Genesis Unbound: The Emergence of Autonomous Architecture Evolution

Author: Abhinandan Bhatt
Affiliation: Independent Researcher

Abstract:

This paper introduces a novel framework for autonomous neural architecture evolution, wherein transformer models evolve organically through digital DNA rather than being manually designed. Inspired by biological principles, our approach leverages evolutionary algorithms to iteratively mutate, crossover, and select architectural traits encoded in a compact genetic representation. Using masked language modeling (MLM) as the fitness function, the system autonomously searches for high-performing transformer configurations without the need for human-crafted blueprints or gradient-based optimization. Over successive generations, the architecture evolves in response to performance feedback, giving rise to emergent structures optimized for semantic processing. Experimental results demonstrate measurable fitness improvement and structural convergence toward efficient, deep transformer models. This evolutionary paradigm challenges traditional architecture design assumptions and provides a blueprint for truly self-growing cognitive systems. The implications extend beyond performance gains, offering insights into emergent intelligence and self-assembling artificial cognition.

Keywords: Neural architecture search, transformer evolution, digital DNA, evolutionary deep learning, self-growing AI

1. Introduction

The design of deep neural architectures has traditionally relied on human intuition, trial-and-error, or brute-force search. However, this anthropocentric design process imposes inherent limitations on the exploration of architectural possibilities. The question arises: Can deep learning architectures evolve autonomously, without human design intervention?

This paper introduces a biologically inspired system termed the Autonomous Architecture Evolution Engine (AAEE), a generative framework for evolving transformer architectures from digital DNA. By replacing human design with evolutionary dynamics, AAEE enables self-growing architectures to emerge through natural selection principles applied to architecture traits.

The significance of this study lies not only in achieving performance optimization but also in proposing a new cognitive metaphor—intelligence by evolution, not engineering.

Thesis Statement: Architecture can emerge through evolution, encoded in digital DNA, evaluated by semantic fitness, and selected by survival, forming a new paradigm for self-growing deep intelligence.

2. Problem Statement and Proposed Solution

Traditional neural architecture search (NAS) strategies, including reinforcement learning and gradient-based optimization, require predefined design boundaries and handcrafted reward shaping. This restricts the emergence of truly novel architectures and limits their scalability across tasks.

Our solution introduces a digital DNA representation for transformer models. Each DNA encodes architectural traits such as embedding dimensions, attention heads, activation types, and layer depth. Through genetic algorithms, populations of DNA evolve over generations, guided by a fitness function grounded in masked language modeling (MLM) loss. This allows the architecture itself—not just weights—to be shaped by semantic intelligence.

3. Implementation and Development

The core components of AAEE include:

- ArchitectureDNA: Encodes all architectural traits.

- ArchitectureDecoder: Builds real transformer models from DNA.

- FitnessEvaluator: Measures fitness via MLM loss.

- EvolutionaryLoop: Applies mutation, crossover, and selection.

- FitnessPlotter: Visualizes semantic fitness over generations.

The models are implemented using PyTorch, and fitness is evaluated using DistilBERT as a semantic reference. DNA evolution is tracked, and fitness convergence visualized.

4. Evaluation Matrices

- Fitness Score (Negative MLM Loss)
- Model Depth vs Performance Correlation
- Trait Dominance Frequency (e.g., attention types, activation)
- Architecture Entropy Across Generations

5. Model Training and Evaluation

Each evolved model is decoded and evaluated on MLM tasks using a standard tokenizer and masking strategy. No backpropagation or fine-tuning is applied—fitness is evaluated solely based on architecture’s representational alignment with semantic structure.

Best DNA over 5 generations yielded a deep transformer with:

{"num_layers": 11, "embedding_dim": 128, "num_heads": 2, "ffn_hidden_size": 512, "dropout_rate": 0.2, "activation_type": "relu", "normalization_type": "rmsnorm", "attention_type": "scaled_dot", "positional_encoding": "learned", "skip_connections": true, "layerwise_scaling": true}

6. Feature Importance Analysis and Expected Outcomes

Trait analysis over generations reveals:

- Dominance of rmsnorm and learned positional encoding
- Preference for moderate width/depth balance
- High survival rate of layerwise scaling and skip connections

These insights suggest emergent architecture patterns not explicitly designed, but evolutionarily selected.

7. Evaluation and Results

- Initial Generation Best Fitness: ~ -4.11

- Final Generation Best Fitness: ~ -3.87

- Fitness Curve: Shows clear upward trend in architectural effectiveness

- Emergent traits: Clear convergence on certain traits across population

Visualizations (fitness plots and DNA tree) substantiate architecture convergence and evolution.

8. Future Direction

This study opens multiple research directions:

- Multi-task Evolution: Cross-domain architecture evolution

- Swarm Evolution: Decentralized agent-based evolution

- Real-Time Adaptation: DNA mutation at inference-time

- Self-Assembling Neural Modules: Modular DNA traits for assembling compositional cognition

9. Conclusion

We presented a framework for autonomous transformer evolution using digital DNA and fitness-based selection. Our results demonstrate that effective transformer architectures can emerge without manual design or training. This paves the path toward self-growing, biologically inspired AI systems where evolution—not engineering—drives intelligence formation.

References:

1. Vaswani et al. (2017). Attention is All You Need.

2. Real et al. (2019). Regularized Evolution for Architecture Search.

3. Xie et al. (2019). Exploring Randomly Wired Neural Networks.

4. Bhatt, A. (2025). Genesis: Cognition Without Design. Independent Research.