Games Without Frontiers: Cybersecurity and Artificial Intelligence

March 10, 2020

AI is being weaponized by threat actors. But our experience with game-playing AI suggests the good guys may hold the upper hand.

Part one of a series.

Five years ago AlphaGo, an artificial intelligence (AI) developed by DeepMind (now owned by Google), defeated a human master champion at the ancient Chinese board game Go. Most of us know AIs have defeated top human players at everything from Chess,

Texas Hold'em, and Ms. Pac-Man to Rubik's Cube, StarCraft and Jeopardy! (where IBM's Watson thrashed legends Ken Jennings and Brad Rutter.

But Go may be the most complex game in the world: it has an estimated 2×10^{170} possible positions (vs. 10^{45} for chess), which is "vastly greater than the number of atoms in the universe…"

Given its ability to dominate so thoroughly in a context that complex, it's no surprise that AI is a hot topic of conversation in many fields. Cybersecurity is one of them, and some of the more alarming analyses predict cybercrime losses in the trillions of dollars in the coming years.

The wild successes of AIs in human games may actually be good news for cybersecurity pros

In a recent #AskOptiv post we addressed some of the challenges posed by the evolution of AI in cybersecurity, specifically noting threat applications for spear phishing, social engineering, and deepfakes. We also discussed the role of AI as a cybersecurity tool.

Machine learning (ML) and AI applied to threat detection is the first step in helping the security industry identify and prevent AI-based attacks. It's assumed that threat actor payloads and attacks, including TTPs, are dynamic and ever-changing. A robust intelligence approach to processing big data, indicators of compromise (IOCs) and their context, coupled with enrichment, reputational data, detonation data and additional context is a huge undertaking. Leveraging ML and AI are essential to the timely and efficient processing of data (in addition to enhancing threat detection).

ML/AI can be used to process vast amounts of data across multiple client environments and tickets in real time, correlating those, providing granular attribution, coupled with orchestration and automation actions like auto-escalate, auto-notify, and auto-defend actions (e.g. take an infected endpoint offline).

So far, so good. The question is whether our assumptions about "dynamic and ever-changing" AIs fully anticipate how the technology might learn and develop novel approaches.

Go

I recently read futurist Amy Webb's *The Big Nine*, an overview of the evolution of AI and the major companies defining its future. It's an interesting book that addresses not only technology and development but the political, economic and social dynamics shaping the future we'll be sharing with intelligent machines.

The section on Go got me thinking. Since that first big win, AlphaGo and its successors have proven nearly unbeatable by even the most elite human competitors and the current iteration, AlphaZero, is uniformly regarded as the greatest player in the world.

But the remarkable thing isn't that an AI beat a human. Along the way the DeepMind team allowed a second instance of AlphaGo Zero using a larger network to self-train few weeks. It wound up beating the most advanced human-trained version of AlphaGo 90% of the time – using completely new strategies.

Humans were holding the machine back. On its own, the AI learned to think in entirely new ways. As Webb explains:

> Once Zero took off on its own, it developed creative strategies that no one had ever seen before, suggesting that maybe machines were already thinking in ways that are both recognizable and alien to us. What Zero also proved is that algorithms were now capable of learning without guidance from humans…

It meant that in the near future, machines could be let loose on problems that we, on our own, could not predict or solve.

What does this mean for cybersecurity?

First, it suggests AI threats may be able to develop innovative new attack tools and techniques that human hackers (and security professionals) never conceived of. Such an evolution would present CISOs and their teams with the most significant challenge they've ever faced.

However, it also means AI security tools can (as we hinted earlier) develop new and improved cyber defense tactics. Once cybersecurity AIs reach the point where they can teach themselves and engage threats on their own – essentially what AlphaGo Zero did – the challenges facing threat actors grow exponentially.

Which is good, right?

Up next in part two: So, who wins?