The New York Times | https://nyti.ms/2Rjetd3

ESSAY

One Giant Step for a Chess-Playing Machine

The stunning success of AlphaZero, a deep-learning algorithm, heralds a new age of insight — one that, for humans, may not last long.

By Steven Strogatz

Dec. 26, 2018

In early December, researchers at DeepMind, the artificial-intelligence company owned by Google's parent corporation, Alphabet Inc., filed a dispatch from the frontiers of chess.

A year earlier, on Dec. 5, 2017, the team had stunned the chess world with its announcement of AlphaZero, a machine-learning algorithm that had mastered not only chess but shogi, or Japanese chess, and Go. The algorithm started with no knowledge of the games beyond their basic rules. It then played against itself millions of times and learned from its mistakes. In a matter of hours, the algorithm became the best player, human or computer, the world has ever seen.

The details of AlphaZero's achievements and inner workings have now been formally peer-reviewed and published in the journal Science this month. The new paper addresses several serious criticisms of the original claim. (Among other things, it was hard to tell whether AlphaZero was playing its chosen opponent, a computational beast named Stockfish, with total fairness.) Consider those concerns dispelled. AlphaZero has not grown stronger in the past twelve months, but the evidence of its superiority has. It clearly displays a breed of intellect that humans have not seen before, and that we will be mulling over for a long time to come.

Computer chess has come a long way over the past twenty years. In 1997, I.B.M.'s chess-playing program, Deep Blue, managed to beat the reigning human world champion, Garry Kasparov, in a six-game match. In retrospect, there was little mystery in this achievement. Deep Blue could evaluate 200 million positions per second. It never got tired, never blundered in a calculation and never forgot what it had been thinking a moment earlier.

For better and worse, it played like a machine, brutally and materialistically. It could out-compute Mr. Kasparov, but it couldn't outthink him. In Game 1 of their match, Deep Blue greedily accepted Mr. Kasparov's sacrifice of a rook for a bishop, but lost the game 16 moves later. The current generation of the world's strongest chess programs, such as Stockfish and Komodo, still play in this inhuman style. They like to capture the opponent's pieces. They defend like iron. But although they are far stronger than any human player, these chess "engines" have no real understanding of the game. They have to be tutored in the basic principles of chess.

These principles, which have been refined over decades of human grandmaster experience, are programmed into the engines as complex evaluation functions that indicate what to seek in a position and what to avoid: how much to value king safety, piece activity, pawn structure, control of the center, and more, and how to balance the trade-offs among them. Today's chess engines, innately oblivious to these principles, come across as brutes: tremendously fast and strong, but utterly lacking insight.

All of that has changed with the rise of machine learning. By playing against itself and updating its neural network as it learned from experience, AlphaZero discovered the principles of chess on its own and quickly became the best player ever. Not only could it have easily defeated all the strongest human masters — it didn't even bother to try — it crushed Stockfish, the reigning computer world champion of chess. In a hundred-game match against a truly formidable engine, AlphaZero scored twenty-eight wins and seventy-two draws. It didn't lose a single game.

Most unnerving was that AlphaZero seemed to express insight. It played like no computer ever has, intuitively and beautifully, with a romantic, attacking style. It played gambits and took risks. In some games it paralyzed Stockfish and toyed with it. While conducting its attack in Game 10, AlphaZero retreated its queen back into the corner of the board on its own side, far from Stockfish's king, not normally where an attacking queen should be placed.

Yet this peculiar retreat was venomous: No matter how Stockfish replied, it was doomed. It was almost as if AlphaZero was waiting for Stockfish to realize, after billions of brutish calculations, how hopeless its position truly was, so that the beast could relax and expire peacefully, like a vanquished bull before a matador. Grandmasters had never seen anything like it. AlphaZero had the finesse of a virtuoso and the power of a machine. It was humankind's first glimpse of an awesome new kind of intelligence.

Garry Kasparov, left, playing against the I.B.M. Deep Blue computer in the sixth and final game of a match in New York in May 1997. The computer's pieces were moved by Joseph Hoane, right, an I.B.M. scientist. Stan Honda/Agence France-Presse — Getty Images

When AlphaZero was first unveiled, some observers complained that Stockfish had been lobotomized by not giving it access to its book of memorized openings. This time around, even with its book, it got crushed again. And when AlphaZero handicapped itself by giving Stockfish ten times more time to think, it still destroyed the brute.

Tellingly, AlphaZero won by thinking smarter, not faster; it examined only 60 thousand positions a second, compared to 60 million for Stockfish. It was wiser, knowing what to think about and what to ignore. By discovering the principles of chess on its own, AlphaZero developed a style of play that "reflects the truth" about the game rather than "the priorities and prejudices of programmers," Mr. Kasparov wrote in a commentary accompanying the Science article.

The question now is whether machine learning can help humans discover similar truths about the things we really care about: the great unsolved problems of science and medicine, such as cancer and consciousness; the riddles of the immune system, the mysteries of the genome.

The early signs are encouraging. Last August, two articles in Nature Medicine explored how machine learning could be applied to medical diagnosis. In one, researchers at DeepMind teamed up with clinicians at Moorfields Eye Hospital in London to develop a deep-learning algorithm that could classify a wide range of retinal pathologies as accurately as human experts can. (Ophthalmology suffers from a severe shortage of experts who can interpret the millions of diagnostic eye scans performed each year; artificially intelligent assistants could help enormously.)

The other article concerned a machine-learning algorithm that decides whether a CT scan of an emergency-room patient shows signs of a stroke, an intracranial hemorrhage or other critical neurological event. For stroke victims, every minute matters; the longer treatment is delayed, the worse the outcome tends to be. (Neurologists have a grim saying: "Time is brain.") The new algorithm flagged these and other critical events with an accuracy comparable to human experts — but it did so 150 times faster. A faster diagnostician could allow the most urgent cases to be triaged sooner, with review by a human radiologist.

What is frustrating about machine learning, however, is that the algorithms can't articulate what they're thinking. We don't know why they work, so we don't know if they can be trusted. AlphaZero gives every appearance of having discovered some important principles about chess, but it can't share that understanding with us. Not yet, at least. As human beings, we want more than answers. We want insight. This is going to be a source of tension in our interactions with computers from now on.

In fact, in mathematics, it's been happening for years already. Consider the longstanding math problem called the four-color map theorem. It proposes that, under certain reasonable constraints, any map of contiguous countries can always be colored with just four colors such that no two neighboring countries are colored the same.

Although the four-color theorem was proved in 1977 with the help of a computer, no human could check all the steps in the argument. Since then, the proof has been validated and simplified, but there are still parts of it that entail brute-force computation, of the kind employed by AlphaZero's chess-playing computer ancestors. This development annoyed many mathematicians. They didn't need to be reassured that the four-color theorem was true; they already believed it. They wanted to understand why it was true, and this proof didn't help.

But envisage a day, perhaps in the not too distant future, when AlphaZero has evolved into a more general problem-solving algorithm; call it AlphaInfinity. Like its ancestor, it would have supreme insight: it could come up with beautiful proofs, as elegant as the chess games that AlphaZero played against Stockfish. And each proof would reveal why a theorem was true; AlphaInfinity wouldn't merely bludgeon you into accepting it with some ugly, difficult argument.

For human mathematicians and scientists, this day would mark the dawn of a new era of insight. But it may not last. As machines become ever faster, and humans stay put with their neurons running at sluggish millisecond time scales, another day will follow when we can no longer keep up. The dawn of human insight may quickly turn to dusk.

Suppose that deeper patterns exist to be discovered — in the ways genes are regulated or cancer progresses; in the orchestration of the immune system; in the dance of subatomic particles. And suppose that these patterns can be predicted, but only by an intelligence far superior to ours. If AlphaInfinity could identify and understand them, it would seem to us like an oracle.

We would sit at its feet and listen intently. We would not understand why the oracle was always right, but we could check its calculations and predictions against experiments and observations, and confirm its revelations. Science, that signal human endeavor, would reduce our role to that of spectators, gaping in wonder and confusion.

Maybe eventually our lack of insight would no longer bother us. After all, AlphaInfinity could cure all our diseases, solve all our scientific problems and make all our other intellectual trains run on time. We did pretty well without much insight for the first 300,000 years or so of our existence as Homo sapiens. And we'll have no shortage of memory: we will recall with pride the golden era of human insight, this glorious interlude, a few thousand years long, between our uncomprehending past and our incomprehensible future.

Steven Strogatz is professor of mathematics at Cornell and author of the forthcoming "Infinite Powers: How Calculus Reveals the Secrets of the Universe," from which this essay is adapted.

A version of this article appears in print on Jan. 8, 2019, Section D, Page 1 of the New York edition with the headline: One Giant Step for a Chess-Playing Machine