Wiktoria Grzesiuk
Uniwersytet SWPS w Warszawie
wi.grzesiuk@gmail.com
https://orcid.org/0009-0000-9931-5683

SZTUCZNA INTELIGENCJA W KONTEKŚCIE CYBERBEZPIECZEŃSTWA

Wstęp

Współczesny rozwój technologiczny nieodmiennie koreluje z procesem cyfryzacji, w ramach czego coraz większe znaczenie zyskują systemy wykorzystujące sztuczną inteligencję[31]. Rosnące uzależnienie od tego sektora wiąże się z eskalacją zagrożeń w obszarze cyberbezpieczeństwa. W 2021 roku aż 22,2% przedsiębiorstw na terenie Unii Europejskiej doświadczyło co najmniej jednego incydentu skutkującego zablokowaniem dostępu, zniszczeniem bądź ujawnieniem danych poufnych[32]. Rok później Polska uplasowała się na niechlubnym drugim miejscu rankingu[33]. Nie ulega wątpliwościom fakt, iż AI odgrywa istotną rolę w kształtowaniu ochrony sfery wirtualnej, jednakże zasadniczym problemem badawczym pozostaje określenie charakteru tej roli – czy AI stanowi narzędzie wzmacniające profilaktykę przed cyberatakami, czy też przyczynia się do ich intensyfikacji?

1 Uczenie maszynowe

Zgodnie z art. 3 ust. 1 aktu w sprawie sztucznej inteligencji[34] „system AI" został zdefiniowany jako układ maszynowy, zdolny do operowania na różnych poziomach autonomii

[31] Dalej: AI.
[32] European Commission, *2023 Report on the state of the Digital Decade*, 2023.
[33] Polska Agencja Prasowa, *Eurostat: polskie firmy w unijnej czołówce pod względem zgłaszanych naruszeń cyberbezpieczeństwa*, https://www.pap.pl/aktualnosci/eurostat-polskie-firmy-w-unijnej-czolowce-pod-wzgledem-zglaszanych-naruszen-0 (dostęp: 15.06.2025 r.).
[34] Rezolucja ustawodawcza Parlamentu Europejskiego z dnia 13 marca 2024 r. w sprawie wniosku dotyczącego rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące

i przystosowany do adaptacyjnego działania po wdrożeniu. Jego fundamentalną cechą jest zdolność do wnioskowania w sposób jawny lub dorozumiany w celu generowania wyników na podstawie danych wejściowych. Wyniki te mogą wywierać wpływ zarówno na środowisko fizyczne, jak i wirtualne. Przyjmują formę m.in. predykcji, rekomendacji bądź decyzji. Ich generowanie opiera się w szczególności na uczeniu maszynowym, czyli na eksperckim systemie zintegrowanym właśnie z AI. System wykorzystuje algorytmy adaptacyjne, gwarantujące komputerom autonomiczne przyswajanie faktów i przystosowanie do otoczenia, bazując na zgromadzonych danych. Uczenie maszynowe działa na zasadzie integracyjnego doskonalenia poprzez doświadczenie. Istnieją trzy główne typy o różnych przeznaczeniach. Jeden z nich stanowi uczenie nadzorowane, gdzie algorytm jest kształcony przy wykorzystaniu zestawu danych treningowych, w którym już są podane oczekiwane wyniki, a za główny cel obiera się umiejętność ich przewidywania. Z kolei drugi typ to uczenie nienadzorowane, budujące algorytm na danych pozbawionych etykiet lub odpowiedzi referencyjnych. W efekcie system samodzielnie ma zacząć wyodrębniać ukryte wzorce. Grupuje informacje na podstawie podobieństwa bądź redukuje ich wymiarowość. Trzecią kategorię stanowi uczenie wzmacniające, podczas którego komputer rozpatruje wcześniej popełnione błędy. W tej sytuacji priorytet stanowi rozwój umiejętności adaptacji w dynamicznie zmieniającym się środowisku[35]. Zastosowanie takowego podziału uczenia maszynowego na poszczególne typy ma zwiększyć produktywność całego systemu eksperckiego, zintegrowanego z AI. Jest on w stanie zmodernizować codzienną pracę, ale również, niestety, taktyki cyberprzestępców.

2 Inteligentna automatyzacja wobec cyberataków

Inteligentna automatyzacja, bazująca na AI i uczeniu maszynowym, zwiększa ogólną efektywność w łańcuchu cyberataków. Znacznie przyspiesza pozyskiwanie danych uwierzytelniających, wykrywanie luk w oprogramowaniu oraz odgadywanie haseł. Atakujący (*hacker*) spędza mniej czasu w systemie ofiary, proporcjonalnie ograniczając szanse na swoje wykrycie, tym bardziej że AI jednocześnie rozpatruje kroki, jakie może podjąć system obronny w odpowiedzi na atak. Dodatkowo umożliwia przeprowadzenie ofensywy na szerszą skalę,

sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii [COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)].

[35] Finnish Transport and Communications Agency Traficom, *The security threat of AI-enabled cyberattacks*, 2022.

czyli namierza wiele celów jednocześnie. Analizuje i eksploruje zasoby danych w skali masowej, w tym zestawienia biznesowe, co przy manualnych działaniach wiązałoby się ze znacznym ryzykiem przypadkowego pominięcia kluczowej informacji. Uczenie maszynowe dąży również do zapewnienia autonomii złośliwym programom, przy czym ma to wzmocnić pozycję cyberprzestępcy m.in. poprzez proces generowania oczekiwanych treści. Przy użyciu ofensywnej zdolności AI do ukrywania się przeciwnicy dostosowują ruch sieciowy i komunikację w taki sposób, aby wydawały się one naturalne i nie wzbudzały podejrzeń. Przykładem jest organizowanie sieci *botnet*, której zdalne sterowanie stanowi jedno z najgroźniejszych przypadków cyberataków. Proces inkorporowania komputerów w ramach takiej działalności jest motywowany intencjami sprawcy, obejmującymi zazwyczaj utworzenie cybernetycznej broni masowego rażenia. Wywołany chaos skutkuje znacznym utrudnieniem bądź całkowitym uniemożliwieniem identyfikacji niebezpieczeństw, a co za tym idzie, zakłóceniem działań obronnych. Taki kamuflaż pozwala na skuteczne przeprowadzenie skanów i eksfiltracji informacji. Generatywne modele uczenia maszynowego mogą wypełniać ewentualne luki w zbiorach danych wywiadowczych oraz pozyskiwać hasła do kont za pomocą narzędzi CeWL (*Custom Word List Generator*), które tworzą spersonalizowane listy słów, bazując na wyrażeniach znalezionych na stronie internetowej. Uwzględniają przy tym parametry zdefiniowane przez użytkownika, np. minimalną liczbę znaków wymaganych do stworzenia hasła[36].

Określenie zakresu możliwości działań jest niezwykle cenne przy przeprowadzaniu z pomocą AI symulacji środowiska ataku, mającego na celu przetestowanie zaplanowanej kampanii przed jej wdrożeniem. Wówczas cyberprzestępcy biorą pod uwagę informacje uzyskane metodą socjotechniki, która polega na uznaniu konkretnego użytkownika za „najsłabsze ogniwo". Posiłkując się manipulacją, *hackerzy* przeprowadzają, za pomocą np. *chatbotów*, interakcje z ludźmi w sposób spersonalizowany. Dodatkowo, naruszając metody uwierzytelniania, AI jest w stanie ukraść konta innych użytkowników lub tworzyć nowe, fałszywe. Wykorzystuje wówczas pobrane zbiory faktów i kreuje *deepfakes*, czyli pozornie autentyczne treści audio, a także wideo poszerzone o synchronizację ruchu ust postaci z mową. Tym sposobem naśladuje ludzkie zachowania, nawet w czasie rzeczywistym[37]. Jednocześnie pokonuje niejawne systemy rejestrowania kluczy, oparte na zachowaniach użytkowników, takich jak wzorce naciśnięć klawiszy bądź ruch gałek ocznych. Na podobnej zasadzie rozwiązywane są testy CAPTCHA (*Completely Automated Public Turing test to*

[36] Ibidem.
[37] Ibidem.

tell Computers and Humans Apart), które w założeniu mają potwierdzić, że użytkownik faktycznie jest istotą ludzką. Próba składa się z dwóch głównych elementów: zniekształconego obrazu z losowo wygenerowaną sekwencją liter lub liczb oraz z pola tekstowego, w którym użytkownik ma za zadanie poprawnie przepisać układ wspomnianych znaków[38]. Cyberprzestępcy używają zaawansowanych narzędzi AI automatycznie rozwiązujących test CAPTCHA najczęściej podczas ataków typu *phishing*. W 2021 r. był to najczęściej występujący rodzaj cyberataku. Zespół Reagowania na Incydenty Bezpieczeństwa Komputerowego (*Computer Security Incident Response Team*; CSIRT NASK) zarejestrował 29 504 incydenty, z czego 25 512 stanowił właśnie *phishing*[39]. Jego „popularność" wynika ze stosunkowo prostego procesu zdobywania potrzebnych danych za pomocą białego wywiadu (*open-source intelligence*; OSINT), który polega na gromadzeniu informacji pochodzących z ogólnie dostępnych białych źródeł, np. serwisów społecznościowych lub publicznych rejestrów. Na ich podstawie AI generuje spersonalizowane treści, zachowując konkretny styl komunikacji, wzbudzając zaufanie ofiary[40]. Najczęściej *phishing* stosuje się w formie wysyłania wiadomości e-mail z prośbą o zalogowanie do konta bankowego bądź portalu społecznościowego. Jeśli potencjalna ofiara postąpi zgodnie z treścią wiadomości, cyberprzestępca przechwyci kluczowe dane. Może on obrać różne cele: od kradzieży pieniędzy lub danych osobowych po przejęcie kontroli nad kontem i rozsianie złośliwego oprogramowania[41]. Inny sposób oszustwa polega na założeniu fałszywego sklepu internetowego, znanego jako *scam*, gdzie przeciwnicy kreują pozornie atrakcyjne oferty o zaniżonych cenach. Nie są one utrzymane tylko w popularnej tematyce modowej, ale również kosmetycznej, a nawet zdrowotnej. Podczas przebywania na tego typu stronach internetowych może mieć miejsce symulacja wystąpienia problemu technicznego. Wówczas *hacker* lub wygenerowany *bot* podszywa się pod pracownika działu IT (*Information Technology*), który wyraża chęć pomocy. Niekiedy inicjuje żądanie dostępu do kamery internetowej, co stanowi okazję do uruchomienia procedur szpiegowskich, czego konsekwencją może być przejęcie kontroli nad systemem[42]. Warto tutaj wspomnieć o sektorze MŚP (Małe i Średnie Przedsiębiorstwa), charakteryzujący się znaczną podatnością na zagrożenie związane ze złośliwym oprogramowaniem *ransomware*.

[38] Ibidem.
[39] Najwyższa Izba Kontroli, *Działania Państwa w zakresie zapobiegania i zwalczania skutków wybranych przestępstw internetowych, w tym kradzieży tożsamości* (Nr ewid. 125/2022/P/21/042/KPB), 2022.
[40] S. Palczewski, *Cyberataki niemożliwe do wykrycia? Rewolucja AI*, https://cyberdefence24.pl/polityka-i-prawo/sztuczna-inteligencja-w-cyberatakachincydenty-niemozliwe-do-wykrycia (dostęp: 15.06.2025 r.)
[41] Portal polskiej Policji, *Internetowi oszuści nie śpią!*, https://policja.pl/pol/aktualnosci/227892,Internetowi-oszusci-nie-spia.html (dostęp: 11.9.2023 r.)
[42] A.I. Ajibade, *Typowe oszustwa internetowe wymierzone w nastolatków*, https://www.internetmatters.org/pl/hub/expert-opinion/common-online-scams-targeting-teenagers/ (dostęp: 15.06.2025 r.)

Jego działanie polega na zablokowaniu użytkownikowi dostępu do danych przechowywanych na zaatakowanym urządzeniu. Jest to oprogramowanie zaprojektowane z zamiarem uzyskania okupu od ofiary, a jednocześnie jest zabezpieczone przed samodzielnym usunięciem przez użytkownika[43].

3 Obronna AI

Mimo rosnącej roli AI w sferze cyberprzestępczości nie stanowi to synonimu nieskuteczności w dziedzinie walki z cyberzagrożeniami. AI nie jest narzędziem wykorzystywanym dla celów wyłącznie nielegalnych, gdyż istnieją programy obronne bazujące właśnie na jej zastosowaniu. Ich działalność zaczyna się w sferze cybernetyki dot. zarządzania systemami sterowania i kontrolą przekazu informacji, których poziom ochrony ma przekład na ogólne cyberbezpieczeństwo. Podmiotem wyspecjalizowanym w tej dziedzinie jest firma *Darktrace*, gdzie funkcję dyrektora generalnego pełni J. Popelka (stan na dzień 4 kwietnia 2025 r.). Przedsiębiorstwo to wykorzystuje zaawansowane technologie AI w celu zabezpieczania infrastruktury informatycznej. Zespół ekspertów *Darktrace* opracował nowatorskie rozwiązanie *Darktrace Antigena* oparte na cyberinteligencji, określanej również mianem „obronnej AI". Działa jako platforma decyzyjna AI, aktywnie zwalczając zagrożenia w czasie rzeczywistym. Analizuje i przyswaja różnice między standardowymi zachowaniami użytkowników a tymi odbiegającymi od normy, przy czym przestrzega rygorystycznych kryteriów[44]. Zaprojektowano 3 kategorie *Darktrace Antigena*. Pierwszą z nich stanowi *Antigena Network,* mająca za zadanie neutralizować ataki ukierunkowane na nielegalne pozyskanie danych osobowych bądź innych informacji zamieszczonych na koncie internetowym, co najczęściej dotyczy ataków typu *ransomware*[45]. Drugą kategorią jest *Antigena Saas* (*Software-as-a-Service*). Mimo podobieństwa do pierwszego rodzaju *Saas* działa w świetle wielowymiarowego przyswojenia cyfrowych zachowań użytkowników w środowiskach hybrydowych i wielochmurowych. Oznacza to, że nie ogranicza się do działań obronnych wobec ataku na jedno konkretne konto, a na sieci operacji biznesowych. Potrafi zneutralizować cyberatak ukierunkowany na kradzież poufnych danych nawet przez administratorów,

[43] „Cyber 360", *Cyberataki w sektorze MŚP z udziałem sztucznej inteligencji*, https://cyber360.pl/blog-o-cyberbezpieczenstwie/cyberataki-w-sektorze-msp-z-udzialem-sztucznej-inteligencji/ (dostęp: 11.9.2023 r.)
[44] Darktrace Antigena, *Darktrace Antigena Product Overview*, https://darktrace.com/legal/darktrace-antigena-product-overview (dostęp: 15.06.2025 r.)
[45] Ibidem.

którzy nadużyli własnych uprawnień[46]. Trzecią, ostatnią, kategorią *Darktrace Antigena* jest *Antigena Email*. Jako technologia neutralizująca określa, czy dana korespondencja elektroniczna należy do tych standardowo występujących między nadawcą a odbiorcą. W przypadku, gdy wykryje zagrożenie reaguje autonomicznie i proporcjonalnie, podejmując unikalną decyzję. Te uznane za złośliwe neutralizuje jeszcze zanim dotrą do skrzynki adresata[47].

4 Polskie instytucje wobec cyberbezpieczeństwa

Na terytorium Rzeczypospolitej Polskiej mieszczą się instytucje państwowe (np. Biuro do Walki z Cyberprzestępczością Komendy Głównej Policji – BdWzC KGP) dokonujące regularnych analiz ryzyka z zamiarem zaplanowania działań ochronnych przeciwko ewentualnym incydentom cyberprzestępczym. Sama analiza tworzona jest przy wykorzystaniu danych przede wszystkim z Krajowego Rejestru Sądowego, Elektronicznego Rejestru Czynności Dochodzeniowo-Śledczych, a także otwartych internetowych źródeł informacji. Każdego dnia personel BdWzC KGP sporządza dokumentację o wykrytych zagrożeniach wraz z ich szczegółowymi opisami. Skompletowane zestawienia są przekazywane do odpowiednich organów, w tym do właściwych wydziałów Komend Wojewódzkich Policji. Pozwala to na identyfikację cech charakterystycznych dla konkretnych zagrożeń oraz identyfikację informacji umożliwiających opracowanie skutecznych strategii reagowania.

Za analizę ryzyka z obszaru cyberprzestępczości odpowiada również Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy (NASK-PIB), która w międzynarodowej współpracy funkcjonuje na terenie Polski, celem zwalczania przestępstw internetowych. W ramach tego dodatkowo działa CERT Polska (*Computer Emergency Response Team Poland*) z zadaniem całodobowego monitoringu trendów w zakresie zagrożeń dla strefy cyfryzacji. Reaguje na zgłoszenia i prowadzi proaktywną wykładnię ruchów w sieci, szacując również prawdopodobieństwo ponownego ich wystąpienia. Stosuje do tego informacje zarówno ze źródeł wewnętrznych, typu dane o występujących cyberatakach we wskazanym okresie czasowym, jak i ze źródeł zewnętrznych, w tym dane od partnerów krajowych lub międzynarodowych. Następnie formułuje wnioski i ewentualne rekomendacje dot. działań prewencyjnych, które zostają udostępnione dla NASK-PIB, a także innym wspólnikom oraz instytucjom odpowiedzialnym za cyberbezpieczeństwo[48].

[46] Ibidem.
[47] Ibidem.
[48] Najwyższa Izba Kontroli, *Działania Państwa w zakresie…* (Nr ewid. 125/2022/P/21/042/KPB), 2022.

5 Krajowy system cyberbezpieczeństwa

Pierwszym aktem prawnym w Polsce regulującym kwestie ochrony w cyberprzestrzeni jest ustawa o krajowym systemie cyberbezpieczeństwa[49]. Zgodnie z nią operatorzy usług kluczowych są zobligowani do wdrożenia programu zarządzania bezpieczeństwem w swoich systemach informacyjnych wykorzystywanych do świadczeń (art. 8 u.k.s.c.). Powiązane jest to z implementacją istotnych procesów, np. systematycznego szacowania ryzyka i dostosowywania do niego środków bezpieczeństwa. Odpowiedzialne za to są m.in. organy regulujące i kontrolujące działania w obszarze technologii informatycznych oraz stosujące, w razie potrzeby, karne sankcje finansowe. Art. 6 ust. 2 u.k.s.c. określa siedem pierwszorzędnych cech dla wskazania progów istotności skutków zakłócających, wyznaczonych przez Radę Ministrów, uzależniających z kolei ocenę wagi konsekwencji nielegalnych incydentów w strefie wirtualnej. Po pierwsze brana jest pod uwagę liczba użytkowników konkretnej kluczowej usługi świadczonej przez dany podmiot – im jest ona większa, tym za bardziej poważne są uznawane skutki. Szacuje się również zależność innych sektorów gospodarki lub społeczeństwa od danego znaczącego świadczenia. W miarę wzmacniania tej relacji ewentualne nieplanowane zdarzenie może wywoływać znacznie silniejsze konsekwencje. W ocenie istotności skutków incydentu cybernetycznego trzecim kluczowym czynnikiem jest analiza wpływu tego zagrożenia na działalność gospodarczą i społeczną oraz bezpieczeństwo publiczne, w tym zagrożenie zniszczenia infrastruktury. Czwartą przesłankę stanowi sam udział podmiotu na rynku. W przypadku, gdy jest on dominującym dostawcą w sektorze, cyberprzestępstwo staje się znaczącym zagrożeniem dla systemu ekonomicznego. Kolejna przesłanka odnosi się do zasięgu obszaru geograficznego, który mógłby zostać dotknięty incydentem. Jeśli usługa kluczowa ma charakter globalny i jej zakłócenie wpłynęłoby na wielu użytkowników na różnych obszarach, negatywne konsekwencje są bardziej znaczące przy ocenie istotności skutku. Szóstym kryterium jest zdolność podmiotu do utrzymywania wystarczającego poziomu świadczenia usługi kluczowej w kontekście niebezpieczeństwa. To oznacza, że podmiot musi być w stanie szybko i efektywnie reagować na nie, celem minimalizacji szkody. Ostatnią i bardziej uogólnioną przesłankę stanowią pozostałe

[49] Ustawa z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa (t.j. Dz. U. z 2023 r. poz. 913), dalej jako: u.k.s.c.

czynniki charakterystyczne dla konkretnego sektora lub podsektora, które należy brać pod uwagę do poprawnej oceny istotności skutków[50].

Na podstawie ww. siedmiu pierwszorzędnych cech Zespół Analiz Strategicznych i Wpływu Nowoczesnych technologii (NASK-PIB) określił hierarchię trzech poziomów bezpieczeństwa, czego celem ma być odpowiednia ocena wpływu incydentu, a tym samym dobór adekwatnych środków ochrony. Za najmniej groźne uznano zdarzenia jedynie naruszające cyberbezpieczeństwo, np. próby uzyskania dostępu do kont użytkowników. Jest to zazwyczaj stadium, w którym incydent nie wymaga nadania klasyfikacji ani raportowania, niemniej może być monitorowany i oceniany pod kątem ewentualnej eskalacji. W momencie, w którym incydent ten jednak wykazuje takowy potencjał, może wymagać bardziej rygorystycznych działań, co z kolei kieruje go ku drugiemu poziomowi w hierarchii określonej przez PIB-NASK. W jego zakresie mieszczą się zdarzenia z obrębu podmiotu publicznego, a także te poważne i istotne. Ich wpływ na sferę wirtualną jest o wiele silniejszy, za czym idzie konieczność nadawania im klasyfikacji oraz raportowania. Dokonuje tego podmiot publiczny lub operator usługi kluczowej w przypadku zagrożenia dla jakości bądź ciągłości realizacji swoich zadań. W przypadku incydentu istotnego, czyli mogącego przyczynić się do pogorszenia jakości świadczeń usług cyfrowych, to dostawca jest zobligowany do nadania mu klasyfikacji i sporządzenia raportu. Ciągły rozwój zagrożenia kwalifikuje zdarzenie jako krytyczne i należące do poziomu trzeciego[51]. Ze względu na tę perspektywę art. 73 u.k.s.c. wskazuje sankcje za zaniechanie bądź niedopełnienie obowiązków wynikających z treści aktu. Operatorzy usług kluczowych lub dostawcy usług cyfrowych mogą zostać obciążeni karą pieniężną w wysokości od 1.000 zł do nawet 1.000.000 zł. Kwalifikacji incydentu jako krytycznego dokonuje właściwy Zespół Reagowania na Incydenty Bezpieczeństwa Komputerowego (*Computer Security Incident Response Team*; CSIRT) poziomu krajowego (art. 26 ust. 3 pkt 6 u.k.s.c.). Art. 2 pkt 1-3 oraz rozdział 6 u.k.s.c. definiują trzy takie grupy, z czego doniosły charakter ma Zespół Reagowania na Incydenty Bezpieczeństwa Komputerowego koordynujący zagrożeniami zgłaszanymi przez podmioty podległe lub nadzorowane głównie przez Ministra Obrony Narodowej (CSIRT MON). Ma zastosowanie głównie w przypadkach związanych z obronnością kraju, co jest powiązane z jego podporządkowaniem Resortowi Obrony Narodowej (RON). Jednocześnie oznacza to

[50] Zespół Analiz Strategicznych i Wpływu Nowoczesnych Technologii Cyber Policy NASK, *Analiza – Ustawa o krajowym systemie cyberbezpieczeństwa*, 2018.
[51] Ibidem.

funkcjonowanie w ramach Dowództwa Komponentu Wojsk Obrony Cyberprzestrzeni (DKWOC)[52], gwarantującego kontrolę nad zasobami z zakresu m.in. budowy systemów IT. CSIRT MON ściśle współdziała z CSIRT GOV, czyli Zespołem Reagowania na Incydenty Bezpieczeństwa Komputerowego podległemu Szefowi Agencji Bezpieczeństwa Wewnętrznego. Adekwatnie reaguje na zdarzenia zgłaszane przez administrację rządową, Narodowy Bank Polski, Bank Gospodarstwa Krajowego oraz operatorów infrastruktury krytycznej. Wobec ataków terrorystycznych największe kompetencje posiada jednak Zespół Reagowania na Incydenty Bezpieczeństwa Komputerowego funkcjonujący w strukturach Państwowego Instytutu Badawczego NASK (CSIRT NASK). W zakresie jego obowiązków mieści się skoordynowane reagowanie na zgłoszenia podmiotów, takich jak operatorzy usług kluczowych lub cyfrowych, a także samorządy terytorialne. Najistotniejszym aspektem jest fakt, iż CSIRT NASK pełni funkcję „ostatniej szansy" dla osób, które doświadczyły skrajnie niekorzystnej sytuacji i nie uzyskały we wcześniejszych krokach oczekiwanej pomocy od innych instytucji[53].

Zgodnie z art. 26 ust. 3 pkt 15 u.k.s.c. każdy z CSIRT poziomu krajowego jest zobligowany do współpracy z organami właściwymi, ministrem właściwym ds. informatyzacji oraz Pełnomocnikiem do Spraw Cyberbezpieczeństwa. W ramach tego są zobowiązani do przekazywania danych na temat incydentów i ryzyka ich wystąpienia podmiotom wchodzącym w skład Krajowego Systemu Cyberbezpieczeństwa (art. 4 u.k.s.c.). Zestawienie dot. problemów, które wystąpiły w ubiegłym roku kalendarzowym, jest kierowane do Pojedynczego Punktu Kontaktowego, prowadzonego przez ministra właściwego ds. informatyzacji. Stanowi ono ważny element umożliwiający współpracę Polski z innymi państwami Unii Europejskiej, a w szczególności z Agencją Unii Europejskiej ds. Cyberbezpieczeństwa (ENISA). Za podtrzymanie kolaboracji międzynarodowych odpowiada również Minister Obrony Narodowej. Kontroluje on współdziałanie Sił Zbrojnych Rzeczypospolitej Polskiej (SZ RP) z właściwymi organami Organizacji Traktatu Północnoatlantyckiego (NATO), UE i innymi organizacjami w zakresie cyberbezpieczeństwa. Jednym ze sprofilowanych części SZ RP są, wspomniane już, Wojska Obrony Cyberprzestrzeni (WOC), które podlegają Dowództwu Komponentu Wojsk Obrony Cyberprzestrzeni. WOC m.in. monitoruje i reaguje na wirtualne zagrożenia, mogące wpływać na stabilność systemów krajowych[54].

[52] Ustawa z dnia 11 marca 2022 r. o obronie Ojczyzny (t.j. Dz. U. z 2024 r. poz. 248 z późn. zm.).
[53] Zespół Analiz Strategicznych i Wpływu Nowoczesnych Technologii Cyber Policy NASK, *Analiza...*, 2018.
[54] Dz. U. z 2024 r. poz. 248 z późn. zm.

Art. 60-63 u.k.s.c. przedstawiają funkcje Pełnomocnika Rządu do Spraw Cyberbezpieczeństwa, który podlega Radzie Ministrów, przy czym powołuje i odwołuje go Prezes Rady Ministrów. W kontekście swojej roli w nadzorowaniu procesu zarządzania ryzykiem krajowego systemu cyberbezpieczeństwa Pełnomocnik uczestniczy w opiniowaniu dokumentów rządowych, a także wydaje rekomendacje wobec stosowania urządzeń informatycznych lub oprogramowania na wniosek CSIRT. Równocześnie wspiera badania naukowe rozwijające technologie, które stanowią jeden z filarów współpracy międzynarodowej w tej dziedzinie. Ponadto art. 64-67 u.k.s.c. wprowadzają instytucję Kolegium do Spraw Cyberbezpieczeństwa z Prezesem Rady Ministrów na czele. Do współtworzących należą ministrowie właściwi ds. wewnętrznych, informatyzacji, spraw zagranicznych, a także Minister Obrony Narodowej, których obecność wynika z potrzeby zapewnienia koordynacji działań pomiędzy konkretnymi sektorami administracji. Kolegium uzupełniają Szef Kancelarii Prezesa Rady Ministrów, Szef Biura Bezpieczeństwa Narodowego oraz minister odpowiedzialny za koordynację działań służb specjalnych. Te osoby stanowią integralną część procesu opracowywania i wdrażania strategii na szczeblu krajowym. W posiedzeniach, oprócz wymienionych osobistości, udział biorą również reprezentanci kluczowych organizacji, np. Dyrektor Rządowego Centrum Bezpieczeństwa, Szef Służby Kontrwywiadu Wojskowego, a także Dyrektor NASK-PIB. Kolegium odpowiada za strategiczne kierunki działań związanych z cyberbezpieczeństwem na poziomie krajowym, a dodatkowo stanowi instytucję opiniodawczo-doradczą względem Rady Ministrów[55].

6 Cyberatak a stan nadzwyczajny

W obliczu narastających zagrożeń informatycznych pod koniec 2022 r. Polska sformułowała swoje stanowisko dot. zastosowania prawa międzynarodowego w cyberprzestrzeni[56]. Podkreśliła wagę Karty Narodów Zjednoczonych[57], zgadzając się m.in. z Unią Europejską, NATO i wieloma państwami spoza tych organizacji co do jej zastosowania w tej sferze. III RP stwierdziła również, że operacje przeprowadzane w wirtualnej rzeczywistości,

[55] Ministerstwo Cyfryzacji, Powstaje Kolegium ds. Cyberbezpieczeństwa, https://www.gov.pl/web/cyfryzacja/powstaje-kolegium-ds-cyberbezpieczenstwa (dostęp: 15.06.2025 r.)
[56] Ministerstwo Spraw Zagranicznych, *Stanowisko Rzeczypospolitej Polskiej dotyczące zastosowania prawa międzynarodowego w cyberprzestrzeni*, https://www.gov.pl/web/dyplomacja/stanowisko-rzeczypospolitej-polskiej-dotyczace-zastosowania-prawa-miedzynarodowego-w-cyberprzestrzeni (dostęp: 15.06.2025 r.)
[57] Karta Narodów Zjednoczonych Statut Międzynarodowego Trybunału Sprawiedliwości i Porozumienie ustanawiające Komisję Przygotowawczą Narodów Zjednoczonych (Dz.U. z 1947 r. Nr 23, poz. 90 z późn. zm.), dalej jako: KNZ.

generujące negatywne konsekwencje na terytorium kraju, bez rozróżnienia na działania kinetyczne bądź cyfrowe, są uznawane za naruszenie zasady suwerenności państwa, np. modyfikacja wyników głosowania parlamentarnego lub przeprowadzenie kampanii dezinformacyjnej.

W przypadku znacznie bardziej zaawansowanych skutków niedozwolona interwencja może zostać uznana nawet za naruszającą zakaz użycia siły na podstawie art. 2 ust. 4 KNZ. Jeżeli dojdzie do śmierci bądź okaleczenia ludzi, a także do uszkodzenia lub zniszczenia mienia znacznej wartości, Polska ma prawo uznać to za napaść zbrojną będącą bezpośrednią przesłanką do wprowadzenia na terenie kraju stanu nadzwyczajnego, a konkretniej stanu wojny przez Sejm Rzeczypospolitej bądź, gdy ten nie jest w stanie się zebrać, Prezydenta Rzeczypospolitej (art. 116 Konstytucji RP[58]). Zgodnie z art. 51 KNZ, zanim Rada Bezpieczeństwa nie podejmie niezbędnych zarządzeń, celem utrzymania międzynarodowego pokoju, Rzeczpospolita ma prawo do samoobrony. Pozwala to na elastyczne wykorzystanie środków odnoszących się zarówno do działań cybernetycznych, jak i tradycyjnie rozumianych sił zbrojnych, o ile są one proporcjonalne do zaistniałej sytuacji. Z kolei, gdy cyberataki stanowią istotne zagrożenie dla dobrobytu i osiągają poziom, który można uznać za napaść zbrojną, przewiduje się możliwość wykorzystania prawa do samoobrony zbiorowej. NATO potwierdziło, iż ochrona przed atakami cybernetycznymi stanowi integralną część funkcjonowania w tym obszarze[59].

Zgodnie z art. 229 Konstytucji RP w przypadku, gdy nie doszło jeszcze do faktycznego cyberataku, a istnieje jedynie zagrożenie jego wystąpienia, Prezydent Rzeczypospolitej na wniosek Rady Ministrów może wprowadzić stan wojenny na części albo na całym terytorium państwa. Założeniem tego jest zabezpieczenie interesów przede wszystkim wewnętrznych. W tym okresie często występuje ograniczenie praw i wolności obywatela, jednak jest to w pełni kierowane i uzasadnione potrzebą zapewnienia ochrony. Innym stanem nadzwyczajnym jest stan wyjątkowy (art. 230 Konstytucji RP) wprowadzany przez Prezydenta RP na wniosek Rady Ministrów na całym terenie kraju lub jego części. Może on trwać nie więcej niż 90 dni, a w uzasadnionych przypadkach zostać jednorazowo przedłużony. Taki krok wymaga akceptacji Sejmu i może obowiązywać do 60 dni. Stan wyjątkowy występuje w sytuacjach rzeczywistego zagrożenia, gdy zwykłe środki konstytucyjne nie są wystarczające do zachowania bezpieczeństwa i porządku publicznego. Z kolei stan klęski

[58] Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r. (Dz.U. Nr 78, poz. 483 z późn. zm.), dalej jako: Konstytucja RP.
[59] Ministerstwo Spraw Zagranicznych, 2022, *Stanowisko Rzeczypospolitej…*

żywiołowej (art. 232 Konstytucji RP) różni się od pozostałych rodzajów, ponieważ jego obowiązywanie nie jest w pełni zależne od ludzkiej woli. Jest to powiązane z wystąpieniem katastrofy naturalnej lub awarii technicznej, której skutki na szeroką skalę zagrażają zdrowiu i życiu ludzi, mieniu bądź środowisku. Do konsekwencji należą m.in. uszkodzenia urządzeń technicznych w wyniku zaawansowanego cyberataku z użyciem AI. W skrajnych przypadkach może wpłynąć to na całe systemy, takich jak kontrola lotów, ratownictwa lub dostarczania wody. Stan klęski żywiołowej wprowadza na terenie całej Polski lub jej części Rada Ministrów na czas nie dłuższy niż 30 dni. Ewentualne wydłużenie stanowi okres do 30 dni za akceptacją Sejmu.

Podsumowanie

AI stanowi narzędzie o dwojakim zastosowaniu w kontekście cyberbezpieczeństwa. Z jednej strony usprawnia działania cyberprzestępców, z drugiej zaś wspiera mechanizmy obronne. Jej rozwój nie tylko znacząco wspomaga ochronę systemów informatycznych, ale też przyczynia się do tworzenia nowych i intensyfikacji obecnych zagrożeń. W konsekwencji rozstrzygnięcie, czy AI stanowi instrument o przeznaczeniu pozytywnym, czy też negatywnym nie jest jednoznaczne.

Analizując wszystkie korzyści, jak i zagrożenia należy stwierdzić, iż skuteczność AI jako narzędzia cyberbezpieczeństwa zależy od kilku kluczowych czynników. Przede wszystkim niezbędne jest odpowiednie zarządzanie ryzykiem oraz opracowanie regulacji prawnych, które ograniczą możliwość wykorzystania jej w celach przestępczych. Instytucje państwowe, sektor prywatny oraz organizacje międzynarodowe powinny podejmować współpracę w zakresie rozwoju systemów zabezpieczeń opartych na nowych technologiach, a także wypracować standardy etycznego jej wykorzystania. Technologie nie powinny być narzędziem wyłącznie reaktywnym – jej potencjał musi zostać ukierunkowany również na prognozowanie zagrożeń oraz projektowanie innowacyjnych mechanizmów zabezpieczających, które będą w stanie przewidywać i neutralizować ataki jeszcze przed ich wystąpieniem. Istotnym aspektem jest również edukacja użytkowników na temat cyfrowej higieny, celem zwiększenia ogólnej odporności na cyberataki. Podnoszenie świadomości społecznej o metodach stosowanych przez cyberprzestępców, w tym zaawansowanych technik wykorzystujących AI, jest niezbędne do skutecznej obrony przed nimi. Inwestowanie w programy

edukacyjne, które promują odpowiedzialne korzystanie z technologii i zrozumienie zagrożeń cybernetycznych jest równie ważne, jak rozwijanie samego sektora informatycznego.

Podsumowując, AI zarówno wzmacnia ochronę przed cyberatakami, jak i przyczynia się do ich intensyfikacji. W związku z tym kluczowe staje się opracowanie strategii wykorzystania jej w sposób odpowiedzialny i kontrolowany, tak aby zmaksymalizować korzyści przy jednoczesnym ograniczeniu ryzyka jej nadużycia przez cyberprzestępców. Tylko holistyczne podejście, łączące nowoczesne technologie z edukacją i regulacjami prawnymi, może stanowić skuteczną odpowiedź na rosnące zagrożenia. W przyszłości rola AI w cyberbezpieczeństwie będzie determinowana nie tylko przez postęp technologiczny, lecz także przez zdolność instytucji do skutecznego zarządzania jej potencjałem oraz wdrażania regulacji ograniczających ryzyka.

STRESZCZENIE

Sztuczna inteligencja w kontekście cyberbezpieczeństwa

W niniejszym artykule poruszono tematykę cyberzagrożeń, które niesie ze sobą rewolucja cyfryzacji związana z rozwojem sztucznej inteligencji na przestrzeni kilku ubiegłych lat. Przedstawiono przykłady technik cyberataków, a także możliwe rozwiązanie oparte na AI. W tekście analizie poddano również stosunek prawny i polityczny wobec problemu dynamicznie zmieniających się standardów technologii wpływających na bezpieczeństwo obywatela.

Słowa kluczowe: CERT; CSIRT; cyberatak; cyberbezpieczeństwo; cyberprzestępstwo; cyberprzestrzeń; cyfryzacja; Krajowa Strategia Cyberbezpieczeństwa; polityka; prawo; sztuczna inteligencja.

SUMMARY

Artificial intelligence in the context of cybersecurity

This article discusses the cyber threats posed by the digitalisation revolution associated with the development of artificial intelligence in recent years. Examples of cyber-attack techniques are presented, as well as a potential AI-based solution. The text also analyses the legal and political attitude towards the problem of dynamically changing technological standards affecting the security of citizens.

Keywords: artificial intelligence, CERT, CSIRT, cyberattack, cybercrime, cybersecurity, cyberspace, digitalization, law, National Cybersecurity Strategy, politics.

BIBLIOGRAFIA

Wykaz aktów normatywnych

Rezolucja ustawodawcza Parlamentu Europejskiego z dnia 13 marca 2024 r. w sprawie wniosku dotyczącego rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

Źródła internetowe

„Cyber 360", *Cyberataki w sektorze MŚP z udziałem sztucznej inteligencji*, https://cyber360.pl/blog-o-cyberbezpieczenstwie/cyberataki-w-sektorze-msp-z-udzialem-sztucznej-inteligencji/ (dostęp: 15.06.2025 r.)

Ajibade A.I., *Typowe oszustwa internetowe wymierzone w nastolatków*, https://www.internetmatters.org/pl/hub/expert-opinion/common-online-scams-targeting-teenagers/ (dostęp: 15.06.2025 r.)

Darktrace Antigena, *Darktrace Antigena Product Overview*, https://darktrace.com/legal/darktrace-antigena-product-overview (dostęp: 15.06.2025 r.)

European Commission, *2023 Report on the state of the Digital Decade*, 2023.

Finnish Transport and Communications Agency Traficom, *The security threat of AI-enabled cyberattacks*, 2022.

Ministerstwo Cyfryzacji, *Powstaje Kolegium ds. Cyberbezpieczeństwa*, https://www.gov.pl/web/cyfryzacja/powstaje-kolegium-ds-cyberbezpieczenstwa (dostęp: 15.06.2025 r.)

Ministerstwo Spraw Zagranicznych, *Stanowisko Rzeczypospolitej Polskiej dotyczące zastosowania prawa międzynarodowego w cyberprzestrzeni*, https://www.gov.pl/web/dyplomacja/stanowisko-rzeczypospolitej-polskiej-dotyczace-zastosowania-prawa-miedzynarodowego-w-cyberprzestrzeni (dostęp: 15.06.2025 r.)

Najwyższa Izba Kontroli, *Działania Państwa w zakresie zapobiegania i zwalczania skutków wybranych przestępstw internetowych, w tym kradzieży tożsamości* (Nr ewid. 125/2022/P/21/042/KPB), 2022.

Polska Agencja Prasowa, *Eurostat: polskie firmy w unijnej czołówce pod względem zgłaszanych naruszeń cyberbezpieczeństwa*, https://www.pap.pl/aktualnosci/eurostat-polskie-firmy-w-unijnej-czolowce-pod-wzgledem-zglaszanych-naruszen-0 (dostęp: 15.06.2025 r.)

Portal polskiej Policji, *Internetowi oszuści nie śpią!*, https://policja.pl/pol/aktualnosci/227892,Internetowi-oszusci-nie-spia.html (dostęp: 15.06.2025 r.)

S. Palczewski, *Cyberataki niemożliwe do wykrycia? Rewolucja AI*, https://cyberdefence24.pl/polityka-i-prawo/sztuczna-inteligencja-w-cyberatakachincydenty-niemozliwe-do-wykrycia (dostęp: 15.06.2025 r.)

Zespół Analiz Strategicznych i Wpływu Nowoczesnych Technologii Cyber Policy NASK, *Analiza – Ustawa o krajowym systemie cyberbezpieczeństwa*, 2018.