Google Cloud

AI and automation in security

So far, you've learned that automation is the use of technology to reduce the need for human interaction to perform common and repetitive tasks. Automation is especially important in a security setting where it can assist analysts with performing a variety of tasks, like monitoring networks, checking logs, and conducting tests. With the rise of artificial intelligence (AI) in the security field, automating tasks becomes even more useful. In this reading, you'll learn more about AI and automation in the security analyst's role. You'll also learn about some of the risks analysts must consider when implementing automation and AI.

AI automates tasks

Automation reduces human error by reliably completing repetitive tasks in the same way every time. Automation is an important tool in the security analyst's toolkit, and when coupled with AI, it integrates into tasks like vulnerability scanning, incident response, and threat detection.

Types of automated AI security tasks

- **Scan code:** Throughout the software development lifecycle, AI can automate tests on code to check that performance and usability requirements are met. This is helpful for security analysts because automated tests can help identify bugs or errors in the code that can lead to security vulnerabilities. AI streamlines this process, helping reveal defects quickly for security analysts to remediate.
- **Respond to security alerts:** Generative AI (GenAI) is useful when responding to security alerts or threats. For example, AI-assisted security tools can help proactively find abnormal network behavior. Then, GenAI can create and deploy a rule that helps block the threat.
- **Detect malware:** Malware is an ever present threat across industries. GenAI helps combat malware with fine-tuned training. The model is fine-tuned with a large amount of malware samples, which allows it to identify and categorize patterns of different types of malware. Then, security analysts can use these patterns to build stronger systems for malware detection. This approach leads to a faster response time for malware threats.
- **Conduct tests:** AI is useful for conducting penetration (pen) tests. A pen test is a process where a red team simulates cyber attacks to identify vulnerabilities in an organization's security infrastructure. Pen testing can involve manual, human-driven tasks, or automated scripts that perform attacks. While pen tests have the ability to identify many issues, it's challenging to detect all of a system's vulnerabilities. AI helps

relieve some of the manual burden of pen testing by automating tasks, like analyzing the system, and using tools to find vulnerabilities, then exploiting those weaknesses.

AI considerations

AI is a helpful tool to use when automating tasks, but security analysts are a necessary component to ensure the AI works appropriately. AI is susceptible to certain risks that require human intervention to fix. Hallucinations, bias, and inaccuracies are all examples of risk that AI might introduce into workflows.

Hallucinations

A hallucination is when an AI model generates words or phrases that don't make sense. For example, if a model is trained with incorrect data, it can lead to hallucinations. And, if a model isn't trained with enough data, or is not given enough constraints, it can also lead to hallucinations.

For example, a security analyst is compiling information like log data and access privileges related to a recent breach. The analyst decides to use a large language model (LLM) application to help create a summarized report that they can present to their team lead. When the analyst reviews the generated summary, they realize the report includes data that's not related to the security incident. This scenario demonstrates how a hallucination can lead to fabricated information. The LLM likely didn't put enough constraints on the training data, and fed the model a surplus of information that is irrelevant to the organization.

Bias

Bias in AI happens when the training data or algorithms produce predictions that reflect societal inequalities. Bias can be introduced in the machine learning process when the model is fed training data. Biased training data put into the model will result in biased predictions as the output. An AI model that's trained to always assume attacker IP addresses have specific signatures or ranges is an example of harmful bias. While training the model, this doesn't cause an issue, since it's a controlled environment. But, in a production environment, this same AI may decide that IPs originating from entire segments of the internet, or from a specific country, are threats. This bias error could effectively shut down the organization's services to a global audience, or prevent their services from working entirely.

Inaccuracies

Data inaccuracies are another risk AI can introduce. LLMs work by using a vast amount of training data to make their predictions. They accomplish this by collecting data from several different sources. Sources that are not adequately evaluated and contain inaccurate data will skew the corresponding training data. A security analyst might use an LLM to help create a

script that automates infrastructure. If the LLM collects data from sources with corrupted scripts, the generated code samples may also contain those defects.

Key takeaways

AI and automation are both tools that security analysts can use to streamline completing repetitive tasks. Automated AI can expedite scanning code, detecting and responding to threats, conducting tests, and more. AI risks illustrate the importance of monitoring the type of training data provided to models. AI has an ability to automate repetitive tasks, but using AI also requires a human element to ensure it works appropriately with an organization's existing system. It's crucial that security analysts apply attention to detail when using AI applications.