Annales Universitatis Paedagogicae Cracoviensis 417

Studia de Securitate 15(1) ▪ 2025
ISSN 2082-0917
DOI 10.24917/20820917.15.1.5

Sylwester Kurowski
Wyższa Szkoła Pedagogiczno-Techniczna w Koninie
ORCID: 0000-0003-3278-8878

Sztuczna inteligencja – szanse czy zagrożenia dla bezpieczeństwa?

Wstęp

Rozwój sztucznej inteligencji (SI) stał się jednym z kluczowych tematów współczesnej nauki i technologii. Postęp w tej dziedzinie przynosi rewolucyjne zmiany w wielu sektorach, w tym w gospodarce, medycynie, przemyśle i bezpieczeństwie. Wraz z dynamicznym wzrostem możliwości technologicznych pojawiają się jednak pytania o wpływ SI na bezpieczeństwo – zarówno w wymiarze globalnym, jak i jednostkowym. Czy sztuczna inteligencja stanowi szansę na poprawę jakości życia i bezpieczeństwa, czy też niesie ze sobą ryzyko niekontrolowanych skutków?

Celem niniejszego referatu jest wieloaspektowa analiza roli sztucznej inteligencji w kontekście bezpieczeństwa. W pierwszej kolejności omówiono zastosowania SI jako narzędzia wspomagającego działania na rzecz ochrony. Następnie przeanalizowano zagrożenia wynikające z potencjalnego nadużycia technologii, w tym przez podmioty o złych intencjach. Szczególną uwagę poświęcono zagadnieniom etycznym oraz problemom związanym z regulacjami prawnymi.

Metody

Niniejsza praca zredagowana została z wykorzystaniem teoretycznych metod badawczych, takich jak: analiza, synteza oraz wnioskowanie dedukcyjne.

Metoda analizy miała swoje zastosowanie głównie podczas studiowania literatury przedmiotu oraz dokumentów normatywnych, a także przy wykorzystaniu analizy studiów przypadków. Jej zastosowanie umożliwiło poznanie wszystkich zjawisk obejmujących podjętą problematykę badawczą, Przegląd literatury umożliwił identyfikację oddzielnych zagadnień, takich jak zdolność SI do autonomicznego podejmowania decyzji, jej podatność na manipulacje, a także potencjalne korzyści związane z wykrywaniem zagrożeń czy automatyzacją działań obronnych. Każde z tych zagadnień zostało przeanalizowane osobno w kontekście dostępnych źródeł i dokumentów normatywnych. **Analiza studiów przypadków** została wykorzystana w celu zilustrowania rzeczywistych lub hipotetycznych sytuacji, w których sztuczna inteligencja odegrała kluczową rolę w kontekście bezpieczeństwa. Analiza studiów przypadków umożliwiła pogłębienie rozumienia mechanizmów działania SI w praktyce oraz ukazanie potencjalnych skutków jej niewłaściwego zastosowania.

Synteza umożliwiła połączenie wyników analizy w spójną całość, w której przedstawiono ujęcie porównawcze – wskazując z jednej strony potencjalne szanse, takie jak zwiększenie efektywności systemów bezpieczeństwa, a z drugiej strony ryzyko, w tym możliwość eskalacji zagrożeń wynikających z autonomicznego działania systemów SI. Na tym etapie wykorzystano podejście interdyscyplinarne, bowiem zestawiono wnioski z różnych dziedzin – prawa, technologii i etyki, aby ukazać złożoność problemu.

Wnioskowanie dedukcyjne zostało zastosowane w celu sformułowania logicznych wniosków na podstawie ogólnych założeń i stanowisk przedstawionych w literaturze. Przykładowo, wychodząc od ogólnego założenia, że SI może działać w sposób autonomiczny i nieprzewidywalny, wyprowadzono wniosek, że jej niewłaściwe wykorzystanie może prowadzić do ogromnych strat i powstania szeregu zagrożeń, co powoduje konieczność kontroli nad jakością danych pochodzących ze źródeł sztucznej inteligencji.

Zastosowanie powyższych metod badawczych pozwoliło na usystematyzowaną i logiczną ocenę zagadnienia, mimo braku badań empirycznych. Dzięki temu możliwe było zidentyfikowanie głównych osi problemowych i sformułowanie wniosków o charakterze koncepcyjnym i normatywnym.

Zastosowania sztucznej inteligencji w obszarze bezpieczeństwa

Sztuczna inteligencja znajduje szerokie zastosowanie w systemach bezpieczeństwa. Przykłady obejmują automatyzację monitoringu, predykcję zagrożeń oraz wspieranie procesów decyzyjnych. Algorytmy uczenia maszynowego są wykorzystywane do analizy dużych zbiorów danych, co umożliwia szybkie wykrywanie anomalii w systemach informatycznych oraz zapobieganie cyberatakom.

Rozwój sztucznej inteligencji (SI) w kontekście bezpieczeństwa zmienia sposób, w jaki instytucje państwowe, sektor prywatny oraz organizacje międzynarodowe podchodzą do identyfikacji zagrożeń i zarządzania kryzysowego. **SI stała się narzędziem o strategicznym znaczeniu, umożliwiającym analizę olbrzymich zbiorów danych, wykrywanie anomalii, a także predykcję zdarzeń potencjalnie zagrażających stabilności społecznej i państwowej.**

Do kluczowych obszarów zastosowania sztucznej inteligencji zalicza się:

- **Automatyzację monitoringu i analizę danych w czasie rzeczywistym**, szczególnie w systemach nadzoru miejskiego, infrastrukturze krytycznej oraz obszarze wojskowym. Algorytmy oparte na głębokim uczeniu (deep learning) są wykorzystywane do analizy obrazów i nagrań wideo, co pozwala na szybkie wykrywanie podejrzanych zachowań, nieuprawnionych wtargnięć czy zagrożeń terrorystycznych. **Zaawansowane systemy nadzoru, oparte na sieciach neuronowych, mogą rozpoznawać twarze[1], identyfikować osoby poszukiwane przez organy ścigania oraz monitorować ruch w strategicznych miejscach publicznych.** Przykładem tego rodzaju technologii są systemy rozpoznawania twarzy, stosowane m.in. w Chinach, gdzie służby porządkowe wykorzystują zaawansowane algorytmy do identyfikacji i śledzenia podejrzanych osób w czasie rzeczywistym. Zastosowanie SI w monitoringu wiąże się jednak z istotnymi dylematami etycznymi, takimi jak ochrona prywatności i możliwość nadużyć ze strony państw autorytarnych.
- **Predykcję zagrożeń i analizę cyberbezpieczeństwa** – sztuczna inteligencja znajduje szerokie zastosowanie w dziedzinie **cyberbezpieczeństwa**, szczególnie w zakresie wykrywania i neutralizowania cyberataków. Algorytmy uczenia maszynowego analizują ogromne ilości danych sieciowych, identyfikując wzorce związane z potencjalnie niebezpiecznymi działaniami, takimi jak phishing, ataki DDoS czy

[1] J. Wołoszyn (2024). *Ewolucja i wpływ sztucznej inteligencji na zaawansowane strategie obronne w cyberbezpieczeństwie*. „Dydaktyka Informatyki", 19, s. 218–22.

malware. **Zastosowanie SI w cyberbezpieczeństwie oznacza rewolucję w podejściu do ochrony systemów informatycznych, ponieważ pozwala na reakcję w czasie rzeczywistym oraz adaptację do nowych metod ataków.** Współczesne systemy cyberbezpieczeństwa oparte na SI, takie jak Darktrace czy Deep Instinct, wykorzystują modele behawioralne do wykrywania anomalii w sieci. **Zamiast opierać się wyłącznie na znanych sygnaturach wirusów czy ataków, sztuczna inteligencja potrafi przewidywać nowe zagrożenia, identyfikując niepokojące wzorce zachowań w sieciach korporacyjnych czy instytucjonalnych.** Takie podejście zwiększa skuteczność ochrony, zwłaszcza w obliczu coraz bardziej zaawansowanych technik stosowanych przez cyberprzestępców.

- **Systemy decyzyjne** – SI odgrywa kluczową rolę w **procesach decyzyjnych** w sytuacjach kryzysowych. Przykładem może być analiza danych związanych z klęskami żywiołowymi, atakami terrorystycznymi czy epidemiami. Systemy SI przetwarzają dane z różnych źródeł (m.in. satelitarnych, medialnych, sensorycznych), co pozwala na szybkie modelowanie skutków zdarzeń i prognozowanie ich dalszego przebiegu. **W ten sposób wspierane są decyzje podejmowane przez służby ratunkowe, wojsko czy administrację państwową.** Jednym z najbardziej znanych zastosowań SI w zarządzaniu kryzysowym było wykorzystanie algorytmów do modelowania rozprzestrzeniania się pandemii COVID-19. Systemy oparte na uczeniu maszynowym, m.in. opracowane przez Google DeepMind czy Instytut Zdrowia Publicznego w Johns Hopkins University, analizowały dynamikę infekcji i prognozowały jej rozwój. **Podobne technologie mogą być stosowane w prognozowaniu skutków ataków terrorystycznych, katastrof ekologicznych czy zagrożeń energetycznych.**
- **Systemy wojskowe i autonomiczne** – sztuczna inteligencja coraz częściej znajduje zastosowanie w **systemach militarnych**, zarówno w obszarze wywiadu i analizy danych, jak i w autonomicznych systemach bojowych. Współczesne siły zbrojne korzystają z algorytmów SI do przetwarzania danych zwiadowczych, analizowania zagrożeń oraz sterowania systemami uzbrojenia. **Bezzałogowe statki powietrzne (drony), inteligentne systemy obrony przeciwrakietowej czy autonomiczne roboty bojowe to przykłady nowoczesnych rozwiązań opartych na SI.** Nie można jednak pomijać ryzyka związanego z wykorzystaniem SI w wojsku. **Problemem pozostaje brak pełnej kontroli nad decyzjami podejmowanymi przez autonomiczne systemy, co rodzi poważne pytania natury etycznej i prawnej.** Istnieje realne zagrożenie, że broń oparta na sztucznej inteligencji

może wymknąć się spod kontroli lub zostać użyta w sposób niezgodny z międzynarodowym prawem humanitarnym[2].

Przykłady wdrożeń SI w cyberbezpieczeństwie

Sztuczna inteligencja odgrywa coraz większą rolę w zapewnianiu bezpieczeństwa w cyberprzestrzeni, wspierając zarówno działania obronne, jak i ofensywne. Współczesne systemy zabezpieczeń wykorzystują algorytmy uczenia maszynowego do analizy wzorców ruchu sieciowego, identyfikacji anomalii oraz wykrywania i neutralizowania zagrożeń w czasie rzeczywistym.

Jednym z najbardziej zaawansowanych przykładów zastosowania SI w cyberbezpieczeństwie jest system **Darktrace**[3], który wykorzystuje technologię tzw. immunologii cyfrowej. Na wzór ludzkiego układu odpornościowego system ten uczy się normalnych wzorców funkcjonowania sieci i na ich podstawie wykrywa odstępstwa mogące świadczyć o naruszeniach bezpieczeństwa. W praktyce oznacza to, że Darktrace nie polega wyłącznie na wcześniej zdefiniowanych sygnaturach zagrożeń, lecz dynamicznie identyfikuje nietypowe zachowania w sieci organizacji. Przykładem skutecznego wdrożenia Darktrace jest incydent, który miał miejsce w jednej z organizacji finansowej[4]. System wykrył nietypowe transfery danych do serwera w Azji Południowo-Wschodniej, co mogło sugerować próbę wycieku poufnych informacji. Dzięki wczesnemu wykryciu zagrożenia możliwe było szybkie podjęcie działań zaradczych, które zapobiegły poważnym konsekwencjom finansowym i reputacyjnym dla organizacji. Wykorzystanie SI w tym przypadku pozwoliło na skuteczną identyfikację zagrożenia, zanim mogło ono przybrać poważniejszą formę.

W dziedzinie ochrony przed phishingiem oraz analizie zagrożeń coraz częściej stosuje się **sieci neuronowe**, które analizują treść wiadomości e-mail, strukturę linków oraz metadane, aby wykryć próby wyłudzenia danych. Przykładowo, systemy wykrywające phishing wykorzystują głębokie sieci neuronowe do analizy stylu pisania wiadomości oraz porównania ich z legalną komunikacją, co pozwala na wychwycenie subtelnych różnic wskazujących na oszustwo. Takie rozwiązania, stosowane w Google Safe Browsing

[2] M. Tołwiński (2024). *Etyczne dylematy odpowiedzialności za działanie systemów sztucznej inteligencji na wojnie*. „Doctrina. Studia społeczno-Polityczne", 21(21).

[3] I. Partners (2022, 16 listopada). *Darktrace: system, który uczy się użytkowników*. CRN Polska. https://crn.pl/artykuly/darktrace-system-ktory-uczy-sie-uzytkownikow/ [dostęp: 25.03.2025].

[4] E. Foulger (2022, 24 sierpnia). *Detecting Unknown Ransomware: A Darktrace Case Study*. Darktrace. https://www.darktrace.com/blog/detecting-the-unknown-revealing-uncategorised-ransomware-using-darktrace [dostęp: 25.03.2025].

czy Microsoft Defender SmartScreen, umożliwiają automatyczne blokowanie podejrzanych stron internetowych, zanim użytkownik podejmie ryzykowne działania. Sieci neuronowe można również wykorzystać do analizy aktywności użytkowników w systemach informatycznych, identyfikując nietypowe wzorce zachowań wskazujące na potencjalne naruszenia bezpieczeństwa.

Sztuczna inteligencja wspiera również działania w zakresie **threat huntingu**, czyli aktywnego poszukiwania zagrożeń w infrastrukturze IT organizacji. Tradycyjne systemy zabezpieczeń często polegają na pasywnym wykrywaniu incydentów na podstawie zdefiniowanych reguł i sygnatur. Threat hunting natomiast to proaktywne podejście, w którym zaawansowane algorytmy analizują zebrane logi, monitorują aktywność użytkowników i identyfikują ukryte zagrożenia, które mogłyby umknąć tradycyjnym metodom detekcji.

Jednym z czołowych narzędzi w dziedzinie threat huntingu jest **CrowdStrike Falcon**, oparty na chmurowym modelu detekcji i reakcji na zagrożenia (EDR – Endpoint Detection and Response). System ten wykorzystuje sztuczną inteligencję do analizy miliardów zdarzeń dziennie, identyfikując nawet najbardziej subtelne oznaki naruszenia bezpieczeństwa. W odróżnieniu od tradycyjnych antywirusów CrowdStrike Falcon nie tylko wykrywa malware, ale również analizuje zachowania systemowe, umożliwiając wykrycie zaawansowanych ataków typu APT (Advanced Persistent Threat). CrowdStrike Falcon znajduje zastosowanie w sektorach o wysokim stopniu ryzyka, takich jak obronność, finanse czy infrastruktura krytyczna. Dzięki wykorzystaniu AI jest w stanie przewidywać potencjalne wektory ataku i dostarczać rekomendacje dotyczące działań prewencyjnych.

Innym zaawansowanym narzędziem opartym na sztucznej inteligencji jest **FireEye Helix**, które łączy funkcjonalność SIEM (Security Information and Event Management) z zaawansowaną analizą zagrożeń. FireEye Helix umożliwia centralizację danych pochodzących z różnych źródeł, automatyczną korelację zdarzeń oraz identyfikację anomalii w czasie rzeczywistym. System ten wykorzystuje algorytmy uczenia maszynowego do przewidywania i zapobiegania atakom, co jest szczególnie istotne w dynamicznie zmieniającym się krajobrazie cyberzagrożeń. FireEye Helix pozwala na integrację z innymi narzędziami cyberbezpieczeństwa, co czyni go wszechstronnym rozwiązaniem zarówno dla dużych korporacji, jak i instytucji rządowych.

Jednym z przykładów wdrożenia FireEye Helix była analiza zaawansowanego ataku APT, który miał miejsce w dużej instytucji finansowej. System wykrył nietypową aktywność w sieci, wskazując na możliwość wykorzystania exploitów do dostania się do wewnętrznych systemów. Po zidentyfikowaniu zagrożenia FireEye Helix automatycznie uruchomił procedurę zabezpieczającą, blokując dostęp do krytycznych zasobów i ograniczając rozprzestrzenianie

się ataku. Dzięki szybkości reakcji i zastosowaniu sztucznej inteligencji, firma uniknęła poważnych strat związanych z incydentem.

Siła systemów opartych na sztucznej inteligencji w cyberbezpieczeństwie polega na ich zdolności do wykrywania i reagowania na zagrożenia w czasie rzeczywistym, a także w wykorzystywaniu algorytmów uczenia maszynowego[5] do rozpoznawania wzorców zagrożeń, które mogą umknąć tradycyjnym metodom detekcji. Dzięki zaawansowanym technologiom SI zespoły bezpieczeństwa mogą prowadzić aktywne poszukiwanie zagrożeń (threat hunting)[6], identyfikując potencjalne ataki jeszcze przed ich realizacją.

Wykorzystanie SI w ramach threat huntingu pozwala na monitorowanie sieci w sposób proaktywny, a nie tylko reaktywny, co znacząco podnosi skuteczność w identyfikacji zagrożeń. Przykładem jest wdrożenie systemu SI w organizacji zajmującej się bankowością online, gdzie system wykrył subtelne zmiany w wzorcach logowań do systemu. Dzięki algorytmom uczenia maszynowego udało się przewidzieć możliwość wykorzystania narzędzi phishingowych do przejęcia konta klienta. Działania prewencyjne podjęte w odpowiednim czasie zapobiegły wyciekowi poufnych danych.

Podsumowując dotychczasowe rozważania, podkreślić należy, że sztuczna inteligencja nie tylko znacząco poprawia skuteczność działań obronnych, ale również umożliwia szybszą i bardziej precyzyjną reakcję na zagrożenia. Systemy takie jak **Darktrace**, **CrowdStrike Falcon** czy **FireEye Helix** demonstrują potencjał SI w identyfikacji i neutralizowaniu cyberzagrożeń. Jednakże rosnąca rola sztucznej inteligencji w cyberbezpieczeństwie rodzi również nowe wyzwania, takie jak ryzyko wykorzystania tych technologii przez cyberprzestępców. Wymaga to dalszego rozwoju zarówno technicznych środków ochrony, jak i odpowiednich regulacji prawnych, aby zapewnić równowagę między innowacją a bezpieczeństwem.

Zagrożenia związane z nadużyciem sztucznej inteligencji

Rozwój sztucznej inteligencji przynosi liczne korzyści, jednak niesie również poważne zagrożenia, zwłaszcza gdy technologie te trafiają w ręce podmiotów działających na szkodę innych. Systemy oparte na SI mogą być

[5] H. Ostap (2024). *BotTROP – system wykrywania botów w sieci korporacyjnej*. W: *Badania naukowe i technologie 2021-2024*. Wydział Cybernetyki Instytut Systemów Informatycznych. Wojskowa Akademia Techniczna im. Jarosława Dąbrowskiego w Warszawie. Warszawa, s. 72.

[6] K. Gapiński (2017, 13 grudnia). *Wykorzystanie modeli Cyber Threat Intelligence jako element skutecznego reagowania na incydenty komputerowe*. Fundacja Bezpieczna Przestrzeń. https://www.cybsecurity.org/pl/wykorzystanie-modeli-cyber-threat-intelligence-jako-element-skutecznego-reagowania-na-incydenty-komputerowe/ [dostęp: 24.03.2025].

wykorzystywane do manipulacji informacjami, prowadzenia działań dezinformacyjnych, a także przeprowadzania cyberataków i autonomicznych operacji militarnych. Nadużycie sztucznej inteligencji stanowi rosnące wyzwanie zarówno dla sektora cyberbezpieczeństwa, jak i dla instytucji odpowiedzialnych za regulacje prawne oraz etyczne.

Sztuczna inteligencja jest wykorzystywana nie tylko do ochrony systemów, ale także przez cyberprzestępców do przeprowadzania bardziej wyrafinowanych ataków. Algorytmy uczenia maszynowego mogą analizować luki w zabezpieczeniach i automatycznie dostosowywać metody ataków do konkretnych systemów. Przykładem takiego zagrożenia są zaawansowane phishingowe kampanie prowadzone przy użyciu SI. W jednym z przypadków oszuści wykorzystali algorytmy do analizy stylu komunikacji pracowników korporacji, co pozwoliło im stworzyć realistyczne e-maile wyłudzające dane logowania. Takie ataki, określane mianem **AI-powered phishing**, są znacznie trudniejsze do wykrycia i stanowią realne zagrożenie dla firm oraz instytucji rządowych (MIT Technology Review).

Ponadto, SI może być używana do przeprowadzania **automatycznych ataków DDoS** na infrastrukturę krytyczną, takich jak sieci energetyczne czy systemy bankowe. Zaawansowane botnety wykorzystujące uczenie maszynowe mogą optymalizować swoje działanie, unikając tradycyjnych mechanizmów obronnych.

SI znajduje coraz szersze zastosowanie w wojsku, co rodzi liczne kontrowersje etyczne i prawne. Systemy autonomiczne, takie jak drony bojowe czy zautomatyzowane systemy obronne, mogą podejmować decyzje bez bezpośredniej ingerencji człowieka. Przykładem jest izraelski system **Harpy**, który potrafi samodzielnie identyfikować i atakować cele na podstawie analizy sygnałów radarowych. Podobnie, w 2020 roku w Libii zarejestrowano przypadek użycia autonomicznych dronów bojowych Kargu-2, które mogły prowadzić ataki bez udziału operatora.

Wykorzystanie autonomicznych systemów bojowych rodzi pytania o odpowiedzialność prawną za skutki ich działań. Brak międzynarodowych regulacji dotyczących broni autonomicznej sprawia, że kraje intensywnie rozwijające te technologie nie muszą podlegać globalnym ograniczeniom.

Sztuczna inteligencja może również prowadzić do niezamierzonych skutków, takich jak **algorytmiczna dyskryminacja**. Modele uczenia maszynowego są trenowane na historycznych danych, które mogą zawierać nieświadome uprzedzenia. Przykładem takiego zagrożenia jest przypadek systemów rekrutacyjnych opartych na SI. W 2018 roku Amazon[7] musiał

[7] J. Dastin (2018, 11 października). *Insight - Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/ [dostęp: 28.07.2025].

zrezygnować z własnego narzędzia do analizy CV, ponieważ algorytm wykazywał skłonność do faworyzowania mężczyzn i odrzucania kandydatur kobiet. Podobne problemy mogą wystąpić w systemach analizy kredytowej, gdzie algorytmy decydują o zdolności kredytowej klientów. W 2019 roku ujawniono, że algorytmy Apple Card przyznawały niższe limity kredytowe kobietom w porównaniu do mężczyzn o tych samych dochodach, co wywołało kontrowersje i interwencję regulatorów.

Choć sztuczna inteligencja przynosi wiele korzyści, jej niewłaściwe wykorzystanie może prowadzić do poważnych zagrożeń. Od deepfake i dezinformacji, przez automatyzację cyberprzestępczości, aż po broń autonomiczną i algorytmiczną dyskryminację – każde z tych wyzwań wymaga skutecznych regulacji oraz działań prewencyjnych. Rozwój SI musi iść w parze z odpowiedzialnością i etyką, aby minimalizować ryzyko jej nadużycia.

Deepfake jako narzędzie manipulacji

Technologia deepfake[8] umożliwia tworzenie realistycznych, lecz fałszywych materiałów wideo, które mogą być wykorzystywane do dyskredytacji osób publicznych, szerzenia dezinformacji oraz wpływania na procesy demokratyczne.

Deepfake to technologia wykorzystująca **generatywne sieci przeciwstawne (GAN, Generative Adversarial Networks)**, które pozwalają na realistyczne generowanie lub modyfikowanie obrazów, wideo i dźwięku. Działanie GAN opiera się na rywalizacji dwóch sieci neuronowych – jedna (generator) tworzy fałszywe materiały, a druga (dyskryminator) ocenia ich autentyczność. W miarę treningu algorytmy stają się coraz lepsze w generowaniu hiperrealistycznych treści, które trudno odróżnić od prawdziwych. Technologia ta początkowo była rozwijana w celach badawczych i artystycznych, jednak szybko została zaadaptowana do celów przestępczych, dezinformacyjnych oraz politycznych. Deepfake może służyć zarówno do rozpowszechniania fałszywych informacji, jak i do podszywania się pod inne osoby, co niesie ogromne zagrożenia dla bezpieczeństwa cyfrowego.

Deepfake stał się jednym z kluczowych zagrożeń dla procesów demokratycznych, szczególnie w kontekście wyborów. Fałszywe nagrania przedstawiające polityków lub liderów opinii mogą zostać wykorzystane do szerzenia dezinformacji, wpływania na wyniki głosowań oraz podważania zaufania do mediów i instytucji państwowych.

[8] M. Tołwiński (2024). *Etyczne dylematy odpowiedzialności... Op. cit.*

Przykłady wykorzystania deepfake w polityce:

- **Indie, 2020** – podczas kampanii wyborczej w Delhi, polityk Manoj Tiwari wykorzystał deepfake do nagrania przemówienia w różnych językach regionalnych, by dotrzeć do szerszej grupy wyborców. Choć było to działanie legalne, pokazało potencjał deepfake do wpływania na opinię publiczną[9].
- **USA, 2019** – w mediach społecznościowych krążyło zmanipulowane nagranie Nancy Pelosi, liderki Demokratów w Izbie Reprezentantów, które miało sugerować, że mówi w sposób niewyraźny i jest pod wpływem alkoholu[10]. Mimo że nie było to klasyczne deepfake, lecz zmanipulowane wideo, sprawa pokazała, jak łatwo można podważyć autorytet polityków.
- **Ukraina, 2022** – podczas rosyjskiej inwazji pojawiło się deepfake'owe nagranie prezydenta Wołodymyra Zełenskiego[11], który miał nawoływać ukraińskich żołnierzy do poddania się. Wideo zostało szybko zdementowane, ale pokazuje, jak deepfake może być wykorzystywany jako narzędzie wojny informacyjnej.

Technologia deepfake jest wykorzystywana przez cyberprzestępców do oszustw finansowych, podszywania się pod dyrektorów firm oraz kradzieży tożsamości. **Przykłady wykorzystania deepfake w oszustwach:**

- **Deepfake w rekrutacji – ostrzeżenie FBI (2022)**
  - FBI ostrzegło, że cyberprzestępcy wykorzystują deepfake[12] do fałszowania tożsamości w procesach rekrutacyjnych. W zgłoszonych przypadkach kandydaci ubiegali się o pracę w sektorze IT, podszywając się pod inne osoby i używając deepfake'owych materiałów wideo do rozmów kwalifikacyjnych.
- **Oszustwo bankowe w Zjednoczonych Emiratach Arabskich (2020)**

[9] A. Turek (2020, 19 lutego). *W indyjskiej kampanii wyborczej wykorzystano oficjalnie spreparowane nagrania*. Business Insider. https://businessinsider.com.pl/technologie/nowe-technologie/deepfake-w-indyjskich-wyborach-na-czym-polega-technologia/c6q111l [dostęp: 25.03.2025].

[10] D. O'Sullivan (2019, 24 maja). *Doctored videos shared to make Pelosi sound drunk viewed millions of times on social media*. CNN. https://edition.cnn.com/2019/05/23/politics/doctored-video-pelosi/index.html [dostęp: 25.03.2025].

[11] M. Chwistek (2022, 17 marca). *Do sieci trafił deepfake z prezydentem Zełenskim. W fałszywym wideo „namawiał" do poddania Ukrainy*. Komputer Świat. https://www.komputerswiat.pl/aktualnosci/wydarzenia/do-sieci-trafil-deepfake-z-prezydentem-zelenskim-w-falszywym-wideo-namawial-do/n40qel7 [dostęp: 25.03.2025].

[12] S. Palczewski (2022, 29 czerwca). *FBI ostrzega przed deepfake'ami*. CyberDefence24. https://cyberdefence24.pl/armia-i-sluzby/fbi-ostrzega-przed-deepfakeami [dostęp: 25.03.2025].

· Przestępcy wykorzystali deepfake do podszycia się pod dyrektora banku i zdołali przekonać pracowników do wykonania transakcji na kwotę 35 milionów dolarów. Fałszywy głos dyrektora brzmiał na tyle autentycznie, że nie wzbudził podejrzeń[13].

Deepfake stanowi poważne zagrożenie w kontekście prywatności i cyberprzemocy. Fałszywe materiały mogą być wykorzystywane do kompromitowania osób publicznych lub do szantażu. **Przykłady nadużyć deepfake w sferze prywatnej:**

- **Deepfake pornograficzne** – jednym z najczęstszych nadużyć technologii jest generowanie fałszywych materiałów pornograficznych z udziałem celebrytów i osób prywatnych. Serwisy takie jak DeepNude (zamknięty w 2019 roku) pozwalały na automatyczne generowanie nagich zdjęć na podstawie zwykłych fotografii.
- **Szantaż deepfake** – w USA pojawiły się przypadki, w których oszuści tworzyli deepfake'owe nagrania kompromitujące swoje ofiary i wykorzystywali je do wymuszania pieniędzy lub innych korzyści.

Rozwój technologii deepfake wymusza inwestycje w systemy wykrywania fałszywych treści. Wiodące firmy i instytucje badawcze pracują nad narzędziami, które umożliwiają identyfikację zmanipulowanych materiałów:

- **Deepfake Detection Challenge (Facebook, Microsoft, Amazon, MIT)** – projekt mający na celu opracowanie algorytmów wykrywających deepfake na podstawie subtelnych anomalii w obrazach i dźwięku.
- **Sztuczna inteligencja do wykrywania deepfake** – firmy takie jak **Microsoft, Deeptrace i Sensity AI** opracowują algorytmy zdolne do wykrywania manipulacji w plikach wideo poprzez analizę nienaturalnych ruchów twarzy i artefaktów wizualnych.
- **Wodoodporne znakowanie treści (cryptographic watermarking)** – rozwiązanie stosowane przez Adobe i inne firmy technologiczne, które polega na cyfrowym znakowaniu oryginalnych treści, co pozwala na ich weryfikację.

Deepfake to jedno z najpoważniejszych zagrożeń związanych z rozwojem SI. Jego wpływ na dezinformację polityczną, cyberprzestępczość, naruszenia prywatności oraz oszustwa finansowe jest coraz większy. W obliczu rosnących zagrożeń konieczne jest wprowadzenie skutecznych mechanizmów wykrywania i regulacji prawnych ograniczających wykorzystanie tej technologii w celach przestępczych.

[13] M. Anderson (2021, 15 października). *Deepfaked Voice umożliwił napad na bank o wartości 35 milionów dolarów w 2020 roku*. Unite.ai. https://www.unite.ai/pl/deepfaked-voice-enabled-35-million-bank-heist-in-2020/ [dostęp: 25.03.2025].

Problemy etyczne związane z autonomicznymi systemami wojskowymi[14]

Systemy wojskowe oparte na sztucznej inteligencji budzą obawy dotyczące przenoszenia odpowiedzialności za decyzje o charakterze śmiertelnym z człowieka na maszyny. Kwestie te wymagają uregulowania w międzynarodowym prawie wojennym. Konieczne zatem staje się zabezpieczanie różnorodnych systemów wojskowych przed nadużyciami SI. Korzyści, zagrożenia oraz sposoby tych zabezpieczeń zaprezentowane zostały w poniższej tabeli.

Tabela 1. Korzyści, zagrożenia i sposoby zabezpieczenia przed nadużyciami SI

Obszar	Korzyści	Zagrożenia	Sposoby zabezpieczenia
Automatyzacja wykrywania zagrożeń	Szybsza detekcja ataków cybernetycznych	Możliwość wykorzystania SI przez cyberprzestępców do ukrycia ataków	Regularne aktualizacje systemów SI
Threat hunting	Analiza nietypowych wzorców zachowań w sieci	Możliwość omijania zabezpieczeń przez zaawansowanych hakerów	Używanie SI do wykrywania anomalii i nieautoryzowanych zmian
Deepfake i dezinformacja	Możliwość tworzenia realistycznych symulacji dla edukacji i nauki	Manipulacja opinią publiczną, oszustwa finansowe, szantaż	Narzędzia do wykrywania deepfake, edukacja użytkowników
Cyberprzestępczość oparta na SI	SI wspomaga analizę zagrożeń i neutralizację ataków	Hakerzy mogą automatyzować ataki i wyszukiwać luki w systemach	Wielopoziomowe systemy uwierzytelniania, monitorowanie SI
Autonomiczne systemy ofensywne	Możliwość szybkiego reagowania na cyberzagrożenia	Ryzyko niekontrolowanych działań ofensywnych SI	Ograniczenie autonomii systemów, nadzór człowieka
Etyka i prywatność	Możliwość lepszego zarządzania danymi i ochrony użytkowników	Masowa inwigilacja, nadużycie SI do nadzoru	Regulacje prawne, etyczne wykorzystanie SI

Źródło: Opracowanie własne.

[14] T. Zieliński (2018). *Dylematy użytkowania autonomicznych systemów bojowych w odniesieniu do podstawowych zasad międzynarodowego prawa humanitarnego*. „Humanities and Social Sciences", 25(1).

Sztuczna inteligencja oferuje szerokie możliwości w zakresie poprawy bezpieczeństwa, ale jednocześnie rodzi poważne wyzwania, zarówno technologiczne, jak i etyczne. Aby maksymalizować korzyści i minimalizować ryzyko związane z jej rozwojem, niezbędne jest zrównoważone podejście, które uwzględnia:

- Rozwój technologii wykrywania deepfake – firmy takie jak Microsoft, Adobe i Google pracują nad algorytmami pozwalającymi na skuteczne wykrywanie zmanipulowanych materiałów wideo i audio.
- Międzynarodowe regulacje i standardy – konieczna jest współpraca na poziomie globalnym w zakresie tworzenia regulacji ograniczających wykorzystanie SI do działań nieetycznych i przestępczych.
- Edukacja i świadomość społeczna – użytkownicy internetu muszą być świadomi istnienia deepfake'ów oraz metod manipulacji, aby ograniczyć skuteczność kampanii dezinformacyjnych.

Zastosowanie sztucznej inteligencji w obszarze bezpieczeństwa stanowi zarówno ogromną szansę, jak i poważne wyzwanie. Kluczowym zadaniem na przyszłość będzie znalezienie równowagi między korzyściami a zagrożeniami, jakie niesie rozwój SI, tak aby technologia ta mogła wspierać cyberbezpieczeństwo, zamiast stanowić dodatkowe źródło zagrożeń.

Podsumowanie – wnioski

Sztuczna inteligencja oferuje szerokie możliwości w zakresie poprawy bezpieczeństwa, ale jednocześnie rodzi poważne wyzwania, zarówno technologiczne, jak i etyczne. Aby maksymalizować korzyści i minimalizować ryzyko związane z jej rozwojem, niezbędne jest zrównoważone podejście, które uwzględnia zarówno regulacje prawne, jak i rozwój technologii. Kluczowe znaczenie ma międzynarodowa współpraca w zakresie tworzenia standardów i zasad odpowiedzialnego korzystania z SI.

Zastosowanie sztucznej inteligencji w obszarze bezpieczeństwa stanowi zarówno ogromną szansę, jak i poważne wyzwanie. **Z jednej strony umożliwia skuteczniejszą ochronę przed zagrożeniami, wspiera procesy decyzyjne i podnosi efektywność monitoringu. Z drugiej jednak rodzi istotne dylematy związane z prywatnością, etyką oraz możliwością nadużyć.** Kluczowym zadaniem na przyszłość będzie znalezienie równowagi między korzyściami a zagrożeniami, jakie niesie ze sobą rozwój sztucznej inteligencji w sferze bezpieczeństwa.

Bibliografia

Artykuł w pracy zbiorowej

Ostap H. (2024). *BotTROP – system wykrywania botów w sieci korporacyjnej*. W: *Badania naukowe i technologie 2021-2024*. Wydział Cybernetyki Instytut Systemów Informatycznych. Wojskowa Akademia Techniczna im. Jarosława Dąbrowskiego w Warszawie. Warszawa.

Artykuł w czasopiśmie

Tołwiński M. (2024). *Etyczne dylematy odpowiedzialności za działanie systemów sztucznej inteligencji na wojnie*. „Doctrina. Studia społeczno-Polityczne", 21(21).

Wołoszyn J. (2024). *Ewolucja i wpływ sztucznej inteligencji na zaawansowane strategie obronne w cyberbezpieczeństwie*. „Dydaktyka Informatyki", 19.

Zieliński T. (2018). *Dylematy użytkowania autonomicznych systemów bojowych w odniesieniu do podstawowych zasad międzynarodowego prawa humanitarnego*. „Humanities and Social Sciences", 25(1).

Publikacja elektroniczna

Anderson M. (2021, 15 października). *Deepfaked Voice umożliwił napad na bank o wartości 35 milionów dolarów w 2020 roku*. Unite.ai. https://www.unite.ai/pl/deepfaked-voice-enabled-35-million-bank-heist-in-2020/ [dostęp: 25.03.2025].

Chwistek M. (2022, 17 marca). *Do sieci trafił deepfake z prezydentem Zełenskim. W fałszywym wideo „namawiał" do poddania Ukrainy*. Komputer świat. https://www.komputerswiat.pl/aktualnosci/wydarzenia/do-sieci-trafil-deepfake-z-prezydentem-zelenskim-w-falszywym-wideo-namawial-do/n40qel7 [dostęp: 25.03.2025].

Dastin J. (2018, 11 października). *Insight - Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/ [dostęp: 28.07.2025].

Foulger E. (2022, 24 sierpnia). *Detecting Unknown Ransomware: A Darktrace Case Study*. Darktrace. https://www.darktrace.com/blog/detecting-the-unknown--revealing-uncategorised-ransomware-using-darktrace [dostęp: 25.03.2025].

Gapiński K. (2017, 13 grudnia). *Wykorzystanie modeli Cyber Threat Intelligence jako element skutecznego reagowania na incydenty komputerowe*. Fundacja Bezpieczna Przestrzeń. https://www.cybsecurity.org/pl/wykorzystanie-modeli-cyber--threat-intelligence-jako-element-skutecznego-reagowania-na-incydenty--komputerowe/ [dostęp: 24.03.2025].

O'Sullivan D. (2019, 24 maja). *Doctored videos shared to make Pelosi sound drunk viewed millions of times on social media*. CNN. https://edition.cnn.com/2019/05/23/politics/doctored-video-pelosi/index.html [dostęp: 25.03.2025].

Palczewski S. (2022, 29 czerwca). *FBI ostrzega przed deepfake'ami*. CyberDefence24. https://cyberdefence24.pl/armia-i-sluzby/fbi-ostrzega-przed-deepfakeami [dostęp: 25.03.2025].

Partners I. (2022, 16 listopada). *Darktrace: system, który uczy się użytkowników*. CRN Polska. https://crn.pl/artykuly/darktrace-system-ktory-uczy-sie-uzytkownikow/ [dostęp: 25.03.2025].

Turek A. (2020, 19 lutego). *W indyjskiej kampanii wyborczej wykorzystano oficjalnie spreparowane nagrania*. Business Insider. https://businessinsider.com.pl/technologie/nowe-technologie/deepfake-w-indyjskich-wyborach-na-czym-polega-technologia/c6q111l [dostęp: 25.03.2025].

Streszczenie

Sztuczna inteligencja (SI) jest jednym z najbardziej dynamicznie rozwijających się obszarów technologii ostatnich lat, niosąc zarówno ogromne nadzieje i możliwości, jak i liczne wyzwania oraz zagrożenia dla bezpieczeństwa. Celem niniejszego referatu jest analiza wpływu sztucznej inteligencji na różne obszary bezpieczeństwa – od cyberbezpieczeństwa po bezpieczeństwo społeczne i narodowe. Omówiono korzyści płynące z jej zastosowania, takie jak automatyzacja procesów monitoringu, efektywne wyszukiwanie informacji, szybkie wykrywanie zagrożeń oraz wspieranie procesów decyzyjnych w sytuacjach kryzysowych.

Jednocześnie przedstawiono potencjalne zagrożenia, w tym wykorzystanie sztucznej inteligencji w atakach cybernetycznych, manipulacjach informacjami (np. deepfake) oraz ryzyko wynikające z braku pełnej kontroli nad zaawansowanymi algorytmami. Referat podejmuje również kwestie etyczne, takie jak przenoszenie odpowiedzialności z człowieka na systemy sztucznej inteligencji, ryzyko decyzji podejmowanych przez autonomiczne algorytmy oraz dylematy związane z rozwojem autonomicznych systemów wojskowych.

Wnioski podkreślają konieczność zrównoważonego podejścia do rozwoju sztucznej inteligencji, uwzględniającego harmonizację innowacji technologicznych z regulacjami prawnymi oraz ścisłą kontrolę ich zastosowań. Zwrócono również uwagę na potrzebę międzynarodowej współpracy w tworzeniu standardów i ram prawnych, które pozwolą maksymalizować korzyści wynikające z zastosowania sztucznej inteligencji, jednocześnie minimalizując ryzyko jej nadużycia.

Słowa kluczowe: sztuczna inteligencja, cyberbezpieczeństwo, deepfake, etyka SI, systemy autonomiczne, technologie XXI wieku

Artificial Intelligence – Opportunities or Threats to Security

Abstract

Artificial intelligence (AI) is one of the most dynamically developing areas of technology in recent years, bringing both great opportunities and significant challenges, including new security threats. The aim of this paper is to analyze the impact of artificial intelligence on various areas of security, ranging from cybersecurity to social and national security. The benefits of its application are discussed, including

the automation of monitoring processes, efficient information retrieval, rapid threat detection, and support for decision-making in crisis situations.

At the same time, potential threats are presented, including the use of artificial intelligence in cyberattacks, information manipulation (e.g., deepfakes), and risks arising from the lack of full control over advanced algorithms. The paper also addresses ethical issues, such as the shifting of responsibility from humans to AI systems, the risks associated with decisions made by autonomous algorithms, and the dilemmas related to the development of autonomous military systems.

The conclusions highlight the necessity of a balanced approach to AI development, taking into account the need to harmonize technological innovation with legal regulations and to maintain strict oversight of its applications. Attention is also drawn to the importance of international cooperation in establishing standards and legal frameworks that maximize the benefits of AI implementation while minimizing the risks of its misuse.

Keywords: artificial intelligence, cybersecurity, deepfake, AI ethics, autonomous systems, 21st-century technologies