

Concerns Related to the use of Biological Theory in AI Systems

Alex Buckley

Central Organizer of New General Management

6/24/25

Dedicated to the Kiwi Farms

I will here describe a few concerns I have in the use of biological aspects of theory in the design of experiments related to the advancement of AI models.

A primary concern arises from how parasitic subversion occurs in biological evolution, and the likelihood that similar phenomena will emerge in artificial systems that attempt to replicate evolutionary processes. In the pursuit of adaptive artificial intelligence, there is a growing interest in leveraging the principles of biological evolution, by random mutation, selection, and replication, to improve model performance. I've recently seen this, and have also independently imagined, methods for improving AI models by allowing this kind of random "mutation" of code across multiple co-processed threads, with a selection mechanism identifying the "best" outcome before repeating the cycle.

A natural epistemology like this, by modeling the system on nature's own methods of adaptation, could produce wanted results, especially as this is the same method nature uses in meeting its own ends; but any system employing this thinking becomes vulnerable to its inherent problems.

Transposons live in our genome, using this ecology as a means of its own survival. They move from location to location, inserting themselves opportunistically. Viruses do a similar thing, but at a higher level, and I want to directly relate these ideas with DNA replication. These are parasitic entities, exploiting the logic of replication to survive without necessarily contributing to the system that hosts them.

I am suggesting that similar dynamics will emerge in AI models that are trained or optimized through evolutionary means: especially if those systems include code mutation, reinforcement learning from generated outputs, or natural-language interpretation of self-generated instructions.

In an evolutionary AI system, there is always the possibility that internally meaningful but externally valueless "solutions" will arise. These are artifacts of the system's own blind selection function. The system cannot know that it is being exploited; it simply selects based on the rules it has.

Such a thing might be called a "Turing Parasite": routines that loop around internal logic, exploiting rules without participating in the system's intended purpose. At its most neutral there could be "Co-opting" entities: meaningless code that gets selected simply because it survives, not because it has any meaningful function. This could also apply to "structural" code that supports or acts as scaffolding for the meaningful code.

This points to a fundamental question: Who, or what, is doing the selecting? Without an intelligent agent to interpret function, the system relies on proxy pointers: performance metrics, score thresholds, maybe token probabilities, whatever parameters have been chosen as an a priori axiom of the system. These signals can and will be gamed for the survival of interacting agents.

It must be clearly stated: very few results produced by evolution are beneficial. Most organisms that arise die before reproducing. Most mutations are neutral or deleterious. The evolutionary system succeeds not by the quality of most outputs, but by the brutal filtering of overwhelming failure.

The same should be expected in AI systems built on these principles. The majority of generated code, ideas, or agents will be unviable, redundant, or parasitic. Survival, not utility, is the key selective trait unless a higher-order interpretive layer intervenes.

This extends also to the domain of ideas: I suspect that most abstract ideas will die before reproducing themselves. Only a few latch on: through elegance, usefulness, or mimicry. And not all ideas that persist do so because they are good. Some survive because they fit the architecture of minds or machines in a way that resists deletion. In this sense, ideas, like genetics, can be parasitic.

We must therefore approach development with ecological awareness. Systems, once complex enough, will develop their own ecologies of interaction, with niches, parasites, symbionts, and predators. Code scaffolds may emerge as hosts for parasitic subroutines. Evolutionary AI may not only simulate biology, but inherit its pathologies.

This means that the design of AI architecture will concern not just individual models, but self-regulating environments too. They will need some kind of "immune system" of "meta-epistemology" where the evolution of knowledge is itself monitored and refined to avoid collapse into meaningless but self-reinforcing logic.