

When AI learns how to learn: DiscoRL and the automation of RL algorithms

Reinforcement learning (RL) sits under a lot of the iconic AI wins of the last decade. But there's an irony in how it's usually built: the agent learns through experience, while the learning rule itself is still mostly handcrafted by humans. A new Nature paper argues that this split is becoming unnecessary — and demonstrates a system that can *discover* a high-performing RL update rule from experience at scale.

The headline result is bold but easy to appreciate: the discovered rule (called **DiscoRL**) outperformed existing learning rules on the classic Atari benchmark, and then held up well on other challenging benchmarks that weren't part of the discovery process. The deeper point is the direction of travel: if learning rules can be learned, algorithm design starts to look less like artisanal engineering and more like something that can compound with compute and diverse experience.

The idea in plain terms

Most RL algorithms differ in how they update an agent's policy and predictions after it takes actions and receives rewards. In this work, instead of choosing that update rule up-front, the researchers represent the rule as a trainable "meta-network" that outputs targets the agent should move toward — effectively learning *how* the agent should update itself.

They then run a population of agents across many environments, and continuously improve the meta-network so that the agents trained under it achieve higher returns. Over time, this process produces a learning rule that is competitive with, and in some cases better than, the manually designed rules the field has relied on.

Why this matters

If this approach scales, it changes the bottleneck. Instead of relying on slow human iteration to invent new update rules, you can search a much larger space of possible algorithms using experience and meta-optimization — and let the resulting rule generalize beyond the environments it was discovered on.

The paper also makes a practical point: the discovered rule improves as discovery uses more diverse and complex environments, which hints at an "algorithm scaling law" style dynamic — better rules as a function of experience diversity, not just model size. That's a big deal for anyone thinking about general-purpose agents.

What stood out (without the math)

A few pieces are worth calling out even without diving into technical details. First, the authors report that DiscoRL surpassed prior approaches on Atari in their setup, and that a variant discovered on a larger and more diverse environment set improved performance on multiple other benchmarks.

Second, their analysis suggests the learned rule develops its own useful internal predictions that don't map cleanly onto standard RL concepts like "value functions," and that these learned predictions end up informing the policy update rather than staying as a side quest. In other words: it's not just rediscovering the same tricks with different knobs.

The bigger takeaway

This is a glimpse of a future where AI systems don't just learn tasks — they also learn the learning machinery that makes them effective. That doesn't mean research becomes automatic, but it does suggest progress may shift from "invent a new rule" to "design a discovery process that reliably produces strong rules."

Contributor: Alessandro Linzi