Cybersecurity and Law

2025 Nr 1(13)

DOI: 10.35467/cal/214545

Deterrence in Cyberspace. The Role of Artificial Intelligence in Building Resilience, Attribution, and Coordinated Response Between Domains

Odstraszanie w cyberprzestrzeni. Rola sztucznej inteligencji w budowaniu odporności, atrybucji i skoordynowanej odpowiedzi między domenami

Tadeusz ZIELIŃSKI
Akademia Sztuki Wojennej
ORCID: 0000-0003-0605-7684
E-mail: t-zielinski@akademia.mil.pl

Abstract

In the face of growing geopolitical rivalry and the increasing complexity of digital threats, cyberspace is transforming into a fully fledged domain of strategic influence. Classical models of deterrence, based on punishment, denial, or reputation, prove insufficient in an environment characterized by the difficulty of attribution, high operational dynamism, asymmetry of capabilities, and a low escalation threshold. Against this backdrop, artificial intelligence assumes particular significance as a tool that could enhance the effectiveness of deterrence strategies.

The article aims to determine the impact of AI applications on deterrence effectiveness in cyberspace through three complementary mechanisms: strengthening the resilience of critical infrastructure, improving attribution processes, and supporting coordinated responses across operational domains. The research question is: How can artificial intelligence enhance the effectiveness of deterrence in cyberspace? The article

employs a literature review and case studies (among others, NATO, Estonia, and the USA).

The findings indicate that integrating AI into digital security systems increases resilience, shortens response times, enhances the credibility of the deterrent message, and enables dynamic escalation management. However, attention is drawn to the risks associated with automated decision-making and the opacity of algorithms.

Keywords

cyber deterrence; artificial intelligence in security; resilience of critical infrastructure; automation of attribution; multi-domain security

Streszczenie

W warunkach rosnącej rywalizacji geopolitycznej oraz narastającej złożoności zagrożeń cyfrowych, cyberprzestrzeń przekształca się w pełnoprawną domenę strategicznego oddziaływania. Klasyczne modele odstraszania, oparte na karze, odmowie lub reputacji, okazują się niewystarczające w środowisku, które cechuje się trudnością atrybucji, wysoką dynamiką operacyjną, asymetrią możliwości oraz niskim progiem eskalacyjnym. Na tym tle szczególnego znaczenia nabiera sztuczna inteligencja jako narzędzie potencjalnie wzmacniające efektywność strategii odstraszania.

Celem artykułu jest określenie wpływu zastosowań AI na skuteczność odstraszania w cyberprzestrzeni poprzez trzy komplementarne mechanizmy: wzmacnianie odporności infrastruktury krytycznej, usprawnianie procesu atrybucji oraz wspieranie skoordynowanej odpowiedzi między domenami operacyjnymi. Problem badawczy sformułowano w postaci następującego pytania: w jaki sposób sztuczna inteligencja może wzmacniać skuteczność odstraszania w cyberprzestrzeni? W artykule zastosowano metodę przeglądu literatury przedmiotu badań oraz studia przypadków (m.in. NATO, Estonia, USA).

Wyniki badania wskazują, że integracja AI w systemach bezpieczeństwa cyfrowego zwiększa odporność, skraca czas reakcji, podnosi wiarygodność komunikatu odstraszającego i umożliwia dynamiczne zarządzanie eskalacją. Zwrócono jednak uwagę na ryzyka związane z automatyzacją decyzji i nieprzejrzystością algorytmów.

Słowa kluczowe

odstraszanie w cyberprzestrzeni; sztuczna inteligencja w bezpieczeństwie; odporność infrastruktury krytycznej; automatyzacja atrybucji; bezpieczeństwo wielodomenowe

Wprowadzenie

Współczesne środowisko bezpieczeństwa międzynarodowego coraz wyraźniej ewoluuje w kierunku złożonej, wielowymiarowej architektury, w której cyberprzestrzeń odgrywa rolę równorzędnej domeny działań strategicznych. Jednocześnie cyberprzestrzeń cechuje się specyficzną dynamiką operacyjną, niskim progiem wejścia, trudnością jednoznacznej identyfikacji agresora oraz wysokim poziomem niepewności poznawczej. Tradycyjne mechanizmy

odstraszania, oparte na koncepcji wzajemnie gwarantowanej destrukcji lub symetrycznej odpowiedzi siłowej, okazują się w tym kontekście niewystarczające. Ich skuteczność podważana jest przez brak przejrzystości, asymetrię środków oraz wieloznaczność sygnałów strategicznych. z tego względu odstraszanie w cyberprzestrzeni staje się nie tylko wyzwaniem operacyjnym, lecz również problemem koncepcyjnym i normatywnym, wymagającym nowego podejścia, uwzględniającego specyfikę tej domeny.

W ostatnich latach coraz większego znaczenia nabiera sztuczna inteligencja (AI) jako komponent architektury bezpieczeństwa cyfrowego. Systemy oparte na algorytmach uczenia maszynowego oraz fuzji danych stanowią nie tylko narzędzie techniczne, ale również potencjalny element wzmacniający efektywność odstraszania. W literaturze przedmiotu oraz dokumentach strategicznych NATO, Unii Europejskiej i państw sojuszniczych, coraz wyraźniej rysuje się tendencja do traktowania sztucznej inteligencji nie tylko jako instrumentu detekcji zagrożeń, ale również jako akceleratora odporności, usprawnionej atrybucji oraz skoordynowanej reakcji wielodomenowej. Przykłady takie jak estoński model cyberodporności, programy typu „Scarlet Dragon" czy rozwiązania rozwijane w ramach sojuszniczych inicjatyw technologicznych (np. DIANA) ukazują operacyjny potencjał AI jako narzędzia odstraszania przez zaprzeczenie, reputację i odpowiedzialną demonstrację zdolności.

W tym kontekście pojawia się pytanie o rzeczywisty wpływ sztucznej inteligencji na efektywność strategii odstraszania w cyberprzestrzeni. Czy AI istotnie wzmacnia wiarygodność i skuteczność działań odstraszających? Czy może prowadzić do efektów odwrotnych – zwiększenia nieprzewidywalności, utraty kontroli nad eskalacją lub rozmycia odpowiedzialności strategicznej? Problematyka ta nabiera szczególnego znaczenia w warunkach skróconego czasu reakcji, wysokiego stopnia automatyzacji oraz pogłębiającej się decentralizacji zarządzania bezpieczeństwem cyfrowym.

Celem niniejszego artykułu jest zidentyfikowanie, w jaki sposób zastosowanie sztucznej inteligencji w systemach cyberbezpieczeństwa może wzmacniać skuteczność strategii odstraszania, ze szczególnym uwzględnieniem trzech komponentów: odporności systemowej, procesu atrybucji oraz skoordynowanej odpowiedzi międzydomenowej. Podejście to opiera się na założeniu, że skuteczne odstraszanie nie wynika wyłącznie z posiadania zdolności ofensywnych, lecz z umiejętności ich odpowiedniego zakomunikowania, integracji z mechanizmami obronnymi oraz nadania im charakteru odstraszającego w percepcji przeciwnika.

W związku z powyższym, ogólny problem badawczy sformułowany został w postaci pytania: w jaki sposób sztuczna inteligencja może wzmacniać skuteczność odstraszania w cyberprzestrzeni poprzez podnoszenie odporności systemów, ułatwianie atrybucji oraz wspieranie skoordynowanej odpowiedzi międzydomenowej?

Na gruncie tak sformułowanego pytania postawiono następującą hipotezę badawczą: zastosowanie sztucznej inteligencji w systemach cyberbezpieczeństwa

istotnie zwiększa skuteczność odstraszania w cyberprzestrzeni, ponieważ umożliwia szybsze wykrywanie zagrożeń, wiarygodną atrybucję źródła ataku oraz integrację reakcji w wielu domenach operacyjnych, co wzmacnia percepcję kosztów i ryzyka po stronie potencjalnego agresora.

Artykuł opiera się na metodzie przeglądu literatury przedmiotu badań oraz studiach przypadków operacyjnych rozwiązań AI wdrażanych w środowisku cyberbezpieczeństwa. Zastosowano również analizę porównawczą strategii odstraszania w dokumentach sojuszniczych oraz opracowaniach eksperckich, a także przegląd aktualnej literatury teoretycznej w zakresie odstraszania cyfrowego, odporności i automatyzacji procesów decyzyjnych oraz raporty specjalistycznych think-tanków.

Struktura artykułu obejmuje trzy zasadnicze części analityczne. W pierwszej przedstawiono ewolucję koncepcji odstraszania w środowisku cyfrowym, ze szczególnym uwzględnieniem ograniczeń tradycyjnych modeli i rosnącej roli mechanizmów opartych na odmowie, reputacji oraz uwikłaniu. Druga część koncentruje się na roli sztucznej inteligencji w zakresie wzmacniania odporności, przyspieszania atrybucji i automatyzacji procesów decyzyjnych. Część trzecia omawia znaczenie skoordynowanej odpowiedzi międzydomenowej w kontekście integracji zdolności AI z działaniami informacyjnymi, dyplomatycznymi i militarnymi. Całość kończy podsumowanie wraz z wnioskami.

Odstraszanie w środowisku cyfrowym: ewolucja koncepcji i wyzwań

Odstraszanie w środowisku cyfrowym stało się jednym z najbardziej złożonych i niejednoznacznych wyzwań dla współczesnej teorii bezpieczeństwa międzynarodowego. W odróżnieniu od relatywnie stabilnych i jednoznacznych reguł rządzących odstraszaniem nuklearnym, cyberprzestrzeń charakteryzuje się głęboką asymetrią sił, trudnością w precyzyjnej identyfikacji sprawców, brakiem ustalonych norm eskalacyjnych oraz nieustanną ewolucją technologii. To środowisko, w którym dominują aktorzy niepaństwowi, operacje poniżej progu wojny, kampanie wpływu oraz działania ukryte. Złożoność ta sprawia, że tradycyjne mechanizmy odstraszania okazują się niewystarczające lub wymagają zasadniczej reinterpretacji.

W klasycznym ujęciu odstraszanie opiera się na racjonalnym rachunku zysków i strat. Zakłada się, że przeciwnik zrezygnuje z ataku, jeśli zostanie przekonany, iż koszt jego działania przewyższy potencjalne korzyści. Skuteczność odstraszania wymaga zatem trzech podstawowych elementów: zdolności (*capability*), wiarygodności (*credibility*) oraz komunikacji (*communication*)[1]. Państwo musi posiadać odpowiednie środki do odwetu, jego groźba musi być wiarygodna w oczach przeciwnika, a sam przekaz odstraszający

[1] R.P. Haffa Jr., The Future of Conventional Deterrence: Strategies for Great Power Competition, „Strategic Studies Quarterly" 2018, t. 12, nr 4, s. 96-97.

powinien być jednoznaczny i zrozumiały. Problem w tym, że wszystkie te założenia napotykają w cyberprzestrzeni istotne ograniczenia. Zarówno zdolności ofensywne, jak i defensywne są trudne do zademonstrowania bez ujawnienia własnych słabości. Wiarygodność jest trudna do budowania w warunkach braku przejrzystości, a komunikacja odstraszająca może zostać zniekształcona przez dezinformację, kanały pośrednie lub asymetrię percepcji.

Pierwsze koncepcje odstraszania cybernetycznego były próbą bezpośredniego przełożenia logiki odstraszania nuklearnego na nową domenę. W szczególności koncentrowano się na odstraszaniu przez karę (*deterrence by punishment*), zakładając, że groźba proporcjonalnego odwetu, nawet o charakterze konwencjonalnym lub kinetycznym, może skutecznie zniechęcić do cyberataków[2]. Taki sposób myślenia był obecny m.in. w amerykańskich strategiach cyberbezpieczeństwa z początku XXI wieku, gdzie podkreślano, że Stany Zjednoczone zastrzegają sobie prawo do reakcji wszelkimi dostępnymi środkami w przypadku poważnego cyberataku. Także NATO w swoich deklaracjach z kolejnych szczytów, począwszy od Warszawy (2016), uznawało cyberprzestrzeń za równorzędną domenę działań operacyjnych, dopuszczając możliwość aktywacji artykułu 5 Traktatu Północnoatlantyckiego w odpowiedzi na działania w cyberprzestrzeni[3]. Jednak z biegiem czasu okazało się, że odstraszanie przez karę w środowisku cyfrowym napotyka kilka fundamentalnych barier.

Pierwszą z nich jest problem atrybucji, czyli przypisania odpowiedzialności za atak. W przeciwieństwie do klasycznego konfliktu zbrojnego, gdzie sprawca ataku jest zazwyczaj znany, cyberprzestrzeń sprzyja ukrywaniu tożsamości i wykorzystywaniu pośredników (*proxy actors*). Atrybucja techniczna wymaga złożonego śledztwa, często opóźnionego, a nawet wtedy może być kontestowana przez stronę przeciwną. W rezultacie, możliwość natychmiastowego i zdecydowanego odwetu, kluczowa dla odstraszania przez karę, jest ograniczona. Co więcej, państwa obawiają się eskalacji w oparciu o niepewne dane, co dodatkowo osłabia ich gotowość do działania.

Drugim ograniczeniem odstraszania przez karę jest problem proporcjonalności i wiarygodności. z uwagi na trudność w dokładnym oszacowaniu skutków cyberataku, zarówno *ex ante*, jak i *ex post*, trudno określić adekwatną odpowiedź. Nieadekwatna reakcja może zostać uznana za eskalacyjną lub nieprzewidywalną, co podważa racjonalność groźby i obniża jej wiarygodność. Dodatkowo, wiele państw, szczególnie o średnim potencjale, nie dysponuje narzędziami ofensywnymi o wystarczającej sile rażenia, co czyni ich groźby odstraszające mało przekonującymi w oczach potencjalnych agresorów[4].

[2] Zob. A. Lupovici, Deterrence through Inflicting Costs: Between Deterrence by Punishment and Deterrence by Denial, „International Studies Review" 2023, t. 25, nr 3.
[3] S. Wiedemar, NATO and Article 5 in Cyberspace, Center for Security Studies, ETH Zürich, 2023, https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse324-EN.pdf, [dostęp: 09.11.2025].
[4] Zob. T.C. Schelling, Arms and Influence, New Haven and London 1966.

Wobec tych ograniczeń, zaczęto poszukiwać alternatywy w postaci odstraszania przez odmowę (*deterrence by denial*). Podejście to zakłada, że celem nie jest odwet, lecz uczynienie ataku nieefektywnym, nieopłacalnym lub wręcz niemożliwym do przeprowadzenia[5]. W praktyce oznacza to wzmacnianie odporności systemów (*resilience*), implementację mechanizmów automatycznego wykrywania i neutralizacji zagrożeń, segmentację sieci, rozwój zdolności do szybkiego reagowania oraz budowanie kompetencji organizacyjnych w zakresie zarządzania incydentami. Przykładem takiego podejścia może być koncepcja „cyfrowej tarczy obronnej" realizowana przez Estonię po atakach z 2007 roku, która obejmuje zarówno infrastrukturę techniczną, jak i edukację społeczną, normy prawne oraz współpracę międzynarodową[6].

Odstraszanie przez odmowę znalazło także swoje odzwierciedlenie w strategiach NATO oraz Unii Europejskiej, w tym w dokumentach takich jak „EU Cybersecurity Strategy"[7] czy „NATO Cyber Defence Policy"[8]. Ich treść wskazuje, że podniesienie poziomu odporności państw członkowskich i instytucji międzynarodowych jest nie tylko celem samym w sobie, ale również istotnym komponentem skutecznego odstraszania. Odporność ta ma charakter wielowymiarowy, ponieważ obejmuje zarówno infrastrukturę krytyczną, jak i społeczeństwo, instytucje demokratyczne oraz gospodarkę. Niemniej jednak, również ten model odstraszania nie jest wolny od wad.

Paradoksalnie bowiem, zbyt silne skupienie na odporności może prowadzić do przekonania, że państwo jest wyłącznie defensywne, niezdolne do odwetu, a tym samym narażone na powtarzające się działania testujące granice tolerancji. z perspektywy psychologii strategicznej, przeciwnik może interpretować brak reakcji nie jako wynik skutecznej obrony, lecz jako przejaw słabości lub braku woli działania. W konsekwencji, odstraszanie przez odmowę wymaga równoległej komunikacji zdolności ofensywnych i determinacji politycznej, aby uniknąć jednostronnego charakteru sygnału strategicznego.

W ostatnich latach coraz częściej podnosi się potrzebę uzupełnienia klasycznych modeli odstraszania o nowe mechanizmy, lepiej odpowiadające charakterowi środowiska cyfrowego. Jednym z nich jest odstraszanie przez uwikłanie (*deterrence by entanglement*), które opiera się na tworzeniu takiej sieci współzależności technologicznych, gospodarczych i normatywnych, że potencjalny agresor narażałby się na szkody wtórne, np. utratę dostępu do technologii, sankcje, wykluczenie z rynku lub reakcję ze strony społeczności

[5] E.D. Borghard, S.W. Lonergan, Deterrence by denial in cyberspace, „Journal of Strategic Studies" 2023, t. 46, nr 3, s. 539-541.
[6] Zob. I. Skierka, When shutdown is no option: Identifying the notion of the digital government continuity paradox in Estonia's eID crisis, „Government Information Quarterly" 2023, t. 40, nr 1.
[7] Zob. B. Farrand, H. Carrapico, A. Turobov, The new geopolitics of EU cybersecurity: security, economy and sovereignty, „International Affairs" 2024, t. 100, nr 6.
[8] Zob. M.P. Efthymiopoulos, NATO Must Adopt a Pre-emptive Approach to Cyber Security, GJIA, https://gjia.georgetown.edu/2024/03/09/nato-time-to-adopt-a-pre-emptive-approach-to-cyber-security-in-new-age-security-architecture/, [dostęp: 09.11.2025].

międzynarodowej[9]. Mechanizm ten bywa porównywany do strategicznego sprzężenia zwrotnego w systemach złożonych – atak na jeden element sieci wywołuje skutki dla wszystkich jej uczestników, co działa odstraszająco. Przykładem może być współzależność między państwami w zakresie łańcuchów dostaw półprzewodników, platform cyfrowych czy norm technologicznych.

Innym uzupełnieniem klasycznego odstraszania jest mechanizm reputacyjny. Odstraszanie przez reputację (*deterrence by reputation*) polega na kształtowaniu wizerunku państwa jako przewidywalnego, konsekwentnego i zdolnego do działania. W środowisku cyfrowym reputacja nie jest jedynie funkcją siły militarnej, lecz także transparentności, przestrzegania norm międzynarodowych, uczestnictwa w inicjatywach kooperacyjnych i jakości cyberdyplomacji. Państwo, które posiada spójną strategię komunikowania swoich wartości, czerwonych linii oraz odpowiedzialności za własne działania, buduje autorytet odstraszający. z drugiej strony, brak reakcji lub reakcja nieproporcjonalna może prowadzić do erozji wiarygodności i zachęcać przeciwników do dalszych działań destabilizujących[10].

Ostatecznie, ewolucja koncepcji odstraszania w cyberprzestrzeni prowadzi do uznania, że skuteczna strategia odstraszania musi mieć charakter złożony, wieloskładnikowy i dynamiczny. Powinna ona integrować różne mechanizmy: odmowę, karę, reputację i uwikłanie, w spójnej architekturze reagowania, wspieranej przez technologie, instytucje i normy. W tym kontekście rosnącą rolę odgrywa sztuczna inteligencja, która może wspierać automatyczne wykrywanie zagrożeń, usprawniać procesy atrybucji, modelować scenariusze eskalacyjne i rekomendować warianty działania w czasie rzeczywistym[11]. Jednocześnie należy pamiętać, że samo wdrożenie technologii nie gwarantuje skuteczności odstraszania, ponieważ w dalszym ciągu kluczowe pozostają kwestie przejrzystości, interoperacyjności, etyki i kontroli nad algorytmami.

Odstraszanie w środowisku cyfrowym wymaga więc nowego paradygmatu strategicznego, który łączy odporność technologiczną z inteligentnym zarządzaniem ryzykiem, elastycznością operacyjną oraz zdolnością do wiarygodnego sygnalizowania woli obrony interesów narodowych. W tym sensie cyberprzestrzeń nie jest jedynie nową domeną konfliktu, lecz także nowym testem dla zdolności państw do adaptacji, współpracy i odpowiedzialnego korzystania z zaawansowanych technologii.

[9] A.F. Brantly, Entanglement in Cyberspace: Minding the Deterrence Gap, „Democracy and Security" 2020, t. 16, nr 3, s. 224-226.
[10] M.J. Mazarr, Understanding Deterrence, [w:] F. Osinga, T. Sweijs (ed.), NL ARMS Netherlands Annual Review of Military Studies 2020: Deterrence in the 21st Century – Insights from Theory and Practice, The Hague 2021, s. 26–27.
[11] Zob. T.V. Paul, Reimagining Deterrence: New Security Threats and Challenges to the Deterrence Paradigm [w:] B. Wasser i in. (ed.), Comprehensive Deterrence Forum: Proceedings and Commissioned Papers, Santa Monica 2018.

Rola sztucznej inteligencji w odporności, atrybucji i automatyzacji decyzji w odstraszaniu w cyberprzestrzeni

W miarę jak cyberprzestrzeń staje się kluczowym środowiskiem rywalizacji międzynarodowej, a państwa i aktorzy niepaństwowi coraz intensywniej wykorzystują jej potencjał ofensywny, tradycyjne podejścia do odstraszania okazują się niewystarczające. W literaturze strategicznej coraz wyraźniej dostrzega się potrzebę adaptacji mechanizmów odstraszania do charakteru działań prowadzonych w przestrzeni cyfrowej, która cechuje się niskim progiem wejścia, wysoką dynamiką operacyjną, trudnością jednoznacznej atrybucji oraz ograniczoną widocznością intencji przeciwnika. W tym kontekście szczególnie doniosłą rolę odgrywa sztuczna inteligencja. Jej zastosowanie w obszarze przetwarzania danych, klasyfikacji sygnatur, identyfikacji zagrożeń oraz automatycznego generowania wariantów działania wprowadza zupełnie nową dynamikę operacyjną. Sztuczna inteligencja skraca czas niezbędny do podjęcia decyzji, a tym samym zwiększa skuteczność reakcji odstraszającej. Jednak równocześnie literatura strategiczna ostrzega przed jej ograniczeniami, obejmującymi m.in. wrażliwość na zanieczyszczone dane, podatność na oszustwa i manipulacje, brak przejrzystości algorytmów oraz ryzyko automatycznego wzmacniania błędnych sygnałów, które mogą prowadzić do poważnych błędów decyzyjnych. Ostateczna skuteczność AI zależy więc od jakości projektowania interakcji człowiek–maszyna, czyli od tego, czy uda się stworzyć pętlę decyzyjną, w której człowiek zachowa kontrolę nad logiką eskalacyjną, ale skorzysta z przewagi informacyjnej i predykcyjnej zapewnianej przez algorytmy[12].

Jednym z pierwszoplanowych zastosowań sztucznej inteligencji w kontekście odstraszania cyfrowego jest wzmacnianie odporności infrastruktury krytycznej na ataki zakłócające i destrukcyjne. Odporność oznacza tu nie tylko zdolność do przetrwania ataku i powrotu do funkcjonowania, lecz także umiejętność jego wczesnego wykrycia i zneutralizowania jeszcze przed osiągnięciem skutków operacyjnych. Sztuczna inteligencja, szczególnie w formie uczenia maszynowego i jego zaawansowanych odmian, takich jak uczenie głębokie, umożliwia implementację systemów detekcji opartych na dynamicznej analizie wzorców[13]. Zamiast polegać wyłącznie na wcześniej zidentyfikowanych sygnaturach zagrożeń, systemy te potrafią wykrywać anomalie w czasie rzeczywistym, co jest szczególnie przydatne w przypadku ataków nieznanych wcześniej (*zero-day attacks*).

[12] Artificial Intelligence (AI) in Cybersecurity: The Future of Threat Defense, Fortinet, https://www.fortinet.com/resources/cyberglossary/artificial-intelligence-in-cybersecurity, [dostęp: 09.11.2025].

[13] J. Lynch, E. Morrison, Deterrence Through AI-Enabled Detection and Attribution, Henry A. Kissinger Center for Global Affairs, https://kissinger.sais.jhu.edu/programs-and-projects/kissinger-center-papers/deterrence-through-ai-enabled-detection-and-attribution/, [dostęp: 09.11.2025].

Przykładem zastosowania takiego podejścia jest amerykański program „Scarlet Dragon", który umożliwia fuzję danych pochodzących z różnych sensorów, w tym danych wywiadowczych, sygnałowych oraz satelitarnych, w jeden wspólny obraz operacyjny[14]. Dzięki wykorzystaniu sztucznej inteligencji system ten był w stanie w czasie rzeczywistym identyfikować potencjalne zagrożenia i proponować środki zaradcze. Podobne rozwiązania są obecnie wdrażane przez NATO w ramach inicjatyw takich jak DIANA (*Defence Innovation Accelerator for the North Atlantic*), które mają na celu integrację zdolności AI z systemami wczesnego ostrzegania i reagowania na poziomie sojuszniczym[15].

W literaturze dotyczącej odstraszania często podkreśla się, że odporność sama w sobie może mieć wartość odstraszającą. Jeżeli potencjalny agresor postrzega atak jako nieskuteczny lub mało opłacalny, ponieważ nie przyniesie oczekiwanego efektu operacyjnego, wówczas może zaniechać działania. Taka logika odstraszania przez odmowę opiera się na założeniu, że odporność systemu obniża atrakcyjność ataku. W tym kontekście sztuczna inteligencja jako narzędzie wczesnego wykrywania i adaptacyjnej obrony, wzmacnia wiarygodność sygnału odstraszającego[16].

Drugim obszarem, w którym AI ma kluczowe znaczenie dla odstraszania, jest proces atrybucji, czyli przypisania odpowiedzialności za przeprowadzenie ataku. W warunkach cyberprzestrzeni atrybucja jest nie tylko trudna technicznie, ale także politycznie wrażliwa. Sztuczna inteligencja może znacząco skrócić czas niezbędny do przeprowadzenia dochodzenia atrybucyjnego, poprzez automatyczne łączenie danych z różnych źródeł: technicznych (adresy IP, malware, logi systemowe), społecznych (media społecznościowe, komunikacja publiczna), wywiadowczych (informacje HUMINT, SIGINT) oraz behawioralnych (charakterystyczne schematy działania). Tego rodzaju fuzja danych, wspomagana przez modele predykcyjne, pozwala na rekonstrukcję łańcucha decyzyjnego i operacyjnego przeciwnika. Przykładem może być wykorzystanie sztucznej inteligencji do identyfikacji grupy APT (*Advanced Persistent Threat*) na podstawie wzorców kodowania, języka używanego w komunikacji i godzin aktywności, co może sugerować określony region geograficzny lub afiliację instytucjonalną[17].

Szybka i wiarygodna atrybucja ma znaczenie odstraszające tylko wtedy, gdy zostaje skutecznie zakomunikowana. Dlatego też wiele państw inwestuje nie

[14] A. Obis, Scarlet Dragon exercises, Maven Smart System paving the way for AI adoption, development, Federal News Network, https://federalnewsnetwork.com/defense-main/2024/08/scarlet-dragon-exercises-maven-smart-system-paving-the-way-for-ai-adoption-development/, [dostęp: 09.11.2025].

[15] NATO, Defence Innovation Accelerator for the North Atlantic (DIANA), NATO, https://www.nato.int/cps/en/natohq/topics_216199.htm, [dostęp: 09.11.2025].

[16] C. Roudani, Cyber Deterrence and Digital Resilience: Towards a New Doctrine of Global Defense, Modern Diplomacy, https://moderndiplomacy.eu/2025/06/18/cyber-deterrence-and-digital-resilience-towards-a-new-doctrine-of-global-defense/, [dostęp: 09.11.2025].

[17] A. Wilner, C. Babb, New Technologies and Deterrence: Artificial Intelligence and Adversarial Behaviour, [w:] F. Osinga, T. Sweijs (ed.), NL ARMS Netherlands Annual Review of Military Studies 2020, The Hague 2021.

tylko w rozwój narzędzi analitycznych, ale także w systemy dyplomacji cyfrowej, które umożliwiają formułowanie przekonujących narracji na forum międzynarodowym. Publiczne przypisanie ataku, nawet bez pełnego ujawnienia dowodów technicznych, może wpłynąć na reputację agresora, zniechęcić innych aktorów do współpracy z nim, a jednocześnie wzmocnić legitymację własnych działań odwetowych. W takim ujęciu atrybucja wspierana przez AI staje się nie tylko narzędziem analitycznym, lecz także komponentem strategii komunikacyjnej i reputacyjnej odstraszania[18].

Trzeci komponent, automatyzacja decyzji, wiąże się z coraz większą potrzebą działania w warunkach skróconego czasu reakcji. W przeciwieństwie do klasycznych konfliktów zbrojnych, działania w cyberprzestrzeni odbywają się w czasie rzeczywistym. Opóźnienie w wykryciu i odpowiedzi na atak może prowadzić do utraty kontroli nad systemem, wycieku danych lub uszkodzenia fizycznej infrastruktury. W takim środowisku konieczne jest zautomatyzowanie pewnych elementów procesu decyzyjnego, począwszy od klasyfikacji zagrożeń, przez wybór scenariuszy odpowiedzi, aż po uruchomienie środków defensywnych lub ofensywnych. AI może tu pełnić rolę wspomagającą (systemy rekomendacyjne), ale również autonomiczną – zwłaszcza w przypadku tzw. systemów reakcji zbliżeniowej (*proximity response systems*), które nie mogą czekać na zgodę człowieka[19].

Jednakże automatyzacja decyzji rodzi szereg problemów prawnych, etycznych i strategicznych. Jednym z największych wyzwań jest brak przejrzystości działania algorytmów, tzw. czarne skrzynki decyzyjne (*black box AI*), których logika działania nie jest w pełni zrozumiała nawet dla ich twórców. W warunkach potencjalnego konfliktu zbrojnego taka nieprzejrzystość może prowadzić do błędnych interpretacji, niezamierzonej eskalacji lub zastosowania nieproporcjonalnej siły[20]. W literaturze pojawiają się również ostrzeżenia przed tzw. eskalacją przez algorytm (*escalation by algorithm*), kiedy systemy autonomiczne obu stron, interpretując działania przeciwnika jako groźne, mogą uruchomić łańcuchy reakcji, których nikt już nie kontroluje. Aby temu zapobiec, konieczne jest projektowanie systemów z wbudowanymi „bezpiecznikami" oraz wyraźne określenie roli człowieka w pętli decyzyjnej (*human-in-the-pp*).

Nie mniej istotne są również problemy legitymizacji użycia zautomatyzowanych środków w ramach strategii odstraszania. Choć AI może przyspieszyć reakcję i zwiększyć skuteczność działań, to z perspektywy prawa międzynarodowego oraz norm etycznych, zwłaszcza w kontekście ewentualnych odpowiedzi kinetycznych, decyzje te muszą być uzasadnione, proporcjonalne

[18] CentroStudiEuropeo, Artificial Intelligence: Redefining Diplomacy in the Digital Age, Hermes – Centro Studi Europeo, https://www.hermescse.eu/en/artificial-intelligence-redefining-diplomacy-in-the-digital-age/, [dostęp: 09.11.2025].
[19] T. Erskine, S.E. Miller, AI and the decision to go to war: future risks and opportunities, „Australian Journal of International Affairs" 2024, t. 78, nr 2, s. 140-143.
[20] K. Roy, Artificial Intelligence, Ethics and the Future of Warfare: Global Perspectives, London 2024, s. 120.

i oparte na zrozumieniu kontekstu. Sztuczna inteligencja nie posiada takich kompetencji poznawczych i aksjologicznych jak człowiek, co oznacza, że jej wykorzystanie w strategii odstraszania wymaga nie tylko rozwiązań technologicznych, ale także ram politycznych i normatywnych[21].

Podsumowując, sztuczna inteligencja w kontekście odstraszania w cyberprzestrzeni pełni funkcje zarówno funkcjonalne, jak i symboliczne. z jednej strony wzmacnia zdolności techniczne w zakresie wykrywania, klasyfikowania i neutralizacji zagrożeń, skracając czas reakcji i zwiększając odporność systemów. z drugiej strony, jej obecność w architekturze odstraszania komunikuje przeciwnikowi, że państwo dysponuje nowoczesnymi, adaptacyjnymi i zintegrowanymi narzędziami reagowania. To połączenie wymiaru operacyjnego i percepcyjnego sprawia, że AI staje się kluczowym czynnikiem w budowie nowoczesnych strategii odstraszania. Jednak jej skuteczność będzie zależeć nie tylko od możliwości obliczeniowych czy dostępu do danych, lecz także od umiejętności harmonijnego włączenia algorytmów w szerszy system bezpieczeństwa, uwzględniający czynniki polityczne, prawne, kulturowe i moralne.

Zintegrowana odpowiedź międzydomenowa i nowe paradygmaty odstraszania

Współczesna dynamika środowiska bezpieczeństwa międzynarodowego coraz wyraźniej wskazuje na konieczność przezwyciężenia ograniczeń tradycyjnych modeli odstraszania, opartych na mechanizmach znanych jeszcze z epoki zimnej wojny. Klasyczne odstraszanie przez odmowę oraz odstraszanie przez karę, choć nadal odgrywają istotną rolę w planowaniu strategicznym państw, okazują się coraz mniej skuteczne w obliczu rosnącej złożoności zagrożeń hybrydowych, asymetrycznych i cyfrowych. Szczególnie wyraźnie widoczne jest to w domenie cyberprzestrzeni, gdzie niejednoznaczność atrybucji, wysoka prędkość operacyjna oraz niskie progi wejścia tworzą środowisko wyjątkowo podatne na eskalację, zakłócenia i działania poniżej progu wojny. Odpowiedzią na te wyzwania jest wyłaniający się paradygmat odstraszania zintegrowanego oraz międzydomenowego, którego zasadniczym celem jest osiągnięcie efektów odstraszających nie poprzez izolowane środki w jednej domenie, ale przez skoordynowane działania obejmujące wiele sfer jednocześnie: cyberprzestrzeń, przestrzeń informacyjną, domenę militarną, dyplomatyczną oraz gospodarczą.

Zintegrowana odpowiedź międzydomenowa to nie tylko koncepcja teoretyczna, ale coraz częściej praktyczna potrzeba, która znajduje odzwierciedlenie w dokumentach strategicznych NATO, USA i państw

[21] J. Vujicic, AI Ethics in Legal Decision-Making Bias, Transparency, And Accountability, „International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering” 2025, t. 14, nr 5, s. 1457.

członkowskich Unii Europejskiej. Opiera się ona na założeniu, że skuteczna reakcja odstraszająca powinna być projektowana jako kompleksowy efekt, powstający w wyniku synchronizacji instrumentów oddziaływania o różnym charakterze, a nie jako pojedyncze działanie w ramach jednej struktury sił[22]. Sztuczna inteligencja odgrywa w tej koncepcji rolę katalizatora, umożliwiając nie tylko automatyzację wykrywania zagrożeń, ale również wspierając proces decyzyjny poprzez analizę danych, predykcję zachowań przeciwnika i rekomendację scenariuszy reakcji. W tym sensie AI staje się nie tyle narzędziem wspomagającym, ile kluczowym komponentem architektury odstraszania, szczególnie w wymiarze czasu reakcji, który w środowisku cybernetycznym jest czynnikiem o znaczeniu krytycznym.

Przykładem wdrażania tego typu rozwiązań są eksperymenty z systemami wczesnego ostrzegania opartymi na fuzji danych z wielu źródeł, od sensorów fizycznych po dane z przestrzeni informacyjnej. Modele te, oparte na uczeniu głębokim, umożliwiają detekcję anomalii, identyfikację kampanii dezinformacyjnych, a nawet przewidywanie potencjalnych ścieżek eskalacyjnych w odpowiedzi na określone bodźce geopolityczne. W połączeniu z systemami przetwarzania brzegowego (*edge computing*), zdolnymi do analizy danych lokalnie i w czasie rzeczywistym, pozwalają one na wyprzedzające działanie i adaptacyjną odpowiedź, stanowiąc istotny element odstraszania dynamicznego[23]. W praktyce oznacza to możliwość przygotowania odpowiedzi w czasie liczonym w sekundach, co z perspektywy przeciwnika znacząco zwiększa niepewność co do rezultatu potencjalnego ataku.

Nowy paradygmat odstraszania zakłada również redefinicję pojęcia proporcjonalności i adekwatności reakcji. W tradycyjnym ujęciu działania odwetowe miały charakter symetryczny: cyberatak odpowiadał cyberatakiem, sankcja sankcją, gest dyplomatyczny gestem dyplomatycznym. W koncepcji odstraszania zintegrowanego kluczowe staje się osiągnięcie efektu odstraszającego nie poprzez symetrię środków, ale przez asymetryczną, a jednocześnie skoordynowaną odpowiedź o wysokim poziomie nieprzewidywalność[24]. Przykładowo, skoordynowany cyberatak na infrastrukturę finansową państwa może spotkać się nie tylko z retorsją w domenie cyfrowej, ale również z presją międzynarodową, operacją informacyjną mającą na celu podważenie legitymacji przeciwnika, a także z demonstracją siły wojskowej w obszarze wrażliwym geostrategicznie. Istotne jest przy tym, że AI może wspierać nie tylko wybór środków reakcji, ale również ocenę ich skuteczności i potencjalnych konsekwencji strategicznych.

[22] J.J. Wirtz, J.A. Larsen, Wanted: a strategy to integrate deterrence, „Defense & Security Analysis” 2024, t. 40, nr 3, s. 371–373.
[23] Zob. D. Sazhin, T. Gandee, C. Stevens, L. Borghetti, Gaps in Deterrence Theory And Artificial Intelligence Solutions, 2025.
[24] P. Pijpers, A. Kraesten, Rethinking Cyber Deterrence: Adapting to the Realities of the Digital Battlefield, „Journal of Strategic Security” 2025, t. 18, nr 1, s. 65-67.

W tym kontekście coraz większego znaczenia nabierają mechanizmy zarządzania eskalacją, które muszą być dostosowane do warunków środowiska automatyzowanego i zdecentralizowanego. Systemy sztucznej inteligencji, odpowiednio zaprojektowane, mogą pełnić funkcję moderatora eskalacyjnego, rekomendując warianty reakcji o niższym poziomie ryzyka, sugerując ścieżki deeskalacyjne lub umożliwiając symulacje skutków alternatywnych decyzji w czasie rzeczywistym. Tego rodzaju podejście, choć wymagające wysokiego poziomu interoperacyjności i zaufania między agendami państwowymi, ma potencjał zwiększenia przejrzystości strategicznej i zredukowania ryzyka niezamierzonego konfliktu[25].

Zintegrowana odpowiedź międzydomenowa zakłada również istotną zmianę w sposobie komunikowania zdolności odstraszających. Tradycyjna doktryna odstraszania opierała się na założeniu, że przeciwnik musi znać nasze możliwości, aby ich się obawiał, przy jednoczesnym zachowaniu wystarczającego poziomu niepewności co do dokładnego zakresu tych zdolności. W środowisku cyfrowym, gdzie percepcja może być kształtowana błyskawicznie przez przekaz informacyjny, media społecznościowe i kampanie wpływu, sztuczna inteligencja może odgrywać rolę „architekta narracji odstraszającej". Poprzez analizę percepcji przeciwnika, monitorowanie reakcji na określone sygnały oraz wspieranie tworzenia spójnych komunikatów strategicznych, AI może wzmacniać efekt odstraszania nie tylko przez zdolności fizyczne, ale również przez kontrolę nad informacyjnym wymiarem konfliktu[26].

Równocześnie, budowa zintegrowanej architektury odstraszania wymaga nowego podejścia do zarządzania relacjami sojuszniczymi i międzyinstytucjonalnymi. Zdolność do wspólnej reakcji między domenami zakłada nie tylko kompatybilność technologiczną, ale również wspólnotę interesów, synchronizację procesów decyzyjnych oraz zbieżność priorytetów strategicznych. Organizacje takie jak NATO, Unia Europejska czy koalicje *ad hoc* tworzone wokół wspólnych zagrożeń cyfrowych, stają przed wyzwaniem zbudowania ekosystemu odstraszania, struktury, która nie tylko koordynuje działania państw członkowskich, ale także wspiera tworzenie wspólnych standardów, dzielenie się informacjami oraz rozwój interoperacyjnych narzędzi AI wspomagających odstraszanie[27].

Podsumowując, zintegrowana odpowiedź międzydomenowa oraz nowe paradygmaty odstraszania w erze sztucznej inteligencji oznaczają przejście od odstraszania liniowego i reaktywnego do modelu sieciowego, adaptacyjnego i proaktywnego. Sztuczna inteligencja staje się nie tylko akceleratorem działań

[25] N. Katagiri, Artificial Intelligence and Cross-Domain Warfare: Balance of Power and Unintended Escalation, „Global Society" 2024, t. 38, nr 1, s. 40-42.
[26] H. Lahmann, B. Custers, B.I. Scott, The fundamental rights risks of countering cognitive warfare with artificial intelligence, „Ethics and Information Technology" 2025, t. 27, nr 4.
[27] M. Kostarakos, European Union and NATO Cooperation in Hybrid Threats, [w:] K.P. Balomenos, A. Fytopoulos, P.M. Pardalos (ed.), Handbook for Management of Threats, Springer International Publishing, Cham 2023, s. 416.

operacyjnych, ale również instrumentem kształtowania przestrzeni strategicznej, w której odstraszanie rozumiane jest jako proces nieustannego dostosowywania, przewidywania i zarządzania ryzykiem. W tak rozumianym systemie, skuteczność odstraszania nie zależy wyłącznie od potencjału siłowego, lecz od zdolności do integracji informacji, koordynacji działań i komunikowania zamiaru, czyli od poziomu inteligencji strategicznej państwa i jego zdolności do działania w środowisku wielodomenowym, złożonym i szybkozmiennym.

Podsumowanie

Integracja narzędzi opartych na sztucznej inteligencji z systemami bezpieczeństwa cyfrowego stanowi istotny czynnik wzmacniający zdolność państw do odstraszania agresorów w środowisku informacyjnym. Kluczowym mechanizmem tego oddziaływania jest przyspieszenie i automatyzacja procesów wykrywania zagrożeń, zwiększenie dokładności atrybucji ataków oraz możliwość skoordynowanego i szybciej uruchamianego reagowania w wielu domenach jednocześnie: od infrastruktury informatycznej, przez przestrzeń informacyjną, po środki o charakterze gospodarczym czy militarnym.

Mechanizmy predykcyjnej analizy zagrożeń (predyktywne rozpoznanie sygnatur ataków, modelowanie zachowań agresorów), oparte na architekturach uczenia maszynowego, umożliwiają wcześniejsze rozpoznanie działań wrogich i ich neutralizację zanim osiągną one etap destrukcji. Oznacza to, że w tradycyjnym modelu odstraszania przez odmowę, zastosowanie AI przekłada się na realne wzmocnienie zdolności defensywnych państwa, a przez to na podniesienie progu opłacalności potencjalnego ataku. Równocześnie, algorytmiczne systemy atrybucji pozwalają skrócić czas identyfikacji źródła ataku, co z kolei zwiększa wiarygodność zapowiedzi reakcji oraz zdolność do jej eskalacji w razie potrzeby. Tym samym, także logika odstraszania przez karę zyskuje na wiarygodności, gdyż agresor ma podstawy sądzić, że jego działanie nie pozostanie anonimowe i bez konsekwencji.

W odniesieniu do państw średniej wielkości, takich jak Polska, Litwa, Estonia czy Finlandia, zastosowanie sztucznej inteligencji w architekturach bezpieczeństwa cyfrowego ma istotne znaczenie systemowe. Ze względu na ograniczone zasoby konwencjonalne, a także niejednokrotnie mniejszą autonomię strategiczną, państwa te nie są w stanie konkurować z potęgami cyfrowymi w wymiarze bezpośredniej projekcji siły. Niemniej, wdrożenie wyspecjalizowanych narzędzi AI do monitoringu zagrożeń, budowy odporności infrastruktury oraz automatycznego generowania scenariuszy reakcji, może znacząco zrównoważyć asymetrię. W praktyce oznacza to, że państwo średniej wielkości, które potrafi efektywnie zintegrować algorytmy uczenia maszynowego z systemami wykrywania i analizy incydentów, zyskuje możliwość odstraszania aktorów zewnętrznych poprzez realne podniesienie kosztów ewentualnej agresji. Co więcej, państwa takie mogą w ramach struktur sojuszniczych (np. NATO)

pełnić rolę „węzłów odporności" (*resilience nodes*), dostarczając w czasie rzeczywistym informacji o zagrożeniach lub wspierając działania w ramach kolektywnej reakcji. Przykładem może być estoński model współpracy z Centrum Doskonalenia Cyberobrony (CCDCOE), który opiera się na integracji zasobów narodowych z mechanizmami wymiany informacji, wspieranymi technologią sztucznej inteligencji[28].

W kontekście dalszych badań zasadne wydaje się pogłębienie analizy zjawisk pogranicza, takich jak wykorzystanie AI w operacjach hybrydowych, w tym szczególnie w konstruowaniu fałszywych narracji, *deepfake'ów* oraz operacjach wpływu (*perception management*). Obszar ten pozostaje relatywnie słabo zbadany, mimo że konsekwencje działań opartych na generatywnej sztucznej inteligencji, takich jak fałszywe wypowiedzi liderów politycznych czy zmanipulowane nagrania wideo, mogą mieć krytyczne znaczenie dla wiarygodności systemu odstraszania, percepcji stabilności oraz reakcji społecznej. Należy również zwrócić uwagę na ryzyko wynikające z automatyzacji decyzji w kontekście odstraszania. Mechanizmy oparte na sztucznej inteligencji mogą niekiedy generować tzw. pętle eskalacyjne, w których błędna interpretacja sygnału przez algorytm prowadzi do niezamierzonej reakcji o wysokim poziomie intensywności. W szczególności zagraża to stabilności strategicznej w relacjach państw wyposażonych w broń nuklearną, gdzie błąd algorytmiczny mógłby mieć konsekwencje egzystencjalne.

Podsumowując, zastosowanie sztucznej inteligencji w odstraszaniu cyfrowym jest kierunkiem nieuniknionym, lecz wymagającym precyzyjnego nadzoru, odpowiedzialnego projektowania i ścisłej integracji z architekturą bezpieczeństwa państwa oraz strukturami wielostronnymi. W nadchodzących latach nie tylko sama zdolność techniczna, ale również przejrzystość, zaufanie społeczne oraz normy regulujące interakcję ludzi i algorytmów staną się filarami skutecznego odstraszania w środowisku cyfrowym.

Bibliografia

Artificial Intelligence (AI) in Cybersecurity: The Future of Threat Defense, Fortinet, https://www.fortinet.com/resources/cyberglossary/artificial-intelligence-in-cyber-security.

Borghard E.D., Lonergan S.W., Deterrence by denial in cyberspace, „Journal of Strategic Studies" 2023, t. 46, nr 3, https://doi.org/10.1080/01402390.2021.1944856.

Brantly A.F., Entanglement in Cyberspace: Minding the Deterrence Gap, „Democracy and Security" 2020, t. 16, nr 3, https://doi.org/10.1080/17419166.2020. 1773807.

CentroStudiEuropeo, Artificial Intelligence: Redefining Diplomacy in the Digital Age, Hermes – Centro Studi Europeo, https://www.hermescse.eu/en/artificial-intelligence-redefining-diplomacy-in-the-digital-age/.

[28] Zob. P. Giordano, NATO Centres of Excellence – Cooperative Cyber Defence (CCD COE), NATO's ACT, https://www.act.nato.int/article/nato-centres-of-excellence-cooperative-cyber-defence-ccd-coe/, [dostęp: 09.11.2025].

Efthymiopoulos M.P., NATO Must Adopt a Pre-emptive Approach to Cyber Security, GJIA, https://gjia.georgetown.edu/2024/03/09/nato-time-to-adopt-a-pre-emptive-approach-to-cyber-security-in-new-age-security-architecture/.

Erskine T., Miller S.E., AI and the decision to go to war: future risks and opportunities, „Australian Journal of International Affairs" 2024, t. 78, nr 2, https://doi.org/10.1080/10357718.2024.2349598.

Farrand B., Carrapico H., Turobov A., The new geopolitics of EU cybersecurity: security, economy and sovereignty, „International Affairs" 2024, t. 100, nr 6, https://doi.org/10.1093/ia/iiae231.

Giordano P., NATO Centres of Excellence – Cooperative Cyber Defence (CCD COE), NATO's ACT, https://www.act.nato.int/article/nato-centres-of-excellence-cooperative-cyber-defence-ccd-coe/.

Haffa Jr. R.P., The Future of Conventional Deterrence: Strategies for Great Power Competition, „Strategic Studies Quarterly" 2018, t. 12, nr 4.

Katagiri N., Artificial Intelligence and Cross-Domain Warfare: Balance of Power and Unintended Escalation, „Global Society" 2024, t. 38, nr 1, https://doi.org/10.1080/13600826.2023.2248179.

Kostarakos M., European Union and NATO Cooperation in Hybrid Threats, [w:] K. P. Balomenos, A. Fytopoulos, P. M. Pardalos (ed.), Handbook for Management of Threats, Cham 2023.

Lahmann H., Custers B., Scott B.I., The fundamental rights risks of countering cognitive warfare with artificial intelligence, „Ethics and Information Technology" 2025, t. 27, nr 4, https://doi.org/10.1007/s10676-025-09868-9.

Lupovici A., Deterrence through Inflicting Costs: Between Deterrence by Punishment and Deterrence by Denial, „International Studies Review" 2023, t. 25, nr 3, https://doi.org/10.1093/isr/viad036.

Lynch J., Morrison E., Deterrence Through AI-Enabled Detection and Attribution, Henry A. Kissinger Center for Global Affairs, https://kissinger.sais.jhu.edu/programs-and-projects/kissinger-center-papers/deterrence-through-ai-enabled-detection-and-attribution/.

Mazarr M.J., Understanding Deterrence, [w:] F. Osinga, T. Sweijs (ed.), NL ARMS Netherlands Annual Review of Military Studies 2020: Deterrence in the 21st Century – Insights from Theory and Practice, The Hague 2021.

NATO, Defence Innovation Accelerator for the North Atlantic (DIANA), NATO, https://www.nato.int/cps/en/natohq/topics_216199.htm.

Obis A., Scarlet Dragon exercises, Maven Smart System paving the way for AI adoption, development, Federal News Network, https://federalnewsnetwork.com/defense-main/2024/08/scarlet-dragon-exercises-maven-smart-system-paving-the-way-for-ai-adoption-development/.

Paul T.V., Reimagining Deterrence: New Security Threats and Challenges to the Deterrence Paradigm, [w:] B. Wasser i in. (ed.), Comprehensive Deterrence Forum: Proceedings and Commissioned Papers, Santa Monica 2018.

Pijpers P., Kraesten A., Rethinking Cyber Deterrence: Adapting to the Realities of the Digital Battlefield, „Journal of Strategic Security" 2025, t. 18, nr 1.

Roudani C., Cyber Deterrence and Digital Resilience: Towards a New Doctrine of Global Defense, Modern Diplomacy, https://moderndiplomacy.eu/2025/06/18/cyber-deterrence-and-digital-resilience-towards-a-new-doctrine-of-global-defense/.

Roy K., Artificial Intelligence, Ethics and the Future of Warfare: Global Perspectives, London 2024.

Sazhin D., Gandee T., Stevens C., Borghetti L., Gaps in Deterrence Theory And Artificial Intelligence Solutions, 2025.

Schelling T.C., Arms and Influence, Yale University Press, New Haven and London 1966.

Skierka I., When shutdown is no option: Identifying the notion of the digital government continuity paradox in Estonia's eID crisis, „Government Information Quarterly" 2023, t. 40, nr 1, https://doi.org/10.1016/j.giq.2022.101781.

Vujicic J., AI Ethics in Legal Decision-Making Bias, Transparency, And Accountability, „International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering" 2025, t. 14, nr 05, https://doi.org/10.15662/IJAREEIE.2025.1404001.

Wiedemar S., NATO and Article 5 in Cyberspace, Center for Security Studies, ETH Zürich, 2023, https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse324-EN.pdf.

Wilner A., Babb C., New Technologies and Deterrence: Artificial Intelligence and Adversarial Behaviour, [w:] F. Osinga, T. Sweijs (ed.), NL ARMS Netherlands Annual Review of Military Studies 2020, The Hague 2021.

Wirtz J.J., Larsen J.A., Wanted: a strategy to integrate deterrence, „Defense & Security Analysis" 2024, t. 40, nr 3, https://doi.org/10.1080/14751798.2024.2352943.