An Overview of Artificial Intelligence Used in Malware

Lothar Fritsch(✉), Aws Jaber, and Anis Yazidi

Department of Information Technology, Faculty of Technology, Art and Design, Oslo Metropolitan University, Oslo, Norway
{lotharfr,awsalzar,anisy}@oslomet.no
https://www.oslomet.no

Abstract. Artificial intelligence (AI) and machine learning (ML) methods are increasingly adopted in cyberattacks. AI supports the establishment of covert channels, as well as the obfuscation of malware. Additionally, AI results in new forms of phishing attacks and enables hard-to-detect cyber-physical sabotage. Malware creators increasingly deploy AI and ML methods to improve their attack's capabilities. Defenders must therefore expect unconventional malware with new, sophisticated and changing features and functions. AI's potential for automation of complex tasks serves as a challenge in the face of defensive deployment of anti-malware AI techniques. This article summarizes the state of the art in AI-enhanced malware and the evasion and attack techniques it uses against AI-supported defensive systems. Our findings include articles describing targeted attacks against AI detection functions, advanced payload obfuscation techniques, evasion of networked communication with AI methods, malware for unsupervised-learning-based cyber-physical sabotage, decentralized botnet control using swarm intelligence and the concealment of malware payloads within neural networks that fulfill other purposes.

Keywords: Information security · Artificial intelligence · Malware · Steganography · Covert channels · Machine learning · Adverse artificial intelligence

1 Introduction

In recent years, AI has been increasingly adopted as part of cyber attack methods. The application of AI on the defender's side has been successfully used in intrusion detection systems and is widely deployed in network filtering, phishing protection, and botnet control. However, the enhancement of the capabilities of malware with the help of AI methods is a relatively recent development.

This article presents the result of a literature survey mapping the state of AI-powered malware. The salient aims of this survey is to map AI-enhanced attacks carried out by malware, to identify malware types that conceal themselves from detection using AI techniques, to get a better understanding of the maturity

E. Zouganeli et al. (Eds.): NAIS 2022, CCIS 1650, pp. 41–51, 2022.
https://doi.org/10.1007/978-3-031-17030-0_4

of those attacks, and to identify the algorithms and methods involved in those attacks (Fig. 1 and Table 1).

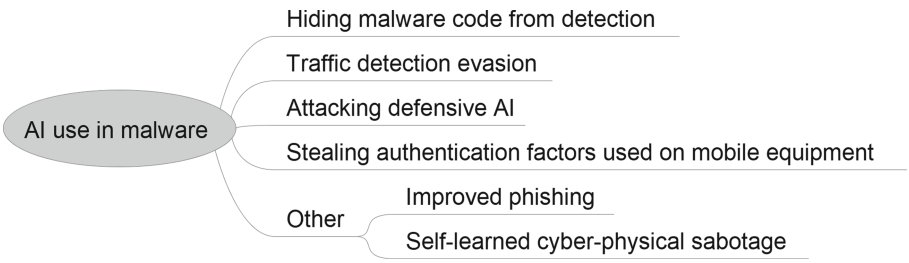

Fig. 1. Uses of AI in malware.

Table 1. Table of acronyms

Acronym	Expanded
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CC	Command and Control
DNN	Deep Neural Network
GAN	Generative Adversarial Neural Networks
ML	Machine Learning

2 Literature Review on AI-Powered Malware

2.1 Literature Search

For assessing the state of the art in AI-supported malware, we performed a literature search using the Google Scholar database of scientific publications. We defined the search criteria as follows. Search keywords were *malware, artificial intelligence, machine learning* combined with *offensive, adversarial, attack, network security, information security*. The resulting articles were checked against inclusion criteria. The resulting article set was then snowballed backward and forward [36]. We limited the backward snowballing range by cutting off snowballing for articles older than 2010. Eligible forms of publications were *scientific articles, conference presentation, pre-prints and technical reports*. For inclusion, articles needed to contain *descriptions of malware functionality based on machine learning or AI functionality*. Both *survey articles* as well as *articles describing demonstrators or specific malware* were included. Our final set of articles were 37 articles.

After collecting the articles, we classified the articles into categories reflecting the specific malware functionality enhanced with AI techniques. Our findings are summarized below.

2.2 Findings

Among the deployed technologies are authentication factor extraction, generation of phishing and malware domain names, adaptive generation of phishing-e-mail, direct attacks against malware detection (code obfuscation, model poisoning) and intrusion detection (generative traffic imitation as well as AI model poisoning attacks). In addition, we found publications describing the successful parsing and controlling of graphical application user interfaces (GUIs). Finally, self-learning malware aimed at sabotage of or through cyber-physical systems was found. In particular, the evasion of detection of malware and the exfiltration of information through covert channels have been recently used in AI-powered malware.

The establishment of covert channels is an established practice for malware distribution, command and control of malware agents, and information exfiltration. Such covert channels intend to bypass intrusion detection, malware detection, and anomaly detection systems.

2.3 Surveys

Our search found 13 survey articles that were either fully or partially present knowledge about AI-enhanced malware (see Table 2). We found ten surveys, two taxonomic articles, and one anecdotal collection of AI attack use cases.

The surveys focus on different perspectives of the offensive use of AI against information security in malware:

- Surveys that summarize the use of AI-enhanced malware for different purposes: Probing, scanning, spoofing, misdirection, execution, or bypass;
- Summary of methods and algorithms used for direct attacks against a defender's AI and ML systems, e.g. evasion attacks, model poisoning, adverse samples.
- Surveys of malware improvements concerning exfiltration, code permutation, automation, and reverse engineering with AI;
- Surveys on generative networks used for attack and defense;
- Survey on stegomalware, where AI is used to hide malware in images;
- Several surveys taxonomiz offensive AI in malware into categories: intelligence, evasion, target selection, attack automation, generating malware, hiding malware, combining attack techniques, adjusting features, automating attacks at high speed.

Table 2. Surveys and taxonomies

Paper	Malware class (purpose)	AI capability used (algorithm, goal)
[18]	Malicious uses of AI: Probe, Scan, Spoof, Flood, Misdirect, Execute, Bypass	Survey with both use cases, algorithms mentioned and references to prototypes
[34]	AI Exfiltration and intelligent malware background	Various sources for exfiltration, permutation of code, reverse engineering of functionality, automation
[17]	Attack opportunities for AI attacks in COVID-19 themed fraud	Attack cases and known implementations
[25,26]	Systematic taxonomy of adversarial attacks against ML	Detailed analysis of attack goals, algorithms, threat model
[21]	Attacks on ML in Training and Inference phase	Poisoning, Evasion, Impersonation, Inversion, Summary of algorithms
[3]	Use of generative networks in attack and defense	Describes various application areas and attacks
[5]	Stegomalware - hiding malware in images (evasion)	Large survey over algorithms and their performance
[11]	Weaponizing code, use cases and risks	Issues of control, deployment, Proliferation of AI cyberweapons
[32]	AI techniques in malware	Evasion, Autonomy, Anti-AI, Bio-inspired
[38]	AI-empowered cyberattacks	Including malware capabilities and references
[20]	Speculative taxonomy of malware with AI	Various purposes: intelligence, Evasion, Target selection, Attack automation, Generating malware, Hiding malware, Combining attack techniques, Adjusting features, Automating attacks at high speed,
[10]	Anecdotal enumeration of AI attack use cases	No algorithms or methods mentioned

2.4 AI-Enabled Attacks on Authentication Factors

Four articles described attacks against authentication factors on mobile devices'. The devices' sensors (microphone, accelerometer) were used in combination with AI models with the intention of extracting PINs, passwords, and patterns. The articles are listed in Table 3. We found two categories of AI weaponization against authentication factors:

– Prediction of PINs and passwords using accelerometer sensors in phones and wearables;
– Analysis of phone microphone records to generate PIN and credit card numbers from touch tones;

Table 3. Password extraction or prediction

Paper	Malware class (purpose)	AI capability used (algorithm, goal)
[27]	Smartphone PIN prediction using smartwatch motion sensors	Random forest classifier
[28]	Soundcomber: Extraction of PIN and credit card numbers through mobile phone microphones	Speech and touchtone analysis based on model
[30]	PIN skimmer: prediction of PIN codes using smartphone sensors	Prediction model in mobile malware
[23]	Password extraction through mobile device accelerometer	Classifier, random forest, 46 features

2.5 Techniques for Hiding Malware Code from Detection

AI is frequently used for hiding malware code from detection. The eleven articles listed in Table 4 show these approaches:

1. Hiding malware code as payload inside AI models fulfilling other functions, e.g., neural networks for face recognition;
2. Code perturbation for detection evasion automated with learning algorithms and prediction;
3. Code generation with Generative Adversarial Networks that blackbox-test filters for successful evasion;
4. Attacking AI systems for malware detection through attacks against the learning function (presentation of malicious samples, model poisoning, gradient attacks);
5. Sandbox detection in order to evade detection in sandboxed environments.

Table 4. Code detection evasion

Paper	Malware class (purpose)	AI capability used (algorithm, goal)
[15]	Hiding malware in Deep Neural Network (DNN)	Demonstrates how malware bytes can be hidden in neural networks without loss of DNN performance
[35]	EvilModel 2.0: Hiding malware - systematic experiments on model performance	Neural networks
[31]	Deeplocker: Hiding and targeting malware in neural networks	DNN, evasion, personalized biometric trigger
[12,13]	Proposal: Improved malware performance; some background on vulnerability prediction	Unsupervised learning, learning and evasion techniques are suggested (decision tree, Bayes)
[15]	Malware code obfuscation	A Turing-complete evolutionary algorithm able to generate completely new code, evaluated with Jacquard Similarity
[16]	Generating malware that bypasses malware filter	Generative Adversarial Network (GAN) with a substitute detector to fit the black-box malware detection system
[1]	Malware binary detection evasion	Prototyped code obfuscation with reinforcement learning tested against antivirus software
[19]	Detection evasion through gradient attack	Model poisoning in DNN through malicious samples
[6]	Evasion of malware detection based on OS API calls	Feature set manipulation using bidirectional feature selection, forward feature addition
[24]	Sandbox detection from within malware	Two methods using decision trees and neural networks

2.6 Evading Network Traffic Detection

Hiding malware's communication traffic is published in four articles (see Table 5). AI and specifically unsupervised learning, is deployed against intrusion detection systems. Demonstrators described in the articles hide probing and infiltration traffic as well as command and control traffic. One noteworthy article deploys swarm intelligence in order to coordinate Botnet agents without a centralized command server.

Table 5. Evasion of network intrusion detection

Paper	Malware class (purpose)	AI capability used (algorithm, goal)
[14]	Evasion: Perturbation of network traffic against learning IDS	Stochastic approximation and adaptive random search
[22]	Evasion of malware command and control traffic from detection	Generated adversarial samples
[33]	Evasion of network intrusion detection	GANs perturbate traffic patterns
[4]	Botnet coordination without hierarchical CC servers	Multi-agent-swarm using stigmeric communication model

2.7 Other AI Deployment

Table 6 lists the miscellaneous applications of AI in the malware context. We found six articles describing enhanced capabilities in the areas of phishing, Application control and sabotage. AI is used for creating phishing domain names that evade detection in anti-phishing-systems. One spear phishing demonstrator extracts social media sentiments using AI in order to turn them into phishing e-mail-text, learning which topics are susceptible of currently provoking most reaction from the targets.

An interesting application of image recognition is malware that can understand graphical user interface elements with AI with the goal of finding out which GUI elements it can control to execute functionality.

Finally, undetectable sabotage in cyber-physical systems has been demonstrated in two cases: i) A surgical robot which - injected with malware - can learn how to modify its actions similar to normal actions in order to hurt patients. ii) The second demonstration case showed how to AI can learn to manipulate smart house technology in ways that will be hard to notice. Such AI-empowered sabotage is envisioned to be used against variable targets, dramatically leveraging the preparation effort of cyber sabotage.

3 Discussion of Findings

The presented survey investigated the use of artificial intelligence (AI) techniques and of machine learning (ML) for the improvement of malware capabilities. We found surveys and literature that describe a variety of deployments of AI in the malware context:

Table 6. Miscellaneous AI applications in malware

Paper	Malware class (purpose)	AI capability used (algorithm, goal)
[29]	Spear phishing on social media	Phishing text generation with GAN models learning trendy topics from social media
[2]	Generation of undetectable phishing domain URLs (evasion)	GAN to construct a deep learning based Domain generation algorithms (DGA) that is designed to intentionally bypass a deep learning based detector.
[37]	Malware controlling GUI elements	AI-based object recognition
[7]	Cyber-physical attacks through hidden malicious behavior	Self-learning attack strategies, disguising, failure injection
[8]	Demonstrator: surgery robot with hidden malicious behavior	Failure injection, learning, disguising
[9]	Malware attacks on surgical robot and home automation	Statistical learning, payload generation and attack planning

- Direct sabotage of defending AI or ML algorithms;
- Detection evasion through intelligent code perturbation techniques;
- Detection evasion through learning of traffic patterns in case of scanning systems, communication or connection to command and control infrastructures;
- Black-box-techniques bypassing intrusion detection using generative networks and unsupervised learning;
- Direct attacks predicting passwords, PIN codes;
- Automatic interpretation of user interfaces for application control;
- Self-learning system behavior for undetected automated cyber-physical sabotage;
- Botnet coordination with swarm intelligence, removing need for command and control servers;
- Sandbox detection and evasion with neural networks;
- Hiding malware within images or neural networks.

We conclude that AI deployed to either improve or hide malware poses a considerable threat to malware detection. Code obfuscation, code behavior adaption, as well as learned communication detection evasion potentially bypass existing malware detection techniques.

Offensive deployment of AI within malware improves malware performance, including methods such as selection of targets, extracting authentication factors, enabling the automated and fast generation of highly efficient Phishing messages, and swarm-coordinated action planning.

We consider AI-enhanced malware to be a serious risk for information security, which should be thoroughly investigated.

Acknowledgements. The work leading to this article was partially sponsored by OsloMET's AI Lab.

References

1. Anderson, H.S., Kharkar, A., Filar, B., Evans, D., Roth, P.: Learning to evade static PE machine learning malware models via reinforcement learning. arXiv:1801.08917 [cs] (2018)
2. Anderson, H.S., Woodbridge, J., Filar, B.: DeepDGA: adversarially-tuned domain generation and detection, pp. 13–21 (2016). https://doi.org/10.1145/2996758.2996767
3. Bauer, L.A., Bindschaedler, V.: Generative models for security: attacks, defenses, and opportunities (2021). http://arxiv.org/abs/2107.10139
4. Castiglione, A., De Prisco, R., De Santis, A., Fiore, U., Palmieri, F.: A botnet-based command and control approach relying on swarm intelligence. J. Netw. Comput. Appl. **38**, 22–33 (2014). https://www.sciencedirect.com/science/article/pii/S1084804513001161
5. Chaganti, R., Ravi, V., Alazab, M., Pham, T.D.: Stegomalware: a systematic survey of malwarehiding and detection in images, machine learningmodels and research challenges (2021). https://arxiv.org/abs/2110.02504v1
6. Chen, L., Ye, Y., Bourlai, T.: Adversarial machine learning in malware detection: arms race between evasion attack and defense. In: 2017 European Intelligence and Security Informatics Conference (EISIC), pp. 99–106 (2017)
7. Chung, K., Kalbarczyk, Z.T., Iyer, R.K.: Availability attacks on computing systems through alteration of environmental control: smart malware approach. In: ICCPS 2019: ACM/IEEE 10th International Conference on Cyber-Physical Systems, pp. 1–12 (2019). https://doi.org/10.1145/3302509.3311041
8. Chung, K., et al.: Smart malware that uses leaked control data of robotic applications: the case of Raven-II surgical robots. In: 22nd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2019), pp. 337–351 (2019). https://www.usenix.org/conference/raid2019/presentation/chung
9. Chung, K., et al.: Machine learning in the hands of a malicious adversary: a near future if not reality. In: Game Theory and Machine Learning for Cyber Security, pp. 289–316 (2021). https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119723950.ch15
10. CISOMAG: artificial intelligence as security solution and weaponization by hackers. CISO MAG Cyber Security Magazine (2019). https://cisomag.eccouncil.org/hackers-using-ai/
11. Cobb, S., Lee, A.: Malware is called malicious for a reason: the risks of weaponizing code. In: 2014 6th International Conference on Cyber Conflict (CyCon 2014), pp. 71–84 (2014)
12. Easttom, C.: Integrating machine learning algorithms in the engineering of weaponized malware. In: ECIAIR 2019, European Conference on the Impact of Artificial Intelligence and Robotics, pp. 113–121 (2019)
13. Easttom, C.: A methodological approach to weaponizing machine learning. In: The 2019 International Conference, pp. 1–5 (2019). http://dl.acm.org/citation.cfm?doid=3358331.3358376
14. Fladby, T., Haugerud, H., Nichele, S., Begnum, K., Yazidi, A.: Evading a machine learning-based intrusion detection system through adversarial perturbations, pp. 161–166 (2020). https://doi.org/10.1145/3400286.3418252
15. Gaudesi, M., Marcelli, A., Sanchez, E., Squillero, G., Tonda, A.: Malware obfuscation through evolutionary packers, pp. 757–758 (2015). https://doi.org/10.1145/2739482.2764940

16. Hu, W., Tan, Y.: Generating adversarial malware examples for black-box attacks based on GAN (2017). http://arxiv.org/abs/1702.05983
17. Jaber, A.N., Fritsch, L.: COVID-19 and global increases in cybersecurity attacks: review of possible adverse artificial intelligence attacks. In: 2021 25th International Computer Science and Engineering Conference (ICSEC), pp. 434–442, November 2021. https://doi.org/10.1109/ICSEC53205.2021.9684603
18. Kamoun, F., Iqbal, F., Esseghir, M.A., Baker, T.: AI and machine learning: a mixed blessing for cybersecurity. In: 2020 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–7 (2020)
19. Kolosnjaji, B., et al.: Adversarial malware binaries: evading deep learning for malware detection in executables. In: 2018 26th European Signal Processing Conference (EUSIPCO), pp. 533–537, September 2018. https://doi.org/10.23919/EUSIPCO.2018.8553214
20. Kubovič, O., Košinár, P., Jánošík, J.: Can artificial intelligence power future malware? Technical report, ESET (2018)
21. Liu, Q., Li, P., Zhao, W., Cai, W., Yu, S., Leung, V.C.M.: A survey on security threats and defensive techniques of machine learning: a data driven view. IEEE Access **6**, 12103–12117 (2018)
22. Novo, C., Morla, R.: Flow-based detection and proxy-based evasion of encrypted malware C2 traffic, pp. 83–91 (2020). https://doi.org/10.1145/3411508.3421379
23. Owusu, E., Han, J., Das, S., Perrig, A., Zhang, J.: Accessory: password inference using accelerometers on smartphones, pp. 1–6 (2012). https://doi.org/10.1145/2162081.2162095
24. Pearce, W., Landers, N., Fulda, N.: Machine learning for offensive security: sandbox classification using decision trees and artificial neural networks. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) SAI 2020. AISC, vol. 1228, pp. 263–280. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52249-0_18
25. Rosenberg, I., Shabtai, A., Elovici, Y., Rokach, L.: Adversarial machine learning attacks and defense methods in the cyber security domain. ACM Comput. Surv. **54**(5), 108:1–108:36 (2021). https://doi.org/10.1145/3453158
26. Rosenberg, I., Shabtai, A., Elovici, Y., Rokach, L.: Adversarial machine learning attacks and defense methods in the cyber security domain - supplementary material. ACM Comput. Surv. **54**(5), 1–36 (2021). https://doi.org/10.1145/3453158
27. Sarkisyan, A., Debbiny, R., Nahapetian, A.: WristSnoop: smartphone pins prediction using smartwatch motion sensors. In: 2015 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–6 (2015)
28. Schlegel, R., Kapadia, A., Wang, X.: Soundcomber: a stealthy and context-aware sound trojan for smartphones. In: Proceedings of the Network and Distributed System Security Symposium (NDSS) (2011)
29. Seymour, J., Tully, P.: Generative models for spear phishing posts on social media. Technical report, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA (2018). http://arxiv.org/abs/1802.05196
30. Simon, L., Anderson, R.: PIN skimmer: inferring pins through the camera and microphone, pp. 67–78 (2013). https://doi.org/10.1145/2516760.2516770
31. Stoecklin, M.: DeepLocker: how AI can power a stealthy new breed of malware. Technical report, IBM (2018). https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/
32. Thanh, C.T., Zelinka, I.: A survey on artificial intelligence in malware as next-generation threats. Mendel **25**(2), 27–34 (2019). https://doi.org/10.13164/mendel.2019.2.027. https://mendel-journal.org/index.php/mendel/article/view/105

33. Usama, M., Asim, M., Latif, S., Qadir, J., Ala-Al-Fuqaha: Generative adversarial networks for launching and thwarting adversarial attacks on network intrusion detection systems. In: 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), pp. 78–83 (2019)
34. Varney, A.: Analysis of the impact of artificial intelligence to cybersecurity and protected digital ecosystems. Technical report, October 2021
35. Wang, Z., Liu, C., Cui, X., Yin, J.: EvilModel 2.0: hiding malware inside of neural network models. arXiv:2109.04344 [cs] (2021)
36. Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, EASE 2014, pp. 1–10. Association for Computing Machinery, New York, May 2014. https://doi.org/10.1145/2601248.2601268
37. Yu, N., Tuttle, Z., Thurnau, C.J., Mireku, E.: AI-powered GUI attack and its defensive methods, pp. 79–86 (2020). https://doi.org/10.1145/3374135.3385270
38. Zouave, E., Gustafsson, T., Bruce, M., Colde, K.: Artificially intelligent cyber-attacks. Technical report, FOI-R-4947-SE, Totalförsvarets forskningsinstitut FOI, March 2020