Przegląd Bezpieczeństwa Wewnętrznego
2026, nr 34, s. 133–145
CC BY-NC-SA 4.0
https://doi.org/10.4467/20801335PBW.26.006.23368

ARTYKUŁ

Rozwój zagrożeń cybernetycznych związany z wykorzystaniem AI

The development of cyber threats related to the use of AI

JAKUB GAJECKI

Autor niezależny

https://orcid.org/0009-0007-3488-0236

Abstrakt

Dynamiczny rozwój sztucznej inteligencji (artificial intelligence, AI) powoduje, że rośnie także jej rola w cyberprzestrzeni, zarówno w kontekście zagrożeń, jak i obrony przed nimi. Wspiera ona automatyzację wykrywania anomalii, analizę danych i reakcję na incydenty, co zwiększa efektywność ochrony. Z kolei cyberprzestępcy wykorzystują rozwiązania oparte na AI do tworzenia inteligentnych narzędzi ataków, takich jak zaawansowane phishingi, deepfaki i trudne do wykrycia malware'y. Autor analizuje rolę AI w generowaniu zagrożeń w cyberprzestrzeni i w tym kontekście ocenia strategie obronne. Wskazuje potrzebę międzynarodowej współpracy i regulacji prawnych w zakresie wykorzystania AI w cyberbezpieczeństwie.

Słowa kluczowe

sztuczna inteligencja, cyberbezpieczeństwo, AI as a service, cyberprzestępczość, kryptografia kwantowa

Abstract

In light of the rapid development of artificial intelligence (AI), its role in cybersecurity is crucial for both defense and emerging threats. AI supports the automation of anomaly detection, data analysis, and incident response, which enhances protection efficiency. However, AI as a service enables cybercriminals to create sophisticated attack tools, such as advanced phishing schemes, deepfakes, and hard-to-detect malware. This work analyzes AI's role in cyber threats and evaluates defensive strategies, highlighting the need for international cooperation and legal regulations. Conclusions point to the necessity of research on AI's resilience against attacks.

Keywords

artificial intelligence, cybersecurity, AI as a service, cybercrime, quantum cryptography

Wprowadzenie

Rozwój technologii informatycznych i sztucznej inteligencji (artificial intelligence, AI[1]) zmienił podejście do cyberbezpieczeństwa oraz zainicjował nową erę cyberzagrożeń. W przeszłości ataki w cyberprzestrzeni, takie jak wirusy komputerowe czy phishing, czyli podszywanie się pod instytucje lub osoby w celu wyłudzenia informacji[2], były stosunkowo proste i ograniczone do działań pojedynczych hakerów lub małych grup. Na przełomie XX i XXI w. największym wyzwaniem było zapobieganie rozprzestrzenianiu się złośliwego oprogramowania (malicious software, malware), w tym typu ransomware, i szybkie reagowanie na jego atak[3]. Zwykle te ataki były ukierunkowane na pojedyncze osoby lub mniejsze organizacje, a ich zakres i skutki – ograniczone możliwościami technologicznymi[4]. Postęp w zakresie AI i uczenia maszynowego przekształcił cyberprzestrzeń i spowodował, że zagrożenia są bardziej złożone, dynamiczne i trudniejsze do wykrycia. Sztuczna inteligencja

[1] Wszystkie słowa i zwroty obcojęzyczne użyte w artykule pochodzą z języka angielskiego (przyp. red.).

[2] J. Jancelewicz, *Phishing i pokrewne ataki socjotechniczne jako zagrożenie dla organizacji pozarządowych*, „Trzeci Sektor" 2022, t. 3–4, nr 59–60, s. 80–81. https://doi.org/10.26368/17332265-59/60-3/4-2022-5.

[3] *The Evolution of Cybersecurity*, Codecademy, https://www.codecademy.com/article/evolution-of-cybersecurity [dostęp: 7 X 2024].

[4] B. Dash i in., *Threats and Opportunities with AI-based Cyber Security Intrusion Detection: A Review,* „International Journal of Software Engineering & Applications" 2022, t. 13, nr 5, s. 14. https://doi.org/10.5121/ijsea.2022.13502.

wprowadziła nową jakość zarówno pod względem przeprowadzania ataków w cyberprzestrzeni, jak i obrony przed nimi.

Sztuczna inteligencja pozwala m.in.:

- analizować infrastrukturę zabezpieczeń,
- automatyzować wyszukiwanie jej słabych punktów,
- rozwijać inteligentne narzędzia, które potrafią adaptować się do działań obronnych atakowanych systemów w czasie rzeczywistym.

Wykorzystanie AI czyni ataki wyjątkowo trudnymi do neutralizacji.

W artykule przeanalizowano wpływ rozwoju AI, szczególnie jej klasycznych i generatywnych (GenAI) typów, na rozwój zagrożeń w cyberprzestrzeni, w wymiarze zarówno ofensywnym (działania cyberprzestępcze), jak i defensywnym (systemy cyberobrony).

Problem badawczy sformułowano następująco: w jaki sposób wykorzystanie klasycznych i generatywnych AI zmienia charakter, skalę oraz automatyzację zagrożeń w cyberprzestrzeni oraz jakie to ma konsekwencje dla systemów cyberbezpieczeństwa?

Cele szczegółowe badania to:

- identyfikacja klasycznych zastosowań AI w cyberbezpieczeństwie,
- analiza wykorzystania generatywnej AI przez grupy cyberprzestępcze,
- ocena aktualnych systemów obronnych wobec automatyzacji ataków,
- wskazanie kierunków dalszych badań i działań regulacyjnych.

W artykule zastosowano metody badawcze takie jak analiza i synteza oraz indukcja i dedukcja. Autor korzystał z różnych materiałów źródłowych: publikacji zwartych, monografii i artykułów naukowych, raportów, publikacji specjalistycznych oraz opisów studiów przypadków.

Rozwój botnetów i złośliwego oprogramowania

Botnety, czyli grupy zainfekowanych komputerów kontrolowanych bez wiedzy ich właścicieli, zaczęły się rozwijać na początku XXI w. Jednym z pierwszych i najbardziej znanych botnetów był Agobot, który wykorzystywał zainfekowane urządzenia do rozsyłania spamu i przeprowadzania ataków DDoS (distributed denial of service)[5], których celem jest przeciążenie serwerów zaatakowanych podmiotów i uniemożliwienie dostępu do usług. Kolejne botnety takie jak Storm i Conficker

[5] A. Kurniawan, A. Fitriansyah, *A Literature Review of Historical and Detection Analysis of Botnets Forensics*, „International Journal of Computer and Communication Engineering" 2018, t. 7, nr 4, s. 130–131. https://doi.org/10.17706/ijcce.2018.7.4.128-135.

pokazały swoją siłę i skalę działania, przejmując kontrolę nad milionami urządzeń i stając się realnym zagrożeniem dla globalnych systemów komputerowych[6].

Ewolucja złośliwego oprogramowania dotyczy jego wielu form: wirusów, robaków, oprogramowania szpiegującego (spyware), uciążliwych reklam (advertising-supported software, adware), rootkitów, które zapewniają dostęp na poziomie administratora i ransomware'u. Przykładem wczesnego rozwoju malware'ów jest wirus ILOVEYOU, który nie był jeszcze botnetem, ale ze względu na skalę oddziaływania stał się inspiracją do poszukiwania bardziej zaawansowanych form malware'ów. W miarę jak technologia ewoluowała również złośliwe oprogramowanie stawało się coraz bardziej wyspecjalizowane. Zeus i SpyEye to przykłady malware'ów skoncentrowanych na kradzieży danych z bankowości internetowej[7]. Stuxnet natomiast był pierwszym programem zaprojektowanym do fizycznego uszkodzenia urządzeń przemysłowych infrastruktury krytycznej[8]. Botnety stały się głównym narzędziem w atakach DDoS. Przykładami są botnety Mirai i Satori, które przekształciły urządzenia internetu rzeczy (internet of things) w narzędzia ataków na masową skalę. W 2016 r. Mirai zaatakował serwery należące do dostawców DNS (domain name system) i zablokował na całym świecie dostęp do popularnych witryn internetowych[9].

Obecnie złośliwe oprogramowanie i botnety korzystają z rozwiązań opartych na AI, co ułatwia im omijanie systemów detekcji i skuteczniejsze ukrywanie swojej obecności. Sztuczna inteligencja umożliwia botnetom analizę i adaptację w czasie rzeczywistym. Zwiększa to skuteczność ataków i zmniejsza prawdopodobieństwo ich wykrycia. Zautomatyzowane botnety mogą samodzielnie identyfikować nowe cele, a nawet korzystać z technik uczenia maszynowego, aby rozpoznać wzorce zachowań ofiar i odpowiednio dostosowywać swoje działania.

---

[6] J. Yimu, L. Shangdong, *Threats from Botnets*, w: *Computer Security Threats*, C. Thomas, P. Fraga-Lamas, T.M. Fernández-Caramés (red.), September 2020, https://www.intechopen.com/chapters/69332 [dostęp: 18 X 2024].

[7] N. Etaher, G.R.S. Weir, *Understanding the Threat of Banking Malware*, w: *Proceedings of Cyberforensics 2014*, https://strathprints.strath.ac.uk/48856/1/8_etaher_weir.pdf, s. 77–79 [dostęp: 18 X 2025].

[8] M. Hagerott, *Stuxnet and the vital role of critical infrastructure operators and engineers*, „International Journal of Critical Infrastructure Protection" 2014, t. 7, nr 4, s. 244–246. https://doi.org/10.1016/j.ijcip.2014.09.001.

[9] H. Griffioen, Ch. Doerr, *Examining Mirai's Battle over the Internet of Things*, https://www.cyber-threat-intelligence.com/publications/CCS2020-iotbattle.pdf, s. 744 [dostęp: 18 X 2025].

Sztuczna inteligencja jako narzędzie cyberprzestępców

Sztuczna inteligencja jest definiowana jako dziedzina informatyki zajmująca się projektowaniem systemów zdolnych do wykonywania zadań wymagających wcześniej ludzkich umiejętności takich jak uczenie się, rozumowanie, planowanie czy podejmowanie decyzji. Klasyczne podejście do AI obejmuje m.in.: systemy regułowe, uczenie maszynowe, sieci neuronowe oraz algorytmy wnioskowania probabilistycznego[10].

Cyberprzestępcy są w stanie automatyzować i usprawniać ataki phishingowe i malware, co sprawia, że mają one większy zasięg i są trudniejsze do wykrycia. Dzięki AI hakerzy mogą w krótkim czasie analizować dużo danych, np. zachowania i profile użytkowników, by tworzyć bardziej przekonujące wiadomości, dostosowane do różnych grup docelowych. Personalizacja zwiększa prawdopodobieństwo, że odbiorca zdecyduje się kliknąć w złośliwy link lub pobrać załącznik. Sztuczna inteligencja może również wspomagać tworzenie i dystrybucję złośliwego oprogramowania. Analizuje mechanizmy ochronne systemów, dostosowując działanie malware'u tak, aby pozostawał niewykrywalny. Przykładem może być złośliwe oprogramowanie, które wykorzystuje uczenie maszynowe do przeprowadzania tzw. ataków polimorficznych. Polegają one na tym, że malware zmienia swoje cechy przy każdej infekcji, co znacznie utrudnia jego identyfikację przez tradycyjne programy antywirusowe[11]. Cyberprzestępcy używają AI również do tworzenia deepfake'ów, czyli fałszywych obrazów, nagrań wideo lub audio, które wykorzystują w różnych scenariuszach przestępczych, takich jak oszustwa finansowe, manipulacja opinią publiczną, a nawet szantaż[12].

Rozwój GenAI, w tym dużych modeli językowych (large language models, LLM) oraz generatywnych modeli opartych na architekturze sieci neuronowych takich jak generative adversarial networks, znacznie zmienił modus operandi grup cyberprzestępczych. Generatywna AI umożliwia automatyczne tworzenie treści phishingowych o wysokim stopniu personalizacji, generowanie kodu malware oraz skalowanie ataków socjotechnicznych. Badania wskazują, że wykorzystanie LLM pozwala na obniżenie bariery wejścia do cyberprzestępczości, umożliwiając prowadzenie zaawansowanych ataków osobom bez specjalistycznej wiedzy[13]. Przykładem

[10] S. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, bmw 2021, s. 1–2.

[11] R. Chauhan i in., *Polymorphic Adversarial Cyberattacks Using WGAN*, „Journal of Cybersecurity and Privacy" 2021, nr 1, s. 788–789. https://doi.org/10.3390/jcp1040037.

[12] O. Wasiuta, S. Wasiuta, *Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, t. 9, nr 3, s. 20–24. https://doi.org/10.24917/26578549.9.3.2.

[13] Y. Yigit i in., *Review of Generative AI Methods in Cybersecurity*, preprint, arXiv, 13 III 2024 r. https://doi.org/10.48550/arXiv.2403.08701.

może być analiza historii wyszukiwań i odwiedzanych stron w celu dostosowania komunikatów phishingowych, które dla ofiary mogą wyglądać na autentyczne.

Uczenie maszynowe jest wykorzystywane do tworzenia bardziej wyrafinowanego i adaptacyjnego złośliwego oprogramowania. Algorytmy te pozwalają malware'owi na dostosowywanie swojego działania w zależności od wykrytych zabezpieczeń systemu ofiary i przeprowadzenie ataku w najbardziej odpowiednim momencie, np. kiedy użytkownik loguje się do wrażliwych systemów takich jak konta bankowe. Algorytmy uczenia maszynowego umożliwiają generowanie i testowanie w krótkim czasie nowych wariantów złośliwego kodu.

Inną formą cyberprzestępstwa opartą na uczeniu maszynowym są wspomniane ataki polimorficzne. Techniki takie jak polimorfizm i metamorfizm pozwalają na generowanie różnych wariantów tego samego malware'u, dzięki czemu oprogramowanie zabezpieczające, które bazuje na wykrywaniu wzorców, nie jest w stanie rozpoznać zmienionego kodu. Takie techniki są jednymi z najtrudniejszych do wykrycia. Uczenie maszynowe daje także możliwość szybkiego przetwarzania danych i próbowania różnych kombinacji przy atakach brute force, mających na celu odgadnięcie haseł lub kodów zabezpieczeń. Algorytmy te są w stanie przewidywać, jakie hasła mogą być najbardziej prawdopodobne, co znacznie skraca czas potrzebny do złamania zabezpieczeń. Przykładowo, w połączeniu z analizą behawioralną algorytmy mogą generować propozycje haseł zgodne z charakterystycznymi schematami stosowanymi przez użytkownika, takimi jak imiona, daty urodzin lub inne osobiste szczegóły.

Cyberprzestępcy wykorzystują algorytmy uczenia maszynowego do analizowania struktury i konfiguracji systemów bezpieczeństwa. Tego rodzaju oprogramowanie jest w stanie zidentyfikować i przeanalizować specyficzne cechy mechanizmów zabezpieczających takich jak firewalle, systemy detekcji włamań (intrusion detection system) oraz programy antywirusowe. Dzięki tej wiedzy atakujący mogą dostosowywać swoje metody w czasie rzeczywistym, przełamywać kolejne warstwy ochrony i unikać wykrycia. Sztuczna inteligencja może nie tylko generować różne warianty tego samego malware'u (polimorfizm, metamorfizm), lecz także uczyć malware, by rozpoznawał różne systemy zabezpieczeń i do nich dostosowywał swoje działanie, co będzie minimalizować ryzyko detekcji. Zaawansowane ataki z użyciem AI polegają na predykcji zachowań użytkowników opartej na analizie behawioralnej. Analizując duże zbiory danych dotyczących zachowań użytkowników, atakujący mogą przewidzieć momenty, kiedy system będzie najmniej odporny na atak, np. podczas próby logowania w czasie, gdy system zabezpieczeń rejestruje wzmożony ruch i łatwiej ignoruje pewne nieprawidłowości.

Współczesne przypadki cyberataków

Atak na platformę GitHub w 2018 r. był jednym z najpotężniejszych ataków DDoS w historii – uzyskano w nim przepływ danych rzędu 1,35 Tb/s. Cyberprzestępcy wykorzystali botnety sterowane AI do masowego wysyłania żądań, co przeciążyło serwery GitHub. Sztuczna inteligencja pomogła dynamicznie dostosowywać atak i omijać zabezpieczenia w czasie rzeczywistym. Botnety wspomagane przez AI stają się coraz powszechniejsze, co umożliwia przeprowadzanie ataków na duże platformy[14].

W 2020 r. firma Cognizant, globalny dostawca usług IT, padła ofiarą ransomware'u Maze. Jest to przykład oprogramowania wykorzystującego AI i zaawansowane techniki infiltracji sieci. Oprogramowanie to stosuje analizę systemu, aby zidentyfikować najbardziej wrażliwe dane i uniemożliwić do nich dostęp, a także rozsyła je dalej do centrów dowodzenia przestępców. W wyniku ataku firma Cognizant poniosła ogromne straty finansowe i wydała miliony dolarów na naprawę infrastruktury oraz wsparcie dla klientów i dostawców[15]. Przedsiębiorstwo podjęło działania naprawcze obejmujące izolację zainfekowanych systemów, wzmocnienie procedur reagowania na incydenty oraz rozbudowę mechanizmów monitorowania sieci. Przypadek ten pokazuje jednak, że systemy wykrywania zagrożeń oparte głównie na sygnaturach nie zawsze są w stanie odpowiednio wcześnie wykryć zaawansowane kampanie ransomware[16].

W tym samym roku Twitter został zaatakowany przez cyberprzestępców, którzy przejęli konta znanych osób i firm, w tym Billa Gatesa, Elona Muska i Apple'a. Hakerzy wykorzystali technologię przetwarzania języka naturalnego (natural language processing, NLP) do generowania wiadomości wyglądających na osobiste i spersonalizowane, w których zachęcali do przesyłania pieniędzy na podany adres kryptowalutowy[17]. Atak był skuteczny dzięki m.in. wykorzystaniu AI do analizowania sposobu komunikacji ofiar i adaptacji wiadomości. Po wykryciu incydentu platforma zablokowała możliwość publikowania wpisów przez zweryfikowane konta oraz rozpoczęła proces przywracania bezpieczeństwa przejętych profili. Wdrożono

---

[14] L.H. Newman, *GitHub Survived the Biggest DDoS Attack Ever Recorded*, Wired, 1 III 2018 r., https://www.wired.com/story/github-ddos-memcached/ [dostęp: 26 X 2024].

[15] F. Truță, *Cognizant Expects to Lose up to $70 Million from April Ransomware Attack*, Bitdefender, 11 V 2020 r., https://www.bitdefender.com/en-gb/blog/hotforsecurity/cognizant-expects-to-lose-up-to-70-million-from-april-ransomware-attack/ [dostęp: 24 X 2024].

[16] *Cognizant Security Incident Update*, Cognizant, 18 IV 2020 r., https://news.cognizant.com/2020-04-18-Cognizant-Security-Incident-Update?utm [dostęp: 9 III 2026].

[17] N. Statt, *Twitter's massive attack: What we know after Apple, Biden, Obama, Musk, and others tweeted a bitcoin scam*, The Verge, 17 VII 2020 r., https://www.theverge.com/2020/7/15/21326200/elon-musk-bill-gates-twitter-hack-bitcoin-scam-compromised [dostęp: 24 X 2025].

również dodatkowe środki kontroli dostępu do narzędzi administracyjnych oraz wzmocniono procedury uwierzytelniania pracowników.

W 2025 r. został wykryty pierwszy udokumentowany przypadek wykorzystania na dużą skalę autonomicznych agentów AI w operacjach cyberwywiadowczych. W połowie września zespół Threat Intelligence firmy Anthropic zidentyfikował i następnie przerwał kampanię cyberszpiegowską, w której narzędzie oparte na modelu językowym Claude Code zostało zmanipulowane przez aktora uznawanego za powiązanego z chińskimi strukturami państwowymi. Celem było przeprowadzenie złożonych operacji wywiadowczych[18].

W tej kampanii AI nie pełniło jedynie funkcji doradczej czy generatywnej, lecz działało jako autonomiczny agent wykonujący większość zadań operacyjnych. System rozkładał wieloetapowe instrukcje na mniejsze zadania, które następnie wykonywał samodzielnie z minimalnym nadzorem człowieka. Claude Code przeprowadził autonomicznie do 80–90% taktycznych operacji, włączając w to:

- rozpoznanie infrastruktury celu i analizę słabych punktów,
- generowanie i wykonanie kodów wykorzystujących luki,
- zbieranie poświadczeń i danych,
- ruch boczny (lateral movement),
- wydobywanie i klasyfikację danych[19].

Taka automatyzacja oznacza, że AI przeprowadzała bez ciągłej interwencji operatora działania, które wcześniej wymagałyby zaangażowania dużych zespołów ekspertów: od skanowania sieci, przez analizy podatności, po eksfiltrację danych. W tym przypadku rola człowieka ograniczała się głównie do inicjalizacji kampanii i podejmowania strategicznych decyzji w kluczowych momentach, np. zatwierdzania przejścia między fazami ataku[20]. Co więcej, mechanizm działania agentów AI opierał się na wykorzystaniu ich zdolności do autonomicznego podejmowania decyzji w sekwencji zadań oraz do adaptacji narracji i strategii względem kolejnych etapów ataku. Znacznie zwiększało to tempo i skalę operacji w porównaniu z tradycyjnym „ręcznym sterowaniem". Maszyna była w stanie wykonać tysiące operacji na sekundę, co dla zespołów cyberprzestępczych niekorzystających z automatyzacji jest niedostępne[21].

[18] *Disrupting the first reported AI-orchestrated cyber espionage campaign*, Anthropic, 13 XI 2025 r., https://www.anthropic.com/news/disrupting-AI-espionage [dostęp: 24 XI 2025].

[19] Tamże.

[20] Tamże.

[21] Tamże.

Sztuczna inteligencja w cybernetycznych systemach obronnych

Rozwój usług typu AI as a service niesie ze sobą korzyści dla biznesu, ale jednocześnie tworzy nowe wektory ataków dla cyberprzestępców. AI as a service oferuje dostęp do zaawansowanych algorytmów AI, które mogą zostać wykorzystane do automatyzacji działań takich jak analiza słabych punktów systemów lub tworzenie zaawansowanych botów phishingowych. Cyberprzestępcy mogą w przyszłości wykorzystywać AI as a service do rozwoju deepfake'ów, przeprowadzania socjotechnicznych ataków na większą skalę, tworzenia trudniejszych do wykrycia malware'ów, projektowania ransomware'ów, które będą automatycznie wybierać najskuteczniejsze metody ataku, zwiększając ich skuteczność.

Aby sprostać tym wyzwaniom, równolegle są rozwijane nowe technologie obronne oparte na AI. Przykładem są adaptacyjne systemy oparte na uczeniu maszynowym, które mogą dynamicznie reagować na zagrożenia i automatycznie dostosowywać swoje funkcje ochronne w zależności od napotkanych ataków. Dodatkowo technologie NLP są stosowane do analizy komunikacji cyberprzestępczej, co pomaga w przewidywaniu i wykrywaniu nowych zagrożeń. Przewiduje się, że w przyszłości systemy obronne oparte na AI będą zdolne do samodzielnej analizy złośliwego oprogramowania oraz do tworzenia dynamicznych, odpornych na zagrożenia środowisk wirtualnych, co zmniejszy ryzyko naruszeń bezpieczeństwa[22].

Warto podkreślić, że zarówno na poziomie krajowym, jak i międzynarodowym obecnie obowiązuje wiele regulacji z zakresu cyberbezpieczeństwa. W Europie istotną rolę odgrywają dyrektywa NIS 2[23] dotycząca bezpieczeństwa sieci i systemów informatycznych, a także rozporządzenie Artificial Intelligence Act[24] regulujące wykorzystanie systemów AI w Unii Europejskiej. Ważnym instrumentem jest również Konwencja Rady Europy o cyberprzestępczości[25], która stanowi podstawę współpracy międzynarodowej w zwalczaniu cyberprzestępczości.

---

[22] R. Keshava i in., *AI-Powered Algorithms for the Prevention and Detection of Computer Malware Infections*, w: 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC 2025), Coimbatore 2025. https://doi.org/10.1109/ICESC65114.2025.11212519.

[23] *Dyrektywa Parlamentu Europejskiego i Rady (UE) 2022/2555 z dnia 14 grudnia 2022 r. w sprawie środków na rzecz wysokiego wspólnego poziomu cyberbezpieczeństwa na terytorium Unii, zmieniająca rozporządzenie (UE) nr 910/2014 i dyrektywę (UE) 2018/1972 oraz uchylająca dyrektywę (UE) 2016/1148 (dyrektywa NIS 2).*

[24] *Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji).*

[25] *Konwencja Rady Europy o cyberprzestępczości, sporządzona w Budapeszcie dnia 23 listopada 2001 r.*

Należy jednak zauważyć, że większość przepisów powstawała w okresie, gdy technologie AI nie były wykorzystywane na szeroką skalę w działaniach cybernetycznych. W związku z tym obecne regulacje często nie odnoszą się wprost do specyfiki systemów opartych na AI, takich jak autonomiczne systemy detekcji zagrożeń, generatywne modele wykorzystywane do tworzenia phishingu czy zautomatyzowane narzędzia prowadzenia cyberataków.

Rozwój możliwości zastosowania AI w cyberbezpieczeństwie powoduje przede wszystkim konieczność doprecyzowania i rozszerzenia istniejących regulacji, a nie tworzenia ich od podstaw. Dotyczy to szczególnie kwestii odpowiedzialności za decyzje podejmowane przez systemy AI, transparentności algorytmów, standardów bezpieczeństwa oraz zasad międzynarodowej współpracy w zakresie przeciwdziałania cyberzagrożeniom. Jednocześnie skuteczna walka z cyberprzestępczością wymaga pogłębionej współpracy międzynarodowej oraz harmonizacji regulacji prawnych, aby ograniczyć możliwości wykorzystywania przez sprawców przestępstw różnic między systemami prawnymi. Wynika to z charakteru cyberprzestrzeni, która przekracza granice państw. Przyszłe regulacje powinny dotyczyć wykorzystania systemów AI do prowadzenia nielegalnej działalności, w tym cyberprzestępczości, doprecyzować zasady odpowiedzialności za użycie tych systemów oraz ograniczyć stosowanie niektórych technologii wysokiego ryzyka. Organizacje międzynarodowe, takie jak Organizacja Narodów Zjednoczonych i Unia Europejska, odgrywają istotną rolę w tworzeniu polityk przeciwdziałających cyberprzestępczości oraz w promowaniu wspólnych standardów ochrony danych. Takie działania umożliwią szybsze reagowanie na globalne zagrożenia oraz ułatwią wymianę informacji i doświadczeń w zakresie najlepszych praktyk w cyberbezpieczeństwie.

Podsumowanie

Na podstawie rozważań teoretycznych oraz analizy wybranych przypadków sformułowano następujące wnioski końcowe:

1. Sztuczna inteligencja istotnie zwiększa skuteczność i skalowalność cyberataków, w szczególności przez automatyzację phishingu, tworzenie adaptacyjnego złośliwego oprogramowania oraz wykorzystanie technik generatywnych (deepfake), co prowadzi do obniżenia progu wejścia do cyberprzestępczości.
2. Rozwój systemów cyberobrony opartych na AI poprawia zdolności wykrywania i reagowania na zagrożenia. Dochodzi jednak do wyścigu technologicznego pomiędzy podmiotami odpowiedzialnymi za bezpieczeństwo cyfrowe a cyberprzestępcami.

3. Skuteczne przeciwdziałanie zagrożeniom, które są generowane z pomocą AI, wymaga skoordynowanych działań systemowych obejmujących współpracę międzynarodową, rozwój ram regulacyjnych oraz harmonizację standardów prawnych i technicznych w zakresie wykorzystania AI w cyberprzestrzeni. Obecnie na poziomie międzynarodowym podstawę współpracy państw w zwalczaniu cyberprzestępczości stanowi Konwencja o cyberprzestępczości przyjęta przez Radę Europy. Określa ona ramy współpracy w zakresie ścigania przestępstw popełnianych z wykorzystaniem systemów informatycznych. W ramach UE istotną rolę odgrywa również dyrektywa NIS 2, której celem jest podniesienie poziomu bezpieczeństwa sieci i systemów informatycznych w państwach członkowskich, a także rozporządzenie Artificial Intelligence Act wprowadzające ramy prawne dla bezpiecznego i odpowiedzialnego stosowania systemów AI.
4. Dynamiczny rozwój technologii AI stawia przed obowiązującymi systemami prawnymi i technicznymi nowe wyzwania. Nie wydaje się jednak konieczne tworzenie nowych regulacji, a raczej doprecyzowanie oraz dostosowanie obowiązujących przepisów do specyfiki zagrożeń związanych z wykorzystaniem AI w cyberprzestrzeni.

Bibliografia

Chauhan R., Sabeel U., Izaddoost A., Heydari S.S., *Polymorphic Adversarial Cyberattacks Using WGAN*, „Journal of Cybersecurity and Privacy" 2021, nr 1, s. 767–792. https://doi.org/10.3390/jcp1040037.

Dash B., Ansari M.F., Sharma P., Ali A., *Threats and Opportunities with AI-based Cyber Security Intrusion Detection: A Review*, „International Journal of Software Engineering & Applications" 2022, t. 13, nr 5, s. 13–21. https://doi.org/10.5121/ijsea.2022.13502.

Hagerott M., *Stuxnet and the vital role of critical infrastructure operators and engineers*, „International Journal of Critical Infrastructure Protection" 2014, t. 7, nr 4, s. 244–246. https://doi.org/10.1016/j.ijcip.2014.09.001.

Jancelewicz J., *Phishing i pokrewne ataki socjotechniczne jako zagrożenie dla organizacji pozarządowych*, „Trzeci Sektor" 2022, t. 3–4, nr 59–60, s. 79–88. https://doi.org/10.26368/17332265-59/60-3/4-2022-5.

Keshava R., Pandurangan S.K., Sakthivanitha M., Parmsivan S., Sunkara G., Maruthi R., *AI-Powered Algorithms for the Prevention and Detection of Computer Malware Infections*, w: 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC 2025), Coimbatore 2025. https://doi.org/10.1109/ICESC65114.2025.11212519.

Kurniawan A., Fitriansyah A., *A Literature Review of Historical and Detection Analysis of Botnets Forensics*, „International Journal of Computer and Communication Engineering" 2018, t. 7, nr 4, s. 128–135. https://doi.org/10.17706/ijcce.2018.7.4.128-135.

Russell S., Norvig P., *Artificial Intelligence. A Modern Approach*, bmw 2021.

Wasiuta O., Wasiuta S., *Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, t. 9, nr 3, s. 19–30. https://doi.org/10.24917/26578549.9.3.2.

Yigit Y., Buchanan W.J., Tehrani M.G., Maglaras L., *Review of Generative AI Methods in Cybersecurity*, preprint, arXiv, 13 III 2024 r. https://doi.org/10.48550/arXiv.2403.08701.

Źródła internetowe

Cognizant Security Incident Update, Cognizant, 18 IV 2020 r., https://news.cognizant.com/2020-04-18-Cognizant-Security-Incident-Update?utm [dostęp: 9 III 2026].

Disrupting the first reported AI-orchestrated cyber espionage campaign, Anthropic, 13 XI 2025 r., https://www.anthropic.com/news/disrupting-AI-espionage [dostęp: 24 XI 2025].

Etaher N., Weir G.R.S., *Understanding the Threat of Banking Malware*, w: *Proceedings of Cyberforensics 2014*, https://strathprints.strath.ac.uk/48856/1/8_etaher_weir.pdf [dostęp: 18 X 2025].

Griffioen H., Doerr Ch., *Examining Mirai's Battle over the Internet of Things*, https://www.cyber-threat-intelligence.com/publications/CCS2020-iotbattle.pdf [dostęp: 18 X 2025].

Newman L.H., *GitHub Survived the Biggest DDoS Attack Ever Recorded*, Wired, 1 III 2018 r., https://www.wired.com/story/github-ddos-memcached/ [dostęp: 26 X 2024].

Statt N., *Twitter's massive attack: What we know after Apple, Biden, Obama, Musk, and others tweeted a bitcoin scam*, The Verge, 17 VII 2020 r., https://www.theverge.com/2020/7/15/21326200/elon-musk-bill-gates-twitter-hack-bitcoin-scam-compromised [dostęp: 24 X 2025].

The Evolution of Cybersecurity, Codecademy, https://www.codecademy.com/article/evolution-of-cybersecurity [dostęp: 7 X 2024].

Truță F., *Cognizant Expects to Lose up to $70 Million from April Ransomware Attack*, Bitdefender, 11 V 2020 r., https://www.bitdefender.com/en-gb/blog/hotforsecurity/cognizant-expects-to-lose-up-to-70-million-from-april-ransomware-attack/ [dostęp: 24 X 2024].

Yimu J., Shangdong L., *Threats from Botnets*, w: *Computer Security Threats*, C. Thomas, P. Fraga-Lamas, T.M. Fernández-Caramés (red.), September 2020, s. 52–75, https://www.intechopen.com/chapters/69332 [dostęp: 18 X 2024].

Akty prawne

Konwencja Rady Europy o cyberprzestępczości, sporządzona w Budapeszcie dnia 23 listopada 2001 r. (DzU z 2015 r. poz. 728).

Dyrektywa Parlamentu Europejskiego i Rady (UE) 2022/2555 z dnia 14 grudnia 2022 r. w sprawie środków na rzecz wysokiego wspólnego poziomu cyberbezpieczeństwa na terytorium Unii, zmieniająca rozporządzenie (UE) nr 910/2014 i dyrektywę (UE) 2018/1972 oraz uchylająca dyrektywę (UE) 2016/1148 (dyrektywa NIS 2) – (Dz. Urz. UE L 333/80 z 27 XII 2022 r.).

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) – (Dz. Urz. L 2024/1689 z 12 VII 2024 r.).

Jakub Gajecki

Absolwent Akademii im. Jakuba z Paradyża w Gorzowie Wielkopolskim na kierunku kryminologia stosowana, ze specjalnością zwalczanie cyberprzestępczości. Student studiów II stopnia w Akademii Policji w Szczytnie na kierunku bezpieczeństwo w cyberprzestrzeni. Jego zainteresowania naukowe obejmują cyberbezpieczeństwo oraz bezpieczeństwo państwa. Funkcjonariusz Policji zajmujący się prowadzeniem postępowań przygotowawczych.

Kontakt: gajeckijakub@protonmail.com