LINA YAO

# HOW ARTIFICIAL INTELLIGENCE **TRANSFORMS CYBERSECURITY**

by **Lina Yao**, Scientia Associate Professor at UNSW

As cyberattacks grow in volume and complexity, artificial intelligence (AI) is helping under-resourced security operations analysts stay ahead of threats.

By curating threat intelligence from millions of research papers, blogs and news stories, AI can provide instant insights to help cut through the noise of thousands of daily alerts, drastically reducing response times and mis/dis information on the internet, etc. The latest advancements in AI can take cybersecurity to a new level, and boost relevant research and application development.

According to the Australian Cyber Security Centre's (ACSC) Annual Cyber Threat Report July 2019 to June 2020, in Australia alone there are, on average, more than six cyberattack incidents every single day, and most of them have moderate or substantial impacts.

ACSC says it received 59,806 cybercrime reports in the 12 months to June 2020, almost one every 10 minutes. It says the true figure is probably much larger, because cybercrime in Australia is underreported. Notably, the attacks were mostly targeted at large organisations.

Governments and businesses are making every effort to protect themselves, but the volume of attacks can be overwhelming for security analysts and professionals. And there will always be new and unforeseen attacks and threats, such as the notorious ransomware attacks of the past two years that paralysed countless computers and even IoT devices.

A security paradigm that is purely responsive will fail to provide adequate protection. It can resolve issues only after they have been discovered, by which time, damage is likely to have already been done.[1] Without long-term vision, only identified and confirmed threats can be dealt with. New ones will not be addressed.

## MACHINE LEARNING IS HOT

Machine learning is a hot topic in artificial intelligence, and is capable of extracting valuable insights from existing knowledge, such as recordings of experiences, and identified threats or attacks.

Machine learning has proved to be very effective in detecting variants of existing malware, attacks and

threats, no matter how deep the malicious code or attack patterns are hidden.

Data-driven machine learning powered by deep neural networks can learn the activity patterns or tendencies of individuals in an organisation. Given sufficient time or sufficient data it can develop an understanding of patterns and tendencies that may be too complicated or subtle for human cognition.

This enables machine learning to respond rapidly to threats, such as a link in a phishing email, a malware payload, or attacking network traffic. A system powered by machine learning is able to continuously monitor an entire system and provide a real-time threat response.

Some of the most successful applications of AI to cybersecurity have been to provide predictive protection. For example, modern malware may be hard to detect solely by examining its code and its behaviour. [2]

*"A security paradigm that is purely responsive will fail to provide adequate protection. It can resolve issues only after they have been discovered, by which time, damage is likely to have already been done.[1] Without long-term vision, only identified and confirmed threats can be dealt with. New ones will not be addressed."*

In recent years, few shot and lifelong machine learnings are attracting increasing attention, which equips AI with human-like ability of Learning to Learn and enables the AI systems to quickly learn and generalize to new tasks from very limited data.

## ANTICIPATING ATTACKS WITH AI

An AI-based malware detection system [3] has been able to detect malware while it is downloading, and so prevent it from being installed and executing on the target system.

Another example is data breach prediction with AI. Liu et al [3] modelled this as a binary prediction problem, based on historical data and observations, to determine whether a system is likely to face such an attack in the near future.

What's fascinating is that this was done with no access to the client's internal networks: data were only collected externally. Furthermore, there are also applications to make fine-grain predictions that identify the risk associated with specific business information. This would enable a business to adjust resource allocation and prioritise protection so as to minimise the impact of an attack.

However, many solutions assume the input data fed into their algorithms are clean with no noise or errors. Such assumptions can be exploited by attackers, who may poison the input by providing counterfeit malicious incident reports, or creating a fake honeypot network for the algorithm, which can mislead its predictors and sabotage its learning. This is referred to as adversarial machine learning. It is critical that it be addressed.

Work is also underway to develop adversarial machine learning that will provide security to the machine learning itself.[5][8][9] This is key to successfully applying machine learning to cybersecurity.

## MACHINE LEARNING UNDER ATTACK

In general, there are two common types of attacks on machine learning: poisoning attacks which attack the learning during the training, and evasion attacks which attack the inferencing stage of the machine learning process.

There is another kind of attack called model stealing. This either tries to figure out the internal structure of the machine learning model or to extract the sensitive data the model has been trained on.

Another major research project we are conducting aims to develop robust predictive machine learning models that will detect and defend against false/misinformation spread over the Web via social media.

Such techniques are initially being developed against disinformation like fake news, fake reviews and clickbait, which can be used for cyberattacks, nefarious business operations and political subversion, creating social tension.[4] [5] [10]

Also, AI-powered false information can be even harder to distinguish from legitimate information than false information created by humans. Researchers need to develop methods to alleviate and address such misuse of AI technologies.

Much of the current work on proactive AI for cybersecurity is providing results that are too ambiguous, so few developments are finding practical application.

More detailed security recommendations with specific actions are needed for practical applications, and this could be the subject of future research. This may also lead to another research topic. A standalone security recommendation powered by comprehensive recommender systems, especially on critical services, can sometimes be hard to trust.[8] [12] An explainable system that differs from explainable AI, is preferable. It should provide explanations and visualised reasoning processes for intermediate risks and explain why the actions it suggests can minimise such risks, and at what costs.

Also, some reports suggest that, just as developments in AI technology can be applied for security, they can also be weaponised for malware and attacks, making these harder or even impossible to detect.

It may not be possible to prevent AI being used for nefarious activities, but it should be possible to prevent its impacts.

***www.linayao.com/***

***insdata.org/beta/***


References

[1] B. Morel, " Artificial intelligence and the future of cybersecurity," in The 4th ACM workshop on Security and artificial intelligence (AISec '11), Chicago, Illinois, USA, 2011.

[2] Sun, Nan, Jun Zhang, Paul Rimba, Shang Gao, Leo Yu Zhang, and Yang Xiang, "Data-driven cybersecurity incident prediction: A survey.," IEEE communications surveys & tutorials, vol. 2, no. 21, pp. 1744-1772, 2018.

[3] B. J. Kwon, J. Mondal, J. Jang, L. Bilge, and T. Dumitras, "The Dropper Effect: Insights into Malware Distribution with Downloader Graph Analytics," in The 22nd ACM Conference on Computer and Communications Security (CCS'15), Denver, Colorado, USA., 2015.

[4] Yang Liu, Armin Sarabi, Jing Zhang, and Parinaz Naghizadeh, Manish Karir, Michael Bailey, Mingyan Liu, "Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents," in The 24th USENIX Security Symposium (USENIX Security '15), Washington, D.C., USA, 2015.

[5] Abraham, Tamas, Olivier de Vel, and Paul Montague, "Adversarial Machine Learning for Cyber-Security: NGTF Project Scoping Study," Defence Science and Technology Group, Australia, 2018.

[6] Xianzhi Wang, Quan Z. Sheng, Lina Yao, Xue Li, Xiu Susie Fang, Xiaofei Xu and Boualem Benatallah, "Truth Discovery via Exploiting Implications from Multi-Source Data," in The 25th ACM Conference on Information and Knowledge Management ( CIKM 2016), Indianapolis, USA, 2016.

[7] Dong, Manqing, Lina Yao, Xianzhi Wang, Boualem Benatallah, Chaoran Huang, and Xiaodong Ning, "Opinion fraud detection via neural autoencoder decision forest," Pattern Recognition Letters, no. 132 , pp. 21-29, 2020.

[8] Yuanjiang Cao, Xiaocong Chen, Lina Yao, Xianzhi Wang and Wei Emma Zhang. Adversarial Attack and Detection on Reinforcement Learning based Recommendation System. The 43rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020). Xi'an, China, July 25-30, 2020.

[9] Zhe Liu, Lina Yao, Lei Bai, Xianzhi Wang and Can Wang. Spectrum-Guided Adversarial Disparity Learning. The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2020). Research Track. (KDD 2020), San Diego, CA, USA, August 23 - 27, 2020.

[10] Zhe Liu, Lina Yao, Xianzhi Wang, Lei Bai and Jake An. Are You a Risk Taker? Adversarial Learning of Asymmetric Cross-Domain Alignment for Risk Tolerance Prediction. International Joint Conference on Neural Networks (IJCNN 2020), Glasgow, UK, July 19 - 24, 2020

[11] Bin Guo, Yasan Ding, Lina Yao, Yunji Liang and Zhiwen Yu, The Future of Misinformation Detection: New Perspectives and Trends ACM Computing Surveys (CUSR) , 2020

[12] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. . Deep Learning based Recommender System: A Survey and New Perspectives ACM Computing Surveys (CUSR) , 2019