HADMÉRNÖK
VÉDELEMINFORMATIKA

21. évfolyam (2026) 1. szám 25–39.
DOI: 10.32567/hm.2026.1.2

Zoltán Kovács[1]

The Use of Artificial Intelligence in Cyberattacks, Part 3

Phases 5–7 (+1) of the Cyber Kill Chain Model, Challenges, Trends, Outlook

Abstract

The first part of the series of articles provided a basic understanding of the fundamental concepts of artificial intelligence and its relevant subfields, and demonstrated that the Cyber Kill Chain (CKC) model is suitable for achieving the main objective of the article, despite all its limitations, i.e. it can be used to review what AI-supported tools attackers can already use in the various phases of a cyberattack and how these tools help them. The second part of the series examined the latter in phases 1–4 of the CKC, while this article, the third part of the series, shows how attackers can use AI in the last three plus one phases of the CKC and how it helps them achieve their goals. In addition, this article highlights the challenges attackers face when using AI and presents the trends that can be observed in cyberattacks. Finally, this article makes recommendations for the direction of further research, namely, to examine the possibilities of using AI-supported defence solutions in the individual phases of CKC and to compare the possibilities of the attacking and defending sides in order to develop effective response capabilities.

Keywords: artificial intelligence, cybersecurity, cyberattack, Cyber Kill Chain, C2, maintaining a persistent presence, exfiltration, system disruption

Introduction

The second part of this series of articles examined how attackers can use AI in phases 1–4 of the Cyber Kill Chain (CKC) model and how it helps them achieve their goals. One of the main findings of this article was that although in 2024, major cybersecurity companies still wrote in their annual studies that attackers typically rely on artificial

[1] Senior Lecturer, Ludovika University of Public Service, e-mail: zkovacs.24@gmail.com

intelligence in the first two phases of the CKC, i.e. reconnaissance and weaponisation, based on an examination of the first four phases, it can be said that over the past 1.5–2 years, with the development of AI, its malicious use has also advanced significantly. This can be measured in several ways. On the one hand, attackers have further refined previously known AI-assisted attack methods, making them even more effective. Examples include the production of extremely realistic deepfake content and the preparation and execution of convincing phishing attacks that bridge language gaps and are written with perfect grammar. On the other hand, AI support has also appeared in forms of attack that were previously unknown or only marginally encountered and were therefore little known to defenders. Examples include the integration of defence evasion techniques into malicious code, the creation of polymorphic[2] and metamorphic[3] malware, intelligent exploit generation, autonomous exploit execution and optimisation, and even AI support for acoustic side-channel attacks.

Based on the above, it can be concluded that if so many new developments have occured in the past 1.5–2 years, even in the early stages of CKC, when the forms of AI support for attacks and their usage were more familiar, then it is worthwhile to continue the research and examine the AI-supported attack options used by attackers in the later stages of CKC. The third part of this series of articles examines the current possibilities available to attackers in the last three (plus one) phases of CKC if they wish to use AI-supported tools for their attacks and shows how these tools can help attackers achieve their goals. It examines the challenges and trends that are emerging in the use of AI in attacks and then makes recommendations for the direction of further research.

The application of AI in the individual phases of the Cyber Kill Chain

The capabilities of artificial intelligence, and thus its use, significantly enhance attackers' capabilities in cyberattacks, enabling them to accelerate, automate and refine their execution. The second part of this series of articles presented in detail how attackers could utilise artificial intelligence in the first four phases of the CKC (i.e. the cyberattack chain) and what specific technologies they could use to achieve their malicious goals. This article examines the last three (plus one) phases of the CKC from the same perspective.

[2] Polymorphic malware: malicious software that can change its own code with each infection in order to make it more difficult to detect by traditional defence tools, such as antivirus software. Polymorphic malware changes its original code so that each time it infects a system, it has a different, unique code fragment and binary code structure, while its malicious function remains unchanged. Malware often includes a mutation engine that automatically generates new encryption keys and algorithms (ITszótár.hu 2025).

[3] Metamorphic malware: in addition to the characteristics of polymorphic malware, it is also capable of changing, rewriting, translating and editing its own code, so that each new version differs from the previous one, all without using an encryption key. This is a more advanced technique than polymorphic malware, which uses a key to modify the code. Metamorphic malware is more difficult to detect and identify because its code is constantly changing, it has no constant part that would identify it, and it does not return to its original form; each distributed version differs from the previous one. This rewriting may involve changing the order of the code, adding unnecessary instructions (so-called *garbage code*), or replacing existing instructions with instructions that have equivalent functionality but different code (ITszótár.hu 2025).

Phase 5: Installation

After successful exploitation, the installation phase follows. In this phase, the attacker installs malicious software (e.g. backdoor, rootkit, virus) to ensure a persistent presence on the compromised system. AI can help attackers here, for example by automating the installation of backdoors and helping to maintain a persistent hidden presence.[4]

- *Malware configuration, installation and privilege escalation:* AI can help attackers not only in creating the malware they use, but also in installing it, configuring it during operation, and obtaining higher privilege levels. They are able to collect system information, search key folders of the operating system, dynamically control malware, etc. AI can also facilitate privilege escalation, for example by using LLM design to generate exploit chains. These allow attackers to penetrate deeper into the network after the initial intrusion, install the prepared malicious code, and then be able to execute their final goal.[5]
- *Rootkit and bootkit*[6] *optimisation:* AI can optimise rootkits (malware hidden deep within the system that conceals its presence and activity, allowing attackers to access compromised machines and the data stored on them undetected) and bootkits (malware that compromises the system boot process, load before the operating system, and are able to mask their presence from the operating system and the applications installed on it, such as security software for example, antivirus programmes). This is because AI is capable of recognising kernel-level security mechanisms and making modifications to the malicious code that are least likely to trigger alarms from security devices and software, thus ensuring a persistent, deeply embedded presence in the compromised system.[7]
- *Hidden file system operations and defence against forensic investigations:* AI can help malware successfully installed on the target system to maintain a persistent presence and achieve its goals by performing the necessary operations on the file system (e.g. creating or modifying files, manipulating system logs or metadata) that leave as little digital footprint as possible. This can make it significantly more difficult to detect malicious activity and identify the source of the attack. This capability significantly increases the ability to maintain the hidden nature of attacks and makes it difficult to identify attackers.[8]
- *Intelligent, persistent presence maintenance mechanisms:* AI can analyse the target system configuration, installed software and user accounts in order to select the least detectable and most resistant mechanisms for the malicious code to maintain its presence. With the help of AI algorithms, malware can learn which registry keys, scheduled tasks, services, or file system modifications appear the least suspicious, and can automatically adapt if defence systems detect and

[4] Darktrace 2024.
[5] De Pasquale et al. 2024; Toulas 2025.
[6] A rootkit or bootkit is a malicious code that inserts itself into the boot process of a computer and is able to perform operations that would not be permitted by the operating system.
[7] Kewat 2025; Lee et al. 2025.
[8] Jain–Chhabra 2014; Pappachan et al. 2024.

attempt to remove them. AI can analyse the target system's configuration, installed software, and user accounts in order to select the least detectable and most resilient mechanisms for the malicious code to maintain its presence. With the help of AI algorithms, malware can learn which registry keys, scheduled tasks, services, or file system modifications appear the least suspicious, and can automatically adapt if defence systems detect and attempt to remove them. Factors such as the self-modifying capabilities of malware (polymorphic and metamorphic malware), the use of obfuscated code, the generation of false telemetry or log data to evade defence tools, the use and maintenance of persistent, multiple backdoors against system reboots or authentication changes, or the use of fileless malware techniques, as described earlier, also contribute to maintaining a persistent presence. In addition, AI-driven malware may be able to repair itself if certain components are deleted or if the persistence points necessary for its continued existence are modified.[9]

Based on the above, it can be said that AI can help increase the effectiveness of attacks even in the fifth phase of the CKC, i.e. the installation phase. AI enables malicious software to become more autonomous and resilient, gain higher privileges more quickly, and evade security measures, making it difficult to remove and ensuring its persistent presence within the compromised network. This attack capability also strengthens the necessity to shift the defence paradigm from relying on static signatures to AI-supported behavioural analysis and continuous monitoring to detect and eliminate advanced threats.

Phase 6: Command and Control (C2)

In the Command and Control (C2) phase of CKC, the attacker establishes an encrypted and hidden communication channel with the malware installed on the compromised system, enabling remote control and data exfiltration from the target system. AI enables autonomous management of C2 channels, allowing the malware to operate independently. The AI-controlled use of the C2 channel can dynamically adapt to network defences, thereby increasing the resilience of C2 communications against detection and helping to maintain a persistent connection between the control server and the malicious code running on the target system.[10]

- *Adaptive and covert C2 communications:* AI-based C2 modules are capable of dynamically changing their communication patterns (e.g. protocols, ports, timings, data formats) in order to avoid detection by security systems during network traffic analysis. With the help of machine learning (ML), the C2 agent can learn which network traffic patterns appear legitimate on a given network (e.g. web browsing, DNS queries, cloud-based services, etc.) and can blur or embed malicious communications within them. For example, AI-driven

[9] Schröer et al. 2025.
[10] Lockheed Martin s. a.; Powell 2025.

C2 modules may be capable of using automatic Domain Generation Algorithms (DGA), which, on the one hand, evaluate and classify DNS queries from compromised hosts using the Generative Adversarial Networks[11] (GANs) already mentioned in the previous article, and on the other hand, generate a large number of rapidly changing domain names, making it significantly more difficult for defence systems to effectively filter them out using blacklists. But AI-supported C2 modules are also capable of hiding C2 communications by embedding them in legitimate protocols using steganography. This makes it difficult or impossible for defence systems to detect such communications, as C2 communications are more difficult to track or, if detected, more difficult to block.[12]

- *Decentralised and flexible C2 infrastructure:* AI can help attackers manage the distributed and dynamically changing C2 infrastructure they create, which is more resistant to disconnections and blockages, enabling them to carry out effective attacks. AI can autonomously switch C2 servers as needed, use peer-to-peer (P2P) networks, or embed C2 capabilities into legitimate cloud services (e.g. Google Drive, Dropbox, etc.), making it significantly more difficult to detect and neutralise the attacker's infrastructure. AI-controlled botnets are capable of building and maintaining highly adaptive and difficult-to-eliminate C2 infrastructure, which they can use to launch extremely sophisticated, coordinated, and targeted attacks with unpredictable or difficult-to-predict appearances, changes and further evolution.[13]
- *Intelligent detection avoidance during network traffic monitoring:* The AI components controlling C2 communications may be able to learn from the operation of network anomaly detection systems (NADS). This includes the use of adversarial machine learning techniques, where AI makes minimal modifications to the communication pattern that are just enough to outsmart the defensive AI models, but do not interfere with the C2 functionality used for the attack.[14]

In summary it can be said that AI also plays a significant role in the command-and-control phase of CKC, where it can greatly assist attackers in achieving their objectives. AI enables the development, operation and maintenance of more covert, adaptive and autonomous C2, can reduce the digital footprint left behind by attackers, and can increase resistance to the activities of defenders.

[11] GANs consist of two neural networks (generator and discriminator) that compete with each other and ultimately produce better output (results) than if only one network were operating and providing results for a given question or task.
[12] Powell 2025; Anderson et al. 2016.
[13] Wang et al. 2022; Frisoni 2020.
[14] Abbadi–Lachkar 2024; Wang et al. 2023.

Phase 7: Actions on Objectives

This is the final phase of the Cyber Kill Chain, where the attacker achieves the ultimate goals of the attack. AI can significantly increase the effectiveness of the attack in this phase, as well as its speed or, in some cases, its destructive potential. AI can also effectively support the evasion of defence systems (e.g. dynamic malware behaviour) and, in the case of ransomware, it can help estimate the value of data, but it can also automate the collection of additional authentication data required for access, lateral movement, or the destruction or wiping of data in the target system and the disabling of systems.[15]

- *Optimised exfiltration:* AI can also help attackers by automatically identifying the most valuable and sensitive data within the network (e.g. intellectual property, personal data, financial information, critical configuration files, etc.), prioritise the information marked for acquisition, and optimise the routes and timing of the data to be exfiltrated so that the exfiltration of the data can be as inconspicuous as possible. To circumvent anomaly detection used on compromised networks, AI can use reinforcement learning to find out what network bandwidth usage or data packet size is considered normal and then *regulate* the extraction of the desired data accordingly.[16]
- *Intelligent execution of DDoS attacks and adaptation to defensive operations:* AI-based distributed denial-of-service (DDoS) botnets are capable of adapting in real time to defence mechanisms and changes in the infrastructure of their victims. AI can learn which attack vectors (e.g. volumetric targeting the network layer, application-level targeting the application layer) are most effective against a given target and dynamically modify the type of attack, traffic intensity, or source IP address rotation to achieve maximum results.[17]
- *Automated ransomware management:* AI can accelerate either the spread of ransomware within a compromised network or its operation, i.e. the encryption of data on target machines, by automating the identification of the most valuable files and systems that are worth encrypting for the attacker. AI-based chatbots are able to automate negotiations with ransomware victims based on a profile of the victim previously prepared with the help of AI, optimising the amount of ransom that can be obtained and the payment deadlines for the attacker, while minimising the manual effort required by the attackers. The use of AI-based chatbots further reduces the risk of the attacker being caught, as in the event of a possible *red-handed* arrest, the police will only find machines and not people.[18]
- *System disruption and integrity damage:* As mentioned earlier, AI can help attackers identify critical system components and/or data within a compromised network. This not only helps attackers extract or encrypt data, but can also optimise destructive operations (e.g. deleting data, disabling systems,

[15] Microsoft Security 2025.
[16] Al-Azzawi et al. 2025; Aboze 2025.
[17] Sangfor Technologies 2025; Utter 2024.
[18] Cherepanov–Strýček 2025; Trinckes s. a.

modifying firmware, etc.). All this is done in a way that causes the greatest possible damage while making it difficult for defenders to recover quickly. AI may also be able to develop strategies to maximise chaos, such as preparing operations aimed at disrupting recovery processes or damaging the integrity of backups, and may even be able to assist in their execution.[19]

Overall, it can be said that even in the final phase of CKC, i.e. the action on objectives, AI can significantly assist attackers in carrying out cyberattacks. The use of AI accelerates and optimises the final stage of a cyberattack by maximising the damage caused and financial gain, while minimising the time available for defenders to detect and respond. This means that the impact of successful intrusions is likely to be more severe and widespread.

Summary of AI use in the seven phases of Cyber Kill Chain

Based on the above, it can be said that the use of artificial intelligence in each phase of the Cyber Kill Chain significantly enhances the capabilities of attackers, automates, accelerates, and refines cyberattacks. Table 1 below summarises the use of AI in each phase of the CKC.

Table 1: The application of AI in the phases of the Cyber Kill Chain – Summary

CKC phase	**Main objective of the CKC phase**	**AI-enhanced attack capabilities and abilities**	**AI sub-branches that can be used for attacks**
1. Reconnaissance	Gathering information about the target	• Automated OSINT • Target profiling • Network mapping • Vulnerability detection • Human behaviour analysis • Data recovery from acoustic signals	ML, NLP, DL
2. Weaponisation	Preparation of malicious tools (exploit + payload)	• AI-based malware generation (polymorphic/metamorphic) • Exploit selection • Intelligent exploit generation • Generation of content that can be used for personalised social engineering attack • Integration of defence evasion techniques	Generative AI (GANs, LLMs), RL, DL, NLP
3. Delivery	Delivery of cyber weapons to their target	• Optimised delivery channels and timing • Delivery hiding • Automated password cracking • Execution of intelligent phishing campaigns • Supply chain attacks	ML, Generative AI, NLP

[19] Boutin 2025; Raz et al. 2025.

CKC phase	Main objective of the CKC phase	AI-enhanced attack capabilities and abilities	AI sub-branches that can be used for attacks
4. Exploitation	Exploiting vulnerabilities to gain access	• Independent exploit execution and optimisation • Application of intelligent defence evasion techniques • Application of polymorphism and metamorphism	RL, ML, DL
5. Installation	Providing persistent presence on the compromised system	• Malware configuration, installation, privilege escalation • Rootkit/bootkit optimisation • Hidden file system operations and defence against forensic investigations • Intelligent persistence mechanisms	ML, Adversarial ML
6. Command and Control (C2)	Establishment of encrypted communication channels	• Adaptive and hidden C2 communication • Decentralised and flexible C2 infrastructure • Avoidance of detection during network traffic monitoring	ML, Adversarial ML
7. Actions on Objectives	Achievement of the ultimate objectives of the attack	• Optimised data exfiltration • Execution of intelligent DDoS attacks • Automated ransomware management • System destruction	ML, RL

Source: compiled by the author

Phase +1: Monetisation

Although the monetisation phase is not part of Lockheed Martin's original model, many experts have added it to the model. Phase +1 focuses on the attacker's activities to monetise the information obtained from a successful attack. AI can also optimise processes and assist attackers in this phase.[20]

Ransomware and ransom payments: AI can assist attackers not only in creating and distributing ransomware, but also in managing ransom payments for stolen data and conducting negotiations themselves.[21]

Selling data on the dark web: AI can help evaluate and sell stolen data on the dark web, thereby optimising profits for attackers.[22]

Fraud and financial manipulation: AI-driven fraud, such as financial fraud committed using deepfakes (e.g. fake CEO calls), can cause significant financial losses. AI-based systems such as FraudGPT are capable of creating personalised business email compromise (BEC) emails and fraudulent websites, increasing their credibility and the success rate of social engineering attacks.[23]

[20] Microsoft Security 2025.
[21] Holden 2023.
[22] Tejedor 2025.
[23] Humphries 2026.

Overall, it can be concluded that phase +1 of CKC is closely related to the previous phases, as most of the elements described above have already appeared in them. At the same time, it is worth focusing on the possibilities described here during defence planning, i.e. it may be worthwhile to look outside the protected networks, to the dark web, in order to break or track attacks. Of course, in certain cases, defenders will need to seek the help of the appropriate authorised organisations (e.g. the police), which may again be an advantage for attackers.

Challenges and trends in the use of AI in cyberattacks

The above series of articles has shown how attackers can use AI in the various phases of a CKC and how it can increase the effectiveness of an attack. It can therefore be concluded that the use of AI in cyberattacks heralds new trends and, at the same time, significant challenges. However, these new challenges do not only affect defence professionals, but also pose significant problems for attackers, requiring serious consideration on their part. It can be argued that the use of AI will become indispensable on the offensive side, given the need to increase the effectiveness of attacks and the spread of evolving AI-based defence tools. In addition, competition is also expected to increase among attackers, as those who use AI-supported solutions will be more successful and able to obtain more information and thus more money.

At the same time, in addition to the challenges, thanks to the rapid development of AI, significant new opportunities will emerge in the near future that could further enhance the effectiveness of attackers. At the end of this series of articles, it is worth reviewing these opportunities alongside the challenges in order to be even better prepared for defence.

Challenges in applying AI in attacks

As explained above, although AI offers enormous opportunities for attackers, they may encounter a number of obstacles when carrying out sophisticated AI-based attacks. These challenges typically include the following:

- *Computing resources required for AI:* Running and training advanced AI models, especially deep learning and reinforcement learning systems, requires significant computing resources (primarily specialised hardware with multiple GPU cards) and energy consumption. This can hinder and significantly limit the capabilities of average or less well-funded attack groups. At the same time, *AI-as-a-Service* (AIaaS) platforms and easy and relatively inexpensive access to cloud-based computing resources can significantly reduce this barrier, making these advanced attack capabilities available to the aforementioned attack groups with less capital.[24]

[24] Timilehin 2023; Mailchimp s. a.

- *Data quality and quantity:* The effectiveness of AI models depends largely on the quality and quantity of the data used for training. The accuracy, completeness, consistency, and timeliness of the latter are essential for AI to work effectively. In other words, data quality has a critical impact on the performance of machine learning models. AI models trained and operating on poor-quality, incomplete, or distorted data may be insufficiently effective or produce poor-quality results that are easily detectable by defenders. This can make it difficult for attackers to collect a sufficient amount of relevant data that is appropriate for their purposes to train targeted AI models.[25]
- *Attribution and traceability:* Although AI-driven attacks make it significantly more difficult for defenders to understand and trace attacks due to dynamically changing infrastructure and automated decision-making, this can also cause difficulties for attackers. This is because it can be challenging for them to fully control and track the behaviour of their own AI systems used for attacks. Losing control can lead to unwanted side effects during the attack. In such cases, attackers' AI tools may leave digital footprints (e.g. specific characteristics of the AI model used, patterns derived from training data, etc.) that defenders may be able to detect and even trace back.[26]
- *The AI arms race between attackers and defenders:* The increasing use and sophistication of AI in attacks is triggering an arms race between attackers and defenders. As attackers use AI in their attacks, defenders must rely on advanced AI-based detection and response solutions to develop and maintain effective defences. This results in a continuous cycle of innovation, where capabilities on both sides must evolve rapidly and systems on both sides must constantly adapt to each other's *learned* strategies. This can also place a heavy burden on attackers.[27]

Future trends in the use of AI in attacks

Despite the challenges described above, attackers will certainly continue to develop and use their AI-based capabilities in order to carry out their cyberattacks more effectively. The development of AI-based cyberattacks is expected to lead to the following worrying trends:

- *Fully autonomous cyberattack agents:* In the future, we can expect the emergence of fully autonomous AI-based attack systems that will be capable of independently planning a cyberattack, gather the necessary information, execute complex cyberattacks, and maintain a continuous presence, all without human intervention, based solely on an initial target or mission assigned to them. These autonomous cyberattack agents will have self-learning capabilities and will be able to adapt to defence systems.[28]

[25] Dilmegani 2026.
[26] Ragho–Chaudhari 2025; Prasad et al. 2025.
[27] Gil–Williams 2025; Erdész 2023.
[28] Moraes 2025; Schippers 2025.

- *The increasing use of AI-as-a-Service (AIaaS) for malicious purposes:* The availability of tools and services that can be used for AI-based cyberattacks will increase on the black market, enabling even less skilled or less technically knowledgeable attackers to launch highly sophisticated and effective AI-based attacks. This trend means that advanced attack techniques are no longer the preserve of nation states or large cybercriminal groups but are becoming more widely available. This phenomenon makes advanced attack capabilities possible and accessible to a wide range of cybercriminals, as it significantly lowers the threshold for entry into advanced cybercrime.[29]
- *Further refinement of social engineering attacks with the help of AI:* The use of generative AI, particularly deepfake audio and video technologies, already makes it possible to create extremely convincing and realistic content that can be used for social engineering attacks, and then to carry out effective attacks with the help of this content. On the one hand, this content will become even more authentic in the future, making it even more difficult to recognise that it is not real but produced with the help of AI, allowing attackers to imitate senior executives or trusted sources even more convincingly. On the other hand, in the future, it will be easier to produce such content with even less knowledge, even with the help of publicly available generative AI models, which will also lower the threshold for entry into advanced cybercrime.[30]
- *Adversary AI, or AI against AI:* Just as defence systems are increasingly using and developing AI-based defence capabilities to identify threats and manage incidents, responding to them as quickly as possible, the same is happening on the offensive side, i.e. attackers are also increasingly developing AI-based attack capabilities and will use AI-based techniques more widely when carrying out attacks. Thanks to these continuous developments on the attacker side, which include the creation and refinement of machine learning models against defence solutions, the adversary AI used in attacks is becoming increasingly sophisticated, for example, in the areas of evading defence AI systems and deliberately misleading defence algorithms.[31]
- *Cyber warfare with AI-based tools:* State-sponsored cyberattack groups are expected to devote significant human and financial resources in the near future to the development of AI-based cyber weapons and systems that will enable them to carry out large-scale, targeted, and covert attacks with much greater efficiency than is currently possible, even against targets with advanced defence systems including critical infrastructure. The rapid development and spread of AI attack capabilities will accelerate the pace of the changes outlined above, exacerbate threats, and lower the barrier to entry in this segment, which will jeopardise cybersecurity more than ever before. This development also means that future conflicts will shift even more to cyberspace, where AI-based offensive (and, on the other side, defensive) tools can provide a strategic advantage.[32]

[29] European Union Agency for Law Enforcement Cooperation 2025; Matejic–Wilson 2024.
[30] IProov 2023; Babaei et al. 2025.
[31] Gil–Williams 2025; CrowdStrike 2025.
[32] Anthropic 2025; Crichton et al. 2024.

Conclusion

This series of articles examined the malicious uses of artificial intelligence from the perspective of cyberattacks, building its content around the Cyber Kill Chain model. Based on the above, it can be concluded that artificial intelligence is not just the next step in technological development in cybersecurity, but a revolutionary event that brings fundamental changes in both its dynamics and quality. Its significance could perhaps be compared to the developmental leap of the first industrial revolution. The rapidly developing capabilities of AI in the areas of automation, adaptation, and complex pattern recognition enable attackers to carry out cyberattacks with unprecedented sophistication, using AI in every phase of the Cyber Kill Chain, i.e. the attack. From accelerating reconnaissance to generating advanced malware, through covert C2 communication, to effectively reaching attack objectives, AI can significantly increase the effectiveness, speed and covert nature of cyberattacks. On the one hand, by making advanced attack capabilities available to a wide range of people at low cost and with less and less knowledge required to use them, AI lowers the threshold for entry into cybercrime. On the other hand, it creates the possibility of fully autonomous cyberattack agents, fundamentally changing the cyber threat landscape.

Defenders must also be prepared to detect and counter AI-assisted attacks, and they must also use AI-based systems to develop effective defences. This phenomenon is resulting in a constantly escalating, spiral-like *AI arms race* between attackers and defenders, where AI-assisted capabilities are expected to develop exponentially on both sides, as is AI itself.

Those who work in incident management, i.e. cyber defence, are already encountering the fight against AI in the current landscape, but in the cybersecurity environment of the near future, this will increasingly become part of everyday life. In this new environment, defence systems will not only have to detect and respond to threats but also outsmart the AI models used by attackers.

In order to develop effective cyber defence, identify necessary areas for further development and recognise points for moving forward, it is advisable to conduct further studies and research. In this context, it is necessary to examine, on the one hand, what defence tools are currently used by cybersecurity experts in the various phases of CKC, which of these are already available today as AI-supported solutions, and what capabilities they currently have. On the other hand, it is also necessary to assess where and how artificial intelligence can help defenders in each phase of the CKC in the future. Following this, it is advisable to review AI-based defence trends and then compare the capabilities of the attacker and defender, demonstrating where and how effective response capabilities can be developed using artificial intelligence in response to attacker AI capabilities.

References

Abbadi, Driss – Lachkar, Abdelkader (2024): Cyber Threats in the Age of Artificial Intelligence: Exploiting Advanced Technologies and Strengthening Cybersecurity. *International Journal of Science and Research Archive,* 13(1), 2576–2588. Online: https://doi.org/10.30574/ijsra.2024.13.1.1961

Aboze, John B. (2025): A Comprehensive Guide to Data Exfiltration. *Lakera,* 21 May 2025. Online: www.lakera.ai/blog/data-exfiltration

Al-Azzawi, Mays – Doan, Dung – Sipola, Tuomo – Hautamäki, Jari – Kokkonen, Teri (2025): Red Teaming with Artificial Intelligence-Driven Cyberattacks: A Scoping Review. *ArXiv*. Online: https://doi.org/10.48550/arXiv.2503.19626

Anderson, Hyrum S. – Woodbridge, Jonathan – Filar, Bobby (2016): DeepDGA: Adversarially-Tuned Domain Generation and Detection. *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security,* 13–21. Online: https://doi.org/10.1145/2996758.2996767

Anthropic (2025): *Disrupting the First Reported AI-orchestrated Cyber Espionage Campaign.* Online: www.anthropic.com/news/disrupting-AI-espionage

Babaei, Reza – Cheng, Samuel – Duan, Rui – Zhao, Shangqing (2025): Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis. *Journal of Sensor and Actuator Networks,* 14(1). Online: https://doi.org/10.3390/jsan14010017

Boutin, Jean-Ian (2025): *ESET APT Activity Report Q2 2025 – Q3 2025.* Online: www.welivesecurity.com/en/eset-research/eset-apt-activity-report-q2-2025-q3-2025/

Cherepanov, Anton – Strýček, Peter (2025): First *Known AI-powered Ransomware Uncovered by ESET Research.* Online: www.welivesecurity.com/en/ransomware/first-known-ai-powered-ransomware-uncovered-eset-research/

Crichton, Kyle et al. (2024): *Securing Critical Infrastructure in the Age of AI.* Center for Security and Emerging Technology. Online: https://doi.org/10.51593/20240032

CrowdStrike (2025): *CrowdStrike 2025 Global Threat Report.* Online: www.crowdstrike.com/en-us/resources/infographics/global-threat-report-2025/

Darktrace (2024): *Navigating a New Threat Landscape: Breaking Down the AI Kill Chain.* Online: www.darktrace.com/resources/navigating-a-new-threat-landscape

De Pasquale, Giulio – Grishchenko, Ilya – Iesari, Riccardo – Pizzaro, Gabriel – Cavallaro, Lorenzo – Kruegel, Christopher – Vigna, Giovanni (2024): *ChainReactor: Automated Privilege Escalation Chain Discovery via AI Planning.* Online: www.usenix.org/system/files/usenixsecurity24-de-pasquale.pdf

Dilmegani, Cem (2026): AI Data Quality in 2026: Challenges & Best Practices. *AIMultiple,* 27 March 2026. Online: https://research.aimultiple.com/data-quality-ai/

Erdész, Viktor (2023): *A mesterséges intelligencia alkalmazása a katonai nemzetbiztonsági hírszerzésben* [The Use of Artificial Intelligence in Military and National Security Intelligence]. Budapest: Katonai Nemzetbiztonsági Szolgálat.

European Union Agency for Law Enforcement Cooperation (2025): *IOCTA. Internet Organised Crime Threat Assessment 2025*. Luxembourg: Publications Office of the European Union. Online: https://doi.org/10.2813/4926508

Frisoni, Daniele (2020): *Potential Impact of Artificial Intelligence to C2 Systems.* Joint Air Power Competence Centre. Online: www.japcc.org/essays/potential-impact-of-artificial-intelligence-to-c2-systems/

Gil, Luis – Williams, Beth (2025): AI vs. AI: The Race Between Adversarial and Defensive Intelligence. *CrowdStrike,* 4 August 2025. Online: www.crowdstrike.com/en-us/blog/ai-vs-ai-cybersecurity-arms-race/

Holden, Alex (2023): Contending with Artificially Intelligent Ransomware. *ISACA,* 1 September 2023. Online: www.isaca.org/resources/news-and-trends/isaca-now-blog/2023/contending-with-artificially-intelligent-ransomware

Humphries, Russ (2026): The Dark Side: How Threat Actors Are Using AI. *Connect Wise,* 3 March 2026. Online: www.connectwise.com/blog/the-dark-side-how-threat-actors-are-using-ai

IProov (2023): Understanding the Different Types of Generative AI Deepfake Attacks. *iProov,* 30 November 2023. Online: www.iproov.com/blog/generative-ai-attack-types-explained

ITszótár.hu (2025): Metamorf és polimorf kártevők: Ezen kártékony szoftverek működésének magyarázata [Metamorphic and Polymorphic Malware: An Explanation of How this Malicious Software Works]. *ITszótár.hu,* 15 May 2025. Online: https://itszotar.hu/metamorf-es-polimorf-kartevok-ezen-kartekony-szoftverek-mukodesenek-magyarazata/

Jain, Anu – Chhabra, Gurpal S. (2014). Anti-Forensics Techniques: An Analytical Review. *2014 Seventh International Conference on Contemporary Computing (IC3),* 412–418. Online: https://doi.org/10.1109/IC3.2014.6897209

Kewat, Rajnish (2025): What Are AI-Generated Rootkits and How Do They Threaten Enterprise Systems? *Cyber Security Training Institute,* 8 August 2025. Online: www.cybersecurityinstitute.in/blog/what-are-ai-generated-rootkits-and-how-do-they-threaten-enterprise-systems

Lee, Junho – Kwon, Jihoon – Seo, HyunA – Lee, Myeongyeol – Seo, Hyungyu – Jung, Jinho – Koo, Hyungjoon (2025): *BootKitty: A Stealthy Bootkit-Rootkit Against Modern Operating Systems.* Online: www.usenix.org/conference/woot25/presentation/lee

Lockheed Martin (s. a.): *Cyber Kill Chain.* Online: www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

Mailchimp (s. a.): *AI as a Service: The Latest Business Model.* Online: https://mailchimp.com/resources/ai-as-a-service/

Matejic, Nicole – Wilson, Chris (2024): Crimes of Influence: Generative Artificial Intelligence-led Crime as a Service. *The Commonwealth Cyber Journal,* 2, 75–95.

Microsoft Security (2025): *What Is the Cyber Kill Chain?* Online: www.microsoft.com/en-us/security/business/security-101/what-is-cyber-kill-chain

Moraes, Marco Túlio (2025): AI Agents: The New Frontier of Cybercrime Business Must Confront. *World Economic Forum,* 18 June 2025. Online: www.weforum.org/stories/2025/06/ai-agent-cybercrime-business/

Pappachan, Princy – Adi, Novi S. – Firmansyah, Gerry – Rahaman, Mosiur (2024): Deep Learning-Based Forensics and Anti-Forensics. In Abd El-Latif, Ahmed A. – Tawalbeh, Lo'ai – Mohanty, Manoranjan – Gupta, Brij B. – Psannis, Konstantinos E. (eds.): *Digital Forensics and Cyber Crime Investigation.* Boca Raton: CRC Press.

Powell, Evan (2025): Invisible C2 – Thanks to AI-powered Techniques. *Deep Tempo,* 14 March 2025. Online: https://medium.com/deeptempo/invisible-c2-thanks-to-ai-powered-techniques-462ef2624c8a

Prasad, Nilantha – Diro, Abebe – Warren, Matthew – Fernando, Mahesh (2025): A Survey of Cyber Threat Attribution: Challenges, Techniques, and Future Directions. *Computers & Security,* 157. Online: https://doi.org/10.1016/j.cose.2025.104606

Ragho, Soni R. – Chaudhari, Narendra (2025): Artificial Intelligence in Digital Forensics: A Review of Cyber-Attack Detection Models and Frameworks. *Journal of Information Systems Engineering and Management,* 10(57s). Online: https://jisem-journal.com/index.php/journal/article/view/12402

Raz, Md – Udeshi, Meet – Charan, Sai P. V. – Krishnamurthy, Prashanth – Khorrami, Farshad – Karri, Ramesh (2025): Ransomware 3.0: Self-Composing and LLM-Orchestrated. *Arxiv*. Online: https://doi.org/10.48550/arXiv.2508.20444

Sangfor Technologies (2025): AI DDoS: How Artificial Intelligence Is Changing the Face of Cyber Attacks. *Sangfor,* 9 September 2025. Online: www.sangfor.com/blog/cybersecurity/ai-ddos-attacks

Schippers, Raymond (2025): AI 2030: The Coming Era of Autonomous Cyber Crime. *Check Point,* 24 October 2025. Online: https://blog.checkpoint.com/executive-insights/ai-2030-the-coming-era-of-autonomous-cyber-crime/

Schröer, Saskia L. – Pajola, Luca – Castagnaro, Alberto – Apruzzese, Giovanni – Conti, Mauro (2025): Exploiting AI for Attacks: On the Interplay between Adversarial AI and Offensive AI. *ArXiv*. Online: https://arxiv.org/html/2506.12519

Tejedor, Jesús (2025): A Dangerous Alliance: The New Dark Web + AI Marketplace. *Telefónica,* 5 June 2025. Online: https://telefonicatech.com/en/blog/a-dangerous-alliance-how-ai-is-reshaping-the-dark-web-economy

Timilehin, Oladoja (2023): *Defending the Digital Horizon: Artificial Intelligence in Cybersecurity Warfare*.

Toulas, Bill (2025): LameHug Malware Uses AI LLM to Craft Windows Data-theft Commands in Real-time. *Bleeping Computer,* 17 July 2025. Online: www.bleepingcomputer.com/news/security/lamehug-malware-uses-ai-llm-to-craft-windows-data-theft-commands-in-real-time/

Trinckes, Jay (s. a.): AI Data Breach: Understanding their Impact and Protecting Your Data. *Thoropass*. Online: www.thoropass.com/blog/ai-data-breach

Utter, Jamison (2024): The Machine War Has Begun: Cybercriminals Leveraging AI in DDoS Attacks. *A10 Networks,* 24 September 2024. Online: www.a10networks.com/blog/the-machine-war-has-begun-cybercriminals-leveraging-ai-in-ddos-attacks/

Wang, Yulong – Sun, Tong – Li, Shenghong – Yuan, Xin – Ni, Wei – Hossain, Ekram – Poor, Vincent H. (2023): Adversarial Attacks and Defenses in Machine Learning-Empowered Communication Systems and Networks: A Contemporary Survey. *IEEE Communications Surveys & Tutorials*, 25(4), 2245–2298. Online: https://doi.org/10.1109/COMST.2023.3319492

Wang, Zhi – Liu, Chaoge – Cui, Xiang – Yin, Jie – Liu, Jiaxi – Wu, DI – Liu, Qixu (2022): DeepC2: AI-Powered Covert Command and Control on OSNs. In Alcaraz, Cristina – Chen, Liqun – Li, Shujun – Samarati, Pierangela (eds.): *Information and Communications Security.* Cham: Springer, 394–414. Online: https://doi.org/10.1007/978-3-031-15777-6_22