Beyond the Hype: Where Artificial Intelligence Excels in Cybersecurity

An examination of the state of AI across 7 key use cases

Vina Nguyen

Executive Summary

Cybersecurity talent has been difficult to retain and recruit, and the outlook does not seem much better. To solve this challenge and the reasons behind it, we look to artificial intelligence (AI) to perform what our teams lack the capacity to do. However, AI is not the silver bullet that some might claim. The key to making AI work for you is to understand what AI is capable of, the use cases for which it's best suited, and how to compensate for its limitations. In this white paper, we examine the state of AI across seven use cases selected to reflect the cybersecurity lifecycle of a typical enterprise operation.

Introduction

The challenge of finding cybersecurity talent has only worsened over time. From 2013 to 2022, the number of unfilled jobs jumped from 1 million to 3.5 million worldwide—a gap of 350% in nearly 10 years.[1] With known hurdles such as alert fatigue, disparate tools, and attacks around the clock, it's no wonder this talent gap is so hard to fill. Add the rise of Ransomware-as-a-Service and an increasing number of threat actors with nation-state ability, and cyber defense becomes an uphill battle.

To keep up with cybersecurity threats, we look for answers in artificial intelligence (AI) technology. If AI can be accurate, faster, and more cost-effective than the status quo, it's an option worth considering. To determine the worth of investment, we ask: What use cases can AI actually solve? How good is AI at doing what it's intended to do? What are the alternatives to AI?

Arriving at answers can be difficult to discern amid the noise of marketing and media. Some enthusiastic vendors say AI is the one-stop solution while skeptical others say AI is just being hyped up—again. After all, starting with the 1950s, AI endured multiple cycles of "giddying hype and humiliating collapse" before rapidly advancing in the

Contents

2000s with the speed of graphics processing units (GPUs), when techniques long theorized were finally able to come to fruition.[2] Such techniques include the foundations of deep learning, which today powers drug discovery[3], self-driving cars[4], and automated malware creation[5].

How do these advancements translate to protecting the enterprise? AI's stock value can vary widely depending on the use case and its performance. In this paper, we define AI in relation to cybersecurity, map our selected use cases to a framework, cover the capability of AI per use case, and recommend paths forward for AI integration in cyber defense, both short-term and long-term.

Artificial Intelligence Defined

In cybersecurity and other various industries, AI is an overloaded term. In many cases, what's called AI is really machine learning (ML), a subset of AI. This

distinction matters because AI can imply humans are no longer needed to guide the technology, while the other assumes humans (i.e., cybersecurity professionals) are still part of the loop. Google describes it well: "While artificial intelligence encompasses the idea of a machine that can mimic human intelligence, machine learning does not. Machine learning aims to teach a machine how to perform a specific task and provide accurate results by identifying patterns."[6]

If the goal of artificial intelligence is to replicate a human with reason, the goal of machine learning is to learn a specific task with data. Specific tasks can include seeing (computer vision) and reading (natural language processing). ML models learn how to do one task very well by learning patterns from a corpus of data and improving performance the more data it receives. Most of the use cases in this paper are applications of ML, and we will use AI/ML going forward to refer to the two terms.

The NIST Cybersecurity Framework: Our Use Cases

To ensure a holistic view, we map our seven use cases to the National Institute of Standards and Technology (NIST) Cybersecurity Framework, as seen in Table 1. (Keep in mind that the NIST Cybersecurity Framework is evolving—check for version 2.0 in 2024).[7] We group the functions "Identify" and "Protect" together to refer to preemptive measures aimed to reduce an attack surface. Similarly, we group "Respond" and "Recover" together as the functions of analyzing an incident and returning to business operations often go hand in hand.

For each use case, we describe the current capability, limitations, and potential for AI/ML. We note if AI/ML has nearly assumed a task, provides a limited but necessary benefit, or is still in its infancy. We also note any non-AI/ML methods that effectively solve the use case as well.

Identify and Protect

The first two functions of the NIST Cybersecurity Framework are "Identify" and "Protect". We've grouped the two in this paper to refer to any preventative measures aimed at reducing an organization's attack surface. While these functions encompass everything from identifying key assets to establishing policies, we go over two specific use cases where AI/ML could provide the most benefit in reducing labor-intensive activity.

Source Code Analysis

Unlike civil engineering, which leaves little room for error, software engineering accepts that bugs in code are part of the deal. However, when bugs become entry points for an attacker (such as a buffer overflow that allows remote exploitation), they become vulnerabilities that should be patched before deployment.

Research in using AI/ML to detect vulnerabilities in source code has significantly grown since 2015[8] and has only recently moved from prototypes to production. Gitlab, for example, recently released its ML-based static analysis tool to the public in 2022[9]. The biggest limitation in advancing this research has been generating a large enough, realistic data set that AI/ML can learn from. Factors like the

Table 1. Use Cases Mapped to NIST Cybersecurity Functions

NIST Cybersecurity Function	Use Cases
Identify and Protect	Source Code Analysis
	Vulnerability Prioritization
Detect	Spam Detection
	Endpoint and Network Security
	Data and Event Correlation
Respond and Recover	Incident Analysis and Reporting
	Mitigation and Recovery

programming language chosen also add to the complexity; Gitlab's tool, for example, works for only JavaScript and TypeScript.

Industry-leading alternatives include static application security testing (SAST) tools based on formal methods, an approach of rule-based mathematical modeling vs. pattern-based machine learning. While an essential tool in safety-critical fields like medical devices and aircraft navigation, a common issue in formal-based SASTs is the time to compute and a significant proportion of false positives. In a 2023 journal review of 6 popular SAST tools against 47 code bases, researchers found "little to no agreement among the tools and a low degree of precision."[10] While AI/ML is by no means better yet, the performance of current tools leaves room for new approaches.

Vulnerability Prioritization

Vulnerability scanning is a mostly automated process that refers to checking a system against a database of known vulnerabilities. These vulnerabilities include those found in source code, network device configurations, and database compliance checks. While vendors may have their own databases, most rely on the National Vulnerability Database (NVD) maintained by the U.S. National Institute of Standards and Technology. The NVD currently indexes over 200,000 vulnerabilities, each assigned with a Common Vulnerabilities and Exposures Identifier (CVE ID).[11]

Once the scanning is done, however, the real work begins. The number of vulnerabilities output by these scanning tools can range from 100s to 1000s of vulnerabilities. To prioritize what needs patching first, one factor that analysts use is vulnerability severity. The industry standard for measuring the severity of a CVE ID is the Common Vulnerability Scoring System (CVSS), which was launched in 2005. CVSS is an equation that produces a total score based on categorical metrics manually assigned by analysts.[12] For example, one metric is "attack complexity" which can be "low" or "high". An analyst currently decides this rating for each vulnerability by using their subject matter expertise.

> *A 2023 journal review of six popular static application security testing tools found minimal consensus and low precision. The performance of current tools leaves room for new AI/ML approaches.*

As a complement to the manually-scored CVSS, in 2019 researchers proposed a new metric for vulnerability severity: the Exploit Prediction Scoring System (EPSS), an AI/ML-derived probability of exploitation.[13] While EPSS outperforms CVSS on accuracy, efficiency, and coverage[14], like the CVSS, it still relies on known vulnerabilities with published CVE IDs. For software vendors or bug bounty programs who are seeking to find zero-days, neither the EPSS or CVSS provides added value.[15]

Using the EPSS provides another perspective to complement standard scores like the CVSS, assuming you know a vulnerability exists. If you're hoping AI/ML can find and assess unknown vulnerabilities to prevent zero-day exploits, the performance of current models makes it unsuitable for production, and future research is needed to advance it.[16]

Detect

The third function of "Detect" in the NIST Cybersecurity Framework refers to detecting any suspicious behavior that may indicate an attack. The bulk of AI/ML in cybersecurity falls under this function, and logically so, as attacks follow certain patterns that are a good fit for ML's pattern-recognition capability. While the maturity of AI/ML in this function exceeds that of vulnerability identification, note that with the exception of spam

detection, the significant proportion of false positives in these use cases typically necessitates having subject matter experts on hand to interpret and sort through the results.

Spam Detection

Detecting spam has become so commonplace that we may forget AI/ML powers it. Since the 2000s where ML techniques first were applied to spam,[17] Google now blocks an extra 100 million spam messages a day as of 2019 with ML.[18] With some models performing at >95% accuracy,[19] AI/ML advancements in spam detection are focused primarily on optimization. In an environment where phishing is an ever-popular attack vector, increasing 61% from 2021 to 2022,[20] AI/ML plays an important role as an automated first defense in email security. Most organizations have this capability built into their systems through their email vendor; to augment it, staff training can reduce the success of a phishing email that escaped the spam filter.

Endpoint and Network Security

AI/ML supports endpoint and network security, but not at the level of automation in email security. We've grouped these domains together because at its core, AI/ML is doing the same functions of behavioral analytics and anomaly detection, applied to endpoint and network environments.

> *While AI/ML approaches for detection are better than none, the complexity of endpoint and network environments creates a significant proportion of false negatives, requiring a talented team to interpret the results.*

Endpoint security solutions like Endpoint Detection and Response (EDR) monitor anything on the host (device) to detect the unusual. Data monitored include system logs, processes, file activity, user actions, and network traffic coming in/out of that device. This approach is the same one used for User and Entity Behavior Analytics (UBA/UEBA), with data collected per user or entity vs. endpoint. Providers like Blackberry, CrowdStrike, and Microsoft are applying AI/ML to understand an individual user's behavior (files accessed, applications used) and device environment (type, configuration) to develop customized models of what's normal.[21] Ongoing research focuses on expanding the data set to include web browsing behavior.[22]

The same techniques—identifying what's normal and what's not—are used to identify potential intrusions in network security via solutions like network intrusion detection systems (NIDS). Instead of system logs or files, these AI/ML models analyze network traffic and logs from network infrastructure devices like switches, routers, and firewalls.

Considering that adversaries use AI/ML as well, having AI/ML methods for detecting an attack is probably better than none. AI/ML detection still produces too many false positives though, leaving room for improvement. A 2020 survey of NIDS concludes that one of the biggest obstacles to higher performance is the lack of a data set reflective of modern networks and attack techniques.[23] In the field of EDR, where half of the alerts signaled by EDR tools are false positives,[24] alert fatigue becomes the reason for burnout of staff and lack of retention.[25]

While AI/ML approaches are better than manually sorting through computer data, they aren't stand-alone solutions. The complexity of the endpoint and network environments creates a significant proportion of false negatives, which requires a talented team to sort through the results. In the meantime, researchers continue to work on optimizing the technology.

Data and Event Correlation

With all the data and security events coming in from endpoints, networks, logs, and cloud data, the next use case would be how to correlate all the data to threat intelligence and detect an actual incident across an enterprise. Add in the fact that the tools that input data and events talk in different languages, and it takes a highly skilled professional to interpret the data and correlate them to a higher-level story.

This idea of correlation across all normalized data is Gartner's definition of the latest security evolution in Extended Detection and Response (XDR), a term coined in 2018.[26] From industry leaders offering their own XDR solution[27] to startups promoting an open XDR platform[28], vendors are betting that their AI/ML-based solutions will help organizations respond faster than an already stretched-thin cybersecurity team would. So far, the results show promise; organizations with XDR shortened a data breach by 29 days, a 10% increase from those without XDR.[29] As the market and technology mature, AI/ML could solve one of the biggest obstacles to enterprise security.

Respond and Recover

The two functions of "Respond" and "Recover" from the NIST cybersecurity framework relate to summarizing what happened, figuring out what to do, and recovering to sustain business operations. AI/ML technology here is emerging but has great potential.

Incident Analysis and Reporting

With the release of ChatGPT in November 2022, this game-changing generative AI/ML technology was quickly applied to use cases across multiple industries. One of the main downsides, however, was the concern that ChatGPT-generated text looked and sounded correct but was factually inaccurate. Now, this may be less of a concern, as researchers have optimized its performance through training on very specific use cases.

Enter ChatGPT for security. Microsoft[30] released an AI assistant in March 2023, and both Recorded Future[31] and SentinelOne[32] released one of their own in April 2023. With prompts like "Create a single PowerPoint slide outlining the incident and the attack chain"[33] and "Are companies in my industry being targeted by ransomware attacks?"[34], these technologies claim to remove the burden of often highly-technical search. As these technologies are for limited preview only, the impact remains to be seen. If these assistants can maintain accuracy, be transparent about their limitations, and provide workarounds that compensate for these limitations, these generative AI assistants will be transformative for the cybersecurity industry.

Mitigation and Recovery

After an incident is detected and analyzed, an organization needs to mitigate its effect and return to normal business operations. This may include patching a vulnerability or moving to backup systems. Beyond a rule-based playbook, AI/ML is needed here to reason—to weigh all the options in a changing environment and decide which action makes the most sense.

In 2016, this ability came to life on the grand stage. The Defense Advanced Research Projects Agency (DARPA) hosted the Cyber Grand Challenge, the first fully autonomous Capture The Flag (CTF) competition, where AI/ML systems hack each other's systems to gain assets and patch vulnerabilities in real-time[35]. The winning team played in the annual DefCon CTF Finals later that year, ending last but still within the score of the other human teams.

As David Brumley, Professor at Carnegie Mellon University and member of the winning team notes, the challenge was much more about security; "it was about making decisions that operated within a

confined space to make sure that it wasn't just the most secure but also maintained performance and functionality."[36] The technology still has some way to go from here to operation, but it is an ongoing research field with promise.

Summary

AI/ML plays an important role in giving us accuracy and speed when the amount of data is simply not possible to process, especially with the lack of talent we face today and tomorrow. In Table 2, the AI/ML capability we've assessed is listed alongside each use case mentioned here.

- **Low**: Probably better to use non-AI/ML techniques in the short-term. Technology is still in its infancy.

Note that each organization will have its own tolerance for risk to adopt newer (less capable) technologies first or wait until the market settles on standards and the research optimizes performance.

Recommendations

For the short-term, expect to use AI/ML side-by-side with security teams, particularly for use cases like endpoint detection, where their limitations in high false positive rates must be managed by

Table 2. Summary of AI/ML Capability by Use Case

NIST Cybersecurity Function	Top Use Cases	AI/ML Capability
Identify and Protect	Source Code Analysis	Low
	Vulnerability Prioritization	Medium
Detect	Spam Detection	High
	Endpoint and Network Security	Medium
	Data and Event Correlation	Low*
Respond and Recover	Incident Analysis and Reporting	Low*
	Mitigation and Recovery	Low

*AI/ML to Watch in 2023

We've labeled AI/ML capability of high, medium, or low based on how accurate and fast AI/ML is, or rather, how much they make life easier for a human analyst.

- **High**: AI/ML automates nearly everything with low false positives, and compensating for its mistakes takes minimal overhead. Advancements here are mainly on optimization.
- **Medium**: Definitely a benefit vs. no AI/ML; having an expert to sort through false positives or other limitations of the AI/ML would leverage these solutions the best. Advancements here are on improving accuracy.

cybersecurity professionals who can interpret and analyze the results. Alternatively, outsource this work to a managed provider who will do it for you—and well. For use cases like spam detection where AI/ML operates at high capability, adopt them if not already, and compensate for what it doesn't do through staff training.

For the long-term, keep an eye out on emerging technologies, especially for use cases that represent your biggest blockers. We note in Table 2 our top technologies to watch: (1) XDR for data correlation and (2) generative AI assistants for incident response. The market has seen first-of-a-kind prototypes for both these technologies in the last few years, and we expect to see multiple iterations in the upcoming months. If successful, they will have

an enormous impact on reducing the time and burden in a security operations workflow.

Keep in mind that the use cases here aren't exhaustive; AI/ML is also being researched for fields such as adversarial AI[37] and threat actor attribution.[38] The advancement of AI/ML could change rapidly, and keeping track of the state of AI/ML will keep you informed of opportunities to stay ahead of the game in cyber defense.

References

[1] Companies are desperate for cybersecurity workers—more than 700K positions need to be filled | Fortune
[2] What the history of AI tells us about its future | MIT Technology Review
[3] Deep learning in drug discovery: an integrative review and future challenges | SpringerLink
[4] End-to-End Deep Learning for Self-Driving Cars | NVIDIA Technical Blog
[5] I Built a Zero Day Virus with ChatGPT | Forcepoint
[6] AI vs. Machine Learning: How Do They Differ? | Google Cloud
[7] Cybersecurity Framework | NIST
[8] Machine Learning for Source Code Vulnerability Detection: What Works and What Isn't There Yet (computer.org)
[9] Leveraging machine learning to find security vulnerabilities | The GitHub Blog
[10] A critical comparison on six static analysis tools: Detection, agreement, and precision - ScienceDirect
[11] NVD - NVD Dashboard (nist.gov)
[12] CVSS v3.1 Specification Document (first.org)
[13] The EPSS Model (first.org)
[14] Predictive Vulnerability Scoring System - Black Hat USA 2019 | Briefings Schedule
[15] Probably Don't Rely on EPSS Yet (cmu.edu)
[16] A review of Machine Learning-based zero-day attack detection: Challenges and future directions - ScienceDirect
[17] A Plan for Spam (paulgraham.com)
[18] Spam does not bring us joy—ridding Gmail of 100 million more spam messages with TensorFlow | Google Workspace Blog
[19] MAKE | Free Full-Text | A Survey of Machine Learning-Based Solutions for Phishing Website Detection (mdpi.com)
[20] Phishing attacks are increasing and getting more sophisticated (cnbc.com)
[21] Where CISOs rely on AI and machine learning to strengthen cybersecurity | VentureBeat
[22] User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user - ScienceDirect
[23] Network intrusion detection system: A systematic study of machine learning and deep learning approaches - Ahmad - 2021 - Transactions on Emerging Telecommunications Technologies - Wiley Online Library
[24] Half of Alerts Signaled by EDR Tools Are False Alarms; Lack of Personnel Prevents Rapid Detection and Response (bitdefender.com)
[25] Information overload, burnout, talent retention impacting SOC performance | CSO Online
[26] What Is Extended Detection and Response? (blackberry.com)
[27] XDR- Extended Detection and Response - Palo Alto Networks
[28] Know Everything about XDR, Extended detection and response (stellarcyber.ai)
[29] Cost of a data breach 2022 | IBM
[30] With Security Copilot, Microsoft brings the power of AI to cyberdefense - Stories
[31] Recorded Future Announces World's First AI for Intelligence
[32] Message from SentinelOne
[33] Microsoft Security Copilot | Microsoft Security
[34] Introducing Recorded Future AI: AI-driven intelligence to elevate your security defenses
[35] Mayhem, the Machine That Finds Software Vulnerabilities, Then Patches Them - IEEE Spectrum
[36] The Cyber Grand Challenge and the Future of Cyber-Autonomy | USENIX
[37] Adversarial AI and the dystopian future of tech | VentureBeat
[38] Going ATOMIC: Clustering and Associating Attacker Activity at Scale | Mandiant

About the Author

Vina Nguyen is a freelance writer and cybersecurity subject matter expert specializing in machine learning, software analysis, and security operations. She has over 10 years of technical experience as a former computer security researcher at places including the U.S. Department of Defense and the Johns Hopkins University Applied Physics Laboratory. She currently provides copy and content services to technology companies in cybersecurity and artificial intelligence.

www.vinawrites.com | vina@vinawrites.com