Artificial intelligence: New cyber risks, new security capabilities
Artificial intelligence: New cyber risks, new security capabilities

Artificial intelligence (AI) — once thought to be the holy grail for protecting the digital realm and a treasure that would require immense computing power and be economically infeasible — has proved to be surprisingly accessible and affordable in recent years. While AI is proving highly useful for fighting cyber crime, in the hands of malicious actors, AI also can be used for nefarious purposes, raising the question: Is AI a panacea or is it a powerful tool that — in the wrong hands — can be turned against us? And what can today’s enterprises do about it?

Although weaponizing AI and machine learning for cyber security is still in the early stages, most well-funded organizations today are able to build autonomous defensive capabilities that incorporate AI and machine learning to protect their security systems and applications. But unfortunately, they’re not alone in the cyber universe. Cyber attackers have also grown in their understanding of AI and are using it daily to uncover vulnerabilities and cause havoc. This white paper pulls back the curtain on the role of AI and machine learning in cyber crime and cyber security. How serious are AI-based threats? How can AI help improve an organization’s security posture?

Introduction: The rise of autonomous attacks

Not long ago, the computing world had an AI security-related awakening it is still trying to fully decipher.

The Defense Advanced Research Projects Agency (DARPA) had organized the first “Cyber Grand Challenge” competition at DEFCON 2016, one of the world’s largest hacker conventions. Seven teams competed to have their computing machines battle in 96 rounds of a “Capture the Flag” (CTF) game. This was a historic competition, as it was the first time machines were poised to hack each other while at the same time defending themselves from attacks coming from the other machines ... all without human intervention during the competition.

Like any CTF competition, these machines were provided new code filled with bugs and security vulnerabilities that the machines not only had to figure out how to patch, but also how to use to attack the other systems. For the first time, the world witnessed how the world of cyber security could leverage artificial intelligence (AI), not only to help improve and defend against security attacks, but also, when used in a nefarious manner, to offensively conduct autonomous attacks on vulnerable systems.

The CTF competition proved that computing requirements and financial challenges were no longer roadblocks to autonomous security attacks.

It also showed that it is possible for a malicious threat actor to build systems that can conduct attacks with little human involvement, threatening the cyber security posture of companies and public sector agencies around the world. Since then, cyber security professionals have been working to understand how to best use AI and machine learning to combat cyber crime.
Cyber risk in the age of AI and machine learning

To understand the impact of AI in cyber security, it is important to fully understand what AI is. As coined in 1955 by American scientist John McCarthy, the term artificial intelligence refers to the ability of a computer program or machine (entity) to learn, think and make decisions like a human being. This is an extremely broad definition, and even the machines taking part in the CTF competition would not completely qualify. Machines may be mimicking how a human being may think, but they are still programmed to behave in a certain way to perform learning and decision making. It is not an innate ability, and machine learning is the capability programmed into these machines to make them possess AI behavior.

While the term “machine learning” is often used interchangeably with “AI” it is important to note that they are completely different. It is possible — if not realistic — to achieve AI without machine learning. One way is to write billions of lines of code and generate decision trees to emulate the AI behavior, but hardcoding the intelligence is largely impractical. A more efficient way is to teach the AI entity to learn and make decisions by exposing it to large amounts of training data and by using algorithms to teach it to recognize patterns. In more advanced cases, the AI entity may also be taught to modify the algorithms to derive different knowledge from the same training data.

Machine learning is the most practical way to build AI capabilities, and in the last 3 years, it has become the building block of every AI engine.

In its essence, machine learning is engineered to recognize patterns at scale. For example, an AI entity can be supplied with numerous sets of photos that depict buses in various colors, shapes and sizes. Eventually the AI entity will build a machine learning model that is a three-dimensional rectangular entity with lots of rectangular holes and four circle-shaped wheels, which would probably be a bus.

Pattern recognition can also be applied for activity-based patterns, which is where security events largely fall. In building machine learning models to interpret security events, the AI entity should analyze datasets and look for and understand the following patterns:

• Who is logging into or currently accessing the system?
• What is being accessed in the system?
• From where is it being accessed?
• When is it being accessed?
• How is it being accessed?

By combining the answers to these five questions, the AI entity can establish a transaction behavior and then match this behavior to past behaviors. If the behaviors are not statistically consistent, an anomaly will be deemed to have occurred, and in the security context, this could indicate a security incident taking place.

As an example: A staff member could log in from the same workstation every day at 9 a.m. and log out at 5 p.m., and throughout these hours, for an entire year, access only the file server and word processor files. The AI entity can easily establish the baseline that any activities occurring with these patterns are deemed normal
behavior. However, if this staff member starts accessing spreadsheet files at 9 p.m. from a workstation they do not normally access, the activity will be deemed an anomaly and will be flagged immediately.

In some situations, various users may log in and out at different times of the day. Traditional rule-based threat-detection techniques cannot keep up with such diverse variations, as the human administrator would have to anticipate and individually define every user’s behavior to establish their normal login pattern. Such a method is certainly not conceivable in a large environment, where it is impractical to predict every user’s behavior manually.

**How bad guys and good guys exploit machine learning**

Most cyber security controls today that advocate the use of machine learning use one or all of three general categories of machine learning algorithms:

- **Supervised learning** “trains” an AI engine to understand patterns of behavior through large amounts of data in sets preselected by human operators for the specific task at hand.

- **Unsupervised learning** allows an AI engine to develop specific pattern learning models based on datasets that are not classified or categorized.

- **Reinforcement learning** helps the AI engine learn through interaction with the environment and by receiving rewards for performing optimal actions, such as improving performance or yielding better data quality.

All three machine learning methods are well-suited to developing specific AI capabilities and are equally relevant for both offensive and defensive cyber security. Unfortunately, they are also well-suited for use by malicious actors.

An AI engine may use supervised learning to understand patterns from a large number of binary applications that are known to have code vulnerabilities related to a specific class, such as buffer overflow vulnerabilities. Once trained, the AI engine can be used to look for that class of vulnerabilities, regardless of how the executed code mimics certain specific behavior. These techniques may be seem useful for cyber security professionals who want to improve and secure an application, but they are also useful for an adversary who wants to look for vulnerabilities quickly rather than manually.

Likewise, both an ethical security professional and a malicious adversary can use unsupervised learning to develop fuzzing techniques for automated testing that look for vulnerabilities that would be difficult to find manually within a feasible timeframe. (Fuzzing involves inputting a huge amount of random data, or fuzz, to try to make a test system crash.)

**Reinforcement learning** is really the interesting area for cyber security offensive and defensive techniques. The machines participating in the “Cyber Grand Challenge” all used reinforcement learning to optimize their resources for the best chance to win the competition. They found that since using computing power to attack other machines comes at a cost, and that cost is much greater than using the same computing power for defense, from a rewards perspective, it is better to spend the time defending oneself than to attack others. Using this logic, organizations may use machine learning to look for optimal choke points to stop offensive attacks coming in from
outside, rather than dissect and respond to every conceivable combination of attacks that may try to get past their existing defenses.

How machine learning detects cyber threats

Machine learning is highly suited for cyber security threat management activities that might slip by traditional security tools.

Machine learning excels at looking for “unknown unknowns,” the events that humans do not know that they are not aware of — including novel attack behaviors. Consequently, machine learning is well-suited for detecting zero-day vulnerability exploitation, unusual network lateral movements and access to data at unusual hours.

All these activities can occur in large numbers, and it would be time-consuming and tiring for a human operator to search manually for suspicious security behavior.

Machine learning automates the learning quickly and excels at analyzing highly repeated patterns in large numbers and at mapping relations between data points. For example, it can easily map a human user ID to an IP address accessing certain data with specific access privileges. In a world where data and logs are growing exponentially, it is harder and harder to look for patterns manually. Machine learning becomes instrumental in identifying usage patterns and ensuring that the usage patterns remain statistically consistent. If the usage pattern is determined to be an anomaly, chances are that it could be a security concern.

Malicious activities can occur in different patterns, which means traditional security information and event management (SIEM) tools that use correlation rules to look for specific attack behavior will often fail. As such, modern security operations centers should supplement the use of rule-based controls, such as SIEM, anti-malware and firewalls, with machine learning-capable controls, such as endpoint protection platforms and user and entity behavior analytics (UEBA), to detect anomalies that are otherwise hard for human operators to anticipate.

The limits of machine learning for security management

Despite the potential of machine learning, there is a limit to its value to threat management. While machine learning is good for identifying patterns and thus generates high-fidelity information that can help the human operator to focus immediately on unusual behavior, it cannot authoritatively be used to determine whether that behavior is malicious. For example, if a sensitive file is accessed after office hours by a staff member who is authorized to access it, the event cannot be determined to be a security issue unless the human operator manually investigates it. Machine learning can never provide contextualization to an event, but it can quickly bring unusual patterns to the human operator’s attention so that investigations can begin with minimal delay.

Since machines take time to “learn” and be “trained,” machine learning controls are ineffective in the first few weeks of use while they develop models to understand baseline behavior. During this time, if the environment is already tainted, the machine learning models will deem malicious activities to be normal behavior. Even if the environment is not tainted, rule-based controls are essential components to protect the environment while the complementing machine learning controls attempt to understand it and build a baseline behavior for spotting anomalies.
Conclusion: The marriage of AI and human intelligence

Since machine learning is strongly tied to artificial intelligence, and machine learning capabilities are starting to become pervasive in contemporary security controls, will AI become mainstream in managing security threats in the future? Unfortunately, it may take a few years.

In the meantime, humans can take comfort in the fact that they’re still more intelligent than machines. Let’s go back to DEFCON 2016. The winning machine, Mayhem, won the “Cyber Grand Challenge” competition and was promptly sent to compete with real human competitors in the CTF game. No surprise, it ended up last on the scoring board. Offensive and defensive artificial intelligence may seem tangible today, but at its current maturity, it is still not on par with human intelligence.

Whether AI is more of a threat than a cure perhaps comes down to human intelligence — for in truth, there is no “intelligence” in any AI engines, regardless of how many attack mechanisms they learn or how many anomalies they detect. A key aspect of intelligence is the ability to conceive new ideas or make predictions, and AI engines are incapable of doing either. Human contextualization will always be an essential step for building new machine learning models in AI engines that anticipate new security attack techniques.

When used in a narrow, prescriptive manner to detect security attacks that exhibit specific behaviors, AI engines will continue to be extremely well-suited to performing security orchestration activities — by both upstanding companies and nefarious actors.

However, the deciding factor in the role of AI in security — whether “the good machines” or “the bad machines” win — still lies in how we apply human intelligence to specific security needs. When organizations consistently up their game and strive to understand the latest technologies, AI and machine learning can effectively help them stay ahead of the threats and improve their security posture.

About the author

TM Ching is the chief technology officer for Security at DXC Technology. TM is chiefly responsible for security thought leadership as well as research and development activities for DXC Security worldwide. He works closely with vendors and professional bodies to identify technological evolutions or disruptions on the horizon. TM also develops roadmaps for both clients and DXC Security to achieve service readiness to meet the threat landscape changes in the next 12 to 36 months.

Learn more: www.dxc.technology/security

About DXC Technology

DXC Technology (NYSE: DXC) helps global companies run their mission critical systems and operations while modernizing IT, optimizing data architectures, and ensuring security and scalability across public, private and hybrid clouds. With decades of driving innovation, the world’s largest companies trust DXC to deploy our enterprise technology stack to deliver new levels of performance, competitiveness and customer experiences. Learn more about the DXC story and our focus on people, customers and operational execution at www.dxc.technology.

© 2020 DXC Technology Company. All rights reserved.