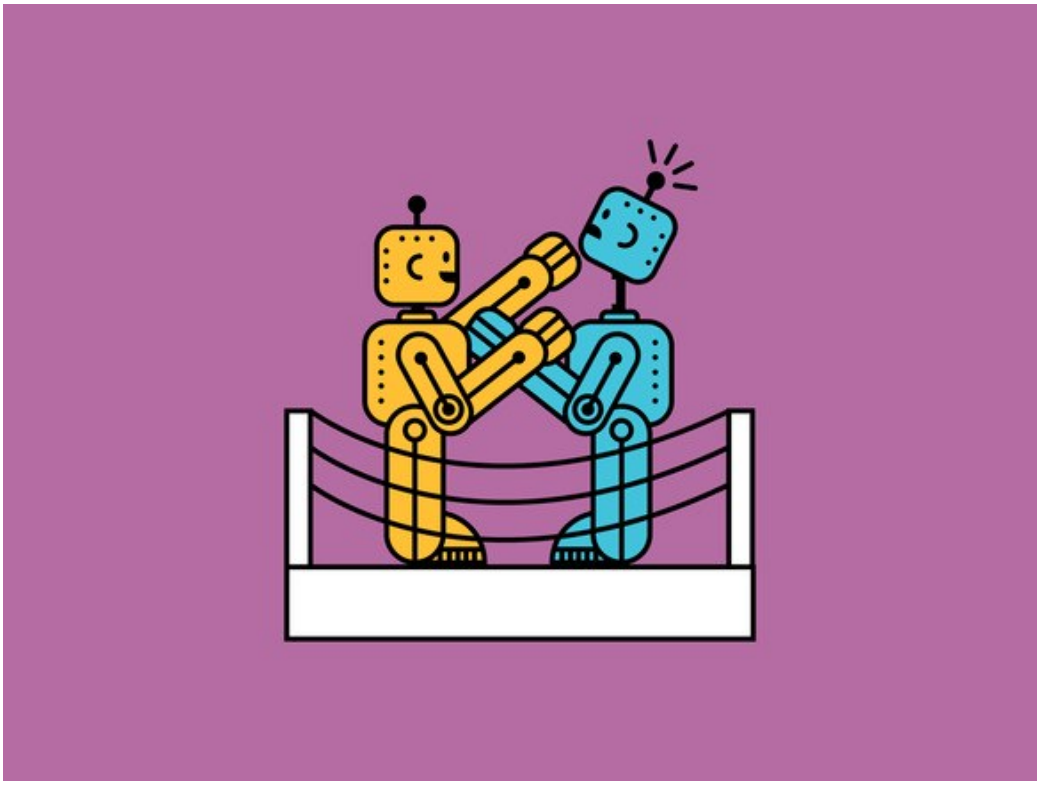
LILY HAY NEWMAN SECURITY 04.29.18 07:00 AM

AI CAN HELP CYBERSECURITY—IF IT CAN FIGHT THROUGH THE HYPE



There are a ton of claims around AI and cybersecurity that don't quite add up. Here's what's really going on.

ALYSSA FOOTE

WALKING THE ENORMOUS exhibition halls at the recent RSA security conference in San Francisco, you could have easily gotten the impression that digital defense was a solved problem. Amidst branded t-shirts and water bottles, each booth hawked software and hardware that promised impenetrable defenses and peace of mind. The breakthrough powering these new panaceas? Artificial intelligence that, the sales pitch invariably goes, can instantly spot any malware on a network, guide incident response, and detect intrusions *before they start*.

That rosy view of what AI can deliver isn't entirely wrong. But what next-generation techniques actually do is more muddled and incremental than marketers would want to admit. Fortunately, researchers developing new defenses at companies and in academia largely agree on both the potential benefits and challenges. And it starts with getting some terminology straight.

"I actually don't think a lot of these companies are using artificial intelligence. It's really training machine learning," says Marcin Kleczynski, CEO of the cybersecurity defense firm Malwarebytes, which promoted its own machine learning threat detection software at RSA. "It's misleading in some ways to call it AI, and it confuses the hell out of customers."

Rise of the Machines

The machine learning algorithms security companies deploy generally train on large data sets to "learn" what to watch out for on networks and how to react to different situations. Unlike an artificially intelligent system, most of the security applications out there can't extrapolate new conclusions without new training data.

Machine learning is powerful in its own right, though, and approach is a natural fit for antivirus defense and malware scanning. For decades AV has been signature-based, meaning that security companies identify specific malicious programs, extract a sort of unique fingerprint for each of them, and then monitor customer devices to ensure that none of those signatures appear.

Machine learning-based malware scanning works in a somewhat similar manner —the algorithms train on vast catalogues of malicious programs to learn what to look for. But the ML approach has the added benefit of flexibility, because the scanning tool has learned to look for characteristics of malware rather than specific signatures. Where attackers could stymie traditional AV by making just slight alterations to their malicious tools that would throw off the signature, machine learning-based scanners, offered by pretty much all the big names in security at this point, are more versatile. They still need regular updates with new training data, but their more holistic view makes a hacker's job harder.

"The nature of malware constantly evolves, so the people who write signatures for specific families of malware have a huge challenge," says Phil Roth, a data scientist at the machine learning security firm Endgame, that has its own ML-driven malware scanner for Windows systems. With an ML-based approach, "the model you train definitely needs to reflect the newest things that are out there, but we can go on a little bit of a slower pace. Attackers often build on old frameworks or use code that already exists, because if you write malware from scratch it's a lot of effort for an attack that might not have a large payoff. So you can learn from all the techniques that exist in your training set, and then recognize patterns when attackers come out with something that's only slightly new."

Similarly, machine learning has become indispensable in the fights against spam and phishing. Elie Bursztein, who leads the anti-abuse research team at Google, notes that Gmail has used machine learning techniques to filter emails since its launch 18 years ago. But as attack strategies have evolved and phishing schemes have become more pernicious, Gmail and other Google services have needed to adapt to hackers who specifically know how to game them. Whether attackers are setting up fake (but convincing-looking) Google Docs links or tainting a spam filter's idea of which messages are malicious, Google and other large service providers have increasingly needed to lean on automation and machine learning to keep up.

As a result, Google has found applications for machine learning in almost all of its services, especially through an ML technique known as deep learning, which allows algorithms to do more independent adjustments and self-regulation as they train and evolve. "Before we were in a world where the more data you had the more problems you had," Bursztein says. "Now with deep learning, the more data the better. We are preventing violent images, scanning comments, detecting phishing and malware in the Play Store. We use it to detect fraudulent payments, we use it for protecting our cloud, and detecting compromised computers. It's everywhere."

At its core, machine learning's biggest strength in security is training to understand what is "baseline" or "normal" for a system, and then flagging anything unusual for human review. This concept applies to all sorts of ML-assisted threat detection, but researchers say that the machine learning-human interplay is the crucial strength of the techniques. In 2016, IBM estimated that an average organization deals with over 200,000 security events per day.

Machine learning's most common role, then, is additive. It acts as a sentry, rather than a cure-all.

"It's like there's a machine learning assistant that has seen this before sitting next to the analyst," says Koos Lodewijkx, vice president and chief technology officer of security operations and response at IBM Security. The team at IBM has increasingly leaned on its Watson computing platform for these "knowledge consolidation" tasks and other areas of threat detection. "A lot of work that's happening in a security operation center today is routine or repetitive, so what if we can automate some of that using machine learning or just make it easier for the analyst?" Lodewijkx says.

The Best Offense

Though many machine learning tools have already shown promising results in providing defense, researchers almost unanimously warn about the ways attackers have begun to adopt machine learning techniques themselves. And more of these types of attacks are on the horizon. Examples already exist in the wild, like hacking tools that use machine vision to defeat Captchas.

Another present threat to machine learning is data poisoning. If attackers can figure out how an algorithm is set up, or where it draws its training data from, they can figure out ways to introduce misleading data that builds a counter-

rrative about what content or traffic is legitimate versus malicious. For example, attackers may run campaigns on thousands of accounts to mark malicious messages or comments as "Not Spam" in an attempt to skew an algorithm's perspective.

In another example, researchers from the cloud security firm Cyxtera built a machine learning-based phishing attack generator that trained on more than 100 million particularly effective historic attacks to optimize and automatically generate effective scam links and emails. "An average phishing attacker will bypass an AI-based detection system 0.3 percent of the time, but by using AI this 'attacker' was able to bypass the system more than 15 percent of the time," says Alejandro Correa Bahnsen, Cyxtera's vice president of research. "And we wanted to be as close as possible to how an actual attacker would build this. All the data was data that would be available to an attacker. All the libraries were open source."

Researchers note that this is why it is important that ML systems are set up to encourage "human in the loop," so systems aren't sole, autonomous arbiters. ML systems "should have the option to say 'I have not seen this before' and ask help from a human," says Battista Biggio, an assistant professor at the University of Cagliari, Italy, who studies machine learning security. "There's no real intelligence in there—it's inferences from data, correlations from data. So people should just be aware that this technology has limitations."

To this end, the research community has worked to understand how to reduce the blind spots in ML systems so they can be hardened against attacks on those weaknesses. At RSA, researchers from Endgame released an open source threat data training set called EMBER, with the hope that they can set an example, even among competing companies, to focus on collaboration in security ML. "There are good reasons that the security industry doesn't have as many open data sets," Endgame's Roth says. "These kinds of data might have personally identifying information or might give attackers information about what a company's network architecture looks like. It took a lot of work to sanitize the EMBER dataset, but my hope is to spur more research and get defenders to work together."

That collaboration may be necessary to stay ahead of attackers using machine learning techniques themselves. There's real promise behind machine learning in cybersecurity, despite the overwhelming hype. The challenge is keeping expectations in check.

Machine vs Machine

- IBM Watson went to work in cybersecurity nearly two years ago
- MIT has also been working on intelligent solutions
- And if you're still curious about what artificial intelligence can and can't do, read our full guide here

RELATED VIDEO

SECURITY

How to Protect Yourself After a Massive Corporate Hack

It seems like every time you turn around there's a new breach of personal information. Follow these steps to minimize the damage.

#ARTIFICIAL INTELLIGENCE #MACHINE LEARNI ECURITY #ARTIFICIAL INTELLIGENCE

VIEW COMMENTS

SPONSORED STORIES

POWERED BY OUTBRAIN

ICE POP
[Pics] Photoshop Didn't Exist Back Then, So Yes This Is Real

THE CHEAT SHEET
The Most Corrupt States in America to Live In

BNY MELLON WEALTH MANAGEMENT
Maintain Your Family Estate's Wealth With These 3 Key Points

GLASSESUSA
Glasses-Wearers Are Going Crazy Over This Website

MY ANTIVIRUS REVIEW
Top 10 Antivirus For Mac Users. #1 Is Free. (2018)

MORE SECURITY

WIRED25

Robert Mueller Has Already Told You Everything You Need To Know

EMILY DREYFUSS

SECURITY ROUNDUP

Kanye's Password Literally Couldn't Be Worse

EMILY DREYFUSS

MALWARE

Fake Adobe Flash Installers Come With a Little Malware Bonus

BRIAN BARRETT

BREACHES

How Facebook Hackers Compromised 30 Million Accounts

LILY HAY NEWMAN

SOCIAL MEDIA

How to Check If Your Facebook Account Got Hacked

BRIAN BARRETT

BREACHES

No One Can Get Cybersecurity Disclosure Just Right

LILY HAY NEWMAN

GET OUR NEWSLETTER

WIRED's biggest stories delivered to your inbox.

Enter your email

SUBMIT

FOLLOW US ON PINTEREST

See what's inspiring us.

FOLLOW

SUBSCRIBE	ADVERTISE
SITE MAP	PRESS CENTER
FAQ	ACCESSIBILITY HELP
CUSTOMER CARE	CONTACT US
SECUREDROP	T-SHIRT COLLECTION
NEWSLETTER	WIRED STAFF
JOBS	RSS

CNMN Collection