TRENDS

AI and Security

Four Considerations and Approaches for Making Progress on DX

Ryoichi Sasaki
Professor Emeritus, and Advisor and Visiting Professor of Center for Research and Collaboration,
Tokyo Denki University

Introduction

Many companies are pursuing digital transformation (DX) as a means of creating new business value. The use of artificial intelligence (AI) is a crucial factor in achieving this DX, with security issues being among the obstacles to doing so successfully. Accordingly, if DX is to succeed, there is a need not only to enhance the application of AI itself, but also to consider what problems arise when AI and security come together. This article addresses these problems, and what to do about them, in terms of the following four perspectives:

(1) Attacks using AI
(2) Attacks by AI
(3) Attacks on AI
(4) Countermeasures using AI

Four Perspectives to Consider when AI and Security are Combined and Research and Development Required

(1) Attacks using AI

Cyberattacks by perpetrators using AI are expected to become more frequent in the future. In particular, it is anticipated that malware equipped with AI functionality will become a genuine threat in the near future. This may well involve intrusions by a number of small AI-equipped malware agents of different types that work together to optimize how they mount an attack on the system environment in which they find themselves. At the very least, it is important that research into future countermeasures take account of these developments.

(2) Attacks by AI

The most significant negative impact that AI could have on humanity would be the emergence of an AI with capabilities that exceed those of human beings, ultimately leading to their extermination. In practice, however, most current AI research does not deal with "general-purpose AI" but with "special-purpose AI" for specific applications, and most AI researchers see the possibility of AI rebelling against humanity as too unlikely to take seriously. The author's response to this is to note that, while the potential for AI to turn against humans may be minimal, humans have very poor risk perception as demonstrated by the cognitive biases associated with risk that have been observed in psychology. Moreover, even in special-purpose applications of AI, there is a strong possibility that any action against humans from areas like AI weaponry would be very difficult to roll back. This makes it important that experimental and other work be undertaken to make sure that any abnormal behavior arising from AI can be reliably detected and stopped.

(3) Attacks on AI

It is also necessary to consider the problems that might arise from attacks on AI systems. The following are recognized as the main forms of attack.

(a) Noise-based attacks that cause already trained models to produce incorrect results: As the accuracy of decision-making and prediction degrades when the data being used contains noise, this can be exploited to produce incorrect results.

(b) Attacks that deliberately use biased training data for machine learning to induce incorrect decisions: a US IT company used crowd-sourcing for AI learning in its development of a chatbot that used this technology to interact with people autonomously. Unfortunately, when malicious users teamed up to input numerous instances of discriminatory opinions, it took no more than a day for the chatbot to start repeating those same opinions.

While studies have already begun on how these attacks are perpetrated and how to counter them, this is an area that calls for more advanced research.

(4) Countermeasures using AI

This is the approach of using AI in security countermeasures. AI is already widely used for this purpose, as can be seen from the results of academic literature search services or by searching for relevant products on the web. Examples include malware detection, log monitoring and analysis, traffic monitoring and analysis, security evaluation, spam detection, and information leaks.

The author has experience in such applications of AI, including an automatic system that uses machine learning for identifying the command and control (C&C) servers in targeted attacks and an intelligent network forensics system using rule-based systems and Bayesian networks. These applications have demonstrated the viability of using AI in security countermeasures. However, given that the characteristics of cyber-attacks tend to change over time, issues that need to be addressed include how to obtain adequate data covering different periods of time.

Future Approaches to Combining AI and Security

Always finding itself one step behind attackers, the as yet unrealized dream of cybersecurity research is to develop some means of getting ahead of attackers. While more in-depth research is still needed into the four perspectives described above, it will likely also be important to find effective ways of combining (4) countermeasures using AI with approaches based on the other three perspectives. When AI-based countermeasures [(4), above] are applied to attacks that themselves use AI [(1), above], for example, it might be possible for countermeasures to be put in place before new types of attacks occur. That is, this use of AI to investigate different threats might allow anticipatory defenses to be implemented. One form of research that might enable this would be to establish testbeds where AI techniques for concealing and detecting malware can be pitted against one another in a competitive format so that potential new threats can be identified through the simulation of attack and defense.

Graduated from the University of Tokyo and joined Hitachi, Ltd. in 1971. He worked on the research and development of security technology and other network management systems at Hitachi's Systems Development Laboratory. He was appointed a professor at Tokyo Denki University in 2001. Following positions that included Fellow Professor at the Center for Research and Collaboration of the Research Institute for Science and Technology and Director of Cyber Security Laboratories, he took up his current position in 2020. Other past appointments include President of the Japan Society of Security Management and Cyber Security Advisor to the Cabinet Office. He holds a doctorate in engineering from the University of Tokyo.

He has numerous publications, including as the author of "How to Deal with IT Risk" (Iwanami Shoten, Publishers), co-author of "IT Risk Science —Beyond the Information Security—" (Kyoritsu Shuppan Co., Ltd.), and editor of "Cyber Risk Management in a Connected World: Supply Chain Strategies for Society 5.0" (Toyo Keizai Inc.).

Conclusions

Along with boosting the deployment of AI itself, comprehensive research and development that combines both AI and security plays an important role in achieving greater progress on DX. The author expects Hitachi to promote even better comprehensive research and development. He also expects Hitachi will utilize the technologies and other outcomes of this work to enable new collaborative creation with customers using Lumada.