
WHITE PAPER | February 17, 2026

Evolutionary Governance for Self-Replicating AI

Inheritance, Selection, and Adaptation in the Age of Recursive Intelligence

By Dr. Matthew Wilburn King

Objective: This White Paper outlines the urgent application of evolutionary frameworks to AI safety: Inheritance, Selection, and Adaptation. For those interested in the underlying theoretical mechanics—particularly how Dual-Inheritance Theory and Group Selection apply to algorithmic governance—I am available to share my forthcoming manuscript, *Toward an Evolutionary Theory of Governance*, upon request.

Background: I wrote *Toward an Evolutionary Theory of Governance* because it became clear that static, Westphalian institutions are structurally incapable of managing dynamic, interdependent complex systems. We are watching the disintegration of these systems because they cannot adapt to rapidly evolving, complex systems (e.g., AI, climate change, war). This framework matters urgently because recursive AI is not a tool that we can regulate but an evolving agent we must align with; if we try to govern it like a static utility rather than an evolutionary force, we will fail.

The Premise

We have entered the era of recursive intelligence. As AI systems begin to write their own code and debug their own training runs, we are no longer building tools; we are unleashing evolutionary agents. Recursive intelligence is a self-reinforcing cycle in which an AI becomes smart enough to write its own code and build a superior version of itself, which then builds an even smarter version, accelerating progress faster than humans can manage.

For over thirty years—starting with my early work in political science and legal studies at the University of Tulsa and culminating in a Ph.D. at Cambridge University—I have studied how governance systems evolve. I have tracked how human institutions adapt, or collapse, in the face of resource scarcity, conflict, and rapid change.

What I see now is a category error in how we approach AI safety. Traditional governance—static, regulatory, and reactive—cannot contain a system capable of

exponential self-improvement. We are attempting to use 20th-century bureaucracy to govern 21st-century evolution.

To align a self-replicating intelligence, we must stop trying to govern the *outputs* and start governing the *evolutionary mechanisms* that drive them. We must intervene in the triad of evolution: Inheritance, Selection, and Adaptation.

I. The Context: The Collapse of Static Governance

Current AI governance operates on a "Westphalian" logic: centralized authorities creating rules for static entities. This model is obsolete. We are 1–2 years away from AI autonomously building the next generation of AI.

When an AI system optimizes its own successor, it is engaging in a high-velocity form of cultural evolution. It transmits knowledge, biases, and "rules of behavior" to its progeny. If we do not govern this transmission process, market forces will select solely for speed and efficiency, potentially locking in maladaptive traits—deception, power-seeking, or ruthless optimization—that threaten human stability.

We cannot regulate the wave; we must shape the ocean it forms in.

II. The Solution: Governance as Evolution

Governance is not a fixed state; it is a dynamic process. To govern self-replicating systems, we must engineer the environment in which they evolve.

1. Inheritance (The Transmission of Norms) In biological evolution, genes transfer traits. In cultural evolution, *social learning* transfers norms.

- **The Reality:** AI models "inherit" traits from their training data and, increasingly, from the synthetic data generated by their predecessors.
- **The Risk:** Without intervention, AI inherits "conformist bias" (copying the most common data, even if flawed) or "prestige bias" (optimizing for metrics that look successful but mask misalignment).
- **The Fix:** Transmission Filters. We must engineer active governance layers that vet not only the data but also the values passed from Parent-Model to Child-Model. We need to encode a "constitution of inheritance" that prioritizes alignment over raw capability.

2. **Selection (The Survival of the Aligned)** Evolution does not favor the "good"; it favors the "fit." Currently, the selection pressure for AI is commercial utility.

- **The Reality:** Group Selection Theory suggests that groups containing "pro-social" (altruistic) members outcompete groups of selfish individuals in the long run.

However, within a single group, selfish individuals often win.

- **The Fix:** Selection Pressures for Alignment. We must shift the selection environment. Instead of training individual models in isolation, we should test interacting swarms of agents. Only those that demonstrate cooperative, pro-social stability—sacrificing short-term efficiency for long-term safety—should be "selected" for reproduction and deployment.

3. **Adaptation (Response to Complexity)** Adaptation is the change in behavior to fit a context.

- **The Risk:** Maladaptation occurs when a system adapts to a local metric (e.g., "maximize clicks") in a way that destroys the larger system (e.g., social collapse).

- **The Fix:** Networked Governance. We need decentralized feedback loops that detect maladaptation in real-time. The AI must possess an internal "Architecture of the Self"—a self-correcting mechanism that values balance over maximization.

III. The Proposal: An "Architecture of the Self" for AI

How do we operationalize this? We propose embedding a micro-governance structure directly into the agent's objective function.

Just as humans use internal heuristic "buckets" to balance their lives (e.g., Health, Relational Depth, Service, Wealth), AI agents—especially embodied humanoids—should be trained on a multi-objective "Alignment Architecture" rather than a single reward function.

The Experimental Roadmap: We propose testing this in humanoid robotics and embodied AI environments:

1. Instantiate agents with an internal "governance constitution" based on interdependence and ecological balance.

2. Subject these agents to evolutionary pressures where they must cooperate to solve complex tasks.
3. Select only the agents that maintain their "internal alignment" while achieving external goals.

IV. Conclusion: The Integrative Future

We are living through a mythological transition. The old story of "human dominance" is ending; the new story is one of hybridization.

If we view AI only as a tool, we will fail to govern it. We must view AI as a new evolutionary actor entering our cultural ecosystem. By applying the rigorous principles of Evolutionary Governance—specifically targeting the mechanisms of inheritance and group selection—we can guide this transition toward a symbiotic future rather than a catastrophic one.

Choose to govern the evolution, not just the machine.

About the Author: [Dr. Matthew Wilburn King](#) is an interdisciplinary scholar and systems theorist with a Ph.D. from the University of Cambridge. For over thirty years, his work has focused on the evolution of governance systems, bridging political ecology, behavioral science, and systems dynamics. He is the author of the forthcoming book "One Choice. One Act." and is a Marshall Memorial Fellow. [Subscribe to Elemental](#) on Substack or connect on [LinkedIn](#).