# Artificial Intelligence Cyberattacks in Red Teaming: A Scoping Review

Mays Al-Azzawi[1], Dung Doan[2], Tuomo Sipola[3(✉)], Jari Hautamäki[3], and Tero Kokkonen[3]

Institute of Information Technology,
Jamk University of Applied Sciences,
Jyväskylä, Finland
[1]ab0168@student.jamk.fi
[2]aa7785@student.jamk.fi
[3]{tuomo.sipola, jari.hautamaki, tero.kokkonen}@jamk.fi

**Abstract.** Advances in artificial intelligence are creating possibilities to use these methods in red team activities, such as cyberattacks. These AI attacks can automate the process of penetrating a target or collecting sensitive data while accelerating the pace of carrying out the attacks. This survey explores how AI is employed in cybersecurity attacks and what kind of targets are typical. We used scoping review methodology to sift through articles to find out AI methods, targets, and models that red teams can use to emulate cybercrime. Out of the 470 records screened, 11 were included in the review. Multiple cyberattack methods can be found to exploit sensitive data, systems, social media user profiles, passwords, and URLs. The use of AI in cybercrime to build versatile attack models poses a growing threat. Additionally, cybersecurity can use AI-based techniques to offer better protection tools to deal with those problems.

**Keywords:** artificial intelligence, red team, red teaming, cyberattack, cybersecurity

## 1 Introduction

The landscape of cybersecurity has undergone an enormous change in the last few years. One phenomenon that stands out is the possibility of artificial intelligence simulating human behavior. The behavior of artificial intelligence in cybersecurity can lead to dangerous situations in terms of security. Using AI as a method for attacks has developed in tandem with the development of attack methodologies and AI capabilities. Only a few cases are reported, and simulating human acts has become more feasible in the last few years.

The term red teaming originates from the military domain as a way to role-play adversaries or assess vulnerabilities [12]. The term red team also originates from widely used military symbols such as APP-6 by NATO or MIL-STD-2525 by U.S. Department of Defense, where the hostile (and suspect) identity is indicated with a red color [15,1]. In the context of cybersecurity, U.S. National Institute



The original article appeared as: Mays Al-Azzawi, Dung Doan, Tuomo Sipola, Jari Hautamäki and Tero Kokkonen. "Artificial Intelligence Cyberattacks in Red Teaming: A Scoping Review." In: *Good Practices and New Perspectives in Information Systems and Technologies. WorldCIST 2024.* Ed. by Álvaro Rocha, Hojjat Adeli, Gintautas Dzemyda, Fernando Moreira, and Aneta Poniszewska-Marańda. Vol. 1. Lecture Notes in Networks and Systems 985. Cham, Switzerland: Springer, 2024, pp. 129–138. https://doi.org/10.1007/978-3-031-60215-3_13

of Standards and Technology (NIST) defines a red team as follows: *"A group of people authorized and organized to emulate a potential adversary's attack."* [5] The red teams improve enterprise security by demonstrating the impacts of successful attacks [5]. In the context of cybersecurity, the term red team is used in cybersecurity exercises and in security testing. In cybersecurity exercises, red teams (RT) simulate the threat actors of the exercise scenario by executing cyberattacks against blue teams (BT), which are defending their assets [11,3,24,9,19]. In security testing, the red team is the group of security testers.

AI red teaming can be understood as an activity from two different perspectives. Several large technology companies use red teaming to expose weaknesses and vulnerabilities in their systems [18,27]. Another aspect is the use of AI to carry out attacks, which can be targeted against technical systems. On the other hand, in social engineering-type attacks, AI is used as a stepping stone to advanced persistent threat (APT) attacks by searching for suitable victims that can be targeted by AI-generated ghost messages [6,17]. The advantage of AI specifically in such attacks is the ability to enable mass attacks using phishing techniques to open attack vectors to multiple targets instead of manual attacks. For example, AI-generated phishing messages in target language create persuasive attack vectors. AI-based solutions are built to make operations more effective. Automating the process of planning attacks for automated cybersecurity testing scenarios could save time and effort [26]. As new artificial intelligence technologies have become more prevalent, automation is easier to implement, although its impact on work and society should be studied [22].

In order to investigate the use of AI for cyberattacks for red teaming, we carried out a scoping review. To examine how AI can be used for cyberattacks, red team actions, and hacking, our research questions were the following:

- *RQ1:* What AI attack methods are there?
- *RQ2:* What are the targets of such attacks?

Next, this paper describes the used scoping review methodology in Section 2, including a figure of the review protocol. The results of the review are presented in Section 3 with two tables summarizing the main findings. Finally, a conclusion is provided in Section 4.

## 2 Methodology

We used the scoping review method [14] to search the academic Finna[1] library database and Google Scholar[2] in order to define the scope of our topic. The review considered the following keywords: ' defensive mission', 'AI-enabled cyber operations', 'AI-augmented cyber defenses', 'national defense postures', 'poisoning attacks', 'offensive cyber operations', 'Cyber activities', 'AI cyber operations', 'AI cyber defense', 'AI cyber attack', 'AI red teaming', 'AI-enabled cyber

[1] https://janet.finna.fi/
[2] https://scholar.google.com/

campaigns', and 'cyber attacks'. In the initial stage, we identified 471 articles (and some book chapters) by screening their titles and abstracts within the 2015–2023 timeframe, found at the time of the research in mid-2023. We included articles written in English with available abstracts. During the second phase of the research, a more involved analysis of these articles was conducted. This analysis included reading the articles closely and concentrating on the topic at hand to precisely determine their content and classify them as directly relevant to addressing the research questions (RQ1, RQ2). We used the following criteria to find answers:

1. Is there a description of an attack method?
2. What was the target of the attacks?
3. How was the attack conducted?
4. What was the cyber-attack methodology used?

The result of the second stage of the research yielded 11 articles related to the subject matter. In the third stage of the research, we composed summaries, which also involved addressing the aforementioned questions when applicable. This comprehensive analysis of the included studies enabled us to gather information on the utilization of AI in red teaming. The review process is detailed by the PRISMA flow chart [13] in Figure 1.

## 3 Results

### 3.1 Attack methods

The literature review encompassed studies published from 2015 to 2023 in which we identified various cyberattack methods. The following techniques were documented in those studies. *Classification methods:* decision tree, convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), support vector machine (SVM), support vector classification (SVC), deep neural network (DNN), least squares support vector machine (LS-SVM), natural language processing (NLP), one-versus-all (OVA), double deep Q-network (DDQN), advantage actor-critic (A3C) regularized least-squares classification (RLSC), domain generation algorithms (DGA). *Regression methods:* generative adversarial network (GAN), random forest (RF), multilayer perceptron (MLP), gradient boosting regression trees (GBRT), artificial neural network (ANN), logistic regression, generalized likelihood ratio test (GLRT), *Clustering strategies:* k-means clustering, restricted Boltzmann machine (RBM), particle swarm optimization (PSO), genetic algorithm (GA), deep autoencoder (DAE), Lagrangian firefly algorithm (LFA), *Other specific methods:* nonsymmetric deep autoencoder (NDAE), cycle-GAN, combining TensorFlow object detection and a speech segmentation method with convolutional neural network (TOD+CNN), k-nearest neighbors (KNN), reinforcement learning (RL), gray wolf optimization (GWO), random weight network (RWN), ML-based approach named MLAPT, software-defined networking (SDN), and singular value decomposition (SVD).

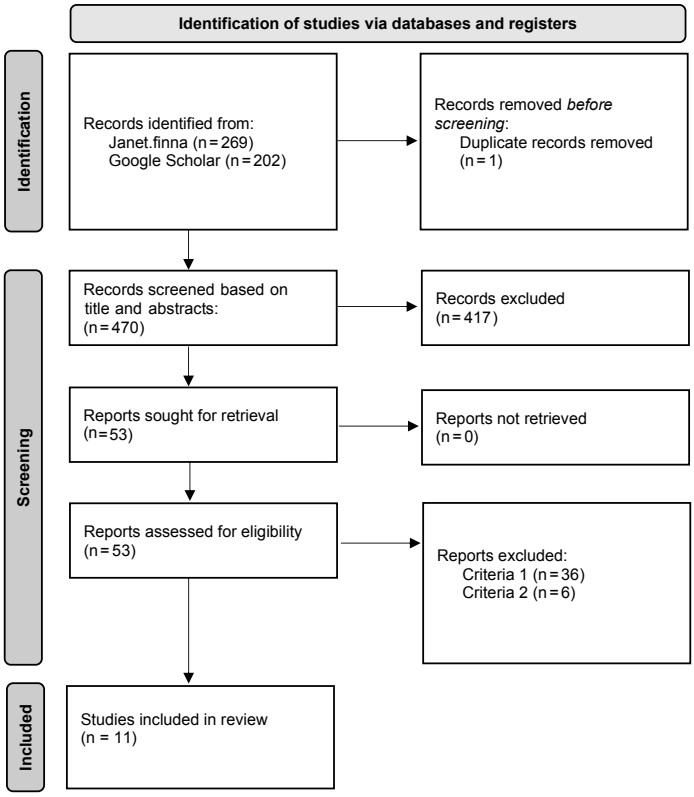


**Fig. 1.** Review protocol

Among these methods, LSTM was the most frequently used, appearing in 5 of the reviewed articles, while GANs and SVM were employed in 4 studies each. Additionally, CNN, RNN, KNN, MLP, and DNN were each featured in three of the reviewed articles. Other methods were referenced only once or twice. For a list of the attack methods in the reviewed articles, refer to Table 1.

### 3.2 Attack targets

Furthermore, we identified common targets that cyberattackers typically aim at (see Table 2 for tabulation of targets), including:

- General data, such as health data, personal data, and sensitive data, including financial and government data, were the most frequently targeted, appearing in 4 of the reviewed articles. [2] [20] [28] [25]
- URLs: Attackers also frequently targeted URLs, with 3 instances in the reviewed articles. [10] [28] [7]
- Social media user profiles: This category was the target in 2 of the sources. [10] [7]
- Passwords: Passwords were a target in 2 sources. [28] [7]
- Details of systems: Details of systems were targeted in one article. [25]

### 3.3 Summaries

The use of AI has been identified as a cyberattack method and recognized as a potential risk. However, Clinton only presents AI as a hacking method, and we did not find any other specific attack methods. [4]

Ward et al. have defined artificial intelligence as a new technology used by hackers and have mentioned "poison" attacks utilizing machine learning algorithms. They also discuss automated vehicles and the potential for high-risk attacks on vehicle systems. However, during their discussion of AI hacking methods, no specific attack methods were mentioned. [23]

From 2015 to 2018, the articles about AI-hacking did not mention any attack methods, and targets were mainly data and sensitive data.

Yamin et al. focused on raising awareness about the use of artificial intelligence as an attack method and assessed its impact on military operations. They employed GANs and Nash equilibrium to describe the attack methods. The targets of these attacks included traffic signs, medical image data, facial image data, digital recommendation systems, CT-scan data, speech and audio data, as well as network intrusion detection systems. The attacks were carried out using malicious AI algorithms designed to manipulate data to evade benign AI algorithm classifiers. The methodologies employed in these cyberattacks included DeepHack, DeepLocker, GyoiThon, EagleEye, Malware-GAN, UriDeep, Deep Exploit, and DeepGenerator. [25]

The article by Kaloudi et al. investigates AI's threat to SCPS. It explores how AI can be used as a malicious tool, emphasizing its potential to increase

**Table 1.** Methodologies found in the reviewed articles.

| Author | Pistono and Yampolskiy | Brundage et al. | Kaloudi and Li | King et al. | Truong et al. | Zouave et al. | Yamin et al. | Wang et al. | Guembe et al. | $\sum$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | 2016 | 2018 | 2020 | 2020 | 2020 | 2020 | 2021 | 2022 | 2022 | |
| Reference | [16] | [2] | [8] | [10] | [20] | [28] | [25] | [21] | [7] | |
| Classification | | | | | | | | | | |
| Dec. tree | | | | | | x | | | | 1 |
| CNN | | | | | x | x | | | x | 3 |
| RNN | | | x | | | x | | | x | 3 |
| LSTM | | | x | | x | x | | x | x | 5 |
| SVM | | | | | x | x | | x | x | 4 |
| SVC | | | | | | x | | | x | 2 |
| DNN | | | x | | | | | x | x | 3 |
| LS-SVM | | | | | x | | | | | 1 |
| NLP | | | | | | x | | | | 1 |
| OVA | | | | | | x | | | | 1 |
| DDQN | | | | | | | | x | | 1 |
| A3C | | | | | | | | x | | 1 |
| RLSC | | | | | | x | | | | 1 |
| DGA | | | | | | x | | | | 1 |
| Regression | | | | | | | | | | |
| GANs | | | | | | x | x | x | x | 4 |
| RF | | | | | | x | | | x | 2 |
| MLP | | | | | x | x | | | x | 3 |
| GBRT | | | | | | x | | | x | 2 |
| ANN | | | | | x | | | | | 1 |
| Log. reg. | | | | | | x | | | | 1 |
| GLRT | | | | | x | | | | | 1 |
| Clustering | | | | | | | | | | |
| k-means | | | x | | | | | | | 1 |
| RBM | | | | | x | | | | | 1 |
| PSO | | | | | x | | | | | 1 |
| GA | | | | | x | | | | | 1 |
| DAE | | | | | x | | | | | 1 |
| LFA | | | | | x | | | | | 1 |
| Other | | | | | | | | | | |
| NDAE | | | | | x | | | | | 1 |
| CYCLE-GAN | | | | | | | | | x | 1 |
| TOD+CNN | | | | | | | | | x | 1 |
| KNN | | | | | x | x | | | x | 3 |
| RL | | | x | | | | | | | 1 |
| GWO | | | | | x | | | | | 1 |
| RWN | | | | | x | | | | | 1 |
| MLAPT | | | | | x | | | | | 1 |
| SDN | | | | | | | | x | | 1 |
| SVD | | | | | | | | x | | 1 |

**Table 2.** Attack targets found in the reviewed articles.

| Author | Pistono and Yampolskiy | Brundage et al. | Kaloudi and Li | King et al. | Truong et al. | Zouave et al. | Yamin et al. | Wang et al. | Guembe et al. | $\sum$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | 2016 | 2018 | 2020 | 2020 | 2020 | 2020 | 2021 | 2022 | 2022 | |
| Reference | [16] | [2] | [8] | [10] | [20] | [28] | [25] | [21] | [7] | |
| Data, sensitive data | | x | | | x | x | x | | | 4 |
| URLs | | | | x | | x | | | x | 3 |
| Social media user profiles | | | | x | | | | | x | 2 |
| Password | | | | | | x | | | x | 2 |
| Systems | | | | | | | x | | | 1 |

attack speed and success rates. Attack methods discussed include k-means clustering, RNN, LSTM, RL, and DNN. Case studies involve k-means clustering for phishing messages, RNN for deceptive reviews, LSTM for phishing URLs, RL for autonomous learning attacks, and DNN for cyberattacks. The paper also examines cyberattack methodologies, including DeepLocker, repurpose attacks, DeepHack, Deep-Phish, review attacks, and SNAP_R. [8]

Guembe et al. address the growing concern of AI-powered cyberattacks and provide insights into how AI can be maliciously utilized in such attacks. They employ various attack methods, including CNN, GAN, RNN, LSTM, SVC, SVM, cycle-GAN, TOD+CNN, RF, MLP, GBRT, KNN, and DNN. The targets of these attacks encompass public social media profiles, passwords, and URLs. The attacks are executed through techniques such as password guessing/cracking (brute-force attacks), intelligent captcha manipulation, smart abnormal behavioral generation, AI model manipulation, and the generation of sophisticated fake reviews. The cyberattack methodologies employed by the authors include DeepLocker, DeepHack, PassGAN, and HashCat. [7]

Truong et al. provide an insightful overview of how artificial intelligence can be leveraged in cybersecurity, both for offensive and defensive purposes. They employ a diverse set of attack methods, including SVM, RBM, MLP, KNN, CNN, PSO, GA, DAE, ANN, LS-SVM, NDAE, GWO, RWN, LFA, MLAPT, LSTM, and GLRT. The targets of these attacks encompass user identities, financial credentials, and sensitive data from large corporations, security agencies, and government organizations. These attacks serve various purposes, including detecting or categorizing malware, identifying network intrusions, countering phishing and spam attacks, mitigating Advanced Persistent Threats (APTs), and identifying domains generated by domain generation algorithms (DGAs). [20]

Articles in 2020 showed different attack methods, such as GANs, CNN, RNN, LSTM, SVM, and SVC, aimed at attacking sensitive data, social media user profiles, passwords, and URLs.

The article by Zouave et al. explores the possibilities and applications of AI throughout various stages of a cyberattack. The authors employ a wide range of attack methods, including RNN, LSTM, NLP, GAN, KNN, logistic regression, SVC, decision tree, RF, gradient boosting regression tree, SVM, MLP, RLSC, OvA, CNN, and DGA. These attacks target URLs, individuals' personal data

in search of relationships, passwords, captchas, and domains. The attacks are executed by creating deceptive URLs to evade automated detection, generating conversations that include harmful links and attachments, attempting password guessing and brute forcing, stealing passwords, solving captchas, and generating numerous random fake domains. The authors utilize cyberattack methodologies such as the DeepPhish algorithm, PassGAN, Torch RNN, Deeptcha, AGDs, and DeepDGA. [28]

In the article by Wang et al. the exploration focuses on poisoning attacks in machine learning, particularly within the context of automated vehicles. The authors utilize various attack methodologies harnessing AI techniques. These include deep learning and deep neural networks (DNN), known for their outstanding performance in recognition tasks like image classification and computer vision. Additionally, other methods are discussed, such as Generative Adversarial Networks (GAN), LSTM, SDN, DDQN (Deep Double Q-Network), Advantage Actor-Critic (A3C), SVM (Support Vector Machine), and singular value decomposition (SVD). [21]

The article by Brundage et al. provides a summary of workshop findings and the authors' conclusions on forecasting, preventing, and mitigating the detrimental impacts of malicious AI use. The targets included sensitive information or financial assets of individuals, specific members of crowds, and historical patterns of code vulnerabilities. The attacks were executed through various methods, such as spear phishing attacks, imitation of human-like behavior, facial recognition, the generation of custom malicious websites/emails/links, visual impersonation of another person in video chats, and the use of drones or autonomous vehicles to deliver explosives and cause accidents. Furthermore, the attackers were engaged in discovering new vulnerabilities and developing code to exploit them. However, no specific methods for these activities were mentioned in the report. [2]

In their article, King et al. introduced the term "AI-Crime" (AIC) to address two key questions regarding the threats posed by AI in criminal activities and potential solutions to mitigate these threats. However, the article does not specify the methods employed in these AIC activities. The primary target of these activities is social media users, particularly through the use of phishing links. [10]

The research paper by Pistono and Yampolskiy focuses on publishing papers related to malicious exploits and discusses the use of software with malicious capabilities, including truly artificially intelligent systems such as artificially intelligent viruses. The paper also introduces the term "Hazardous Intelligent Software" (HIS) to describe the use of intelligence in a malicious context. It highlights that intelligent systems can potentially become malevolent in various ways. However, the paper does not mention specific AI attack methods. [16]

## 4 Conclusion

In today's rapidly evolving digital landscape, cybercriminals are continuously adapting and enhancing their attack strategies, with a particular focus on lever-

aging AI-driven techniques. Our results indicate that primary targets (RQ2) include personal data as well as sensitive information held by governments, organizations, and individuals, spanning URLs, passwords, and critical systems. Furthermore, the results show that to achieve their malicious goals (RQ1), cybercriminals can exploit a wide array of machine learning methods, falling into distinct categories: Classification, Regression, and Clustering.

These categories contain various technologies used for attacks. *Classification Techniques:* They include a multitude of machine learning algorithms such as decision trees, CNN, RNN, LSTM, SVM, SVC, DNN, LS-SVM, NLP, OVA, DDQN, A3C, RLSC, and DGA. These methods enable cybercriminals to classify and categorize data, often to identify vulnerabilities or potential targets. *Regression Methods:* In this category, we find techniques such as GANs, RF, MLP, GBRT, MLP, ANN, Logistic Regression, and GLRT. These approaches are employed to predict and estimate various variables, ranging from password guessing to system security breaches. *Clustering Strategies:* Cybercriminals also rely on clustering methods such as k-means clustering, RBM, PSO, GA, DAE, and LFA. Clustering helps them identify patterns within data, which can be exploited for nefarious purposes.

Cybercriminals employ sophisticated methodologies like the DeepPhish algorithm, PassGAN, Torch RNN, and Deeptcha. These tools aid them in tasks such as cracking passwords, phishing attacks, and infiltrating secure systems. As the threat landscape continues to evolve, it is imperative for the security research community, government agencies, and cybersecurity experts to remain vigilant and well-prepared against AI-based attacks. Red teaming using these AI-based attacks could reveal vulnerabilities to novel attacks. Effective countermeasures and proactive strategies must be developed to address the growing challenges posed by AI-driven cyberattacks.

## Acknowledgements

This research was partially funded by the Resilience of Modern Value Chains in a Sustainable Energy System project, co-funded by the European Union and the Regional Council of Central Finland (grant number J10052). The authors would like to thank Ms. Tuula Kotikoski for proofreading the manuscript.

## References

1. Department of defence interface standard, common warfighting symbology. Standard MIL-STD-2525C, United States of America, Department of Defence (2008)
2. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., et al.: The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228 (2018)
3. Brynielsson, J., Franke, U., Tariq, M.A., Varga, S.: Using Cyber Defense Exercises to Obtain Additional Data for Attacker Profiling. In: 2016 IEEE Conference on Intelligence and Security Informatics (ISI), pp. 37–42 (2016). DOI 10.1109/ISI.2016.7745440

4. Clinton, L. (ed.): Cybersecurity for business. Kogan Page, London, England (2022)
5. Computer Security Resource Center (CSRC) of National Institute of Standards and Technology (NIST): The glossary of terms and definitions extracted verbatim from nist's cybersecurity- and privacy-related publications. URL https://csrc.nist.gov/glossary/term/red_team. Accessed: 15 September 2023
6. Ghafir, I., Prenosil, V.: Advanced persistent threat and spear phishing emails. In: M. Hrubý (ed.) Proceedings of the International Conference Distance Learning, Simulation and Communication 'DLSC 2015', pp. 34–41. University of Defence, Brno, Czech Republic (2015)
7. Guembe, B., Azeta, A., Misra, S., Osamor, V.C., Fernandez-Sanz, L., Pospelova, V.: The emerging threat of ai-driven cyber attacks: A review. Applied Artificial Intelligence **36**(1), 2037,254 (2022)
8. Kaloudi, N., Li, J.: The AI-based cyber threat landscape: A survey. ACM Computing Surveys (CSUR) **53**(1), 1–34 (2020)
9. Kick, J.: Cyber exercise playbook (2014). URL https://www.mitre.org/news-insights/publication/cyber-exercise-playbook. Accessed: 15 September 2023
10. King, T.C., Aggarwal, N., Taddeo, M., Floridi, L.: Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. Science and engineering ethics **26**, 89–120 (2020)
11. Kokkonen, T., Puuska, S.: Blue team communication and reporting for enhancing situational awareness from white team perspective in cyber security exercises. In: O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy (eds.) Internet of Things, Smart Spaces, and Next Generation Networks and Systems, pp. 277–288. Springer International Publishing, Cham (2018)
12. Longbine, D.F.: Red Teaming: Past and Present. School of Advanced Military Studies, Fort Leavenworth, Kansas (2008)
13. McGowan, J., Straus, S., Moher, D., Langlois, E.V., O'Brien, K.K., Horsley, T., Aldcroft, A., Zarin, W., Garitty, C.M., Hempel, S., Lillie, E., Özge Tunçalp, Tricco, A.C.: Reporting scoping reviews—PRISMA ScR extension. Journal of Clinical Epidemiology **123**, 177–179 (2020). DOI 10.1016/j.jclinepi.2020.03.016. URL https://doi.org/10.1016%2Fj.jclinepi.2020.03.016
14. Munn, Z., Peters, M.D., Stern, C., Tufanaru, C., McArthur, A., Aromataris, E.: Systematic review or scoping review? guidance for authors when choosing between a systematic or scoping review approach. BMC medical research methodology **18**, 1–7 (2018)
15. NATO Standardization Office (NSO): Nato standard app-6, nato joint military symbology. Standard Edition D, Version 1, North Atlantic Treaty Organization (NATO) (2017)
16. Pistono, F., Yampolskiy, R.V.: Unethical research: how to create a malevolent artificial intelligence. In: Proceedings of Ethics for Artificial Intelligence Workshop (AI-Ethics-2016), pp. 1–7 (2016)
17. Renaud, K., Warkentin, M., Westerman, G.: From ChatGPT to HackGPT: Meeting the cybersecurity threat of generative AI. MIT Sloan Management Review (2023). Reprint #64428
18. Smith, J., Theisen, C., Barik, T.: A case study of software security red teams at Microsoft. In: 2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 1–10. IEEE (2020). DOI 10.1109/VL/HCC50065.2020.9127203
19. Sommestad, T., Hallberg, J.: Cyber security exercises and competitions as a platform for cyber security experiments. In: A. Jøsang, B. Carlsson (eds.) Secure IT

Systems, pp. 47–60. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). DOI 10.1007/978-3-642-34210-3_4
20. Truong, T.C., Diep, Q.B., Zelinka, I.: Artificial intelligence in the cyber domain: Offense and defense. Symmetry **12**(3), 410 (2020)
21. Wang, C., Chen, J., Yang, Y., Ma, X., Liu, J.: Poisoning attacks and countermeasures in intelligent networks: Status quo and prospects. Digital Communications and Networks **8**(2), 225–234 (2022)
22. Wang, W., Siau, K.: Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. Journal of Database Management **30**(1), 61–79 (2019). DOI 10.4018/JDM.2019010104
23. Ward, D., Wooderson, P.: Automotive Cybersecurity: An Introduction to ISO/SAE 21434, p. 106. SAE International (2021)
24. Wilhelmson, N., Svensson, T.: Handbook for planning, running and evaluating information technology and cyber security exercises. The Swedish National Defence College, Center for Asymmetric Threats Studies (CATS) (2014)
25. Yamin, M.M., Ullah, M., Ullah, H., Katt, B.: Weaponized AI for cyber attacks. Journal of Information Security and Applications **57**, 102,722 (2021)
26. Yuen, J.: Automated Cyber Red Teaming. DSTO Defence Science and Technology Organisation, Edinburgh, Australia (2015)
27. Zhou, W.C., Sun, S.L.: Red Teaming Strategy: Huawei's Organizational Learning and Resilience, pp. 299–317. Springer International Publishing, Cham (2020). DOI 10.1007/978-3-030-47579-6_13
28. Zouave, E., Bruce, M., Colde, K., Jaitner, M., Rodhe, I., Gustafsson, T.: Artificially intelligent cyberattacks. Swedish Defence Research Agency, FOI, Tech. Rep. FOI (2020)