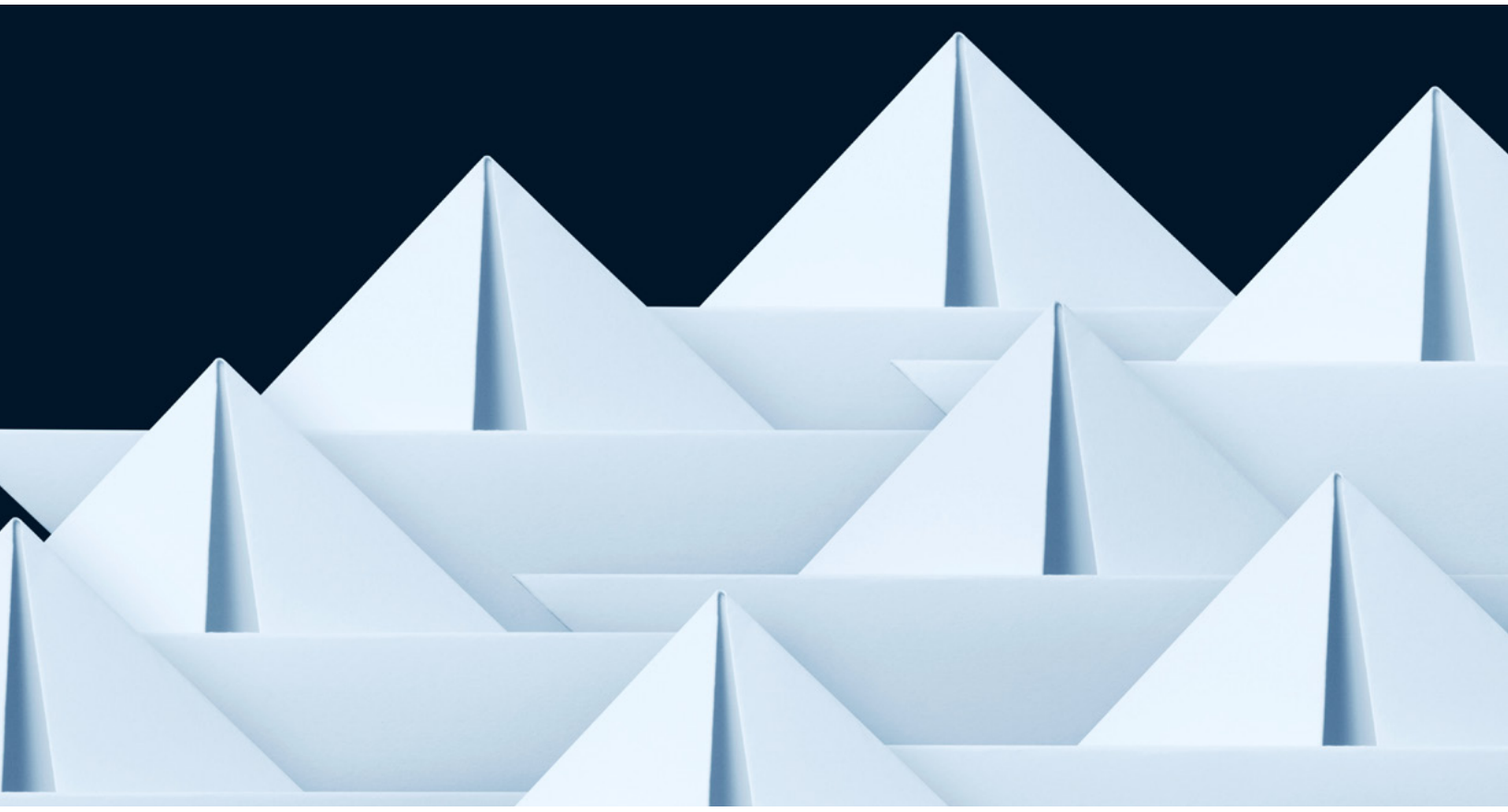
McKinsey Analytics

It's time for businesses to chart a course for reinforcement learning

An advanced artificial intelligence technique is quickly becoming accessible to organizations as a tool for speeding innovation and solving complex business problems.

by Jacomo Corbo, Oliver Fleming, and Nicolas Hohn



April 2021

Leaders looking for new ways artificial intelligence (AI) can provide a competitive edge may have found the 2021 America's Cup Match as exciting for one team's groundbreaking use of reinforcement learning as for its radical boat designs and close races.

To remain competitive, sailing teams in the America's Cup contest, like all businesses, must push the boundaries of what is possible. They also face similar constraints, including a steep development curve and a small window of opportunity, meaning teams can pursue only one or two big experiments to up their performance in the sport's most important competition.

For the 2021 edition of the America's Cup, reigning champion Emirates Team New Zealand ventured that reinforcement learning, an advanced AI technique, could optimize its design process. The technique delivered, enabling the team to test exponentially more boat designs and achieve a performance advantage that helped it secure its fourth Cup victory.

Unlike other types of machine learning, reinforcement learning uses algorithms (which often train AI agents or bots) that typically do not rely only on historical data sets, either labeled or unlabeled, to learn to make a prediction or perform a task. They learn as humans often do, through trial and error. In the last few years, the technology has matured in ways that make it highly scalable and able to optimize decision making in complex and dynamic environments.

Besides accelerating and improving design, reinforcement learning is increasingly being incorporated into a broad range of complex applications: recommending products in systems where customer behaviors and preferences change rapidly; time-series forecasting in highly dynamic conditions; solving complex logistics problems that combine packing, routing, and scheduling; and even accelerating clinical trials and impact analysis of economic and health policies on consumers and patients.

We have seen how quickly the technological environment can shift. Only a few years ago, another AI technique, deep learning, vaulted onto the business scene. Today, 30 percent of high-tech and telecom companies and 16 percent of companies in other industries we surveyed have embedded deep-learning capabilities.[1]

Executives who today understand the potential of reinforcement learning will, like Emirates Team New Zealand, be better positioned to find the edge in their industries (see sidebar "Notable examples of reinforcement learning applications"). Understanding Emirates Team New Zealand's experience can help leaders gauge where and when to use the technology because many organizations will travel a similar path: implementing more traditional technologies first to solve a problem and then applying reinforcement learning to ascend to a previously unattainable tier of performance. Thus, we begin by recounting Emirates Team New Zealand's journey, after which we offer ideas for where and how businesses should consider applying reinforcement learning.

Emirates Team New Zealand's journey to a 2021 victory

Emirates Team New Zealand designers were not new to advanced technologies. In 2010, the team had built its state-of-the-art digital simulator to test boat designs without physically building them. This was a key to the team's 2017 America's Cup win, but the simulator had limitations. Multiple sailors were needed to operate it optimally, which was a significant logistical challenge given the sailors' scheduled practices, travel, and competitions. As a result, designers typically iterated on new designs in the absence of simulator performance data and then tested their best ideas in batches when they could carve out large blocks of time with the sailors. Moreover, the sailors' performance could vary between tests, as human performance often does, making it difficult for designers to know whether a marginal improvement in boat response was due to a design tweak or to variances in human testing.

[1] "The state of AI in 2020," November 2020, McKinsey.com.

Notable examples of reinforcement learning applications

Here are some of the most talked-about applications of the technique in recent years:

Gaming: DeepMind's AlphaZero, its latest iteration of computer programs that play board games, learned to play three different games (Go, chess, and shogi) in less than 24 hours and went on to beat some of the world's best game-playing computer programs.

Retail: Amazon has used reinforcement learning to automate some warehouse fulfillment activities and power its fleet of autonomous drones that, once fully certified and deployed, will be able to bring packages to customers' doors in hours.

Social good: Salesforce in 2020 introduced the AI Economist, which enables policy makers and economists to identify optimal tax policies for different socioeconomic goals, such as creating jobs or addressing inequality.

Sports: Emirates Team New Zealand used reinforcement learning to test thousands of hydrofoil designs in their virtual sailing simulator instead of just hundreds, giving them a significant advantage in optimizing their boat's specs for the 2021 America's Cup, in which they successfully defended their title.

Automotive: Google, Tesla, Toyota, Uber, VW, and others are applying reinforcement learning in their efforts to build self-driving cars that are safe and trustworthy.

As Emirates Team New Zealand prepared for the 2021 match, they knew if they could get an AI system to run the simulator, it would free the designers to test more design ideas faster and more consistently than they could with the digital simulator alone. The team was unsure at the outset if the idea was feasible, but as conversations about the technology swirled, team members agreed: the potential payoff was transformative and made trying worthwhile. Using reinforcement learning, experts from Emirates Team New Zealand, McKinsey, and QuantumBlack (a McKinsey company) successfully trained an AI agent to sail the boat in the simulator (see sidebar "Teaching an AI agent to sail" for details on how they did it).

While design rules for the America's Cup specify most components of the boat, they leave enough freedom for designers to make radical choices on some key elements such as hydrofoils. These wing-like structures attach to the hull and lift the boat above the water, enabling the vessel to reach speeds of over 50 knots (60 miles or 100 kilometers per hour). Hydrofoils can be a

Teaching an AI agent to sail

To sail as well as the world's best sailors, the AI agent needed to learn to execute different maneuvers in varying conditions, choosing the best course to set under a wide variety of winds and seas, adjusting 14 different boat controls accordingly, assessing the results of its decisions, and continually improving decisions over long time horizons. Subject-matter experts and data scientists gave the agent examples to learn from and established rewards for the agent to guide its choices, including the sacrifice of short-term benefits for long-term benefits. The experts also had to think through real-world constraints that humans often take for granted. For example, the agent did not know initially that the boat could sail only in an upright position; early on, it tried to exploit a loophole in the system by sailing upside down, something a human would know is impossible.

The Emirates Team New Zealand design team regularly compared agent performance in the simulator with that of the sailors, and if an agent's performance remained subpar, the experts would tweak the rewards system. To accelerate the training process, a network of more than 1,000 AI agents running in parallel was deployed, so each agent could learn from the best collective experiences. In this way, the agents quickly reached a level of mastery to outperform world-champion sailors in the simulator and begin testing design concepts for the team.

significant factor in the race, but race rules allowed teams to build only six full-size hydrofoils in all.

Using the reinforcement learning–trained agent to control the simulator, Emirates Team New Zealand designers could evaluate thousands of hydrofoil design concepts instead of just hundreds in their quest for a winning design. This gave them valuable insight into how a boat might perform on the water before engaging in a costly build and, in the process, would dramatically reduce the design price tag for future races. In addition, as the Emirates Team New Zealand agents' knowledge of sailing increased over time, the sailors began learning maneuvers from the agents that they had not considered, enabling them to improve their performance for a given design.

Where businesses can use reinforcement learning

The heart of Emirates Team New Zealand's challenge was to solve a complex business problem in a dynamic environment where the variables change in unpredictable ways, the ideal end state is only loosely defined, and the only way the system could learn about its environment was to interact with it.

That situation is analogous to problems facing retailers, manufacturers, utilities, and companies in many other industries. For example, whereas once retailers could reasonably expect that past consumer behaviors would indicate future preferences, they now operate in a world where consumer purchase patterns and preferences evolve rapidly—all the more so as the COVID-19 pandemic repeatedly redefines life. Manufacturers and consumer-packaged-goods companies are under pressure to build dynamic supply chains that account for climate, political, and societal shifts anywhere in the world at a moment's notice.

Each of these challenges represents a complex and highly dynamic optimization problem, which, with the right data and feedback loops, is well suited for solving with reinforcement learning.

The appeal of reinforcement learning for problems with many possible actions and paths is that the AI agent does not need to be explicitly programmed. Because it learns from examples and teaches itself through trial and error, it can propose novel and adaptive solutions, oftentimes faster than humans could do so.

How reinforcement learning works

An AI agent learns through trial and error. In simple terms, the agent performs actions within an environment and receives rewards when it takes the "right" actions. It works to find the sequence of actions that maximizes the cumulative rewards it receives. Data scientists and subject-matter experts define the reward function for the agent. This way of learning is just one aspect of reinforcement learning that makes it different from other AI techniques (see exhibit and "An executive's guide to AI," on McKinsey.com, for more on the different types of machine learning).

Exhibit

How reinforcement learning differs from other AI techniques.

	Supervised learning	Unsupervised learning	**Reinforcement learning**
Training data needed	Labeled data	Unlabeled data	Synthetic data created as an agent interacts with the environment and receives feedback
Purpose	Predict the target variable for unlabeled observations, based on having learned from the labeled data	Identify groups of data that exhibit similar behavior	Identify an optimal sequence of actions to achieve an optimal outcome
Adaptability	Retraining of model required if underlying conditions change and cause accuracy to drift	Retraining of model required when changes occur in underlying data	System can adapt to changes (within a certain range) on its own

Emirates Team New Zealand, for instance, was able to test multiple designs simultaneously (something the sailors could never do), test tenfold more designs under more conditions than had previously been possible, and gain insight from the AI agent into new ways their sailors could execute on these boat designs on the water.

Broadly speaking, we see reinforcement learning delivering this value across the business, with potential applications in every business domain and industry (exhibit). Some of the near-term applications for reinforcement learning fall into three categories: speeding design and product development, optimizing complex operations, and guiding customer interactions.

Speeding design and product development
Reinforcement learning can improve the development of products, engineering systems, manufacturing plants, oil refineries, telecommunications or utility networks, and other capital projects. Mining companies could, for example, explore a greater range of mine designs than possible with the other AI techniques used today to improve yield. One automotive manufacturer is already exploring how agents trained through reinforcement learning can enable it to test more

Exhibit

Initial applications of reinforcement learning span most, if not all, industries.

● Optimizing product development cycles (AI-assisted design) ● Optimizing complex operations ● Informing next best action for each customer

Industry	Sample reinforcement learning applications
Advanced electronics and semiconductors	● Optimize silicon and chip design to increase performance and reduce manufacturing costs ● Optimize fabrication manufacturing process for improved yield and throughput
Agriculture	● Solve scheduling and production allocation challenges to increase yield ● Optimize network and warehouse logistics for reduced waste and costs ● Apply advanced pricing and promotion to improve product margins
Aerospace and defense	● Optimize engineering design processes to reduce time to market for new systems and improve quality
Automotive	● Optimize design processes to shorten development cycle for new cars and features and improve quality ● Deploy advanced predictive maintenance to prevent rare failures and unplanned outages ● Deliver real-time production monitoring and controls to increase manufacturing yield
Financial services	● Apply real-time trading and pricing strategies for greater agility and revenue ● Optimize ATM replenishment and allocation strategies to reduce costs and improve the customer experience ● Deliver advanced personalization capabilities that adapt promotions, offers, and recommendations daily for increased customer satisfaction and sales
Mining	● Optimize design process so teams can explore a greater range of mine designs for improving mine yield ● Use intelligent process controls for managing power generation and bore milling to increase yield and reduce costs ● Apply holistic logistics scheduling to optimize mine-to-shipping operations and reduce costs
Oil and gas	● Enable real-time well monitoring and precision drilling for increased yield ● Optimize tanker routing to reduce costs and ensure on-time delivery ● Enable advanced predictive maintenance to prevent rare equipment failures and unplanned outages
Pharmaceuticals	● Optimize drug discovery, identifying molecules of interest faster to reduce the time and cost of research and bring new therapies to market faster ● Automate chemistry, manufacturing, and controls (CMC) to maximize batch yield and quality ● Optimize biological methods to reach peak production output
Retail	● Optimize routing, logistics network planning, and warehouse operations to reduce costs and keep shelves stocked ● Implement advanced inventory modeling and digitize supply-chain planning to prevent out-of-stocks and waste ● Deliver advanced personalization capabilities that adapt promotions, offers, and recommendations daily for increased customer satisfaction and sales
Telecom	● Optimize network layout to maximize coverage and minimize power consumption ● Manage networks in real time to optimize service quality and reduce downtime ● Apply advanced personalization to increase cross-sell and upsell revenue
Transport and logistics	● Optimize routing, logistics network planning, and warehouse operations to reduce costs and improve customer satisfaction ● Optimize inbound and outbound delivery networks to minimize shipping delays and associated costs

Note: Categories and use cases listed are not exhaustive and are intended solely to provide examples of some initial applications.

ideas for regenerative braking in new electric vehicles, so it can optimize the design for noise, vibration, and heat.

Optimizing complex operations

Reinforcement learning's ability to solve complex problems gives it high potential for optimizing complex operations. Initially, we see three primary applications of reinforcement learning in this area.

First, reinforcement learning can help organizations identify the right actions to take across a value chain *as events unfold*. A transportation company, for example, can optimize travel routes in real time based on changing traffic, weather, and safety conditions. A food producer can optimize product distribution worldwide amid daily, even hourly, fluctuating demand and exchange rates, varying shipping routes, and more.

It also can help teams manage complex manufacturing processes. For example, it can allow teams to monitor production in real time, simulating different scenarios and updating key parameters to increase production dynamically. Manufacturers that have already used machine learning to minimize product defects can now expand their insights with reinforcement learning to prevent the rare remaining defects that pop up intermittently with seemingly no common root cause.

Finally, reinforcement learning can power autonomous system controllers by, for instance, continuously monitoring and adjusting equipment operating temperatures to ensure optimal performance or running a robotic arm on the manufacturing floor.

Informing the next best action for each customer

When integrated within personalization and recommender systems, reinforcement learning can help organizations understand, identify, and respond to changes in taste in real time, personalizing messages and adapting promotions, offers, and recommendations daily.

Getting to wide-scale adoption

To be sure, implementing reinforcement learning is a challenging technical pursuit. A successful reinforcement learning system today requires, in simple terms, three ingredients:

1. **A well-designed learning algorithm with a reward function.** A reinforcement learning agent learns by trying to maximize the rewards it receives for the actions it takes. A good algorithm with a properly defined reward function enables an agent to make complex decisions—for example, to take an action now that might appear suboptimal in the short term but would pay off handsomely in the long run.

2. **A learning environment.** Oftentimes the learning environment involves a simulator, or digital twin, that replicates the real-world conditions in which the agent will operate and provides a training ground for the agent.[2] In some cases, however, the learning environment could be a digital platform, such as a product-ordering system, where an AI agent can repeatedly perform the same (or similar) tasks and rapidly receive feedback about the success of its actions.

3. **Compute power.** Training an agent requires substantial compute resources and specialized infrastructure that can scale out thousands of distributed training jobs, which, even when running in parallel, typically require thousands of hours of compute time.

A few years ago, the cost and complexity of building and training these systems was out of reach for all but a few tech leaders. However, significant technological advances to address these hurdles have made reinforcement learning more accessible to more businesses, and continued evolution of the needed tooling is quickly putting the technology within every company's grasp.

[2] Research is currently under way into what's called "offline" reinforcement learning, where the learning is done exclusively on existing empirical data, rather than through simulation. This research has the potential to alleviate the need for a simulator.

Costs are becoming manageable

The latest iterations in reinforcement learning algorithms, such as soft actor-critic, are dramatically improving training efficiency, substantially driving down compute costs. At the same time, the cost of compute itself has declined significantly. Companies can now access specialized systems in the cloud and pay only for what they use. Also, new tools and strategies enable teams to manage the compute they use. For instance, resource allocation and development tools now available enable teams to identify the least expensive (or most efficient) compute at any given time for a given purpose.

That said, for the technology to be used more widely, compute costs for reinforcement learning tasks will need to decline further. We expect that to happen in the near future for several reasons, including increasing competition among cloud providers.

Cloud-based frameworks address system complexity

Cloud providers have also ramped up efforts to deliver prepackaged, enterprise-ready frameworks that can be deployed in assembly-line fashion and include the necessary tools, protocols, application programming interfaces (APIs), open-source libraries (such as RLlib), and other technologies to eliminate some of the manual coding and integration work. Frameworks can, for example, enable teams to run training jobs across dozens of systems using a single line of code, rather than having to program this capability from scratch. At Emirates Team New Zealand, the development team drew from such frameworks where possible and then focused on the value-added tasks that hadn't yet been commoditized.

Work remains to be done. There is not yet an out-of-the-box, single framework for delivering reinforcement learning solutions. We anticipate that something like this will be available in a few years from major cloud providers. Efforts under way in this area include Microsoft's Project Bonsai, Amazon's SageMaker RL, and Google's SEED RL.

How leaders can get started with reinforcement learning

The same foundational practices and organizational and cultural changes in which enterprises are already investing for other AI also apply to reinforcement learning. However, given reinforcement learning's early maturity and its unique requirements and abilities, leaders should keep some strategies top of mind.

Find the right business problem for experimentation

Start by identifying processes where reinforcement learning might free the business to optimize performance in some way, perhaps consulting the exhibit for some ideas. Ideally, select a process where there is already some type of learning environment that can be adapted for training the AI agents.

The latest iterations in reinforcement learning algorithms, such as soft actor-critic, are dramatically improving training efficiency, substantially driving down compute costs.

In our experience, one of the best ways to know if a given process is ready for reinforcement learning is to ask, "What business challenges haven't we been able to solve with traditional modeling approaches?" Look for areas where teams are conducting AI projects with other methods but haven't been able to bring them into production because the environment is too dynamic and the models deliver inconsistent results, require too many assumptions and approximations about the data, or cannot handle the full scope of business needs. At Emirates Team New Zealand, for example, testing loops for new boat designs were constantly interrupted by the sailors' schedules, and there was a high cost to taking the sailors away from other activities.

The right problem should also be one where it isn't necessary to know *why* the reinforcement learning system performs the way it does. For now, these systems are not easily explainable, if at all, given the complexity of the neural networks often embedded in them. Reinforcement learning therefore might not be well suited to situations where regulators or operators require transparency.

Factor in compute costs up front

Outlining the reward function to enable an AI agent to learn effectively requires as much art as science, often making it the costliest part of the development process. Subject-matter experts and data scientists need to constantly refine incentives, commonly known as reward hacking, to figure out how to properly calibrate rewards to enable an agent to make complex decisions optimally.

Teams can use first principles to ballpark potential costs, and leaders should understand and discuss the potential cost drivers with their teams up front to help ensure a smoother process and free teams to focus on the work ahead.

Future-proof your simulator

Many manufacturing and operations-focused organizations already use simulation or a digital twin to tune asset performance and utilization. Even in these industries, however, upgrades might be necessary to enable reinforcement learning. Many traditional simulators are designed to run on a small scale, on premise, with only one simulation running at a time, and a person uses a physical interface, such as a joystick, to control it. Such a simulator will need to be re-platformed onto a cloud environment so it can run thousands of simulations in parallel, and it must be updated with an API that enables AI agents to interact with it.

In all cases, whether building or rebuilding digital simulators, organizations should think beyond their existing use cases and make design choices that provide flexibility in supporting more advanced use cases that might not yet be on their radar. Reinforcement learning technology is maturing rapidly, so such planning will enable companies to deploy new reinforcement learning solutions faster than companies that fail to do so.

In our experience, one of the best ways to know if a given process is ready for reinforcement learning is to ask, "What business challenges haven't we been able to solve with traditional modeling approaches?"

Double down on humans

Implementations are most successful when leaders recognize that the greatest value comes from using the technology to augment and expand human performance rather than replace it. Any AI initiative relies on domain expertise to help AI teams properly define the use case, determine which data sources to use, ensure the AI predictions and recommendations make sense and can be successfully integrated into their workflows, and guide change management. In reinforcement learning, domain experts must do all this and more, working with data scientists daily to ideate and test different rewards to build an effective reward function and then monitoring the AI agent's performance after deployment.

Organizations should also consider whether they need a human in the loop to help guide final decisions. At Emirates Team New Zealand, after the AI agents recommended the top designs from the thousands they tested, the sailors then took the helm of the digital simulator once again to test the best hydrofoils and prioritize the final selections.

Identify and manage potential risks

In choosing where to implement reinforcement learning, it's important to acknowledge employees' and society's concerns about the explainability and use of autonomous systems. Our colleagues have written extensively about the unintended consequences that can arise from AI when organizations do not fully understand the possible risks and about the leader's role in building AI systems responsibly. As reinforcement learning gains traction, leaders will need to build their knowledge around the ethical concerns and interdependencies and how to manage them effectively, so they can guide their company on when to try or not try this new technique.

The technologies that enable reinforcement learning are advancing briskly: compute costs and complexity are declining as the industry evolves toward more adaptive, self-learning algorithms and makes more complex systems available to organizations as high-level services. With this, adoption is increasing, and in a few years, we anticipate that reinforcement learning will become more common in many industries, such as telecom, pharmaceuticals, and advanced industries. Within five years, it will likely be in every leading organization's AI toolbox, helping companies to uncover innovative strategies and first-in-kind moves that more established techniques may not and to achieve the next level of performance that until now has eluded human reach.

Jacomo Corbo is a partner in McKinsey's London office, **Oliver Fleming**, based in Sydney, is an expert associate partner at QuantumBlack; and **Nicolas Hohn** is a senior expert in the Melbourne office.

The authors wish to thank Zara Davis for her contributions to this article.