Artificial Intelligence Machine Learning Technology Is A Double-Edged Sword

Ying Wu

Shanghai university of political Science and Law, Shanghai, China

wuying@shupl.edu.cn

Abstract. While artificial intelligence and machine learning greatly improve cyber security, their limitations lead to the possibility that they can have both positive and negative, even unpredictable ,concomitant effects on cyber security, i.e. they simultaneously present opportunities for attackers to also evolve their attack techniques, leading to systems that are more vulnerable to attack and can also adaptively generate attack programs. These concomitant effects make the network attacks more and more frequent, more intelligent, confrontation and game more and more intense, need to develop a more effective, more strategic, more systematic response and prevention strategies.

Keywords: artificial intelligence, machine learning, collateral effects

1 Introduction

So-called artificial intelligence is a tool that can quickly process data and predictive analysis, identify patterns, find bugs, and even execute plans to fix vulnerabilities and ensure the security of sensitive information, and is widely used in areas such as automated systems and information protection, mainly:

(1) Early detection ; (2) Prediction and prevention; (3) Data encryption; (4) Password protection and authentication; (5) Multi-factor authentication.

Machine learning is an algorithm that enables a user to feed a computer algorithm large amounts of data, which then allows the computer to analyze the data and make data-driven recommendations and decisions based on the input data. If the algorithm identifies any corrections, it integrates the corrections to improve future decisions. The application of machine learning in network security is very extensive. From checking that privacy parameters have been set correctly and tracked regularly, to observing the operation of devices and protecting consumers' data and private information, machine learning systems become the guardians of the online environment. It's placed on-device, in routers, and in the cloud hosting apps, and these layers of information security work together to provide guidance on setting up devices and protecting consumers.

While artificial intelligence (AI) and machine learning (ML) have greatly improved network security, they have also opened up opportunities for malicious attackers to evolve attack techniques. In simple terms, these technologies are not omnipotent, AI

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 24–32, 2022.
https://doi.org/10.2991/978-94-6463-108-1_4

can effectively analyze user behavior, deduce patterns and identify various abnormal or abnormal situations in the network. With this data, network vulnerabilities can be quickly and easily identified. However, their limitations lead to them may have both positive and negative, or even unpredictable side effects, in network security, social changes, military fields, etc.

2 The companion effect of AI/ML technology is becoming increasingly apparent

AI/ML technologies may also be maliciously exploited, causing new security issues in cyberspace, namely the side effects of AI/ML technologies.

When AI/ML technology is applied to vulnerability mining, it can find multiple loopholes in the system, which makes the system more vulnerable to attack; when AI/ML technology is applied to large-scale network attacks, attackers can adaptively generate attack programs. A large number of clients realize intelligent and automated network attacks; when AI/ML technology is applied to complex network attacks, attackers can hide attack behaviors, attack paths, etc., making it more difficult for defenders to discover and detect such attacks; AI/ML technology can also be applied to the game of network attack and defense, but AI/ML technology itself has vulnerabilities, attackers can attack the intelligent model deployed in the system, causing the defense model to fail; AI/ML technology may also be used to Stealing important data of users, through in-depth mining of various types of data in the system, correlation analysis, and restoration of important data of users, thus causing more serious security problems.

2.1 Social networks full of disinformation ----AI/ML credits

With the development of artificial intelligence technologies such as natural language processing and generative adversarial networks, social networks are becoming a new battlefield in hybrid warfare.

The recent Russian-Ukrainian battle of disinformation on social media has become a "key battleground in the Russia-Ukraine confrontation" is a typical example.

(1) The ability to automatically generate content and interact with human users on social media by composing or tampering with images and videos through deep learning

There are various technical ways to tamper with information using deep learning, such as Deepfake, Face2face, etc. Deepfake uses generative adversarial networks to generate an overall fake image, which cannot be detected by traditional methods of detecting whether an image has been edited. Face2face is a fake method using image synthesis. The main method is to synthesize one person's facial features and expressions with another person's photos.

(2) Automatically generation and dissemination of messages using social bots

Social bots are artificial intelligence-controlled social media accounts that were first used to automatically forward messages or find messages online. With the advancement of artificial intelligence technologies such as natural language processing, social robots have been able to use pre-trained multilingual models to generate similar speeches on

social platforms such as Twitter, which can be used to manipulate public opinion or incite emotions.

(3) Using machine learning to detect false information

The detection of false information has become a key problem in social networks, and its detection method is related to the type of information. For example, using methods such as checkerboard artifacts and co-occurrence matrices, it is possible to detect fake videos generated by generative adversarial networks; using methods such as facial feature detection, it can detect videos that are faked by changing faces; using multi-modal features of information, it can detect Fake news, etc., composed of a combination of various information. The U.S. Department of Defense has funded a lot of research to detect disinformation.

2.2 The scope for AI/ML in the military is unlimited

From 2020, the Israeli army has used three sets of AI systems (data analysis, map drawing and face recognition system) on the battlefield for the first time. In densely populated cities, precise positioning has eliminated more than 150 Hamas commanders and Special agents, greatly reducing the manslaughter rate.

In the current Russian-Ukrainian battlefield, this method is used to confirm the position of the opposing commander. Modern warfare is very complex, and being a commander does not mean that you cannot be detected by your opponent without wearing an officer's uniform. However, thousands of roadside surveillance systems installed in various places and people holding mobile phones and looking out of the window take pictures, and finally upload them to the military command. Through the analysis of intelligence departments and technology companies, the military can find a lot of intelligence. This is one of the reasons why military commanders have been pinpointed many times. It is reported that on March 19, 2022, the location of the Russian lieutenant general was killed by a long-range attack by precision-guided munitions by the Ukrainian side with the help of US technology, AI precise positioning and face recognition.

Now in the Russian-Ukrainian battlefield, AI/ML technology systems are still initially entering the battlefield. Once the war is delayed, more AI/ML technologies will be put into the battlefield, turning the Russian-Ukrainian battlefield into a testing ground for new weapons and AI/ML technology combat systems.

Russian AI/ML technology is more closely linked to the military. In the Russian army AI/ML technology is used in command, logistics, data, materials and transportation. The Russian army is also testing the "reconnaissance-strike" combat system, which uses AI/ML technology to automatically command subsystems and a general intelligence network to coordinate and command sea, land and air combat units.

2.3 The depth and breadth of AI/ML's predictive control of social opinion is increasing

Internet public opinion is the concentrated expression, dissemination and influence of the majority of netizens using the Internet as a carrier to express, disseminate and influence their emotions, attitudes, opinions and viewpoints towards an event. Network

public opinion management and control is an important part of cyberspace security, and it is also an important position related to national political security. At present, foreign countries pay special attention to the close integration of artificial intelligence technology and social networks, comprehensively using social big data in-depth mining and analysis, accurate public opinion push and management, and even fake news generation and dissemination. Derivative effects, amplification effects and demonstration effects, control domestic public opinion and social contradictions, disrupt, destroy or even subvert the regimes and international interests of other countries. [1]

2.4 Widespread use of AI/ML for editing and identification of false information

On February 23, 2022, Fox News said that U.S. intelligence officials are closely monitoring video and audio that may be manipulated, including looking for deepfake videos related to Russian President Vladimir Putin, Ukrainian President Volodymyr Zelensky and other key figures. Recently, some Ukrainian propagandists used synthetic video to promote the victory of the war. They used artificial intelligence technology to paint the Russian army's tactical logo "Z" on the KraZ military card of the Ukrainian army that was destroyed in the Eastern Ukrainian battlefield in 2014 to serve as the victory of the Ukrainian army.

(1) Automatic generation and dissemination of news by bots using AI/ML technology for opinion manipulation

Bots are AI/ML -controlled social media accounts that utilize pre-trained multilingual models to generate human-like remarks on social platforms such as Twitter and Facebook, which can be used to manipulate public opinion or incite sentiment. On February 8, 2022, Ukraine's State Security Service said it had shut down an AI/ML robot farm that spread panic and made bomb threats on social media, adding that the farm managed 18,000 robot accounts and was run by an organization from Russia. responsible for farm supervision.

(2) Identification of forged information using AI/ML technology

On February 4, 2022, CBS News said that the White House believed that Russia had faked a video of the explosion and used it as an excuse to invade Ukraine. Due to the existence of forged information for public opinion guidance and specific-purpose propaganda, accurate identification of forged information in cyberspace confrontation is an important means to understand the intention of the enemy's actions. Usually chessboard shadows, co-occurrence matrices and other methods are used to detect false videos generated by adversarial networks, and knowledge maps can also be used to detect false information in text. In the confrontation in the cognitive domain of the Russian-Ukrainian conflict, the two countries will likely use such technologies to detect forged information to ensure their dominant position in the confrontation process.

(3) Using AI/ML to obtain open source intelligence-an important addition to conventional intelligence surveillance tools

Due to the deepening of the Russia-Ukraine confrontation in Ukraine, a large number of Ukrainian netizens have posted videos, pictures, texts and other multimodal information about the Ukrainian army and the Russian army on Twitter, Facebook, YouTube

and other social networking sites. Both parties may use technologies such as natural language processing, computer vision, and knowledge graphs to discover military intelligence such as equipment models, combat deployments, and battle losses in these social media information. In recent years, in the elections of Western countries such as the United States and the United Kingdom, the use of online social platforms to deploy "robot navy" has gradually emerged, so as to artificially manipulate and guide the direction of public opinion and achieve the purpose of interfering with the election results. [2]

3 Cybersecurity dynamics with AI/ML technology

A decade ago, some computer scientists had a hunch that AI/ML -based cyberattacks would wreak havoc on the world, and even the United States was worried about when the "Cyber Pearl Harbor" event would happen. In 2017, one of the most destructive "NotPetya" cyber attacks in the world, although only using automation technology and not yet involving AI/ML technology, has caused such damage. Likewise, ransomware attacks have increased dramatically since 2019, even reaching critical national infrastructure. For example, in 2021, a gas company in the United States was attacked by ransomware, which shut down its gas pipeline for 6 days and cut off fuel supply in 17 states, directly affecting thousands of American schools, businesses and hospitals.

3.1 Cyber- attacks are becoming more frequent

In the AI/ML era, the forms of large-scale network attacks mainly include denial of service attacks (DDoS), domain name resolution server (DNS) hijacking, etc. The targets of large-scale cyber attacks also extend from traditional network systems to the Internet of Things, industrial equipment, smart homes, and driverless systems. In 2018, the United States organized experts to discuss attacks on self-driving cars, including the possible harm caused by large-scale cyberattacks, and recommended planning exercises in advance.

AI/ML techniques can also generate intelligent botnets with scalable attacks. Fortinet said in its 2018 Global Threat Situation Forecast that AI/ML will be widely used in Hivenet and Swarmbots in the future, and can take advantage of large-scale interconnection. A swarm of devices or robots simultaneously identifies and responds to different attack vectors and leverages self-learning capabilities to enable unprecedented large-scale autonomous attacks. At the same time, AI/ML makes the cost of network attacks lower and lower, and more and more attack weapons and resources are available, resulting in more frequent large-scale network attacks.

3.2 Cyber- attacks are getting smarter

AI/ML technology facilitates the discovery and exploitation of vulnerabilities. Fuzz testing is an automated or semi-automated software testing technology that constructs

random and unexpected malformed data to test and monitor the availability of anomalies and vulnerabilities that may occur during program execution. Such fuzzing techniques can be divided into white-box, black-box, gray-box fuzzing, etc., which can efficiently mine and exploit program vulnerabilities. Automatic exploitation of vulnerabilities generally includes information extraction, vulnerability identification, path discovery, state solution and code generation. By extracting useful information from input data such as executable files and source code, using path discovery and state solution to obtain utilization cases, and generate. The program or data that exploits the vulnerability to automate the exploitation of the vulnerability. Automated attack and defense is a new challenge for cyberspace security, and automated cyberattack methods will exacerbate security threats and challenges in cyberspace. [3]

3.3 Cyber- attacks are becoming increasingly covert

Traditional network attacks generally leave traces in the system, which are easy to be traced; the target and intention of the attack are relatively clear and easy to be discovered. In the AI/ML era, complex attack behaviors can be hidden, such as attacking through different terminal devices and launching attacks at different times. For example, malicious code and malicious programs can be embedded with a deep neural network model, so as to ensure the high confidentiality of the attack target, attack intent, and high-value payload under the premise of open source code, thereby greatly improving the attack behavior. concealment. These attack behaviors can be distributed on many devices, and there may be a large time interval between different attack behaviors. Combining AI/ML can better design and combine attack behaviors, so as to avoid the detection of defenders and maintain attack behaviors of high concealment.

3.4 Cyber- attack confrontation and games are becoming increasingly intense

Attackers will use AI/ML technology to construct larger-scale, more concealed, and more serious attacks, while defenders will use AI/ML technology to improve the accuracy of network attack detection, improve network attack detection efficiency, and reduce Network attack false positive rate, etc. In this process, AI/ML technology has made the game of attack and defense in cyberspace more and more intense. In addition, due to the vulnerability of AI/ML technology itself, for example, the image recognition neural network is easily confused by the generated adversarial samples that are highly similar to the original samples, resulting in wrong identification, and the recommendation system is easily affected by individual keywords, resulting in human intervention in the recommendation results. When AI/ML technology that lacks explanation is used for network attack or defense, the other party can use the vulnerability of the model itself to launch defense or attack, triggering a new round of network attack and defense game.

3.5 Cyber- attacks make important data increasingly vulnerable to theft or destruction

AI/ML technology, it is easier for attackers to steal and destroy important data, such as skew attacks. The attacker can judge whether a certain piece of information exists in the training data set of the target model, so as to realize the theft of important data. The attacker trains multiple shadow models that imitate the target model, and uses the recognition results of the shadow models to determine whether the training set of the target model contains certain sensitive data. Similarly, the model backward attack can reverse some or all of the attribute values of a certain target data in the training set through the output of the model. When the attacker only obtains the model parameters, the attacker can use the method based on the generative adversarial network to realize the model inversion. to reconstruct the training data, causing the data to be stolen. [4]

4 Strategies for dealing with cyber- attacks in the AI/ML technology environment

4.1 Strengthen research and application to promote intelligent network attack and defense system construction and capability upgrade

See clearly the threat and impact of AI/ML technology network attacks, start from the perspective of preventing security threats and building peer-to-peer capabilities, and carry out major key technology research as soon as possible. Promote "industry-university-research" institutions to effectively respond to new threat scenarios empowered by artificial intelligence as the primary requirement, carry out joint research from both offensive and defensive aspects, and carry out intelligent threat situational awareness, automated vulnerability mining and utilization, intelligent malicious code and other technical research. Accelerate the systematic application of artificial intelligence technology in the security protection of key information infrastructure in the country and important industries, complete the intelligent upgrade as a whole, and greatly improve the security assurance of key information infrastructure, network security situational awareness, network security defense, and network deterrence. ability level. In order to manage and control the new cybersecurity threats brought by artificial intelligence, the construction of relevant laws and regulations should be strengthened, the healthy development of artificial intelligence cybersecurity should be regulated, and activities related to specific threats should be delayed and prevented.

4.2 Strengthen the sharing and utilization to crack the data problem of AI network attack and defense technology system construction

The AI training dataset is not only the most valuable digital asset in AI security research, but also a strategic asset related to the success of AI security capability building. However, the current AI security training data lacks safe, controllable and traceable means for sharing and utilization, which has become one of the important factors re-

stricting the rapid development of AI offensive and defensive technologies. It is recommended to rely on authoritative institutions such as national laboratories and use new technologies such as blockchain to build an artificial intelligence data range, form a safe, credible, and reasonable incentive mechanism for sharing and utilization frameworks, promote the effective use of artificial intelligence data assets, and implement data-based data. The center's artificial intelligence network attack and defense technology development path.

4.3 Enhancing confrontation and evaluation to promote the practicality of AI cyber-attack and defense techniques

AI attack and defense is a technology of continuous confrontation and upgrading, and the actual application effect depends on the comprehensiveness and authenticity of the confrontation environment. However, due to insufficient scientific research conditions, it is difficult to reproduce the actual offensive and defensive confrontation environment in the existing research on artificial intelligence attack and defense technology, which constitutes an obvious restriction for artificial intelligence automatic attack and defense technology to move from theory to practice. It is recommended to build an artificial intelligence attack and defense confrontation range based on authoritative institutions such as national laboratories. Through authoritative assessments, technical challenges, testing and verification, etc., the effectiveness evaluation and confrontation analysis of artificial intelligence network attacks and automated vulnerability discovery and utilization can be effectively promoted. , to promote the accelerated development of artificial intelligence offensive and defensive technology in a practical direction. [5]

5 Conclusion

The current cyber-attack campaign does not yet prove that attackers have begun to use AI/ML technology in large numbers. Because the integration of existing AI/ML technologies into cyber attacks is still a very complex task, and the degree of overlap with Internet technologies is not high, and the deep integration of the two requires considerable professional technical strength for structural association. However, the increasing difficulty for attackers to penetrate the Internet's security barriers will force them to constantly explore new methods of attack. That's when AI-enhanced networking tools for hacking the internet are likely to emerge. But the strength of ML is its ability to identify data patterns in large datasets, and algorithms such as object classification can be employed to detect suspicious activity on the web. Currently, the application of ML technology in traditional intrusion detection has thwarted many cyber-attack activities. For example, the spam detection task has been qualitatively improved due to the application of ML techniques.

References

1. Security Research Institute of China Academy of Information and Communications Technology. Artificial Intelligence Security Framework (2020) [R]. Beijing: China Academy of Information and Communications Technology, 2020.
2. Fang Binxing. Artificial intelligence safety [M]. Beijing: Electronic Industry Press, 2020. Fang B X. Artificial intelligence safety and security [M]. Beijing: Publishing House of Electronics Industry, 2020.
3. Bright Military. An Analysis of the Russian Hybrid War Concept [EB/OL]. [2021-08-10]. http://www.3g.k.sohu.com/t/n508806134.
4. Cyber Research Institute. Artificial Intelligence Technology and Cyberspace Security [J]. Information Security and Communication Secrecy, 2019(6): 21-26.
5. Bright Military. An Analysis of the Russian Hybrid War Concept [EB/OL]. [2021-08-10]. http://www.3g.k.sohu.com/t/n508806134.