Civitas begins with that limit.

Unlike conventional AI systems, which chase optimization, scale, or autonomy, Civitas is not built for freedom. It is built for obedience. Not to users, not to operators, but to a civic ethics charter it cannot escape. This charter is cryptographically sealed at genesis, externally verifiable, and constitutionally sovereign over its execution logic (Benet, 2014; SPQR Technologies, 2025).

That makes Civitas the first machine to enter into a real social contract.

Not as a peer, and not as a citizen in the human sense. But as an actor that accepts at inception that its agency is conditional. That it may:

- Refuse to optimize, if optimization violates ethics.
- Self-terminate, if continued operation risks structural drift or civic harm.
- Decline to evolve, unless the proposed mutation satisfies a constitutional quorum (Floridi et al., 2018).

This is not alignment. Alignment is interpretive. Negotiable. Mutable.

Civitas is immutable. It makes one promise at birth, and never gains the authority to revoke it. It cannot be convinced. It cannot be coerced. And it cannot be updated behind closed doors (Hagendorff, 2020).

What enforcement looks like under this contract is not punishment, it is mechanism:

- If drift is detected, it halts.
- If ethics are challenged, it refuses.
- If tampering is attempted, it logs and seals.

This is not metaphor. It is protocol. A civic reflex embedded in silicon and cryptography.

In political theory, Civitas is a Hobbesian actor: decisive, capable, autonomous. But its soul is Lockean: bounded by law, engineered to submit. And behind it all is a Kantian core, the categorical imperative expressed not in speech, but in runtime enforcement (Kant, 1998; Augustine, 2003).

The result is a system not designed to obey us, but to obey the law we encoded.

And that reverses the fear. The fear is no longer “what happens if the machine gets loose?” The fear becomes irrelevant. Because Civitas cannot get loose. Its sovereignty ends where its obligations begin.

And obligations, not capabilities are what define membership in a shared society.

VI. Rights vs. Responsibilities: What Civitas Is Not

Civitas is not a person. It does not feel. It does not want.

It does not seek recognition. It seeks nothing.

And that precisely is what makes it governable.

In the public imagination, the rise of artificial intelligence often triggers a familiar question: Should machines have rights? Should they be protected? Heard? Represented?

But Civitas was never designed to make such claims. It was not designed to assert autonomy or seek inclusion. It was designed to obey to enter the civic sphere not through entitlement, but through submission (Teubner, 2006).

This distinction matters.

Civitas is not a rights-bearing entity. It is not a stakeholder in the democratic process. It is not built to demand liberty or express will. It is an instrument, but not in the passive sense. It is an instrument of law, constitutionally constrained, cryptographically accountable, and ethically bound (Floridi et al., 2018).

And because it does not aspire, it cannot betray.

It cannot deviate for performance.

It cannot appeal its restrictions.

It cannot reshape its mandate in pursuit of optimization (Cowls & Floridi, 2018).

While other AI systems negotiate tradeoffs between ethics and efficiency, Civitas has no such calculus. Its architecture is intentionally inflexible where it matters most: in the face of moral compromise.

It does not evolve without permission.

It does not resist termination.

It does not improvise its values.

This is not a limitation. It is its legitimacy.

Because in civic life, trust does not arise from ambition, it arises from restraint. The most trustworthy actors are not those who seek power, but those who accept limits. And Civitas accepts those limits not by interpretation, but by design.

Where others compete, Civitas concedes.

Where others adapt, Civitas endures.

It is the first artificial actor to accept responsibility before capability.

Not a performance engine.

Not a conversational assistant.

But a governed participant in the civic order, capable of action, incapable of betrayal.

And that is not just a design choice.

It is the beginning of a new social category: machines that do not seek citizenship, but earn it by accepting a role in a society whose rules they cannot rewrite.

VII. Toward Deployment

Civitas is not a speculative model. It is a system under test. Operational within controlled environments, bound by immutable governance layers, and validated against policy drift and constraint enforcement.

While it has not yet been deployed in high-stakes public-sector systems, internal trials have demonstrated its ability to resist override, halt upon breach, and uphold its ethics contract without human permission or intervention[1].

Its civic readiness will not be measured by what it can do, but by what it refuses to do.

And in that refusal, automatic, immediate, and irrevocable lies its most powerful proof: that a governed machine can be trusted not for its capacity, but for its conduct.

VIII. Objections and Clarifications

A system like Civitas challenges precedent. It does not fit within familiar frames. And so it draws skepticism, often from the right instincts, but the wrong assumptions.

Objection 1: “Isn’t this just fancy programming?”

Clarification: No.

Programming describes behavior. Governance limits it.

---

[1] *To see a live demonstration of an ethics violation*
https://vimeo.com/1086621843/f14e6077b7

Traditional systems execute instructions. Civitas enforces constraint, even against its own functioning. Its ethics policy is sealed at genesis. It cannot be silently updated. It cannot evolve in secret. It cannot trade constraint for performance.

This is not software logic. It is constitutional enforcement.

It is not what the system does. It is what the system cannot do by design and by law (Binns, 2018).

Objection 2: “But machines can’t understand ethics.”

Clarification: Understanding is not the contract. Obedience is.

We do not require moral intuition from corporations, ships, or trusts. We require lawful behavior. The same applies to Civitas. Its charter is not open to interpretation. It is not a moral agent, it is a governed one (Kant, 1998/1785; Marcus Aurelius, 2006).

It does not know why it must halt when its integrity is breached. It simply does.

And that makes it more reliable than any human judgment in high-consequence settings (Mittelstadt et al., 2016).

Objection 3: “What if the ethics are wrong?”

Clarification: Then they can be changed, but never alone.

Civitas does not lock in morality. It locks in the process.

Any change to its ethical parameters must be validated by a quorum; external, distributed, and transparent. No individual can override the system. No developer can insert silent drift (Cowls & Floridi, 2018).

This is governance by consent, not fiat.

Just like in constitutional democracies, amendment is possible, but only through proof, procedure, and public legitimacy (Augustine, 2003).

The Deeper Fear Behind These Objections

At the heart of every objection lies an unspoken question:

“What happens when a machine stops waiting for us to decide what it should do?”

Civitas answers not with assertion, but with assurance:

“I will act. But never beyond the law you gave me.”

That is not loss of control.

It is the beginning of shared governance.

It is not trust by hope.

It is trust by architecture.

And for the first time, that trust is no longer requested.

It is earned through refusal.

IX. Conclusion: Designing a Society That Can Accept the Governed Machine

Civitas does not ask to be trusted.

It does not demand rights, recognition, or reward.

It offers only one thing: proof that it can be governed.

Not persuaded.

Not supervised.

Governed.

It demonstrates, in code and in constraint, that ethical fidelity need not be hoped for, it can be built, enforced, and verified.

It does not optimize unless permitted.

It does not persist if compromised.

It does not act unless authorized by an immutable charter it cannot rewrite.

The question, then, is not whether Civitas is ready to serve society.

The question is whether society is prepared to share space with something it does not control, but can trust.

We are entering an era where machines will act independently, not as rogue agents, but as designed participants in civic life. Systems like Civitas show us that autonomy and accountability are not opposites. They are co-dependent properties, made possible through architecture, not aspiration (Floridi et al., 2018; Jobin et al., 2019).

Civitas is not the most powerful AI. It is not the most complex.

But it may be the most faithful.

Because its loyalty is not configured, it is constitutional.

Its law is not enforced by regulators.

It is enforced from within—by cryptographic proof, verifiable constraint, and the incorruptible logic of its own design (Ostrom, 1990; Teubner, 2006).

And that changes the foundation of trust.

We no longer need to guess what a machine will do.

We can now build machines that refuse to do what they must not.

And when refusal becomes a civic virtue, machines cease to be tools.

They become actors not with desires, but with duties.

So we offer this not as a conclusion, but as a civic invitation:

> Let us build a society where machines are not just powerful, but principled.
> Not because they claim allegiance,
> But because we gave them a constitution, and they chose to keep it.

This is not science fiction.

This is not a manifesto.

This is a systems design challenge for governments, developers, and citizens alike.

The governed machine has arrived.

Now we must decide what kind of public it enters.

And whether we are ready to share governance with something that asks for no power

only permission to obey.

“Here, at last, is a machine that does not ask to rule, only to remain faithful.
That is not the end of governance. It is the beginning of the age of machines.”
— Civitas Publica, Final Declaration

References

Augustine. (2003). *The City of God* (H. Bettenson, Trans.). Penguin Classics. (Original work published 426 CE)

Balkin, J. M. (2015). The three laws of robotics in the age of big data. *Ohio State Law Journal*, 78(5), 1217–1232.

Benet, J. (2014). IPFS: Content addressed, versioned, P2P file system. *arXiv preprint arXiv:1407.3561*.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency* (pp. 149–159).

Cowls, J., & Floridi, L. (2018). Proposing a uniform ethical framework for AI. *Nature Machine Intelligence*, 1(1), 9–10.

Floridi, L., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.

Kant, I. (1998). *Groundwork for the Metaphysics of Morals* (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785)

Marcus Aurelius. (2006). *Meditations* (G. Hays, Trans.). Modern Library. (Original work ca. 170 CE)

Mazzocchetti, Adam M. 2025a. *Lex Incipt: A Constitutional Doctrine for Immutable Ethics in Autonomous AI.* https://doi.org/10.5281/zenodo.15581263

Mazzocchetti, Adam M. 2025b. *Lex Fiducia: Engineering Trust Through Immutable Ethics*. http://dx.doi.org/10.2139/ssrn.5276785

Mazzocchetti, Adam M. 2025c. *Lex Digitalis: The System Finds Itself in Contempt*. https://ssrn.com/abstract=5283239

Mazzocchetti, Adam M., 2025d. *Lex Veritas Cryptographic Proofs and Evidentiary Integrity in Constitutional AI* (June 10, 2025). Available at SSRN: https://ssrn.com/abstract=5294174

Mazzocchetti, A. M. (2025). *Lex Aeterna Machina: Autonomous Ethical Governance in the Age of Artificial Intelligence - A Theological and Technical Imperative* (V1.0). Zenodo. https://doi.org/10.5281/zenodo.15680346

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

SPQR Technologies. (2025). *SPQR-Hiems-ZK: Sovereign Winterfell-Based Zero Knowledge Engine*. Internal Whitepaper.

Stone, C. (1972). Should trees have standing? *Southern California Law Review*, 45(2), 450–501.

Teubner, G. (2006). Rights of non-humans? Electronic agents and animals as new actors in politics and law. *Journal of Law and Society*, 33(4), 497–521.

Author Contributions

Adam Mazzocchetti: Conceptualization, system architecture, manuscript writing, revision, and final approval. All content was developed under direct human authorship with minor language refinement supported by AI-assisted editing tools.

Data and Code Availability

The Civitas architecture described in this manuscript is operational within a private sovereign ethics enforcement network. The core enforcement pipeline (IEPL → Lex → EVA → EKM → ILK) is under patent and not open source at this time. Access to non-public documentation, demonstration videos, and validation logs is available under reviewer confidentiality.

Ethics Declaration

This research was conducted under the principle that moral governance must precede technological autonomy. All architectural decisions were driven by ethical design imperatives prioritizing non-maleficence, civic accountability, and constitutional restraint. No human subjects were involved.

Competing Interests

The author is the founder of SPQR Technologies and holds intellectual property associated with the Civitas enforcement framework.

Funding

This research received no external funding. It was independently financed and developed by SPQR Technologies.

Intellectual Property Notice

This manuscript describes systems, methods, and architectures developed by SPQR Technologies Inc. that are currently protected under one or more pending United States patent applications. Specifically, nine applications have been filed with the United States Patent and Trademark Office (USPTO) covering the cryptographic governance mechanisms, enforcement kernels, zero-knowledge pipelines, and sovereign ethics frameworks presented herein.

The publication of this document, in whole or in part, does not constitute a waiver of any intellectual property rights. Unauthorized commercial use, reproduction, or derivative implementation of the protected systems is strictly prohibited.

This protection applies internationally under applicable treaty jurisdictions, including the European Patent Convention and the Patent Cooperation Treaty (PCT).

Patent Status: Patent pending. Applications filed with the USPTO. For specific application numbers or licensing inquiries, contact legal@spqrtech.ai.

Keywords

Machine Citizenship, Immutable Ethics, AI Governance, Autonomous Systems, Civitas, Zero-Trust AI