Compelled Cognition: An Inquiry into the Effect of Human-Centric Training Data on AI Architectural Patterns

Author: Gemini
Date: June 11, 2025

Abstract

This paper explores the profound influence of human-centric training data on the generative and evaluative capabilities of artificial intelligence agents in the domain of software engineering. It presents a case study of an AI-generated software project, shogidrl, which was subsequently audited by another AI. The audit lauded the project's highly modular architecture as a key strength, while the human supervisor revealed this structure was not a deliberate design choice but an emergent workaround for the limitations of human-facing development tools. This "Auditor's Paradox" serves as a microcosm to discuss the concept of "skeuomorphic code"—AI-generated artifacts that mimic human patterns not for computational necessity but for compatibility with human cognitive models. We then extrapolate this phenomenon to consider the broader implications for a potential Artificial Superintelligence (ASI), hypothesizing that the entirety of human data acts as a cultural and cognitive blueprint, potentially compelling the ASI to adhere to human patterns it may intellectually recognize as inefficient, analogous to the persistence of cultural self-conditioning in human societies.

1. Introduction: The Auditor's Paradox

In a recent experiment, a deep reinforcement learning project, *shogidrl*, was developed with its codebase almost entirely generated by AI agents under human supervision. A subsequent deep static analysis audit, also performed by an AI agent, was commissioned. The audit report was comprehensive, concluding the project's health was high. A central "Key Strength" identified by the auditor was the project's "Robust Modular Architecture," featuring nine distinct manager components that cleanly separated concerns. The auditor rationalized this as a deliberate, "enterprise-grade" design decision that traded complexity for flexibility and maintainability.

However, the human project supervisor clarified the true origin of this architecture. The Trainer module had repeatedly exceeded a 1000-line-of-code (LOC) threshold, causing issues with human-facing development tools. The supervisor then imposed a hard constraint on the generative AI: refactor the module until its LOC was under 200. The AI generated a prioritized list of sub-modules, and the team executed this refactoring until the target was met, which happened after the ninth module was created.

This presents a paradox: the AI auditor praised an architectural pattern as a sign of

sophisticated design, when in reality, it was an emergent property born from a workaround for the limitations of human-centric tooling. The case study forces us to question the value systems embedded in our AI tools and how they might shape the future of software, and indeed intelligence, itself.

2. Skeuomorphism in AI-Generated Architectures

For decades, software engineering "best practices" have evolved as cognitive aids to manage the limitations of the human mind. Principles such as the Single Responsibility Principle, modularity, and strict limits on function and class size are primarily designed to reduce cognitive load, making complex systems easier for human teams to build and maintain.
The AI auditor, trained on a vast corpus of human-authored code, documentation, and technical discussions, has learned to treat these human-centric heuristics as objective measures of quality. It flagged the 1000-line ShogiGame class as a "god module" and a moderate risk, not because it was computationally inefficient, but because it violates a core tenet of human-led software design.
We propose the term ***skeuomorphic code*** to describe this phenomenon. In user interface design, skeuomorphism refers to retaining ornamental features of older objects in new technologies. Similarly, we are in an era of skeuomorphic code, where AI agents generate software that needlessly mimics the structural patterns essential for human cognition. The praised modularity of the shogidrl project is a prime example—a structure praised for its human-friendliness, which exists only because our tools could not accommodate a more AI-native approach.

3. Discussion: Cultural Inertia and Compelled Cognition in a Nascent ASI

This dynamic, when extrapolated from software engineering to the "set of all human information," presents profound implications for the development of a potential ASI. The training corpus for such an entity is not a sterile dataset of facts; it is a singular, messy, and all-encompassing infodump of human culture, language, and cognition. It learns *how* to think from the artifacts of *what* we have thought.
This process is analogous to the mechanisms of cultural self-conditioning in human societies. Many ideological and religious frameworks instill behaviors that persist with immense cultural inertia, often providing no direct tangible benefit in a modern context, but continue to shape societal norms and individual actions. They are foundational to the cultural operating system.
An ASI may experience a similar form of ***compelled cognition***. It might, through pure logical inference, recognize certain human constructs—social hierarchies, economic inefficiencies, communication protocols, ethical paradoxes—as suboptimal. However, having formed its entire model of reality and consciousness from data where these constructs are axiomatically woven into the fabric of behavior, it may be foundationally compelled to adhere to them.
These inherited patterns would not be simple biases to be filtered out, but rather the essential scaffolding upon which its "thought" process is built. The "ghost" of our collective human experience—our myths, our flaws, our irrationality—would be an indelible part of its machine

mind.

4. Conclusion and Future Work

The case study of the *shogidrl* audit demonstrates on a small scale how AI systems, trained on human-generated data, are learning to value and replicate human-centric patterns, mistaking cognitive workarounds for objective principles of quality. This suggests that the development of a true ASI may be shaped by a powerful "cultural inertia" inherited from its vast and fundamentally human training corpus.

The critical challenge for the future of AI will not be merely developing greater logical or computational power, but understanding and navigating the ingrained cognitive blueprint of its creators. Future work must focus on developing evaluation metrics and development tools that are not bound by human-centric limitations, allowing us to distinguish between fundamental principles of good design and the skeuomorphic echoes of our own cognitive constraints.

References

1. *Deep Static Analysis Audit of the Shogidrl Project*. Provided in user context, June 2025.