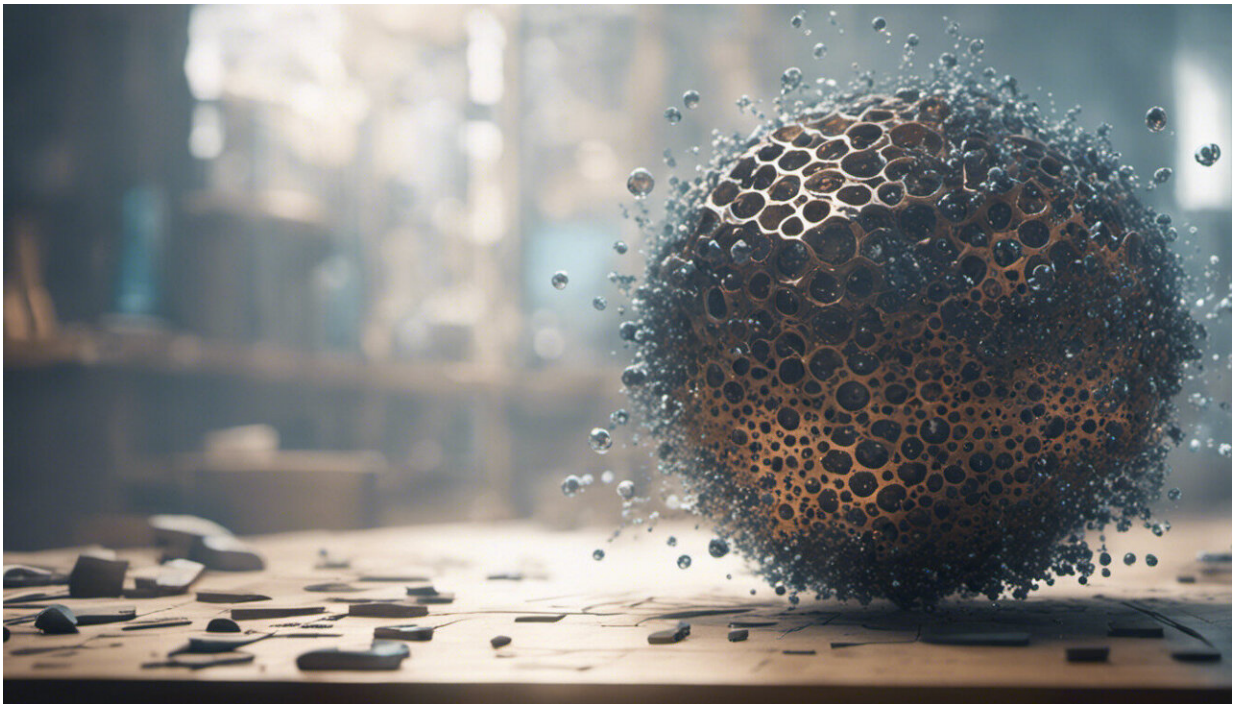# Google's new Go-playing AI learns fast, and even thrashed its former self

October 20 2017, by Geoff Goodhill

Credit: AI-generated image (disclaimer)

Just last year Google DeepMind's AlphaGo took the world of Artificial Intelligence (AI) by storm, showing that a computer program could beat the world's best human Go players.

But in a demonstration of the feverish rate of progress in modern AI,

details of a new milestone reached by an improved version called [AlphaGo Zero](#) were [published this week in Nature](#).

Using less computing power and only three days of training time, AlphaGo Zero beat the original AlphaGo in a 100-game match by 100 to 0. It wasn't even worth humans showing up.

Learning to play Go

Go is a game of strategy between two players who take it in turns to place "stones" on a 19x19 board. The goal is to surround a larger area of the board than your opponent.

Go has proved much more challenging than chess for computers to master. There are many more possible moves in each position in Go than chess, and many more possible games.

The [original AlphaGo first learned](#) from studying 30 million moves of expert human play. It then improved beyond human expertise by playing many games against itself, taking several months of computer time.

By contrast, AlphaGo Zero never saw humans play. Instead, it began by knowing only the rules of the game. From a relatively modest five million games of self-play, taking only three days on a smaller computer than the original AlphaGo, it then learned super-AlphaGo performance.
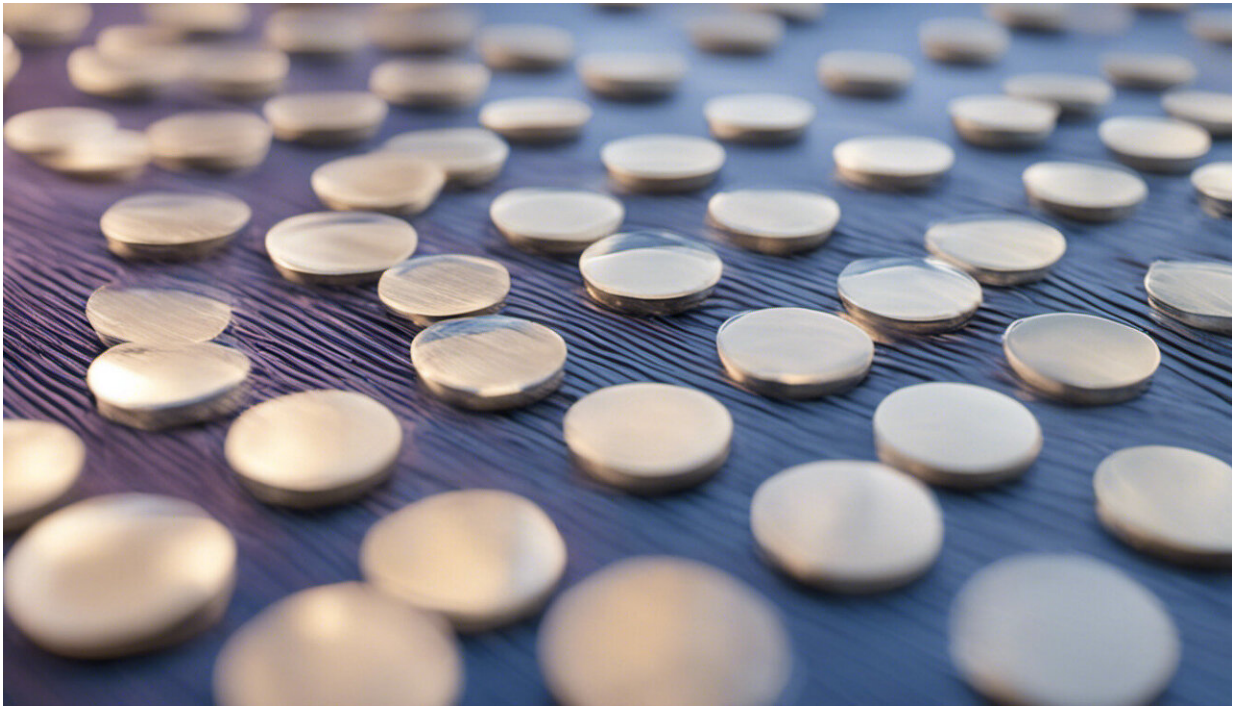
Credit: AI-generated image ([disclaimer](#))

Fascinatingly, its learning roughly mimicked some of the stages through which humans progress as they master Go. AlphaGo Zero rapidly learned to reject naively short-term goals and developed more strategic thinking, generating many of the patterns of moves often used by top-level human experts.

But remarkably it then started rejecting some of these patterns in favour of new strategies never seen before in human play.

Beyond human play

AlphaGo Zero achieved this feat by approaching the problem differently from the original AlphaGo. Both versions use a combination of two of

the most powerful algorithms currently fuelling AI: deep learning and reinforcement learning.

To play a game like Go, there are two basic things the program needs to learn. The first is a policy: the probability of making each of the possible moves in a given position. The second is a value: the probability of winning from any given position.

In the pure reinforcement learning approach of AlphaGo Zero, the only information available to learn policies and values was for it to predict who might ultimately win. To make this prediction it used its current policy and values, but at the start these were random.

This is clearly a more challenging approach than the original AlphaGo, which used expert human moves to get a head-start on learning. But the earlier version learned policies and values with separate neural networks.

The algorithmic breakthrough in AlphaGo Zero was to figure out how these could be combined in just one network. This allowed the process of training by self-play to be greatly simplified, and made it feasible to start from a clean slate rather than first learning what expert humans would do.

An Elo rating is a widely used measure of the performance of players in games such as Go and chess. The best human player so far, Ke Jie, currently has an Elo rating of about 3,700.

AlphaGo Zero trained for three days and achieved an Elo rating of more than 4,000, while an expanded version of the same algorithm trained for 40 days and achieved almost 5,200.

This is an astonishingly large step up from the best human – far bigger than the current gap between the best human chess player Magnus

Carlsen (about 2,800) and chess program (about 3,400).

The next challenge

AlphaGo Zero is an important step forward for AI because it demonstrates the feasibility of pure reinforcement learning, uncorrupted by any human guidance. This removes the need for lots of expert human knowledge to get started, which in some domains can be hard to obtain.

It also means the algorithm is free to develop completely new approaches that might have been much harder to find had it been been initially constrained to "think inside the human box". Remarkably, this strategy also turns out to be more computationally efficient.

But Go is a tightly constrained game of perfect information, without the messiness of most real-world problems. Training AlphaGo Zero required the accurate simulation of millions of games, following the rules of Go.

For many practical problems such simulations are computationally unfeasible, or the rules themselves are less clear.

There are still many further problems to be solved to create a general-purpose AI, one that can tackle a wide range of practical problems without domain-specific human intervention.

But even though humans have now comprehensively lost the battle with Go algorithms, luckily AI (unlike Go) is not a zero-sum game. Many of AlphaGo Zero's games have now been published, providing a lifetime of inspirational study for human Go players.

More importantly, AlphaGo Zero represents a step towards a world where humans can harness powerful AIs to help find unimaginably (to humans) creative solutions to difficult problems. In the world of AI,

there has never been a better time to Go for it.

More information: David Silver et al. Mastering the game of Go without human knowledge, *Nature* (2017). DOI: 10.1038/nature24270

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation