2024 ACM A.M. Turing Award: Richard S. Sutton and Andrew G. Barto for Reinforcement Learning

Matjaž Gams
Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
E-mail: matjaz.gams@ijs.si

Editorial

Abstract: The 2024 ACM A.M. Turing Award (the "Nobel Prize of Computing") was awarded to Andrew G. Barto and Richard S. Sutton "for developing the conceptual and algorithmic foundations of reinforcement learning." Announced on 5 March 2025, the honor not only celebrates nearly five decades of pioneering scholarship but also signals that reinforcement learning (RL) has moved from the periphery of artificial-intelligence research to its very center —most visibly through its role in training large-language models (LLMs).

1 From unfashionable curiosity to mainstream core

In the early 1980s, when Barto and Sutton began formalising how agents learn from trial-and-error reward signals, prevailing AI paradigms favoured rule-based expert systems and supervised pattern recognition. Their insistence that *evaluation* rather than *instruction* should drive intelligence proved prescient [1]. Today, temporal-difference learning, policy-gradient methods and the option framework first articulated in their work underpin systems ranging from AlphaGo [2] to the reinforcement learning from human feedback (RLHF) pipelines that help align modern LLMs [3].

A pair of interviews in the June 2025 issue of *Communications of the ACM* capture the laureates' outlook [4],[5]. "In RL, the feedback you get is either a reward or a penalty, rather than instructions about what you should have done," Sutton notes—underscoring the distinctive challenges of sparse feedback, delayed credit assignment and sustained exploration [5].

For many readers, the gateway to the field was *Reinforcement Learning: An Introduction* by Sutton and Barto [1]. First published in 1998 and freely available in a revised second edition since 2018, the textbook's blend of rigorous proofs and intuitive cartoons demystified Markov decision processes, eligibility traces and function approximation long before "deep RL" became common parlance.

2 Explaining ML through games

An early hardware demonstration of machine learning was Marvin Minsky's SNARC (1951-52), an analog device that let a 'rat' learn a maze by reward signals, foreshadowing today's reinforcement-learning agents. Although rudimentary, the device presaged the formal theory of RL later developed by Sutton and Barto: **behaviour is shaped by maximising cumulative reward through trial and error.**

Modern chess engines embrace the same principle at planetary scale. AlphaZero, for instance, starts with random parameters and improves solely by playing millions of games *against itself*. Each iteration:

- Chooses moves with a combined policy-and-value neural network that estimates both the probability of promising actions and the expected game outcome.
- Guides exploration using Monte-Carlo tree search (MCTS) where network evaluations bias search toward fruitful branches.
- Learns by minimising a temporal-difference loss: the gap between the predicted value and the eventual result (win = +1, draw = 0, loss = –1).

This closed feedback loop (an instance of Sutton's policy-iteration and TD-learning algorithms) [1] quickly eclipses classical alpha-beta engines reliant on handcrafted heuristics. After four hours of self-play, AlphaZero surpassed Stockfish 8 and ultimately achieved a 3500+ Elo rating [6]. As Savage concludes, it has proven to be "a rewarding line of work" [4]. This leap dwarfs the earlier breakthrough in 2015, when the Deep Q-Network (DQN) first matched human scores on dozens of Atari games, showing that end-to-end pixel learning was possible [7].

3 Why the Turing award matters

1. **Conceptual Unity.** RL provides a single mathematical framework that spans robotics, operations research and behavioral neuroscience; dopaminergic prediction-error signals in primate brains can be modeled almost equation-for-equation by temporal-difference learning [1].

2. **Practical Impact**. RL is already saving megawatt-hours of electricity by autonomously optimising Google data-centre cooling loops—cutting energy use by up to 40 % [8]. The same

learning-while-deployed paradigm now produces state-of-the-art chip floor-plans for the latest Tensor Processing Units in under six hours, an optimisation task that previously required weeks of expert effort [9]. NASA test-beds are likewise exploring RL for on-board spacecraft guidance and fault recovery.

3. **Ethical Imperative**. Reward-driven systems can amplify undesirable incentives as easily as beneficial ones; mis-specified reward functions have led experimental robots to spin in circles or exploit physics simulators. Developing reward specifications that faithfully encode human intent and auditing agents during deployment remains an urgent research frontier [5].

Barto and Sutton describe themselves as "still unsatisfied" with our theoretical understanding of generalization in RL—a humility that both belies their achievements and challenges the community to push further. Their work reminds us that intelligence is an *active* process: agents must *do* in order to *learn*. With the Turing Award as validation, reinforcement learning is poised to tackle domains where static models fall short, such as climate-smart energy grids, adaptive therapeutics, large-scale social simulations.

4 Conclusions

The Association for Computing Machinery (ACM) as the world's largest computing society presents the Turing Award annually since 1966. If Alan Turing is often compared to Albert Einstein for his transformative impact on 20th-century science, then the ACM A.M. Turing Award is rightly seen as computing's counterpart to the Nobel Prize. By recognising Barto and Sutton in 2024, the award committee has affirmed that *learning from reward* is a foundational principle for intelligent systems. Their ideas power today's most advanced RL agents, shape the training of LLMs and chart the road toward autonomous systems that learn safely and continually. As they themselves like to remind us, **"the best is yet to come."**

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, Cambridge, MA, 2018.

[2] D. Silver *et al.*, "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, pp. 484–489, 2016.

[3] L. Ouyang *et al.*, "Training Language Models to Follow Instructions with Human Feedback," *Advances in Neural Information Processing Systems* 35, 2022.

[4] N. Savage, "A Rewarding Line of Work," *Communications of the ACM*, vol. 68, no. 6, pp. 8–10, 2025, DOI: 10.1145/3727966.

[5] L. Hoffmann, "Developing the Foundations of Reinforcement Learning," *Communications of the ACM*, vol. 68, no. 6, p. 96, 2025, DOI: 10.1145/3724079.

[6] D. Silver *et al.*, "A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.

[7] V. Mnih *et al.*, "Human-Level Control through Deep Reinforcement Learning," *Nature*, vol. 518, pp. 529–533, 2015.

[8] C. Evans and D. Gao, "DeepMind AI Reduces Google Data Centre Cooling Costs," DeepMind Blog, 2018.

[9] A. Mirhoseini *et al.*, "A Graph Placement Methodology for Fast Chip Design," *Nature*, vol. 594, pp. 207–212, 2021.