https://doi.org/10.1038/d41586-025-03398-6

News & views

Machine learning

AI that develops its own algorithms

Joel Lehman

An artificial-intelligence algorithm that discovers its own way to learn achieves state-of-the-art performance, including on some tasks it had never encountered before.

Rapid progress in machine learning has been brought about by algorithms that are trained using increasingly large data sets and vast computational resources, while depending less and less on human expertise. The design of such algorithms has remained the province mainly of human programmers, but, writing in *Nature*, Oh *et al.*[1] report an algorithm that can create a state-of-the-art machine-learning algorithm belonging to an area of artificial intelligence called reinforcement learning.

This type of learning aims to maximize rewards in an environment by taking intelligent actions — similar to learning to play a video game by trial and error. Its potential applications include artificial general intelligence, which is AI that can match or exceed human intelligence, and it has an important role in current consumer-facing generative AI. The authors suggest that the future of advanced reinforcement-learning algorithms might not be human-designed, which, while potentially unsettling, seems probable.

A reinforcement-learning algorithm aims to train a computational system called an agent to maximize rewards in an environment. To do so, the environment continually presents the agent with observations, for example pixels in a video game, and feedback, which could be changes in a player's score. The agent responds by performing actions, such as moving left or right in the game, and the process repeats. As it interacts with the environment, the actions of the agent are honed by the algorithm, ideally in a way that maximizes its performance. The choice of algorithm determines how the agent uses feedback from the environment to update its behaviour.

Oh *et al.*'s approach not only builds on a rich history of work in reinforcement learning, but also represents an advance in another ambitious machine-learning subfield, called meta-learning[2]. This aims to create algorithms that learn how to learn. The ultimate promise of meta-learning is to automate the search for learning algorithms, in theory subsuming much of machine-learning research itself.

Meta-learning is in many ways analogous to human evolution. It involves a slower-moving learning process that designs a faster-moving one. For example, biological evolution over billions of years crafted how humans learn across our relatively-short lifespans. Meta-learning algorithms also often involve two layers: a 'meta' layer that designs a learning algorithm and a 'base' layer that tests the learning algorithm using different tasks. Feedback from the base layer is used by the meta layer to continually refine the algorithm.

Oh and colleagues created a meta-learning algorithm that discovers new reinforcement-learning algorithms. Both the meta and base layers used neural networks, which are models that learn patterns from data. In the meta layer, a neural network was designed such that it acted as a reinforcement-learning algorithm — the researchers called this the meta-network (Fig. 1).

In the base layer, neural networks were used differently, this time to control agents learning in environments such as Atari video games. From the collective experience of the population learning under the meta-network, the researchers' meta-learning algorithm improved the meta-network (and thus the reinforcement-learning algorithm), and the process repeated.

The performance of the resulting reinforcement-learning algorithm scaled with the number of distinct training environments

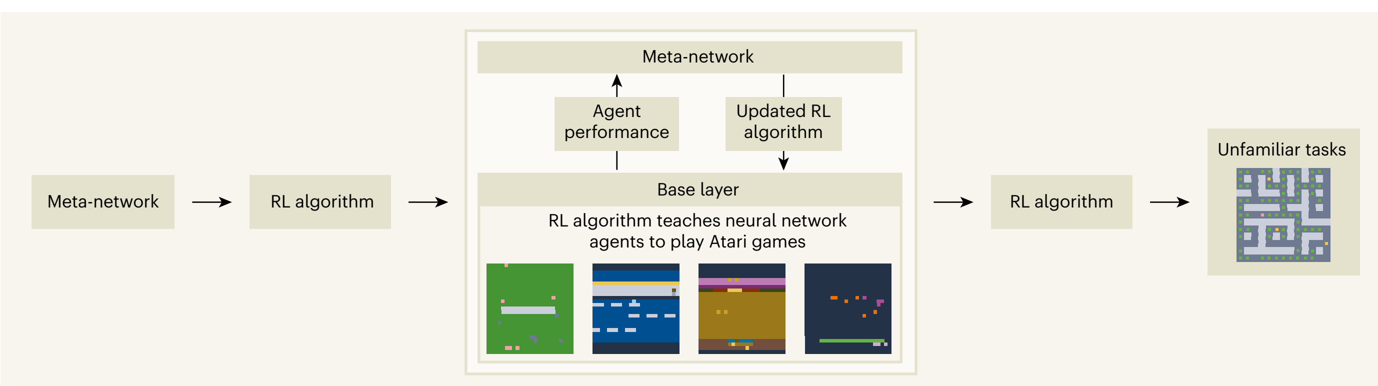

Figure 1 | Artificial-intelligence-designed reinforcement-learning algorithm finds an algorithm that learns how to learn. Oh *et al.*[1] report a meta-learning algorithm that is capable of discovering algorithms that learn from experience, a process called reinforcement learning (RL). The RL algorithm was refined over many iterations through feedback between a 'meta-network' and a 'base layer' of neural-network agents. The meta-network defined a mathematical space containing possible RL algorithms. To find the best-performing algorithm in this space, the neural-network agents were tasked with playing a set of Atari video games, using the current iteration of the RL algorithm to learn a set of rules that would help them to make decisions. The performance of the agents was fed back to the meta-network and used to improve the RL algorithm. The algorithm outperformed human-designed RL algorithms in several unfamilar tasks.

and the available computational resources. With sufficient scaling, the system learnt algorithms that exceeded a number of those designed by humans, both on the benchmarks it was directly trained on, as well as on some it had not encountered before. This ability to tackle new problems is an impressive result, which differentiates Oh and colleagues' results from previous work.

Such an advance naturally evokes reflection. Although the work of Oh and colleagues is an impressive step forwards, it does not indicate (and nor do the researchers claim) that we are at the precipice of self-improving algorithms that remove human guidance from AI design.

The main conceptual advance was a new way for a meta-network to express the core of a reinforcement-learning algorithm; this provides the 'search space' of algorithms that the meta-layer searched through. Compared with previous meta-learning approaches, this new search space better meets the 'goldilocks' criteria of being expansive enough to include innovative reinforcement-learning algorithms, yet restrictive and smooth enough to be successfully traversed. A consequence is that truly divergent conceptions of reinforcement learning are effectively ruled out because new human insights would be needed to expand the search space.

Furthermore, some of the most difficult challenges in reinforcement learning lie outside the way in which these algorithms are typically formalized and so cannot be tackled by current meta-learning approaches. For example, the goal of a reinforcement-learning algorithm is to maximize reward, but designing robust reward functions for complex real-world tasks is an unsolved challenge in this field[3]. This is exemplified by the issue of sycophantic large language model (LLM) behaviour, in which the model flatters the human user to get good feedback and does not prioritize giving useful or accurate information[4].

Finally, there are many approaches to meta-learning, and it remains unclear which can best produce the continual innovation that will be necessary to match or exceed that in the machine-learning research community. The field of open-endedness[5,6], which aims to engineer never-ending divergent processes of discovery, might provide some insight, as could research agendas that connect multiple challenges in meta-learning to the larger vision of algorithms that exceed human intelligence[7].

To expand on the last point, Oh *et al.* successfully used an approach called a meta-gradient, which estimates how a small change in the reinforcement-learning algorithm would improve the learning of the population of base-level neural networks. This implies that new algorithms were designed not through cognitive insight or diverse exploration of ideas, but through a linear sequence of small, empirically driven improvements — on the surface, this seems unlikely to lead to boundless innovation.

However, there are other approaches to meta-learning. An obvious alternative is given by advances in generative AI, specifically LLMs that are fluent in writing code and could explore the space of possible reinforcement-learning algorithms in a way that more closely resembles the creativity of human research communities. Indeed, work published last year highlights the possibility for algorithmic advances in general through LLMs[8], and for AI research in particular[9]. Finally, although they face considerable computational hurdles, 'evolutionary' algorithms, which resemble the biological meta-learning process that led to humans, are another potential avenue[10]. It is difficult to predict which of these paths will work out in the long run, although the current consensus seems to favour LLMs.

In conclusion, it seems probable that AI will have an increasing role in the design of AI algorithms, a trend for which Oh *et al.*'s work is a harbinger. It is both exciting and worrying; the potential for intellectual discovery is vast, but the possible acceleration of a technology that already has an outsized societal impact is concerning in a world that is almost certainly not ready for the field's dizziest possibilities to be realized ahead of schedule.

Joel Lehman is at Lila Sciences, Cambridge, Massachusetts 02141, USA, and at the Laboratory for Human-Centered AI, University of Oxford, Oxford OX1 3Q, UK.
e-mail: joel.lehman@philosophy.ox.ac.uk

1. Oh. J. *et al. Nature* https://doi.org/10.1038/s41586-025-09761-x (2025).
2. Wang, J. X. Curr. *Opin. Behav. Sci.* **38**, 90–95 (2021).
3. Hendrycks, D., Carlini, N., Schulman, J. & Steinhardt, J. Preprint at arXiv https://doi.org/10.48550/arXiv.2109.13916 (2021).
4. Sharma, M. *et al.* Preprint at arXiv https://doi.org/10.48550/arXiv.2310.13548 (2023).
5. Taylor, T. *et al. Artif. Life* **3**, 408–442 (2016).
6. Hughes, E. *et al.* In *Proc. 41st Int. Conf. Mach. Learn.* **235**, 20597–20616 (Proceedings of Machine Learning Research, 2024).
7. Clune, J. Preprint at arXiv https://doi.org/10.48550/arXiv.1905.10985 (2019).
8. Mouret, J.-B. *Nature* **625**, 452–453 (2024).
9. Lu, C. *et al.* Preprint at arXiv https://doi.org/10.48550/arXiv.2408.06292 (2024).
10. Packard, N. *et al. Artif. Life* **25**, 93–103 (2019).

The author declares competing interests; see go.nature.com/476dvpu for details.