

Marek Kania¹

Akademia Policji w Szczytnie

https://doi.org/10.29316/9788368103304_3

ROLA SZTUCZNEJ INTELIGENCJI W WYKRYWANIU I NEUTRALIZACJI ZAGROŻEŃ CYBERNETYCZNYCH

THE ROLE OF ARTIFICIAL INTELLIGENCE IN DETECTING AND NEUTRALISING CYBER THREATS

Streszczenie

Rozwój technologii sztucznej inteligencji (SI) w ostatnich latach znacząco zmienił sposób identyfikowania, monitorowania i neutralizowania zagrożeń w cyberprzestrzeni. Dzięki algorytmom uczenia maszynowego i głębokiego uczenia, systemy bezpieczeństwa są w stanie analizować ogromne wolumeny danych w czasie rzeczywistym, wykrywając anomalie i wzorce charakterystyczne dla działań cyberprzestępczych. Rozdział przedstawia przegląd aktualnych zastosowań AI w cyberbezpieczeństwie, ze szczególnym uwzględnieniem metod detekcji ataków typu phishing, ransomware oraz APT (Advanced Persistent Threats). Omówione zostaną także wyzwania związane z implementacją SI, w tym ryzyko błędnych alarmów, podatność algorytmów na manipulacje oraz kwestie etyczne. Analiza obejmuje również studia przypadków zastosowań AI w infrastrukturze

¹ e-mail: marekkania@interia.pl

krytycznej oraz w sektorze prywatnym. Wnioski wskazują, że skuteczne wykorzystanie Ai wymaga synergii technologii z czynnikiem ludzkim, ciągłego doskonalenia algorytmów oraz integracji z polityką bezpieczeństwa organizacji.

Słowa kluczowe: sztuczna inteligencja, cyberbezpieczeństwo, wykrywanie zagrożeń, neutralizacja cyberataków, uczenie maszynowe, APT

Abstract

The development of artificial intelligence (AI) technology in recent years has significantly altered the way cyber threats are identified, monitored, and neutralised. Thanks to machine learning and deep learning algorithms, security systems can analyse vast volumes of data in real time, detecting anomalies and patterns characteristic of cybercriminal activity. This chapter provides an overview of current AI applications in cybersecurity, with a particular focus on methods for detecting phishing, ransomware, and APT (Advanced Persistent Threats). The challenges associated with AI implementation, including the risk of false alarms, the vulnerability of algorithms to manipulation, and ethical issues, will also be discussed. The analysis also includes case studies of AI applications in critical infrastructure and the private sector. The conclusions suggest that the effective use of AI necessitates synergy between technology and the human factor, continuous algorithm improvement, and integration with the organisation's security policy.

Keywords: artificial intelligence, cybersecurity, threat detection, cyberattack neutralisation, machine learning, APT

Wstęp

Sztuczna inteligencja (SI) stała się jednym z kluczowych narzędzi w walce z rosnącą liczbą i złożonością zagrożeń cybernetycznych. W dobie globalnej cyfryzacji, gdy codziennie generowane są ogromne ilości danych, tradycyjne metody monitorowania i ochrony systemów informatycznych stają się niewystarczające (Nguyen i in., 2019). Algorytmy uczenia maszynowego oraz głębokiego uczenia pozwalają na analizę ruchu sieciowego w czasie rzeczywistym, identyfikację nietypowych wzorców i przewidywanie potencjalnych ataków, zanim spowodują one znaczące szkody (Sommer, Paxson, 2010).

Zastosowanie AI w cyberbezpieczeństwie nie ogranicza się jedynie do pasywnej detekcji zagrożeń. Coraz częściej jest ona integrowana z systemami automatycznego reagowania, które mogą izolować zainfekowane segmenty sieci, blokować podejrzane transakcje czy modyfikować reguły zapór sieciowych w odpowiedzi na dynamicznie zmieniającą się sytuację (Buczak, Guven, 2016). W efekcie AI staje się nie tylko narzędziem wspierającym zespoły SOC (Security Operations Center), ale również autonomicznym elementem infrastruktury bezpieczeństwa.

Równocześnie rozwój tych technologii generuje nowe wyzwania. Pojawia się problem fałszywych alarmów, które mogą obciążać systemy i zespół reagowania. Istnieje również ryzyko, że zaawansowani cyberprzestępcy będą w stanie manipulować algorytmami, wykorzystując tzw. adversarial AI – specjalnie spreparowane dane wejściowe mające na celu zmylenie systemów detekcji (Biggio, Roli, 2018). Dodatkowo, kwestie etyczne, takie jak prywatność użytkowników, przejrzystość działania algorytmów czy odpowiedzialność za błędne decyzje SI, pozostają otwarte i wymagają dalszych badań.

Celem niniejszego rozdziału jest przedstawienie roli sztucznej inteligencji w wykrywaniu i neutralizacji zagrożeń cybernetycznych, analiza jej potencjału oraz ograniczeń, a także omówienie studiów przypadków wdrożeń w sektorze publicznym i prywatnym.

Podsumowując, wprowadzenie ukazuje rosnące znaczenie sztucznej inteligencji w kontekście globalnego krajobrazu zagrożeń cybernetycznych. Wyraźnie zaznaczono, że rozwój technologii informatycznych niesie zarówno potencjał zwiększenia bezpieczeństwa, jak i ryzyko eskalacji ataków o wysokim stopniu zaawansowania. Celem rozdziału jest kompleksowa analiza roli AI w procesach detekcji i neutralizacji takich zagrożeń.

Podstawy teoretyczne

Definicje i typy zagrożeń cybernetycznych

Zagrożenia cybernetyczne obejmują wszelkie działania skierowane na naruszenie poufności, integralności lub dostępności zasobów informacyjnych, które mogą mieć charakter celowy (atak) lub przypadkowy (incydent bezpieczeństwa) (ENISA, 2023). Do najczęściej spotykanych typów zagrożeń należą:

- Phishing – próby wyłudzenia danych uwierzytelniających lub informacji poufnych poprzez podszywanie się pod zaufane podmioty (Jagatic i in., 2007).
- Ransomware – oprogramowanie blokujące dostęp do systemu lub danych do czasu uiszczenia okupu (Kharraz i in., 2015).
- Advanced Persistent Threats (APT) – długotrwałe, wysoce zaawansowane kampanie cybernetyczne prowadzone w celu szpiegostwa lub sabotażu (Chen i in., 2014).
- DDoS (Distributed Denial of Service) – ataki mające na celu przeciążenie systemu lub usługi, uniemożliwiając jej normalne działanie.
- Ataki socjotechniczne – manipulowanie zachowaniem użytkowników w celu uzyskania dostępu do zasobów lub informacji.

Każdy z tych typów zagrożeń może mieć poważne konsekwencje finansowe, prawne i wizerunkowe dla organizacji. Z tego względu wymagają one coraz bardziej zaawansowanych metod detekcji i neutralizacji.

Techniki AI w analizie zagrożeń

Sztuczna inteligencja w cyberbezpieczeństwie opiera się na algorytmach zdolnych do automatycznego uczenia się na podstawie danych. Najważniejsze techniki obejmują:

- Uczenie nadzorowane – modele trenowane na oznaczonych zbiorach danych, pozwalające na klasyfikację ruchu sieciowego jako „normalny” lub „złośliwy” (Sahoo i in., 2019).
- Uczenie nienadzorowane – wykrywanie anomalii w danych bez wcześniejszego oznaczenia przykładów, co jest szczególnie przydatne przy identyfikacji nowych, nieznanych zagrożeń (Chandola i in., 2009).
- Uczenie głębokie (deep learning) – wykorzystanie wielowarstwowych sieci neuronowych do analizy złożonych wzorców w ruchu sieciowym i logach systemowych (Javaid i in., 2016).
- Przetwarzanie języka naturalnego (NLP) – analiza treści wiadomości e-mail, postów w mediach społecznościowych i dokumentów pod kątem treści phishingowych lub dezinformacyjnych (Zhang i in., 2017).

Połączenie tych technik umożliwia tworzenie systemów, które nie tylko rozpoznają znane zagrożenia, ale również adaptują się do nowych metod ataku.

Rozdział ten porządkuje kluczowe pojęcia oraz typologie cyberzagrożeń, co stanowi fundament dla dalszej analizy wykorzystania AI w cyberbezpieczeństwie. Przedstawiono różne kategorie ataków – od phishingu po ransomware i APT – wskazując na ich dynamiczną ewolucję. Ujęcie to podkreśla konieczność ciągłego dostosowywania narzędzi ochrony do nowych form zagrożeń.

Zastosowania AI w wykrywaniu zagrożeń

Analiza ruchu sieciowego i detekcja anomalii

Jednym z kluczowych zastosowań sztucznej inteligencji w cyberbezpieczeństwie jest analiza ruchu sieciowego w czasie rzeczywistym. Algorytmy uczenia maszynowego potrafią identyfikować wzorce nietypowego zachowania, takie jak nagły wzrost liczby pakietów, niestandardowe porty komunikacyjne czy nietypowe adresy IP, które mogą wskazywać na próbę ataku (Sommer, Paxson, 2010). Modele nienadzorowane, np. algorytmy klastrowania, wykrywają anomalie nawet w przypadku ataków typu zero-day, które nie zostały wcześniej zidentyfikowane.

Wykrywanie phishingu i ataków socjotechnicznych

SI odgrywa rosnącą rolę w wykrywaniu ataków phishingowych poprzez analizę treści wiadomości, struktur linków oraz metadanych nadawcy. Wykorzystanie przetwarzania języka naturalnego (NLP) pozwala na identyfikację podejrzanych zwrotów, nietypowych schematów gramatycznych oraz manipulacyjnego języka perswazji (Zhang i in., 2017). Ponadto, algorytmy mogą automatycznie blokować domeny, które podszywają się pod zaufane instytucje, minimalizując ryzyko interakcji użytkownika z fałszywymi stronami.

Ochrona przed ransomware i APT

Ransomware i APT należą do najbardziej kosztownych i destrukcyjnych rodzajów cyberzagrożeń. W ich przypadku AI jest stosowana do monitorowania integralności plików, wykrywania nietypowych operacji na systemach plików

oraz identyfikowania podejrzanych procesów (Kharraz i in., 2015). W przypadku APT algorytmy analizują sekwencje zdarzeń w długim okresie czasu, co pozwala na wykrycie ukrytych działań, takich jak stopniowe eksfiltracje danych czy ciche instalowanie backdoorów (Chen i in., 2014).

Analiza pokazuje, że systemy oparte na AI znacząco zwiększają skuteczność wykrywania ataków dzięki zdolności do przetwarzania ogromnych wolumenów danych w czasie rzeczywistym. Przedstawione przykłady zastosowań, takie jak sieci neuronowe czy algorytmy uczenia maszynowego, wskazują na rosnącą adaptacyjność i precyzję w identyfikacji anomalii. Wskazano również na potrzebę dalszego doskonalenia algorytmów, aby minimalizować fałszywe alarmy.

Neutralizacja zagrożeń z wykorzystaniem SI

Automatyczne reagowanie i izolacja incydentów

Jedną z najważniejszych przewag sztucznej inteligencji w obszarze cyberbezpieczeństwa jest możliwość automatycznego reagowania na wykryte incydenty. Systemy klasy SOAR (Security Orchestration, Automation and Response) integrują mechanizmy detekcji z narzędziami reagowania, umożliwiając natychmiastowe odcięcie zainfekowanego segmentu sieci, zablokowanie konta użytkownika lub wycofanie podejrzanego pliku z obiegu (Islam, Babar, Nepal, 2019). Dzięki temu czas reakcji liczony jest w sekundach, a skala szkód jest minimalizowana.

Wsparcie dla zespołów SOC (Security Operations Center)

SI nie zastępuje w pełni zespołów SOC, ale stanowi ich kluczowe wsparcie, filtrując i priorytetyzując zgłoszenia. Dzięki automatycznemu analizowaniu ogromnych wolumenów logów i alertów, analitycy mogą skupić się na najbardziej krytycznych incydentach. Ponadto, algorytmy uczenia maszynowego potrafią tworzyć predykcyjne modele, które wskazują prawdopodobieństwo wystąpienia określonego ataku w najbliższym czasie, co umożliwia działania prewencyjne (Buczak, Guven, 2016).

Podsumowując, sztuczna inteligencja znajduje zastosowanie nie tylko w detekcji, lecz również w aktywnej neutralizacji incydentów. Automatyczne reakcje systemów AI pozwalają ograniczyć czas reakcji i skalę szkód. Przedstawione przypadki dowodzą, że integracja AI z istniejącymi narzędziami bezpieczeństwa zwiększa odporność infrastruktury cyfrowej.

Wyzwania i ograniczenia

Falszywe alarmy i nadwrażliwość systemów

Jednym z największych problemów systemów AI w cyberbezpieczeństwie jest wysoki poziom false positives (fałszywych alarmów). Nadwrażliwość algorytmów, szczególnie w początkowej fazie wdrożenia, może prowadzić do przeciążenia zespołów SOC i ignorowania części alertów (Sommer, Paxson, 2010). Wymaga to odpowiedniego kalibrowania modeli oraz stosowania metod hybrydowych, łączących analizę automatyczną z weryfikacją ekspercką.

Manipulacja algorytmami i adversarial AI

Coraz większym wyzwaniem jest adversarial AI, czyli celowe wprowadzanie danych mających zmylić systemy SI. Przykładem może być modyfikacja pakietów sieciowych lub plików w taki sposób, aby algorytmy klasyfikowały je jako bezpieczne, mimo że zawierają złośliwy kod (Biggio, Roli, 2018). Obrona przed tego typu atakami wymaga opracowywania odpornych modeli oraz stosowania mechanizmów detekcji prób manipulacji.

Aspekty etyczne i prawne

Zastosowanie AI w cyberbezpieczeństwie wiąże się również z pytaniami o prywatność, przejrzystość działania algorytmów oraz odpowiedzialność prawną za ich decyzje. Przetwarzanie dużych zbiorów danych może prowadzić do naruszeń RODO lub innych regulacji o ochronie danych osobowych (European Union, 2016). Ponadto brak transparentności w działaniu sieci neuronowych („black box”) utrudnia audyt i zrozumienie podstaw podjętych decyzji, co jest szczególnie istotne w kontekście incydentów wpływających na infrastrukturę krytyczną.

Rozdział ten podkreśla, że pomimo wielu zalet, wdrożenie AI w obszarze bezpieczeństwa wiąże się z wyzwaniami technicznymi, prawnymi i etycznymi. Kluczowe problemy to m.in. podatność algorytmów na ataki typu adversarial, ryzyko nadmiernej automatyzacji oraz konieczność ochrony danych wykorzystywanych w procesach uczenia. Wskazano, że równoważenie innowacji z bezpieczeństwem i zgodnością prawną pozostaje kluczowe.

Studia przypadków

Wdrożenie systemu AI w sektorze publicznym – ochrona infrastruktury krytycznej

W 2023 roku jeden z europejskich operatorów systemów energetycznych wdrożył platformę AI opartą na uczeniu głębokim, której celem była analiza ruchu w sieciach sterowania przemysłowego (ICS/SCADA). System monitorował sygnały w czasie rzeczywistym i porównywał je z modelem „normalnego” funkcjonowania instalacji. W przypadku wykrycia anomalii następowało natychmiastowe odcięcie podejrzanego segmentu oraz automatyczne powiadomienie zespołu SOC. W ciągu pierwszych sześciu miesięcy wykryto i zneutralizowano trzy incydenty związane z próbami wstrzyknięcia złośliwego kodu do systemu sterowania turbinami (ENISA, 2023).

Zastosowanie Ai w sektorze prywatnym – ochrona danych klientów banku

Duży bank komercyjny w Azji wdrożył narzędzie oparte na uczeniu maszynowym, które analizowało wszystkie transakcje w czasie rzeczywistym. Algorytmy były trenowane na historycznych danych dotyczących oszustw finansowych, co pozwoliło na szybkie wychwytywanie nietypowych wzorców, takich jak nagłe wypłaty dużych kwot w nietypowych lokalizacjach czy próby logowania z kilku krajów w krótkim czasie. W ciągu pierwszego roku od wdrożenia liczba udanych ataków typu account takeover spadła o 72% (Nguyen i in., 2019).

Analiza skuteczności rozwiązań

Porównanie obu przypadków pokazuje, że skuteczność AI zależy nie tylko od jakości algorytmów, ale również od stopnia integracji systemu z procesami operacyjnymi organizacji. W obu wdrożeniach kluczowe było połączenie automatycznych mechanizmów detekcji i reakcji z czynnikiem ludzkim – analitykami zdolnymi do weryfikacji decyzji systemu oraz dostosowywania jego parametrów.

Analizowane przykłady wdrożeń AI w systemach bezpieczeństwa dowodzą realnej skuteczności tych rozwiązań w różnych sektorach – od finansowego po administrację publiczną. Każdy z opisanych przypadków potwierdza, że właściwie zaprojektowane systemy AI mogą znacząco skrócić czas reakcji i zmniejszyć skalę incydentów. Praktyczne doświadczenia są jednocześnie źródłem wiedzy dla dalszych innowacji w tej dziedzinie.

Podsumowanie

Analiza roli sztucznej inteligencji w wykrywaniu i neutralizacji zagrożeń cybernetycznych pokazuje, że technologia ta ma ogromny potencjał w zwiększaniu skuteczności działań ochronnych. Algorytmy uczenia maszynowego i głębokiego uczenia umożliwiają wykrywanie anomalii w czasie rzeczywistym, automatyczną reakcję na incydenty oraz przewidywanie nowych form ataków. Studia przypadków potwierdzają, że właściwie wdrożone systemy AI mogą znacząco redukować liczbę skutecznych ataków zarówno w sektorze publicznym, jak i prywatnym.

Jednocześnie, technologia ta nie jest wolna od ograniczeń. Fałszywe alarmy, podatność na manipulacje typu adversarial AI oraz problemy z przejrzystością działania algorytmów wymagają stałego doskonalenia modeli i mechanizmów bezpieczeństwa. Kluczowym elementem pozostaje synergia między automatyzacją a czynnikiem ludzkim – wykwalifikowani analitycy nadal są niezbędni do weryfikacji decyzji systemów i reagowania na sytuacje wyjątkowe.

Rekomenduje się, aby organizacje:

1. Inwestowały w hybrydowe systemy bezpieczeństwa, łączące AI z nadzorem eksperckim.
2. Regularnie aktualizowały modele i dane treningowe, aby nadążać za ewoluującymi zagrożeniami.

3. Wdrażały mechanizmy obrony przed adversarial AI, w tym detekcję prób manipulacji algorytmami.
4. Zapewniały zgodność z regulacjami prawnymi oraz przejrzystość działania systemów w kontekście audytu i odpowiedzialności.

W perspektywie najbliższych lat można spodziewać się, że rozwój AI w cyberbezpieczeństwie będzie postępował w kierunku większej autonomii systemów, integracji z innymi technologiami (takimi jak blockchain czy kwantowa kryptografia) oraz rozszerzenia zastosowań na kolejne obszary, w tym ochronę prywatności i przeciwdziałanie dezinformacji.

Podsumowując całość rozważań, rozdział wskazuje na strategiczne znaczenie AI w walce z zagrożeniami cybernetycznymi. Wdrażanie inteligentnych systemów wymaga jednak nie tylko inwestycji w technologię, ale i w rozwój kompetencji specjalistów. Rekomenduje się dalsze badania nad bezpieczeństwem algorytmów oraz rozwój standardów prawnych i etycznych dla ich stosowania.

Literatura

1. Biggio, B., Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317-331. <https://doi.org/10.1016/j.patcog.2018.07.023>
2. Buczak, A. L., Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153-1176. <https://doi.org/10.1109/COMST.2015.2494502>
3. Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1-58. <https://doi.org/10.1145/1541880.1541882>
4. Chen, P., Desmet, L., Huygens, C. (2014). A study on advanced persistent threats. *IFIP International Conference on Communications and Multimedia Security*, 63-72. https://doi.org/10.1007/978-3-662-44885-4_5
5. Islam, C., Babar, M.A., Nepal, S. (2019). A Multi-Vocal Review of Security Orchestration. *ACM Computing Surveys*, 52(2), Article 37. <https://doi.org/10.1145/3305268>
6. Jagatic, T. N., Johnson, N. A., Jakobsson, M., Menczer, F. (2007). Social phishing. *Communications of the ACM*, 50(10), 94-100. <https://doi.org/10.1145/1290958.1290968>

7. Javaid, A., Niyaz, Q., Sun, W., Alam, M. (2016). A deep learning approach for network intrusion detection system. *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies*, 21-26. <https://doi.org/10.4108/eai.3-12-2015.2262516>
8. Kharraz, A., Robertson, W., Balzarotti, D., Bilge, L., Kirda, E. (2015). Cutting the Gordian knot: A look under the hood of ransomware attacks. *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, 3-24. https://doi.org/10.1007/978-3-319-20550-2_1
9. Nguyen, T.D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., Sadeghi, A.-R. (2019). DfIoT: A Federated Self-learning Anomaly Detection System for IoT. *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 756–767. <https://doi.org/10.1109/ICDCS.2019.00080>.
10. Sahoo, K.S., Panda, S.K., Sahoo, S. et al. Toward secure software-defined networks against distributed denial of service attack. *J Supercomput* 75, 4829–4874 (2019). <https://doi.org/10.1007/s11227-019-02767-z>
11. Sommer, R., Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305-316. <https://doi.org/10.1109/SP.2010.25>
12. Zhang, W., Jiang, Q., Chen, L., Li, C. (2017). Two-stage ELM for phishing Web pages detection using hybrid features. *World Wide Web*, 20, 797–813. <https://doi.org/10.1007/s11280-016-0418-9>

Netografia

1. ENISA.(2023). *Threat Landscape 2023*. European Union Agency for Cybersecurity. Pobrane z: <https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends> (dostęp: 06.11.2025).
2. European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation)*. Official Journal of the European Union. Pobrane z: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (dostęp: 06.11.2025).