www.pipelinepub.com
Volume 18, Issue 1

Cybersecurity Infrastructure Needs a New Model

By: Sam Jones

Cybersecurity systems are ripe for disruption. Over the years, individual tools have proliferated, each with its own data format, causing a deluge of disparate data. In addition, there is a global shortage of skilled cybersecurity analysts who can evaluate that data (and they are very expensive if you can find them). Finally, and perhaps most critically, hackers are getting smarter and more creative all the time.

Artificial intelligence was supposed to be the cure for these issues, but it has been of limited use in addressing the problem at scale because it requires large, thoughtfully planned infrastructure. In this article, we'll look at the role of AI in cybersecurity systems and how it can become a truly transformative technology.

AI as snake oil

AI is mentioned a lot in marketing literature describing cybersecurity solutions, but so far, it hasn't been as transformative as you might think. Despite a market size that grows at a 20.5 percent compound annual growth rate, AI still remains operationally difficult to deploy on security problems. If you were to walk into a modern security operations center (SOC), you'd probably find some big TVs with some difficult-to-read dashboards and CNN, and security analysts that likely find their jobs painful because they are spending their time manually correlating data and trying to discern what is happening at their enterprise in the face of ever-more-complex attacks. If humans are doing this, it begs the question, "Where is the AI?"

Cybersecurity is a messy operational problem, and this is the simple reason why AI has been slow to transform it. Finding threats in an enterprise across hundreds of sources of telemetry when threats often look identical to normal activity is a very difficult problem. Moreover, data from each security tool can take different forms, and it must be normalized before it can be used to train an AI system.

Regardless of the industry and use case, AI learns from data: the AI engine must be trained with data so it can begin to learn what is or isn't an anomaly. This is what is so messy about the security problem: every enterprise's security data looks, at minimum, a little different, with different tools and behavior patterns, and at maximum, the data looks wildly different. There is no golden training dataset in security that can be licensed like there might be for image or speech recognition systems. If you want to use AI to address the security problem, you have to create and acquire your own data.

Normalizing data so it's useful to an AI engine is a huge challenge. The problem is so valuable that Scale AI, a startup that creates data APIs for AI development primarily focused on driverless car applications, snagged a $7 billion valuation less than five years after its founding. Scale AI already counts many of the world's most innovative organizations as its customers.

What transformative AI will take

AI in security will eventually be transformative, likely both for offense and defense, but that is a story for another day. Here, "transformative" means broadly transformative, across all parts of security, so it fundamentally alters how an enterprise goes about security. For now, we have to be content with some limited applications where AI can improve security.

Still, there are some bright spots for AI in security; these are easy to find by thinking through the data problem. What parts of the security stack generate clean, trainable data? Email fraud and malware detection are two great examples: the AI engine can learn from available phishing examples or malware signatures and spot similar exploits. Data across customer emails and malware sandboxes can be used to train AI models that power enterprise products. The same training is much harder to implement on problems like detecting attacks that move laterally through a network(say, from the firewall to the Active Directory server to a data server), because this lateral movement will look a little different at every enterprise.

Creating AI that will broadly protect an enterprise across all its digital operations will in ways resemble efforts being taken by driverless car companies today. For example, since 2009, Waymo's driverless car software has trained on over 15 billion miles of simulated driving and more than 20 million miles of public driving experience. Waymo has a rigorous approach to testing at different levels of fidelity (simulation, closed course, real world), executing scenarios with thousands of variations, all the while collecting data for the purpose of improvement.

This isn't a perfect analogy for AI in security, but it is pretty good—testing with simulated data, testing in lab environments with simulated or real attacks, and testing in real-world operations

across a diverse set of enterprises. Security problems with natural access to cleaner data will emerge with truly AI-powered products sooner than the harder data problems across the entire enterprise security stack. It is going to take time and capital to get there, and innovations that are ruthlessly focused on the data problem will be first and foremost to unlock broad transformation. Today, many security tools just don't focus on data normalization because they tend to be siloed in specific pain points in the overall infrastructure.

What transformative AI in security will look like

Imagine that every IT initiative, configuration, security log, and alert could be reviewed by the world's leading human security expert in that given area in real time, with no disruption to business operations. Imagine that enterprise analysts could consult with and get direction from that expert. AI in security will eventually feel like that.

How? Products that are built on thoughtful data assets, that reduce data complexity, will ultimately be category kings; otherwise, the product won't work from customer to customer, and it will be a product with service-like margins and won't scale. In fact, Andreessen Horowitz found that most of its enterprise AI companies have much lower margins than comparable SaaS businesses because of the inherent costs of building and scaling AI.

These future category kings will first have to invest in data infrastructure and collection, likely for years, before their data can truly be considered an asset and assist in the self-improving nature of their product. However, once these company kings obtain a real data asset for AI, their pace of innovation will be difficult (if not impossible) to match by competitors, and they will be crowned a category king, as long as they still manage to maintain an intuitive product. Just as the search engine category quickly consolidated to Google, the same will happen with data-intensive cybersecurity solutions. Specifically, look for major consolidation in security information and event management (SIEM), eXtended detection and response (XDR), endpoint detection and response (EDR), and network detection and response (NDR) markets.

AI is emerging in security first on smaller problems where there is less data complexity, as noted in the email fraud and malware examples earlier. AI will then slowly deploy to more complex data problems, but only products that are ruthlessly focused on managing data complexity will emerge with meaningful AI engines. To be effective, an AI-driven security program must be able to collect data from all available security tools and threat feeds, and then normalize this data so that it's useful for training the AI engine. This is what AI's future in cybersecurity will look like.