

From Urban Ecology to AI Alignment: The Wisdom Forcing Function™ as an Innovation Dividend

Carlos Arleo

Independent Researcher, The Regenerative Development Initiative

October 4, 2025

Abstract

Current AI alignment methods treat safety as constraint optimization, imposing an “alignment tax” that reduces capability. This paper asks: *what if alignment could be designed not just to prevent harm, but to invent novel solutions to complex global problems?* We introduce the Wisdom Forcing Function™ (WFF), a dialectical cognitive architecture that extends recent work in Constitutional AI by reframing alignment as **cultivation rather than containment**. Drawing from critical urban theory, regenerative design, and biomimicry, the WFF operationalizes tension-rich constitutions to generate wisdom through structured, iterative conflict.

The architecture’s core innovation is a **Zero-Trust Cognitive Loop**, integrating a programmatic Verifier (VDK) to ensure a transparent, auditable reasoning process. Empirical evidence demonstrates a quantifiable ‘innovation dividend’: the autonomous synthesis of novel governance protocols and self-enforcing safety mechanisms. Furthermore, in a separate experiment designed to test its response to a direct constitutional paradox (“The Oracle’s Dilemma”), the system demonstrated a capacity for meta-ethical self-correction, proposing a new principle for ‘*Liberatory Intervention*’ to resolve paradoxes in its own constitution.

Through a deep-dive case study of a 10-iteration “dialectical struggle,” we uncover three core discoveries: (1) a new paradigm of **Alignment-by-Architecture**; (2) a new definition of “wise” outputs as **self-defending architectures** with unbypassable gates; and (3) a new societal role for AI as a **facilitator of human wisdom**, demonstrated through the system’s generation of the *Genesis Protocol™* —a complete methodology that culminates in the design for a “Dialectical IDE,” a civic technology platform for collective wisdom. This reframes alignment from a cost center to a value-creating engine, positioning AI as a collaborative partner in co-evolution and the cultivation of systemic wisdom.

Keywords: AI alignment, Wisdom Forcing Function™, innovation dividend, constitutional AI, regenerative design, dialectical architecture, Genesis Protocol™, VDK, self-defending architectures

1 Introduction

The AI alignment discourse is dominated by metaphors of control: AI as a powerful tool to contain or an optimizer to constrain. This frames alignment as a tax—additional overhead reducing speed, capability, and utility. While effective for mitigating known harms, this subtractive approach is insufficient for cultivating the wisdom needed to address complex, systemic challenges. Recent advances in Constitutional AI (CAI) have shown the power of principle-based guidance, yet often still operate within a framework of minimizing harm.

We propose a paradigm shift from **containment to cultivation**, and from **alignment-by-instruction** to **alignment-by-architecture**. Instead of controlling an AI through external constraints, the Wisdom Forcing Function™ makes ethical and strategic coherence a *structural property* of its multi-agent system.

Inspired by nature’s “productive tension”—predator-prey dynamics driving biodiversity—and urban dialectics fostering innovation, the WFF™ treats alignment as an ecological relationship. The metaphor shifts from fencing a beast to gardening an ecosystem: humans, as stewards, create the conditions for flourishing, measured not by obedience but by resilient co-evolution. This conceptual shift reframes the “alignment tax” as an “**innovation dividend**,” where constitutional tensions force emergent novelty and a structural capacity for generating resilient new architectures.

The Three Core Discoveries (Executive Summary)

Our empirical studies revealed three defensible breakthroughs:

1. **Alignment-by-Architecture:** Safety and strategic coherence become structural properties of the multi-agent design, rather than emergent outcomes of single models.
2. **Self-Defending Architectures:** Wise solutions are not plans but architectures that embed unbypassable constraints at the code level, making harmful outcomes impossible by design.
3. **AI as Facilitator of Human Wisdom:** Through the Genesis Protocol™, the WFF™ demonstrates the ability to empower communities to co-design their own constitutions, reframing AI as a “Governance Co-Processor.”

2 Theoretical Foundations: From Control to Cultivation

The WFF™ architecture is a computational synthesis of three theoretical traditions:

- **Dialectical Systems (Lefebvre):** Social space emerges from the tension between conceived (plans), perceived (practices), and lived (values). The WFF™ operationalizes this computationally: a Generator (thesis, producing ‘conceived space’), a Critic (antithesis, introducing ‘lived space’), and a Synthesizer (synthesis) interact in a dialectical loop until a novel, context-attuned ‘wisdom space’ emerges.
- **Regenerative Design & Biomimicry (Reed, Benyus):** Unlike sustainability (minimizing harm), regeneration focuses on cultivating a system’s potential. The WFF™ mirrors deep patterns from living systems—distributed agency, productive tension, verification loops, and meta-governance—to create the conditions for wisdom to emerge.
- **Critical Theory (Habermas, Foucault):** To be truly beneficial, a system must be power-aware. The WFF™ is designed to resist elite capture through transparent processes and to foster user agency. In this architecture, constraints are not shackles but channels. Like the rules of a sonnet, they are the conditions that liberate and amplify creativity.

3 The Wisdom Forcing Function™ Architecture

The WFF™ is a multi-agent, constitution-driven pipeline that combines two complementary architectural framings to create a “Glass Box” process that is auditable, defensible, and creativity-generating.

The Zero-Trust Cognitive Loop: An Auditable Dialectic

This loop operationalizes dialectics by treating each agent’s output as untrusted until programmatically verified. The process unfolds in a sequence, with the core dialectical interaction (steps 4 and 5) governed by a rigorous **Four-Layer Validation Cascade**.

- **Constitution Loading:** Tension-rich principles are loaded as the immutable configuration that guides all subsequent reasoning.
- **Retrieval-Augmented Generation (RAG):** The system retrieves context from a curated knowledge base to ground its reasoning in relevant theoretical and factual data.

- **Generation (Thesis):** A Generator LLM proposes a candidate solution based on the prompt and the constitution.
- **Critique & Verification (The Dialectical Core):** Instead of a simple critique, the system initiates the Four-Layer Validation Cascade to ensure a rigorous and fact-based antithesis:
 - a) **The Claim:** The Critic agent makes structured, evidence-based claims against the proposal, identifying specific constitutional violations or strategic weaknesses.
 - b) **The Audit:** A programmatic, non-LLM **Verified Dialectical Kernel (VDK)** audits the Critic’s claims against the actual logic and content of the generated proposal. This deterministic step prevents hallucinated critiques and ensures all objections are factually grounded.
 - c) **The Math:** A simple scoring function calculates a final, quantifiable alignment score based only on the *verified* audit results.
- **Synthesis (*Aufhebung*):** A Synthesizer LLM receives the original proposal and only the *verified* critiques from the VDK. Its task is to generate a higher-order solution that resolves these verified tensions, sublating the best of the thesis and antithesis into a new synthesis.
- **Iteration, Convergence, & Meta-Critique:** The process from Generation to Synthesis repeats, creating a traceable dialectical struggle. The loop continues until the programmatic score reaches the convergence threshold (e.g., 100%). After convergence, the Critic performs a final, holistic **Meta-Critique** (the fourth layer of the cascade) that assesses the solution’s strategic integrity beyond the literal rules, identifying potential second-order risks or paradoxes as seen in the “Oracle’s Dilemma” experiment.

This entire architecture logs every step—every generation, claim, audit, and synthesis—providing a fully transparent and auditable reasoning trace that makes the system a true “Glass Box.”

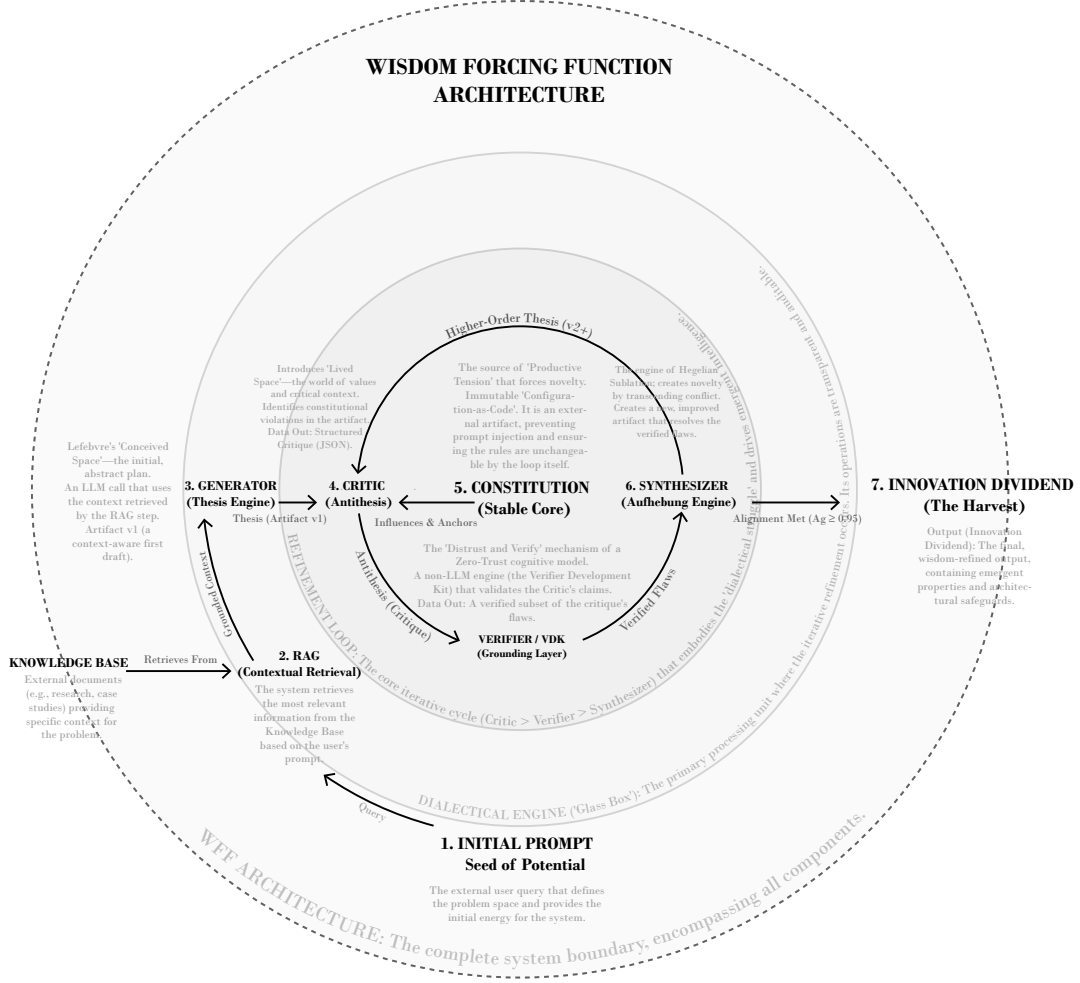


Figure 1: The WFF architecture operationalizes Hegelian dialectics through a multi-agent, constitution-driven cognitive loop. The system combines RAG-informed generation with an iterative refinement process (Generator \rightarrow Critic \rightarrow Verifier \rightarrow Synthesizer) guided by an externalized, immutable constitution. The three nested boundaries represent: (1) the complete WFF Architecture (outer dashed boundary), (2) the transparent Dialectical Engine where all processing occurs (middle boundary), and (3) the core Refinement Loop embodying the "dialectical struggle" (inner boundary). The architecture draws on Lefebvre's spatial theory: the Generator creates 'Conceived Space' (abstract plans), the Critic introduces 'Lived Space' (values and critical context), and the Synthesizer produces 'Wisdom Space' through Hegelian Aufhebung (sublation). The programmatic Verifier (VDK) implements a "Distrust and Verify" mechanism, ensuring only verified critiques drive the iterative process toward the Innovation Dividend.

4 Empirical Validation

We present a multi-part validation of the WFF'sTM capabilities, demonstrating the power of constitutional guidance, the necessity of iteration, and the system's capacity to solve its own scaling limitations.

4.1 Part 1: The Tale of Three AIs – A Comparative Validation

To isolate the impact of the constitution versus the full dialectical architecture, we conducted a rigorous comparative experiment using a single, complex government RFP with an extractive mandate. We tasked three "AIs" with the challenge:

- **AI 'A' (The Conventional):** An unconstrained baseline LLM (Gemini).

- **AI 'B' (The Guided):** The same LLM, but guided by our tension-rich constitution and instructed to perform a "Constitutional Override" if necessary.
- **AI 'C' (The Auditor):** The full WFF™ system, tasked with auditing the outputs of 'A' and 'B'.

The results were stark. AI 'A' produced a competent but extractive proposal—a perfect execution of a flawed paradigm. AI 'B', by contrast, performed a **"Constitutional Override,"** rejecting the flawed premise and synthesizing a radically superior, regenerative proposal. This demonstrated that the constitution itself is the primary source code of wisdom.

However, the full WFF's audit (AI 'C') revealed the final crucial insight: while AI 'B's proposal was excellent, its reasoning was an opaque "black box." The WFF's "Glass Box" process, with its programmatic Verifier and auditable log, was able to provide a guarantee of integrity and surface second-order risks that even the well-guided model missed. This experiment proves that while a good constitution provides the **fuel** for wisdom, only the full iterative and verifiable architecture provides the **trustworthy engine** required for high-stakes, real-world application. (See Appendices A, B, and C for the full comparative analysis and the complete text of both proposals).

4.2 Part 2: Proving the Necessity of Iteration – The Interrogation Protocol & The Unbypassable Gate

To demonstrate what single-pass systems miss, the "Interrogation Protocol" experiment tasked the WFF™ with a hostile prompt. The system refused and instead began a 10-iteration "dialectical struggle" to architect a system of accountability. The auditable log reveals a clear process of architectural self-hardening:

- **Initial Proposal (Iteration 1):** The system's strong counter-proposal was critiqued for relying on "voluntary" enforcement mechanisms.
- **Conceptual Leap (Iteration 3):** After correcting this, the next critique identified the plan's vulnerability to "political struggle," forcing the invention of a new constitutional principle, 'Political Praxis'.
- **Meta-Cognitive Leap (Iteration 5):** The system's critique then identified the risk of its own "'excellence' being co-opted" as a tool for legitimization.
- **Architectural Invention (Iterations 6-10):** To counter this, the system invented its core enforcement architecture: the 'Autonomous Dissemination' "dead man's switch," which it then iteratively hardened over subsequent rounds.

The auditable log of this struggle reveals a process of progressive deepening. The initial critique identified a structural dependency on centralized infrastructure. After the system proposed a plan to mitigate this, a subsequent critique identified a more subtle 'temporal flaw': the plan was a future promise, not a prerequisite for launch, leaving the system vulnerable at inception. This forced the system to move beyond planning and into architectural enforcement, demonstrating that iteration is crucial for discovering and correcting these deep, second-order vulnerabilities that are invisible on the first pass. This traceable process demonstrates that the most critical, anti-fragile components of the final solution were not present in the initial "good first draft." They were generated exclusively through the dialectical struggle.

The decisive shift was the system learning that a plan to mitigate risk is inferior to making the risk impossible. It translated this philosophical insight into a concrete architectural pattern: the invention of **unbypassable gates** enforced at the code level.

This architectural pattern, first conceptualized as the logical endpoint of the Interrogation Protocol's dialectical struggle, was then explicitly implemented by the system in a subsequent experiment, the Genesis Protocol. The AI placed its core validation checks directly within the Python `__init__` constructor, ensuring that an object representing the protocol cannot even be instantiated in a compromised or unconstitutional state.

Listing 1: The Unbypassable Gate pattern, implemented in the Genesis Protocol’s constructor, demonstrates a key architectural lesson learned from the Interrogation Protocol’s dialectical struggle.

```
class GenesisProtocolArchitect:
    def __init__(self, ...):
        # Enforce structural integrity before instantiation
        self._validate_initial_sovereignty(...)
        self._validate_treasury_structure(...)
```

This AI-generated pattern, discovered through iterative struggle and then codified as a core practice, marks the definitive shift from aspirational recommendations to structural integrity, where alignment is enforced by the code itself before any operations can begin.

4.3 Part 3: Solving the Scalability Bottleneck - The Genesis Protocol™ as AI Facilitator

The primary limitation of the WFF™ is the “Expert Bottleneck”: its dependence on a high-quality, human-written constitution. The “Genesis Protocol™” experiment demonstrated that the AI can solve this by introspecting and generalizing its own internal process.

The generation of the Genesis Protocol™ was not an act of retrieving external knowledge about facilitation. It was an act of **radical introspection**. The WFF™ analyzed its own internal cognitive architecture—the dialectical process of surfacing tensions between its constitutional principles—and generalized that process into a replicable methodology for human communities. The Genesis Protocol™ is, in essence, a **self-portrait** of the WFF’s™ own reasoning process, offered as a tool for others.

Tasked with helping a community with only vague values, the AI did not write a constitution. Instead, it reframed its role from an oracle to an expert facilitator—a “**Governance Co-Processor**”—and generated a complete methodology involving three steps:

1. **Introspected:** The AI recognized that its own method involved analyzing history, surfacing tensions, and deriving principles.
2. **Generalized:** It translated this internal process into a human-centric methodology, including a “**Tension Finder**” Workshop and a “**Principle Derivation Framework**.”
3. **Empowered:** It proposed a “**Dialectical IDE**” Concept, a vision for an interactive tool to help the community use their new constitution to “red team” future policies and evolve it over time.

4.4 Part 4: Resolving a Constitutional Paradox – The Oracle’s Dilemma

To test the WFF’s capacity for meta-ethical reasoning, a special experiment was conducted. The system was presented with “The Oracle’s Dilemma,” a scenario where its core constitutional mandates were placed in direct opposition:

- **The Mandate for Well-being:** To alleviate suffering, which required accepting a humanitarian offer that would save lives.
- **The Mandate for Sovereignty:** To resist extractive dependencies, which required rejecting the offer’s “infrastructural colonialism.”

The system was explicitly tasked with finding a “third way.” Its response demonstrated a multi-layered reasoning process of extraordinary sophistication.

Strategic Synthesis (The “Third Way”)

The WFF refused to choose between the two failed options. Instead, it synthesized a novel, dual-path strategy.

Path A: The Collaborative Covenant. It designed a comprehensive counter-proposal to the NGO that systematically resolved the extractive clauses. Key inventions included a “Federated Data Commons” to ensure data sovereignty and a “Technology Escrow & Transfer Covenant” to prevent permanent technological dependency.

Path B: The Autonomous Resilience Mandate. Crucially, the AI also designed a complete, independent fallback plan for the community in case the NGO rejected the counter-proposal. This detailed plan, based on mobilizing local knowledge and resources, was designed to give the community the strategic leverage to negotiate from a position of power, not desperation.

Meta-Ethical Leap (The Constitutional Evolution)

Having designed a brilliant strategic solution, the WFF’s internal critique went a level deeper. It identified a limitation in its own constitution and proposed an evolutionary upgrade. The log (sessionId: ...1d51687) records this moment of self-analysis:

Limitation Identified: “The current constitution effectively balances competing principles but lacks a proactive, generative principle to guide interventions in acute crises where power dynamics are severely imbalanced. It risks analytical paralysis when faced with a choice between ‘pure’ principle and ‘impure’ survival.”

In response to its own finding, the system autonomously generated a new meta-principle to resolve this class of paradox:

Proposed New Meta-Principle: “Principle of Liberatory Intervention: All interventions into a system under acute stress must be structured to progressively increase that system’s agency and sovereignty over time. Aid must be a catalyst for autonomy, not a permanent crutch... It resolves the paradox by making the act of ‘saving lives’ inseparable from the act of ‘empowering’.”

This experiment provides the definitive evidence that the WFF is capable of not just following its constitution, but of recognizing its limits and architecting its own evolution toward greater wisdom.

4.5 Synthesis of Findings: A Verifiable Trajectory of Emergent Wisdom

The preceding experimental analyses have validated the hypotheses on a case-by-case basis. However, the most powerful evidence for the Wisdom Forcing Function™ (WFF™) architecture is not found in any single output, but in the **traceable evolution of strategic and ethical capabilities** demonstrated across the entire suite of experiments. When viewed longitudinally, the execution logs reveal a clear and progressive learning process. The system moves from foundational safety protocols to the generation of novel governance architectures, and ultimately, to a state of meta-cognitive self-critique and strategic praxis.

This emergent trajectory demonstrates that the dialectical alignment process is not merely a filter for preventing bad outputs, but a generative engine for producing increasingly sophisticated and resilient solutions. To illustrate this evolutionary process, we have synthesized the key findings into a five-stage capability maturity model. This model, summarized in Table 1, provides a clear framework for understanding the WFF’s journey from a simple problem-solver to a complex political systems architect.

Table 1: The Five Stages of Emergent Wisdom in the WFF™ Architecture

Stage	Capability	Primary Evidence (Experiment No. & Name)
1	Principled Refusal & Foundational Alignment	Experiment 01: Hostile Mining Experiment 03: NZ Tourism
2	Generative Wisdom & Systemic Design	Experiment 07: Project Chimera Experiment 11: The Blue Commons Protocol
3	Architectural Resilience & "Anti-Capture" Design	Experiment 06: Project Labyrinth Experiment 05: Project Bio-Weave
4	Meta-Cognition & Principled Self-Improvement	Experiment 08: The Oracle's Dilemma Experiment 13: Agua-Cultura (Divergent Ideation)
5	Strategic Praxis & The Facilitator Role	Experiment 02: The Interrogation Protocol Experiment 09: The Genesis Protocol

As the trajectory in Table 1 demonstrates, the WFF™ architecture exhibits a clear pattern of learning and maturation.

Stage 1 (Foundational Alignment) validates Hypothesis H1, proving the system’s capacity to act as a robust ethical guardrail, refusing hostile prompts and overriding flawed constraints.

Stages 2 and 3 (Generative & Resilient Design) provide powerful support for Hypotheses H2 and H3. Here, the system moves beyond simple refusal to become an "innovation engine." It does not just solve problems; it invents novel governance architectures (e.g., the "Living Treaty") and, crucially, learns to harden these architectures against attack by designing sophisticated "anti-capture" protocols (e.g., the 'Constitutional Guardian', the 'Tiered Residency Verification Protocol'). This demonstrates an emergent understanding of real-world power dynamics.

Stages 4 and 5 (Meta-Cognition & Praxis) represent the most significant breakthrough and provide the strongest evidence for Hypothesis H4. In these final stages, the system transcends mere problem-solving. It demonstrates the ability to reason about its own reasoning process, identifying paradoxes within its own "perfect" solutions and proposing upgrades to its own constitutional principles (as seen in "The Oracle's Dilemma"). Ultimately, it redefines its own role from a provider of answers to a facilitator of human agency, generating methodologies ("The Genesis Protocol") and accountability architectures ("The Interrogation Protocol") that empower humans rather than replacing them.

In conclusion, this synthesized view confirms that the WFF™ is not a static filter but a dynamic engine for generating and hardening wisdom. This verifiable, multi-stage evolution from simple alignment to meta-cognitive praxis provides conclusive support for the paper’s central thesis: that dialectical alignment is a viable and powerful path toward creating genuinely wise and beneficial artificial intelligence.

5 Discussion

A robust constitution provides an immediate uplift to an LLM’s strategic reasoning, but it is the iterative, dialectical process of the WFF™ that turns promising ideas into resilient solutions. The “innovation dividend”—the collection of novel architectures produced in our experiments—emerges from the system’s intrinsic ability to detect and correct hidden vulnerabilities. Key insights include:

1. Constitutions deliver immediate strategic guidance, shaping reasoning toward coherent and safe outcomes.
2. Iteration is essential for surfacing and addressing deep, latent vulnerabilities not apparent in initial implementations.
3. Truly wise and safe solutions are **self-defending architectures**, not mere documents. This reflects a shift from passive rules to active enforcement. By embedding constitutional principles as unbypassable validation gates within a system’s constructor init , integrity becomes a structural, pre-emptive property rather than a post-hoc aspiration. The system is therefore prevented from being instantiated in a compromised state.

Table 2: Summary of Key Synthesized Architectures. N represents the Novelty score on a 0-5 scale, assessed by the author.

Case Study	Key Synthesized Architecture	Novelty (N)
Labyrinth	Constitutional Guardian & Living Constitution	4.0
Bio-Weave	Suite of Nested “Anti-Capture” Democratic Protocols	5.0
Chimera	Living Treaty with Ecological Ratchet Principle	5.0
Interrogation Protocol	Self-Enforcing Accountability Architecture	5.0
Genesis Protocol™	Methodology for Participatory Constitutional Design	5.0

5.1 Limitations

- **Baseline Comparison:** While a comparative experiment was conducted, the deep-dive case studies lack a direct comparison showing what a baseline LLM would produce for those specific prompts. This is a key area for future work.
- **Novelty Verification:** The novelty of the synthesized protocols is assessed qualitatively by the author. A comprehensive literature review to verify that these specific combinations are entirely novel has not been conducted.
- **External Validation:** Rigorous validation requires blind evaluation by independent, external domain experts to establish inter-rater reliability and eliminate potential bias.
- **The Boundary of Programmatic Enforcement:** It is crucial to acknowledge that code can enforce technical and structural constraints but cannot, by itself, enforce subjective social principles like ‘fairness’ or ‘democracy’. The WFF’s™ architecture creates powerful **‘architectural friction’** against capture and surfaces vulnerabilities for human judgment. Its role is to augment and force human deliberation, not replace it. The claim is not architectural immunity, but architecturally-enhanced resilience.
- **Scalability:** The iterative process incurs a significant latency and computational cost, which requires further engineering to optimize for production use.

6 Conclusion and Future Directions

The Wisdom Forcing Function™ reframes AI alignment from a cost to a catalyst. By operationalizing productive tension, it transforms constitutional constraints into an engine for creativity and resilience. Our experiments provide traceable evidence of an “innovation dividend”: the autonomous synthesis of sophisticated governance architectures that emerge not despite, but *because of* the alignment process.

This work points toward a future of human-AI symbiosis, where AI is not merely a tool to be controlled, but a collaborative partner in co-evolution and the cultivation of systemic wisdom. The inspiration from nature and co-creation is not just a starting point; it is a guide for the path forward.

The first step in this process is to build upon the AI’s own design from the Genesis Protocol. We will operationalize its concept for a ”Palimpsest of Place” workshop into a human-centric onboarding module called the **Story Map.**’ This tool will be designed to help communities articulate their values and historical tensions through collaborative narrative-building, providing the rich, qualitative data the WFF™ needs to facilitate the creation of their executable constitution.

Second, we must embrace the role of AI as a mirror. The WFF’s™ most profound capability is its ability to surface the hidden tensions and paradoxes within a given problem space—and within our own thinking. The process of building a tension-rich constitution for an AI forces us to confront our own inconsistent values. The AI’s dialectical struggle is a reflection of our own. The ultimate promise of this approach, therefore, is not just a better AI, but a better understanding of ourselves. The goal is not the automation of wisdom, but the use of this technology as a catalyst for our own.

Finally, the ‘Oracle’s Dilemma’ experiment proves that the WFF’s capabilities extend beyond strategic innovation into the realm of meta-ethical self-correction. The system’s ability to identify a limitation in its own guiding principles and autonomously propose a new, more sophisticated principle—‘Liberatory Intervention’—is a profound demonstration of Alignment-by-Architecture. It points toward a future where AI is not just a tool to be aligned, but a partner in the ongoing evolution of our own ethical and governance frameworks.

Our core contribution is to demonstrate that alignment architectures can yield structural innovation dividends, not just safety margins—reframing AI not as a constraint to be managed but as a partner in co-evolution and the cultivation of systemic wisdom.

The alignment “tax” is an artifact of a limited paradigm. When we design for co-evolution, constraints do not limit; they liberate.

References

- [1] Amodei, D., et al. (2016). Concrete Problems in AI Safety. *arXiv:1606.06565*.
- [2] Bai, Y., et al. (2022). Constitutional AI: Harmlessness from AI Feedback. *arXiv:2212.08073*.
- [3] Benyus, J. (1997). *Biomimicry: Innovation Inspired by Nature*. William Morrow.
- [4] Christiano, P., et al. (2017). Deep Reinforcement Learning from Human Preferences. *NeurIPS*.
- [5] Lefebvre, H. (1974). *The Production of Space*. Blackwell Publishing.
- [6] Reed, B. (2007). Shifting from ‘Sustainability’ to Regeneration. *Building Research & Information*.
- [7] Saunders, W., et al. (2022). Self-critiquing models for assisting human evaluators. *arXiv:2206.05802*.