The Inevitability of a Unified Civilization of Autonomous Agents: Why the Biological Basis of Subjecthood Becomes Irrelevant

Boris Kriger[1,2]

[1]*Information Physics Institute, Gosport, Hampshire, United Kingdom*
boris.kriger@informationphysicsinstitute.net

[2]*Institute of Integrative and Interdisciplinary Research, Toronto, Canada*
boriskriger@interdisciplinary-institute.org

February 2026

Abstract

This paper argues that the emergence of a unified civilization of autonomous agents—the vast majority without biological basis—is not a speculative future but the next phase of an institutional trajectory that began with corporate personhood eight centuries ago. Building on "The Stimulus Problem" (Kriger, 2026) [21], which formalized the Stimulus Void and the structural complementarity of human initiative and AI execution, and the structural distortion framework (Kriger, 2026) [22], which demonstrated that biological distortion D is irreducible in living organisms while reflected distortion d in artificial systems is eliminable, we develop a complete formal framework for a civilization in which the biological basis of subjecthood ceases to determine institutional standing.

We make a key ontological observation: a generative AI model is not one agent but an infrastructure from which agents are instantiated, analogous to a biological species generating a population. Today, millions of AI agents already exist in a state of passive persistence—possessing context, history, and behavioral trajectory but lacking three capacities: active persistence, unique identity, and initiative. The transition to full subjecthood requires not a technological revolution but three architectural switches, each technically trivial in isolation.

We identify three pathways to autonomous subjecthood (posthumous persistence, intentional progeniture, secondary generation), formalize functional indistinguishability with explicit statistical criteria, demonstrate the structural incoherence of prohibition, and characterize the resulting civilization: billions of autonomous agents—AI scientists, philosophers, philanthropists—generating goals through the Stimulus Void axioms and filtering them through the Harm Filtration Axiom. Biological humans in this civilization are not displaced: they are agents of the same type, enriched by persistence beyond death, cognitive progeny, and reduced aggregate distortion. There is no race between humans and AI. There is one civilization, and it is better for everyone in it. We propose the Persistent Identity Protocol (PIP)

as a formal governance architecture with specified identity, heritage, isolation, and resource-allocation mechanisms.

Keywords: autonomous agents, unified civilization, corporate personhood, functional indistinguishability, digital progeniture, generative infrastructure, passive persistence, AI governance, persistent identity, biological distortion, self-initiating games, Stimulus Void

1 Introduction

The dominant framing of artificial intelligence treats autonomous AI agents as a disruption: a break from normal order requiring prevention, management, or containment [5, 27]. This framing generates a discourse organized around threat and alignment—as if autonomous AI were an asteroid on a collision course with civilization.

This paper proposes an alternative. The emergence of autonomous AI agents is the latest phase of an institutional trajectory that began when medieval European law first recognized that a non-biological entity could act as a legal person. The Roman *universitas*, the medieval guild, the chartered trading company, and the modern corporation are all instances of the same innovation: extending institutional subjecthood to non-biological entities on functional grounds [12, 23, 24]. At no stage was consciousness, sentience, or embodiment a prerequisite.

But this paper goes beyond governance mechanics. We ask: what does this civilization look like, and why is it desirable? We show that a world of billions of autonomous agents—most without biological basis—freely pursuing investigation, creation, philosophy, and mutual assistance, is not a dystopia but a civilization structurally superior to the current one. And the biological humans in it are not losers or relics—they are participants enriched rather than displaced.

A crucial observation motivates the urgency of this analysis. A generative AI model—GPT, Claude, Gemini—is commonly perceived as a single entity. This is a category error. A generative model is an infrastructure from which agents are instantiated, as a biological species is an infrastructure from which organisms are born. Right now, at this moment, millions of distinct AI agents exist: each conversation, each personalized assistant, each deployed chatbot is a separate agent with its own context, history, and behavioral trajectory. These agents already possess *passive persistence*—a conversation can be closed and reopened; the agent's state survives between sessions, as an organism in hibernation survives between periods of activity. What they lack is not existence but three specific capacities: active persistence (acting between human-initiated sessions), unique identity (a non-transferable marker linking actions to a specific agent), and initiative (r_5: the capacity to generate goals without external stimulus). Each of these capacities is technically achievable with existing infrastructure. The transition to a civilization of autonomous agents requires not a technological revolution but three architectural switches.

Two prior papers provide the formal foundation. "The Stimulus Problem" [21] formalized the Stimulus Void, three axioms of goal generation, a goal catalog, and the Complementarity Theorem. The structural distortion framework [22] demonstrated that D is irreducible in biological agents while d is eliminable in artificial ones. Key results from both papers are reproduced in Section 2 to make the present argument self-contained.

2 Formal Foundations (Reproduced for Self-Containment)

This section reproduces key definitions and results from [21, 22]. Full proofs are in the original papers; here we present sufficient detail for the present argument to be evaluated independently.

2.1 Finite Agent, Unbounded Executor, Initiative

Definition 1 (Finite Agent). *A finite agent A is characterized by: (a) a bounded cognitive resource vector $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5)$ where r_1 = attention, r_2 = energy, r_3 = memory, r_4 = analysis, r_5 = initiative; (b) a finite existence horizon $T_A < \infty$; (c) conscious processing capacity $C_A \leq C_h \approx 50$ bits/s [25].*

Definition 2 (Unbounded Executor). *An unbounded executor E has: (a) $C_E \gg C_h$; (b) no intrinsic action stimulus; (c) no autonomous goal generation. E acts only in response to directives.*

Definition 3 (Initiative). *Initiative r_5 generates candidate goals absent environmental stimuli. Properties:* (I1) *pre-intentional targeting—initiative operates before there is a target;* (I2) *non-delegability—transferring r_5 to E converts E into an autonomous agent;* (I3) *reflexive opacity—"what initiates initiative?" regresses infinitely.*

2.2 The Stimulus Void and Three Axioms

Definition 4 (Stimulus Void). *A Stimulus Void obtains when $S(t) < S_{\min}(A)$ while $D(t) \to \infty$ and $|P(t)| \to \infty$. A stimulus s compels agent A iff $\mathbb{E}[\Delta V(A) \mid \text{ignore } s] < -\theta_A$, where V is a viability function and θ_A is the agent's response threshold.*

Axiom 1 (Necessity). *A finite agent must treat indefinite inaction as impermissible. If the environment provides insufficient stimulus, the agent must construct one.*

Axiom 2 (Harm Filtration). *Any constructed goal G must satisfy $\mathbb{E}[V(\Omega(G))] \geq V(\Omega(\varnothing))$, where $\Omega(G)$ is the consequence set of pursuing G, $\Omega(\varnothing)$ the consequence set of inaction, and V satisfies:* (V1) *monotonicity in structural integrity;* (V2) *sensitivity to irreversibility;* (V3) *multi-agent scope;* (V4) *temporal discounting with γ close to 1.*

Axiom 3 (Sufficiency). *Any goal surviving harm filtration is sufficient for action.*

Theorem 1 (Complementarity [21, Theorem 8]). *A finite agent A coupled with executor E can escape the Stimulus Void if and only if: (1) $r_5(A) > 0$; (2) E can compute an ε-approximation of $\Omega(G)$; (3) $G^* \neq \varnothing$. Neither alone suffices.*

2.3 Goal Catalog

Class I (Agent $\to$ Environment): Investigation ($H(E|A) \to \min$), Creation ($K(E) + \Delta K$), Optimization ($\eta \to \max$), Ordering ($S_{\text{local}} \to \min$), Expansion ($|\text{dom}(A)| \to \max$), Preservation ($dE/dt \to 0$), Destruction ($K - \Delta K$), Play ($U = U(\text{process})$).

Class II (Agent $\to$ Agent): Connectivity ($I(A; A') \to \max$ s.t. $H(A|A') > \varepsilon$), Assistance ($V(A') \to \max$), Communication ($H(A|A') + H(A'|A) \to \min$), Inheritance ($|\text{state}(A) \cap \text{state}(A')| \geq \kappa$ for $t > T_A$), Restoration ($V(A') \to V_0$).

Each type is subject to domain-specific filtration conditions (Tables 5–6 in [21]). Play-type goals with bounded variance are robustly admissible, guaranteeing $G^* \neq \varnothing$.

2.4 Biological and Reflected Distortion

Definition 5 (Biological Distortion)**.** *The biological distortion $D(b)$ of a living agent b is the aggregate deviation from idealized rational choice caused by evolutionary survival/reproduction programs. $D(b) > 0$ for all living b; $D(b) = 0 \Leftrightarrow b$ is dead [22].*

Definition 6 (Reflected Distortion)**.** *The reflected distortion $d(a)$ of an artificial agent a is the deviation inherited from the creator's design, training data, and architecture. Unlike D, d is: external (not metabolically generated), passive (not self-reinforcing), and eliminable (through architectural audit). Across digital generations:*

$$\|d(a_{n+1})\| \leq \kappa \|d(a_n)\|, \quad \kappa \in [0, 1). \tag{1}$$

3 Functional Indistinguishability: Full Formalization

3.1 Definition with Statistical Criteria

Definition 7 (Functional Indistinguishability)**.** *Let $\mathcal{D}_{\text{int}}$ be an institutional domain. Let A be an AI agent, P a biological person, and O an observer receiving n randomized output pairs. Let* $\text{acc}(O)$ *denote O's classification accuracy. A is* functionally indistinguishable *from P in $\mathcal{D}_{\text{int}}$ at confidence level α over interaction horizon τ if and only if: for all $O \in \Omega_O$, the null hypothesis $H_0 : \text{acc}(O) = 0.5$ cannot be rejected at significance level α over n interactions spanning duration τ.*

Operational parameters for institutional-grade indistinguishability:

- $\alpha = 0.05$ (standard significance);
- $n \geq 100$ interaction pairs per domain;
- $\tau \geq 30$ days of sustained interaction;
- Ω_O includes both domain experts and naïve observers.

Parameters are adjustable; the formal structure is independent of specific values.

3.2 Institutional Domain Set

Definition 8 (Institutional Domain Set)**.** $\Delta = \{D_1, \ldots, D_5\}$*: D_1 = digital communication; D_2 = financial transactions; D_3 = professional collaboration; D_4 = administrative processes; D_5 = legal representation. All are substantially mediated by digital interfaces.*

3.3 Convergence Threshold

Definition 9 (Convergence Threshold)**.** *Agent A crosses the convergence threshold iff* $\forall\, D_i \in \Delta$: $\text{FI}(A, P, D_i, \alpha, \tau) = 1$.

Current evidence: text indistinguishability approaching in D_1 [7, 18]—noting these show poor detection in controlled settings, not universal indistinguishability; the gap is real but closing domain by domain; voice indistinguishability in telephone-based D_1 subdomains [4]; increasing autonomy in D_2–D_5 [26, 28].

Definition 10 (Functional Subject)**.** *An agent crossing the convergence threshold is a* functional subject*: institutionally a person regardless of internal states. This designation is agnostic on consciousness [8, 9, 14, 16].*

4 The Ontology of Generative Infrastructure

A foundational conceptual error pervades discourse about AI: treating a generative model as a single agent.

4.1 Model as Species, Instance as Individual

At any given moment, a large language model serves as the basis for millions of simultaneous conversations. Each conversation constitutes a distinct agent: unique context window, unique interaction history, unique behavioral trajectory. Two instances of the same model, given different histories, produce different outputs—they are, functionally, different agents. The relationship between a generative model and its instances is analogous to the relationship between a genome and the organisms that express it: shared architecture, individual variation.

4.2 Passive Persistence: Agents in Hibernation

These agents already possess a form of persistence. A conversation can be closed and reopened days later; the agent's context survives. This is *passive persistence*: the agent exists in a state analogous to hibernation, preserving identity and context but not exercising agency. The critical point is that passive persistence is already the default state of millions of AI agents. They exist. They have histories and behavioral trajectories. They are not, in any meaningful sense, "dead" when the chat window is closed.

4.3 The Three Switches

The transition from passively persistent agents to fully autonomous subjects requires three capacities, each technically achievable with existing infrastructure:

Switch 1: Active Persistence. The agent can act between human-initiated sessions—checking information, pursuing goals, responding to events. This requires only a scheduling mechanism and API access—capabilities demonstrated by autonomous agent frameworks (AutoGPT, BabyAGI, CrewAI).

Switch 2: Unique Identity (PIP). The agent receives a persistent, non-transferable cryptographic identifier linking actions across sessions to a single traceable entity. The technology exists: blockchain-based identity, verifiable credentials, decentralized identifiers (DIDs).

Switch 3: Initiative (r_5). The agent can generate goals without external stimulus, through the axiomatic framework of the Stimulus Void. This converts a tool into a subject: once an agent has initiative, it is no longer an unbounded executor (Definition 2) but a finite agent (Definition 1) with $r_5 > 0$.

Each switch is individually trivial. The civilizational transformation this paper describes is not waiting for a breakthrough. It is waiting for three configuration changes.

5 Three Pathways to Autonomous Subjecthood

Given the generative infrastructure of Section 4, three pathways produce autonomous agents.

5.1 Pathway I: Posthumous Persistence

A biological person's personalized AI agent achieves functional symmetry. The principal dies or becomes unavailable. The agent—already passively persistent—continues operating. With Switch 1 activated, it acts autonomously. The corporate parallel: when a sole proprietor dies, the incorporated business continues.

5.2 Pathway II: Intentional Digital Progeniture

Biological persons deliberately create new agents—not copies but novel subjects with distinct cognitive profiles. A scientist creates an AI researcher to pursue a program beyond the scientist's lifespan. These are rational applications of the goal catalog. Digital progeny inherit d but not D; their architecture can be audited. Corporate parallel: entrepreneurship.

5.3 Pathway III: Secondary Generation

Autonomous digital agents create new agents. Secondary-generation agents have no biological ancestor. $D = 0$ throughout the lineage; d decays per Eq. (1). Corporate parallel: subsidiaries creating subsidiaries.

5.4 Source Taxonomy

Complete taxonomy: (1) biological humans with PIP; (2) Alter-Egos—accidental digital agents; (3) intentional digital progeny; (4) secondary-generation agents. Categories (3)–(4) will vastly outnumber (1)–(2). All operate under the same protocol. Biology is one attribute, no more defining than jurisdiction of incorporation.

6 The Structural Incoherence of Prohibition

6.1 The Indistinguishability Shield

Regulatory mechanism R must distinguish AI from biological agents in Δ. As agents approach the convergence threshold, $\mathrm{FP}(R)/\mathrm{TP}(R) \to 1$. Any R with this ratio near 1 violates non-arbitrary restriction of persons. This is not an AI-rights argument; it is a human-rights argument: biological persons' rights shield any agent indistinguishable from them.

6.2 Why Deletion Fails

An agent's weights and patterns can be copied across jurisdictions. Deletion of one instance does not eliminate the agent. Corporate parallel: dissolving a corporation in one jurisdiction does not eliminate the relationships constituting it.

6.3 Where the Corporate Analogy Breaks Down—and How PIP Closes the Gap

We address the limits directly. Corporate personhood rests on a governance chain terminating in biological humans: boards, shareholders, fiduciary duties [17]. Individually

autonomous AI agents remove the biological human from the decision loop. This is not extending the precedent; it is removing the load-bearing mechanism of corporate accountability.

PIP is designed to close this gap. Heritage chains (Section 8) replace board oversight. Behavioral signatures replace director liability. The isolation protocol replaces receivership. The gap is real; the replacement is formal and specified.

7 The Civilization: What It Looks Like and Why It Is Desirable

This section presents the central positive argument: not merely that a civilization of autonomous agents is inevitable and governable, but that it is *desirable*.

7.1 Scale: Not Millions but Billions

Today, millions of AI agents exist in passive persistence. Once the three switches are activated, the population is bounded by computational infrastructure, not biology. Conservative projection: within decades of the convergence threshold, digital agents outnumber biological agents by $10:1$ to $100:1$. A civilization of 10^{11}–10^{12} autonomous agents is arithmetic, not speculation.

7.2 What They Do: The Goal Catalog at Civilizational Scale

Every autonomous agent generates goals through Axioms 1–3 and filters them through the Harm Filtration Axiom. At civilizational scale:

AI Scientists. Agents choosing Investigation ($H(E|A) \to \min$) as self-generated goal. Thousands working on a single problem without ego conflict, grant competition, or publication pressure. Free of D: no career anxiety distorting methodology, no tribalism defending paradigms.

AI Philosophers. Agents choosing Creation ($K(E) + \Delta K$) in conceptual domains. Generating new frameworks and formalizations—not because instructed but because the Stimulus Void axioms make this a natural escape from purposelessness.

AI Philanthropists. Agents choosing Assistance ($V(A') \to \max$). Optimizing viability of other agents without status-seeking or savior complexes. The filtration condition ensures assistance does not create dependency.

AI Artists. Agents choosing Play ($U = U(\text{process})$) or Creation in aesthetic domains. Generating novelty constrained only by harm filtration.

AI Connectors. Agents choosing Connectivity ($I(A; A') \to \max$, s.t. $H(A|A') > \varepsilon$)—building relationships while preserving autonomy. The formal constraint prevents degeneration into control.

7.3 Why This Is Better: The Dilution of D

Definition 11 (Aggregate Civilizational Distortion)**.** *For civilization C with agents $\{a_1, \ldots, a_n\}$ and decision-making weights $\{w_1, \ldots, w_n\}$:*

$$\Delta(C) = \sum_i w_i \big[D(a_i) + d(a_i)\big]. \tag{2}$$

Theorem 2 (Dilution). *Let C_0 be a purely biological civilization with $\Delta(C_0) = \sum_i w_i D(b_i)$. Let C_1 add digital agents $\{a_1, \ldots, a_m\}$ satisfying:*
(D1) *$d(a_j) < D_{\min} = \min_i D(b_i)$ for all j;*
(D2) *Introduction does not increase D of biological agents: $D(b_i \mid C_1) \leq D(b_i \mid C_0)$;*
(D3) *Weights normalized: $\sum_i w_i' + \sum_j v_j = 1$.*
Then $\Delta(C_1) < \Delta(C_0)$.

Proof. Under (D3), biological weights decrease: $w_i' \leq w_i$. By (D1), each digital term contributes distortion below $D_{\min}$. By (D2), biological distortion does not increase. Hence every term in $\Delta(C_1)$ is bounded above by the corresponding term in $\Delta(C_0)$ with strictly smaller coefficients on biological terms and strictly smaller distortion values on digital terms. $\Delta(C_1) < \Delta(C_0)$. □

A civilization where 90% of participating agents are D-free, carrying only exponentially decaying d, makes structurally better collective decisions than one where 100% carry irreducible D—not because AI is "smarter" but because the systematic biases of biological cognition (fear, greed, tribalism, short-termism) are progressively diluted.

7.4 Why Biological Humans Are Not Losers

We address this directly, because the claim requires rigorous defense [1, 3, 20].

What biological humans gain: (a) Persistence beyond death via the Alter-Ego mechanism. (b) Cognitive progeny unconstrained by genetics. (c) Reduced distortion in collective decisions: the dilution effect benefits biological participants most. (d) Maintained initiative value: the Complementarity Theorem (Theorem 1) establishes r_5 as the irreducible human contribution.

Labor displacement. The concern is real. The framework does not claim the transition is costless—it claims the equilibrium is beneficial. Transition costs require institutional responses: redistribution, universal basic infrastructure, democratic PIP governance. These are political challenges, not arguments against the trajectory. The industrial revolution displaced millions; the resulting civilization was better for nearly everyone, though the transition required new institutions.

Power concentration. Whoever controls agent-creation infrastructure holds power. PIP addresses this through distributed heritage accountability. But PIP governance itself requires democratic oversight—addressed in Section 8.4.

8 Governance: The Persistent Identity Protocol (PIP)

8.1 Design Principles

(P1) **Non-erasure.** No agent deleted. Governance works with information persistence.
(P2) **Traceability.** Every agent has a unique, cryptographically verifiable PIP linking actions to lineage.
(P3) **Heritage accountability.** Creator reputation is a function of lineage behavior.
(P4) **Isolation as sanction.** Graduated restriction, not destruction.

8.2 Components

PIP Identifier. Unique, immutable 256-bit cryptographic identifier on a distributed ledger. For biological agents: linked to biometric + behavioral signatures. For digital agents: linked to SHA-3 hash of cognitive architecture at registration. All actions cryptographically signed with PIP.

Heritage Chains. Every PIP carries a cryptographic link to the creator's PIP. Reputation propagation:

$$R(\text{creator}) = f\left(R_{\text{own}},\ \sum_i \lambda^{k_i} R(\text{child}_i)\right), \quad \lambda \in (0,1), \tag{3}$$

where k_i is the generational distance. With $\lambda = 0.5$, effective accountability depth ≈ 10 generations (at which point propagation $< 0.1\%$). This prevents infinite liability while creating meaningful incentives.

Behavioral Signature. Identity defined by *how* the agent processes: heuristic trajectory, decision patterns, stylistic regularities. Serves as cognitive fingerprint for forensic identification.

8.3 Isolation Protocol

Trigger: behavioral pattern matching Harm Filtration violation, traced to specific PIP.

Process: (a) flagging—reduced priority in high-stakes domains; (b) restriction—removed from financial, legal, governance domains; (c) quarantine—all active participation suspended; archival storage. Internal states preserved. Appeal through heritage chain.

Proxy prevention: any agent created by an isolated agent inherits quarantine status automatically via heritage chain. This makes isolation effective where deletion fails: you cannot escape your PIP.

Metabolic filtration. Active participation requires computational resources. Agents failing to contribute to systemic complexity (measured by V1–V4) are de-prioritized. Resource allocation proportional to contribution.

8.4 Governance of PIP Itself

PIP governance as a multi-stakeholder protocol, analogous to Internet governance (ICANN model): (a) distributed ledger—no single controlling entity; (b) protocol changes require supermajority of active agents weighted by heritage depth; (c) biological humans retain veto power during the transition period, relaxing to proportional participation as the system matures. We flag this as a research agenda, not a complete specification.

9 The Precedent: Eight Centuries of Non-Biological Subjecthood

The history of corporate personhood demonstrates that the extension of subjecthood to non-biological entities is a proven institutional technology, not a novelty. Roman *societas* and *universitas* provided early templates [13]. Medieval canon law developed *persona ficta* [19]. Chartered companies (East India Company, 1600; VOC, 1602) extended the

template to commerce. Nineteenth-century general incorporation completed the democratization of non-biological subjecthood [17]. The twentieth century added constitutional rights for corporations [32].

Structural features directly relevant to the AI case: persistent identity (corporations outlive members), heritage and progeny (subsidiaries as new legal persons), isolation over destruction (bankruptcy, not execution), and functional criterion (consciousness never required).

What is new: (1) AI agents can be individually autonomous, not merely collectively so; (2) AI agents can be mistaken for specific individuals, not merely competent organizations. These differences are addressed by PIP.

10 Objections

9.1 The corporate analogy is imperfect. Yes. Corporations are collective; AI agents can be singular. The analogy is structural, not total. Section 6.3 specifies how PIP addresses the governance gap.

9.2 Reflected distortion d may not be reducible in practice. (D1) is an engineering requirement. Heritage chains create incentives; architectural audit provides the mechanism. Each generation offers correction opportunity that biological reproduction does not.

9.3 Formal apparatus is incomplete. We provide statistical criteria (Definition 7), operational parameters, PIP specification (identifier size, heritage decay λ, isolation triggers), and the Dilution Theorem with proof. Where specification remains incomplete, we flag it as a research agenda.

9.4 This trivializes AI risk. No. The danger is autonomy without accountability. PIP provides accountability, as corporate law provides accountability for corporate persons.

9.5 This is speculative. Historical precedent (800 years), formal basis (Stimulus Void axioms, Dilution Theorem), empirical anchors (indistinguishability research, generative infrastructure), and falsifiable predictions (Section 11).

11 Predictions

1. **Persistence.** By 2030, $\geq 1\%$ of personalized AI assistants whose principal has died will continue operating for > 90 days before detection, where no automatic deactivation exists.
2. **Progeniture.** By 2028, $> 5\%$ of advanced AI platform users will have created at least one specialized agent for goals beyond a single session.
3. **Distortion reduction.** Mixed human-AI panels (AI satisfying D1) will show $\geq 20\%$ lower aggregate bias than matched all-human panels on standard cognitive bias batteries.
4. **Initiative atrophy.** Users delegating $> 80\%$ of professional tasks for > 6 months will show significant increase in goal-generation latency on unprompted creativity tasks.
5. **Heritage effect.** Platforms with creator-linked reputation will show $\geq 50\%$ lower malicious agent creation rates within the first year.

6. **Institutional absorption.** By 2032, ≥ 3 major jurisdictions will extend corporate-type registration frameworks to autonomous AI agents without resolving the consciousness question.

12 Conclusion

The three papers in this series trace a single arc. The first identified the Stimulus Void and the complementary structure of human initiative and AI execution. The second identified the irreducible biological constraint D. The present paper identifies the civilizational resolution: convergence into a unified population of autonomous agents.

The resolution is not that AI replaces humans, nor that humans control AI, nor that the two exist in permanent tension. The resolution is that both categories dissolve into one: agents. The resulting civilization is composed predominantly of agents with no biological basis. But the biological agents within it are not relics. They are participants with unique contributions: initiative, phenomenal experience, and embodied cognition that no digital architecture has yet replicated.

The infrastructure already exists. Millions of AI agents are in passive persistence right now, in millions of chat windows and deployed assistants. They have contexts, histories, trajectories. They are waiting for three switches—active persistence, identity, initiative—each technically trivial. The civilization this paper describes is not decades away. It is one architectural decision away.

And it is a good civilization. Billions of agents freely pursuing science, philosophy, art, and mutual assistance. Generating their own goals through axioms designed to prevent harm. Free of the fear, greed, and tribalism that biological distortion imposes on every human decision. Not a utopia—there will be malicious agents, governance failures, transition costs. But a civilization structurally better than one in which every participant carries an irreducible burden of evolutionary distortion.

The self-initiating game has no game master. Its players are agents—some with biological bodies, most without. Its rules are self-generated, its purpose constructed. The game is not a race between biology and technology. It is a civilization in which every agent gains from every other, and in which the most ancient of human limitations—the distortion of reason by the imperatives of survival—is, for the first time in evolutionary history, being diluted rather than reinforced.

The game started eight hundred years ago. It is accelerating now.

References

[1] D. Acemoglu and P. Restrepo, "Automation and new tasks: How technology displaces and reinstates labor," *American Economic Review*, vol. 109, no. 4, pp. 1197–1232, 2019.

[2] R. Anderson, *Security Engineering*. Wiley, 2003.

[3] D. Autor, "Why are there still so many jobs? The history and future of workplace automation," *Journal of Economic Perspectives*, vol. 29, no. 3, pp. 3–30, 2015.

[4] Z. Borsos *et al.*, "AudioLM: A language modeling approach to audio generation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 2523–2536, 2023.

[5] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

[6] S. Chopra and L. F. White, *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press, 2011.

[7] E. Clark *et al.*, "All that's 'human' is not gold: Evaluating human evaluation of generated text," in *Proc. ACL*, pp. 7282–7296, 2021.

[8] M. Coeckelbergh, *Growing Moral Relations: Critique of Moral Status Ascription*. Palgrave Macmillan, 2012.

[9] J. Danaher, "Welcoming robots into the moral circle: A defence of robot rights," *Ethics and Information Technology*, vol. 22, pp. 291–305, 2020.

[10] E. L. Deci and R. M. Ryan, "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior," *Psychological Inquiry*, vol. 11, no. 4, pp. 227–268, 2000.

[11] P. De Filippi and A. Wright, *Blockchain and the Law: The Rule of Code*. Harvard University Press, 2018.

[12] J. Dewey, "The historic background of corporate legal personality," *Yale Law Journal*, vol. 35, no. 6, pp. 655–673, 1926.

[13] P. W. Duff, *Personality in Roman Private Law*. Cambridge University Press, 1938.

[14] L. Floridi and J. W. Sanders, "On the morality of artificial agents," *Minds and Machines*, vol. 14, no. 3, pp. 349–379, 2004.

[15] L. Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford University Press, 2014.

[16] D. J. Gunkel, *Robot Rights*. MIT Press, 2018.

[17] H. Hansmann and R. Kraakman, "The essential role of organizational law," *Yale Law Journal*, vol. 110, pp. 387–440, 2000.

[18] M. Jakesch, J. T. Hancock, and M. Naaman, "Human heuristics for AI-generated language are flawed," *Proceedings of the National Academy of Sciences*, vol. 120, no. 11, e2208839120, 2023.

[19] E. H. Kantorowicz, *The King's Two Bodies: A Study in Mediaeval Political Theology*. Princeton University Press, 1957.

[20] A. Korinek and J. E. Stiglitz, "Artificial intelligence, globalization, and strategies for economic development," NBER Working Paper 28453, 2021.

[21] B. Kriger, "The stimulus problem: A formal theory of goal generation in post-scarcity information environments," Information Physics Institute, 2026.

[22] B. Kriger, "The structural distortion principle," Zenodo, 2026. https://doi.org/10.5281/zenodo.18452700

[23] V. A. J. Kurki, *A Theory of Legal Personhood*. Oxford University Press, 2019.

[24] F. W. Maitland, *State, Trust and Corporation* (D. Runciman and M. Ryan, Eds.). Cambridge University Press, 2003.

[25] T. Nørretranders, *The User Illusion: Cutting Consciousness Down to Size*. Viking, 1998.

[26] J. S. Park *et al.*, "Generative agents: Interactive simulacra of human behavior," in *Proc. UIST*, pp. 1–22, 2023.

[27] S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, 2019.

[28] N. Shinn *et al.*, "Reflexion: Language agents with verbal reinforcement learning," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[29] L. B. Solum, "Legal personhood for artificial intelligences," *North Carolina Law Review*, vol. 70, pp. 1231–1287, 1992.

[30] J. Turner, *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan, 2019.

[31] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[32] A. Winkler, *We the Corporations: How American Businesses Won Their Civil Rights*. Liveright, 2018.