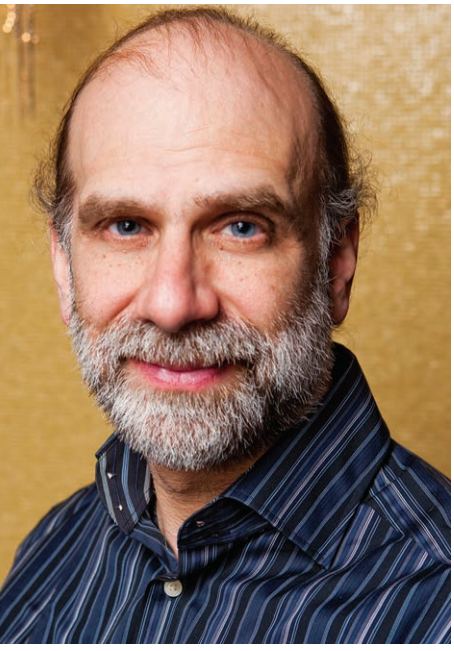

Bruce Schneier
Harvard University

Artificial Intelligence and the Attack/Defense Balance

Artificial intelligence technologies have the potential to upend the long-standing advantage that attack has over defense on the Internet. This has to do with the relative strengths and weaknesses of people and computers, how those all interplay in Internet security, and where AI technologies might change things.

You can divide Internet security tasks into two sets: what humans do well and what computers do well. Traditionally, computers excel at speed, scale, and scope. They can launch attacks in milliseconds and infect millions of computers. They can scan computer code to look for particular kinds of vulnerabilities, and data packets to identify particular kinds of attacks.

Humans, conversely, excel at thinking and reasoning. They can look at the data and distinguish a real attack from a false alarm, understand the attack as it's happening, and respond to it. They can find new sorts of vulnerabilities in systems. Humans are creative and adaptive, and can understand context.

Computers—so far, at least—are bad at what humans do well. They're not creative or adaptive. They don't understand context. They can behave irrationally because of those things.

Humans are slow, and get bored at repetitive tasks. They're terrible at big data analysis. They use cognitive shortcuts, and can only keep a few data points in their head at a time. They can also behave irrationally because of those things.

AI will allow computers to take over Internet security tasks from humans, and then do them faster and at scale. Here are possible AI capabilities:

- Discovering new vulnerabilities—and, more importantly, new types of vulnerabilities—in systems, both by the offense to exploit and by the defense to patch, and then automatically exploiting or patching them.
- Reacting and adapting to an adversary's actions, again both on the offense and defense sides. This includes reasoning about those actions and what they mean in the context of the attack and the environment.
- Abstracting lessons from individual incidents, generalizing them across systems and networks, and applying those lessons to increase attack and defense effectiveness elsewhere.
- Identifying strategic and tactical trends from large datasets and using those trends to adapt attack and defense tactics.

That's an incomplete list. I don't think anyone can predict what AI technologies will be capable of. But it's not unreasonable to look at what humans do today and imagine a future where AIs are doing the same things, only at computer speeds, scale, and scope.

Both attack and defense will benefit from AI technologies, but I believe that AI has the capability to tip the scales more toward defense. There will be better offensive and defensive AI techniques. But here's the thing: defense is currently in a worse position than offense precisely because of the human components. Present-day attacks pit the relative advantages of computers and humans against the relative weaknesses of computers and humans. Computers moving into what are traditionally human areas will rebalance that equation.

Roy Amara famously said that we overestimate the short-term effects of new technologies, but underestimate their long-term effects. AI is notoriously hard to predict, so many of the details I speculate about are likely to be wrong—and AI is likely to introduce new asymmetries that we can't foresee. But AI is the most promising technology I've seen for bringing defense up to par with offense. For Internet security, that will change everything. ■

Bruce Schneier is a security technologist and a Fellow at the Berkman Klein Center for Internet and Society at Harvard University. He's also the chief technology officer of IBM Resilient and special advisor to IBM Security. Contact him via www.schneier.com.

Copublished by the IEEE Computer and Reliability Societies