

AlphaGo and Artificial Intelligence

What board game battles between AI and humans reveal about us

The game of [Go](#) (围棋) originated in China over 2500 years ago, and is one of the oldest board games played today. Behind the simple rules is a deceptively complex game, requiring intuition and strategic thinking. After IBM's [DeepBlue defeated Kasparov](#) in chess 20 years ago, Go was considered the last game in which humans still had the upper hand against computers.

AlphaGo is a computer program developed by Google [DeepMind](#), and it made history by [defeating](#) the European Go Champion Fan Hui in October 2015. In March 2016, it played a [5 game match](#) against one of the world's top ranked players [Lee Se-dol](#), with the winner taking home one million US dollars. The games were streamed live to [60 million viewers](#) in China alone. Despite Lee's confident [pregame prediction](#) that he would sweep the computer, he lost 1-4. As if it were meaningful, AlphaGo is now ranked as a [9-dan pro](#) and [#2 in the world](#), behind China's Go prodigy [Ke Jie](#). The inevitable match with Ke Jie will undoubtedly be exciting to watch.

Should We Worry about Artificial Intelligence?

This apparent loss of humanity's last stand against Artificial Intelligence (AI) has brought another provocative question to the forefront, one that you have undoubtedly heard before: is AI an existential threat to humanity? Let's take a look.

Many heavyweights such as Stephen Hawking, Elon Musk, and Bill Gates have [expressed concern](#) that AI could [spell the end](#) of the human race. First of all, it is worth noting that the type of AI being referring to is [Artificial General Intelligence](#) (strong AI), which is a hypothetical machine that can perform any intellectual task that a human can. It does not currently exist, and [may not ever](#). Strong AI straddles the realms of science and fiction, and the implications are deep and philosophical. However, that discussion is for another day.

What does exist today is AI geared towards a specific, narrow task (weak AI), such as playing Go, driving a car, navigation, even [predicting pregnancies](#). We already use weak AI extensively in our everyday life, and the risks are not only real but often go unnoticed.

What Are Some of the Risks of Weak AI?

AI machines are immune to human shortcomings such as distraction, bias, fatigue, and calculation errors, which makes them perfect for repetitive and well-defined tasks. In fact, they are so good at these specific tasks that we willingly and prematurely delegate our responsibilities to them; we happily embrace every incremental improvement, each step reasonable on its own, until we have eventually delegated our skills away, much like our privacy and liberties. Far from merely being an inconvenience, this loss of skills is often a matter of life and death. For example, many airplanes have crashed because pilots have become [over-reliant on autopilot](#) and lost their flying skills, such as when Asiana Airlines [crashed into the seawall](#) on a clear day in San Francisco in 2013.

Another risk is about responsibility. Let's look at a scenario: A self-driving car is cruising down a 2 lane road, when a large tree suddenly falls and blocks the lane. The car can avoid harm to its passenger by

swerving into the other lane, which is unfortunately occupied by bicyclists. The AI will act based on what it was programmed to do; the passenger's fate is predetermined. The important question is, by whom? Who gets to [play god](#) - the programmer, the passenger, ethicists, or someone else? With AI taking the uncertainty of execution out of the equation, we can no longer hide behind the intentional vagueness that we so want to preserve, for the harm is no longer accidental but premeditated. As we are forced to write down rules and preferences for these difficult situations, we lose our treasured hypocrisies and lay naked our collective biases; biases tacitly accepted but rather left unspoken. We want to cede authority to AI, yet the accompanying responsibilities are far more difficult and uncomfortable to shirk. If we stick our heads in the sand and simply allow AI to make increasingly important decisions for us, we relinquish not only our right of choice, but also our values behind them. We risk relegating responsibility to parties who may not be qualified, and perhaps more appallingly, without our consent.

Going Forward

Throughout history, humans have built and used tools of increasing power, from sticks, energy, abstract thought, information, to simulation of intelligence, which is AI. With this immensely powerful tool of AI we have today, the real risks are not found in the machine but in the mirror. In our reflection we see expanding waistlines because it's easier to drive than walk, and shrinking brains because it's easier to accept a thought than think for ourselves. We crave the conveniences the tool provides, yet pout when we have to read the operating manual; and as the machines have gotten artificially intelligent, we humans have become artificially stupid.

Some mourn the loss of what was considered humanity's last stand in board games against AI. I beg to differ. Even if AlphaGo were unbeatable, the game of Go would not be relegated to the trash bin. Conversely, AlphaGo is already expanding our understanding of the game and opening up possibilities, which is exactly what happened to chess. And in my opinion, a tiny and temporary dent in our imagined superiority is a small price to pay for a profound and lasting improvement. After all, when a teacher is beat by his student, he loses face but gains pride – an emotion not found in the cold silicon hardware and software of AI, but in the warm, squishy wetware between our ears.

Wielded carefully, AI is a powerful tool that will drive the betterment of mankind. It is something to embrace, not with blind fervor but with prudence and reflection. Until the day strong AI emerges, that is what I think AI should be to us; a complement, not a replacement.

Bonus content from the cutting room floor for the web:

What Made This Win So Special?

Unlike the brute force approach to win at chess, AlphaGo had to use a very different approach for Go. It used a neural network, which mimic biological neural processes such as the ones happening inside the human brain. Guided by a "value network" that evaluates the board position and a "policy network" that selects moves, AlphaGo sees the both the battle and the war, much like a human does. It also uses reinforcement learning, in which it learns by playing against a different version of itself.

One thing unusual about Alpha Go is that it actively calculates the probability that a human would make a certain move. This gives it another advantage because not only can it surprise a human player with an unusual move, it can also discount potential opponent moves, however good it may be, if it determines that a human is unlikely to play it. The approach is interesting, as it seems to embody a very rudimentary Theory of Mind; in other words, “I won’t bother looking at this potentially great move you could make, because I don’t think you will play it”. Interestingly, surprising scenarios cutting both ways happened in this match.

On move 37 of game 2, AlphaGo played a brilliant but “inhuman” move, befuddling not only Lee but all of the commentators. In fact, Lee had to leave the room to collect himself, and took another 15 minutes after he returned considering how to respond. He eventually lost the game. Ironically, predicting how humans might play can also be a weakness. On move 78 of game 4, Lee played a genius move, described by many as the “hand of God”. AlphaGo probably failed to consider this move, since it had calculated the probability of being played by a professional player as 1 in 10,000. Failing to realize the power of that move, AlphaGo made some mistakes and only realized it was in trouble several moves later. That cost AlphaGo its only game lost against Lee.

This approach has rather unusual implications. Somewhat insultingly (to the human, not the computer), it chooses not the overall best move but the one more likely to win against a human. Obviously this strategy is not effective against a computer. This also indirectly encourages humans to play unconventionally against the computer, until the day humans can balance the equation by regularly spotting and playing these unusual moves.

References:

1. [Go \(game\)](#), Wikipedia
2. [Deep Blue versus Garry Kasparov](#), Wikipedia
3. [Google DeepMind official site](#)
4. [Mastering the game of Go with deep neural networks and tree search](#), Nature 2016
5. [Google DeepMind Challenge Match Press Release](#), Google DeepMind
6. [Lee Sedol](#), Wikipedia
7. [The Sadness and Beauty of Watching Google’s AI Play Go](#), Wired
8. [S. Korean Go player confident of beating Google’s AI](#), Yonhap News Agency
9. [Google’s AlphaGo gets ‘divine’ Go ranking](#), phys.org
10. [Go Ratings](#), goratings.org
11. [Ke Jie](#), Wikipedia
12. [Elon Musk, Stephen Hawking, Google researchers join forces to avoid ‘pitfalls’ of artificial intelligence](#), The Washington Post
13. [Elon Musk: ‘With artificial intelligence we are summoning the demon.’](#), The Washington Post
14. [Artificial general intelligence](#), Wikipedia
15. [The Doomsday Invention](#), The New Yorker
16. [What’s Even Creepier Than Target Guessing That You’re Pregnant?](#), Slate
17. [Concerns raised about overreliance on autopilot](#), Alaska Dispatch News
18. [Asiana Airlines Flight 214](#), Wikipedia
19. [The Moral Challenges of Driverless Cars](#), Communications of the ACM