

Simplicity Scales

November 30, 2025

Waves flow, whereas trends remain. This plain principle generalizes well in the era of AI, where preprints overflow, but only a few become time-tested. One common merit, as pointed out in The Bitter Lesson¹, of these masterpieces is the spirit of

simple algorithms that scale data with compute.

The core techniques behind the recent breakthroughs illustrate this precept well. Language models stack identical neural blocks that facilitate parallelizable pre-training on data at Internet scale. GRPO, the de-facto option for reinforced fine-tuning, unloads the value model and only steers the actor model with outcome reward. (Un)surprisingly, these recipes generalize to FSD, where scaling the same transformer architecture on video data yields smooth driving actions.

Simple algorithms are elegant in aesthetics. It is the taste embedded in the latent space.²

Simplicity is deliberate reasoning made precise through intuition. Essentially, the complexity of simple algorithms is two-fold: the compression of the search space and the pruning of the dependencies.

Scaling with data is the way to learn without overloaded priors. The rationale has been validated by Moravec’s paradox, where neural nets usually excel in rule-based tasks like Go. In other words, it might be the only recipe for simple algorithms to learn by instinct.

In practice, simplicity is the pursuit of research (thinking), and scaling is the requirement of engineering (action).

Simple Plan is Perfect.

¹“General methods that leverage computation are ultimately the most effective”

²Ilya Sutskever on [Dwarkesh Podcast](#): “research taste: beauty, simplicity, elegance...”