BCG

The Boston Consulting Group

ARTIFICIAL INTELLIGENCE IS A THREAT TO CYBERSECURITY. IT'S ALSO A SOLUTION.

By Ryan Goosen, Anna Rontojannis, Stefan Deutscher, Juergen Rogg, Walter Bohmayr, and David Mkrtchian

In May 2018, the *New York Times* reported that researchers in the US and China had successfully commanded artificial intelligence (AI) systems developed by Amazon, Apple, and Google to do things such as dial phones and open websites—without the knowledge of the AI systems' users. It's a short step to more nefarious commands, such as unlocking doors and transferring money. And while Alexa, Siri, and Google Assistant may be the most widely used AI programs in operation, they are hardly the only ones. It's not hard to imagine cyberthieves targeting a financial institution's AI-controlled customer recognition software or a shady competitor attacking another company's AI pricing algorithm. In fact, more than 90% of cybersecurity professionals in the US and Japan expect attackers to use AI against the companies they work for, according to a survey by cybersecurity firm Webroot.

For people with responsibility for corporate security—everyone from CIOs to CISOs and CROs—AI presents two types of risk that change the nature of their jobs. The first is that criminals, bad state actors, unscrupulous competitors, and inside threats will manipulate their companies' fledgling AI programs. The second risk is that attackers will use AI in a variety of ways to exploit vulnerabilities in their victims' defenses.

Companies are in a cybersecurity arms race. As cybersecurity firm Crowdstrike's *2018 Global Threat Report* makes clear, attackers have easy access to more tools as the lines between state actors and criminal gangs fade. Malware and identity theft kits are easy to find and inexpensive to buy on dark web exchanges. AI-enabled attack kits are on the way, and we can expect that they will be readily available at commodity prices in the next few years.

Yet for all the inherent risk AI presents, part of the answer might lie in harnessing the power of AI itself to strengthen existing cybersecurity set-ups. Our experience shows that companies can begin to protect their systems by integrating AI into their security, starting now.

A New Risk for Companies…

The list of actual AI applications is already long and growing. Faster and more accurate credit scoring for banks, improved disease diagnosis and treatment development for health care companies, and enhanced engineering and production capabilities for manufacturers are just a few examples. A survey in 2017 by BCG and MIT *Sloan Management Review* found that about 20% of companies have already incorporated AI in some offerings or processes and that 70% of executives expect AI to play a significant role at their companies in the next five years.

With all the benefits, however, come substantial risks. For example, machine-learning algorithms and certain other types of AI work by using "training" data to learn how to respond to different circumstances. They then learn by doing, incorporating additional data as they work, refining their approach in an iterative manner. (See "The Building Blocks of Artificial Intelligence," BCG article, September 2017 and "The Big Leap Toward AI at Scale," BCG article, June 2018.) From a security perspective, that methodology presents two challenges.

First, AI systems are generally empowered to make deductions and decisions in an automated way without day-to-day human involvement. They can be compromised, and that can go undetected for a long time. Second, the reasons that a machine-learning or AI program makes particular deductions and decisions are not always immediately clear to overseers. The underlying decision-making models and data are not necessarily transparent or quickly interpretable (although significant effort is under way to improve the transparency of such tools). This means that even if a violation is detected, its purpose can remain opaque. As more machine-learning or AI systems are connected to, or placed in control over, physical systems, the risk of serious consequences—including injury and death—from malevolent interference rises. And we have already seen that while cybersecurity concerns are a consideration in the adoption of AI, especially for pioneers in this field, cybersecurity is of less concern to companies that are lagging behind. (See *Artificial Intelligence in Business Gets Real*, a report by the MIT *Sloan Management Review* in collaboration with the BCG Henderson Institute, Fall 2018.)

Companies' AI initiatives present an array of potential vulnerabilities, including malicious corruption or manipulation of the training data, implementation, and component configuration. No industry is immune, and there are many categories in which machine learning and AI already play a role and therefore present increased risks. For example:

- Financial (credit fraud might be easier, for example)
- Brand or reputational (a company might appear discriminatory)
- Safety, health, and environment (systems might be compromised that control cyberphysical devices that manage traffic flow, train routing, or dam overflow)
- Patient safety (interference might occur in medical devices or recommendation systems in a clinical setting)
- Intervention in, or meddling with devices connected to the Internet of Things (IoT) that use machine learning or AI systems

…And an Opportunity, Too

The good news for companies is that they can tap the power of AI to both upgrade their cybersecurity capabilities and protect their AI initiatives (so long as they layer in appropriate protections to the AI systems being used for defense). Moreover, investments in AI will likely have multiple forms of payback.

For one, companies can build in better protection and the potential to at least stay even with the bad guys. AI not only enhances existing detection and response capabilities but also enables new abilities

in preventative defense. Companies can also streamline and improve the security operating model by reducing time-consuming and complex manual inspection and intervention processes and redirecting human efforts to supervisory and problem-solving tasks. AI cybersecurity firm Darktrace claims that its machine-learning technology has identified 63,500 previously unknown threats in more than 5,000 networks, including zero-day exploits, insider threats, and subtle, stealthy attacks. Consider the number of cyber incidents that the average large bank deals with every day, from the ordinary and innocent (customers mis-entering passwords, for example) to attempted attacks. They need automated systems to filter out the truly dangerous signal from the more-easily-addressed noise. In the medium to long term, companies that invest in AI can offer operational efficiencies and potential operating-expense savings.

To enhance existing cybersecurity systems and practices, organizations can apply AI at three levels.

Prevention and Protection. For some time, researchers have focused on AI's potential to stop cyberintruders. In 2014, the US Defense Advanced Research Projects Agency announced its first DARPA Cyber Grand Challenge, a competition in which professional hackers and information security researchers develop automated systems that can figure out security flaws and develop and deploy solutions in real time. While it is still early days, the future of cybersecurity will likely benefit from more AI-enabled prevention and protection systems that use advanced machine learning techniques to harden defenses. These systems will also likely allow humans to interact flexibly with algorithmic decision making.

Detection. AI enables some fundamental shifts. One is from signature-based detection (a set of static rules that relies on always being up-to-date and recognizing an attack signature) to more flexible and continuously improving methods that understand what baseline, or normal, network and system activity look like. AI algorithms can detect any changes that appear abnormal—without needing an advance definition of abnormal. Another shift is to move beyond classic approaches based on machine learning that require large, curated training datasets. Some companies have employed machine-learning programs in their security systems for several years, and more advanced AI-based detection technologies (such as reinforcement learning and deep neural networks) are now gaining traction, especially in IoT applications. AI can also provide insights into sources of potential threats from internal and external sensors or small pieces of monitoring software that evaluate digital traffic by performing deep packet inspection. Note that for most companies, AI-based detection and potential automated attribution will require careful policy design and oversight to conform with laws and regulations governing data use.

Response. AI can reduce the workload for cybersecurity analysts by helping to prioritize the risk areas for attention and intelligently automating the manual tasks they typically perform (such as searching through log files for signs of compromises), thus redirecting human efforts toward higher-value activities. AI also can facilitate intelligent responses to attacks, either outside or inside the perimeter, based on shared knowledge and learning. For example, today we have technology to deploy semiautonomous, intelligent lures or "traps" that create a duplicate of the environment to be infiltrated to make attackers believe they are on the intended path and then use the deceit to identify the culprit. AI-enabled response systems can segregate networks dynamically to isolate valuable assets in safe "places" or redirect attackers away from vulnerabilities or valuable data. This can help with efficiency as analysts can focus on investigating high-probability signals rather than spending time finding them.

Implementation of automated AI-driven response will require careful design and strategic planning. This will be especially true when it comes to users that should be

isolated or quarantined and systems that work at the digital-physical interface (such as critical links in manufacturing or supply chains, or critical-care medical devices in hospitals or emergency settings).

The Race Is On

Cybersecurity has always been an arms race. In 2016, then-US President Obama talked to *Wired* magazine about his fears of an AI-enabled attacker accessing the US nuclear codes. "If that's its only job, if it's self-teaching and it's just a really effective algorithm, then you've got problems," he said. AI increases attackers' speed, resilience, opportunities, and chances of success. Because AI algorithms are self-learning, they get smarter with each attempt and failure; their endeavors are continuously better informed and more capable. Just as companies can use AI to automate and improve business processes, hackers can automate the identification of vulnerabilities and exploit-writing.

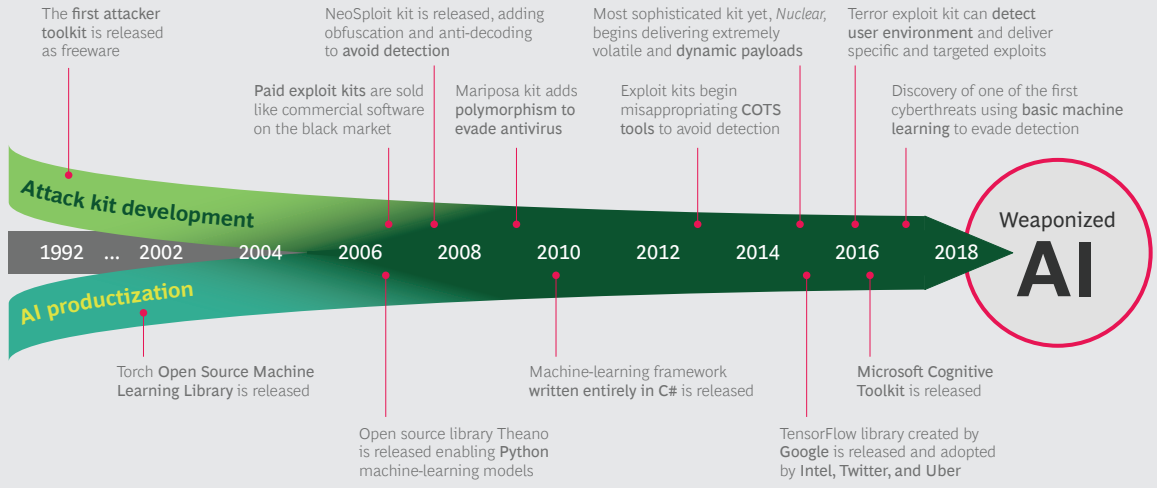
AI algorithms tend to be public, often open-source, software that is widely available on the internet and increasingly easy to use. Just like the software-as-a-service that many companies use, malware-as-a-service is commonplace and a viable business for criminal players. There is even a high degree of competition among cyber-criminal vendors (which can leverage AI and machine learning for competitive advantage) to create superior malware. Moreover, open-source AI libraries and software, which give companies a new source of fast and inexpensive innovation, can also be a source of new vulnerabilities.

In addition, AI can actually help malware avoid detection. Although security companies are increasingly introducing AI features and behavioral analytics into their products, a lot of antivirus and end-point protection software still rely largely on signature-based detection. In response, attackers develop toolkits that obfuscate the nature and sources of the malware, making it harder to recognize by its digital fingerprint.

On the dark web today, anyone can buy a tailor-made virus guaranteed not to be detected by the 10 or 20 or so major antivirus tools. But defensive systems gain knowledge over time. This knowledge could be thwarted by an AI algorithm that adds to the stealthiness of a malware kit over time, masking the malware's identity based on what it learns defense systems are detecting. (See the exhibit.)

Think about this scenario. Attackers often use botnets—global networks of hijacked devices (such as PCs, smartphones, and IoT

Attack Kit Development and AI Productization Will Weaponize AI



Source: BCG analysis.
Notes: Illustrative examples. COTS=commercial, off-the-shelf.

devices)—to do their dirty work. Botnets are effective tools, but they can only do what the attackers direct them to. Suppose, however, that the command-and-control software directing a botnet is replaced by an AI algorithm that enables it to act in a semiautonomous fashion. The botnet now has the ability to learn which of its attacks are working and which aren't, and it teaches itself to become more effective based on its results. Now suppose that when the botnet discovers a vulnerability in a company's system, it is able to craft its own attack, according to its assessment of the highest-potential scheme (data theft, financial theft, or ransomware, for example)—and update its attack toolkit as it goes. AI gives the botnet the capability to direct its own serial interventions—phishing to deposit a payload, for example, then exploiting a software vulnerability to gain access to valuable data, and finally finding a way to exfiltrate the data. It can handle each task systematically and on its own, it does not require someone to steer it from outside—and its self-direction helps it avoid detection. In essence, AI enables botnets to adapt methods and attack toolkits dynamically to operate continually at peak effectiveness and penetrate more hosts.

AI raises the stakes, with an advantage for the attackers. They need to get it right only once to score while defenders need to defend successfully 24/7/365.

A Two-Part Challenge

Companies need to approach AI and cybersecurity from two perspectives: protecting their own AI initiatives and using (AI-enabled) cybersecurity to protect their full set of digital assets (whether AI-enabled or not). Both raise a lot of questions. Here are some of the more urgent questions to consider about the first issue—protecting AI initiatives:

- How are we protecting our AI-based products, tools, and services?
- How do we keep our training data pristine and protect against biased inputs?
- How do we protect the algorithms (or their implementation)?
- Do we have control procedures that stop abnormal events from happening and a Plan B in case we notice that our AI programs are behaving abnormally?
- Do we have the technical and human monitoring capabilities to detect if our AI has been tampered with?
- Have we made conscious decisions about who (or what) can decide and control which capabilities? Did we assign AI systems an appropriate responsibility matrix entry? Do we constrain AI to decision support or expert systems, or do we let AI programs make decisions on their own (and if so, which ones)?
- Do we have the appropriate governance policies and an agreed code of conduct that specify which of our processes or activities are off-limits for AI for security reasons?
- When using AI in conjunction with decisions on cyber-physical systems, do we have appropriate ethical, process, technical, and legal safeguards in place? Do we have compensating controls? How do we test them?
- Have we aligned our cybersecurity organization, processes, policies, and technology to include AI, to protect AI, and to protect us from AI malfunctions?

Some companies might find it useful to adopt a variation of the four-eyes principle—the requirement that certain decisions or actions be approved by at least two people or processes with strict underlying policies for review—and employ redundant systems and implementations or complementary (AI) algorithms that serve as a check on each other. This, of course, comes at an extra cost (similar to the human-based four eyes principle or the redundant flight-control electronics on an airplane), which may affect some initially promising business cases.

In a similar vein, here are some helpful questions for the second issue, leveraging AI in cybersecurity:

- Where is AI being used in our cybersecurity portfolio?
- Is it being used in a manner that creates operational effectiveness, efficiency, or cost reduction (at least in the medium to long term)?
- AI is not a panacea; do we focus sufficiently on educating technicians and end users since, on the one hand, humans are ultimately the key weakness in cybersecurity, and, on the other, they will have to jump in when AI signals an issue has arisen or stops working as expected?

The intelligence may be "artificial," but the risks are all too real. Companies can use powerful new capabilities to enhance their overall cybersecurity efforts and stay even with bad guys in the security arms race. They also need to evaluate how AI is used in their products and services and implement specific security measures to protect against new forms of attack. More and more cybersecurity products will incorporate AI capabilities, and external partners can help integrate this capability into cybersecurity portfolios. Companies can start with an objective assessment of where they stand using the questions outlined above. There is no good reason for delay.

About the Authors

Ryan Goosen is a project leader in the Zurich office of The Boston Consulting Group. You may contact him by email at goosen.ryan@bcg.com.

Anna Rontojannis is a consultant in the firm's Zurich office. You may contact her by email at rontojannis.anna@bcg.com.

Stefan Deutscher is an associate director in BCG's Berlin office. You may contact him by email at deutscher.stefan@bcg.com.

Jürgen Rogg is a partner and managing director in the firm's Zurich office. You may contact him by email at rogg.juergen@bcg.com.

Walter Bohmayr is a senior partner and managing director in BCG's Vienna office and he leads the firm's Cybersecurity practice worldwide. You may contact him by email at bohmayr.walter@bcg.com.

David Mkrtchian is a consultant in the firm's Los Angeles office. You may contact him by email at mkrtchian.david@bcg.com.

11/18