

Cybersecurity and Law

2025 Nr 1(13)

DOI: 10.35467/cal/214603



Between Innovation and Control: Legal Frameworks for the Use of Artificial Intelligence in Armed Conflict

Między innowacją a kontrolą: ramy prawne wykorzystania sztucznej inteligencji w działaniach zbrojnych

Kacper ZDROJEWSKI

Akademia Sztuki Wojennej

ORCID: 0009-0007-5255-0441

E-mail: k.zdrojewski@doktorant.akademia.mil.pl

Abstract

The research aims to analyze and compare the legal and regulatory frameworks, both international and national (Poland), governing the deployment of Artificial Intelligence (AI) in weapons systems. The objective is to determine how existing regulations balance the pursuit of technological advantage against the legal requirements of International Humanitarian Law (IHL), control, and accountability. The methodology employed source document analysis and comparison (treaties, resolutions, strategies, reports), critically assessing the detail of regulations regarding Meaningful Human Control (MHC), attribution of responsibility, and AI explainability within the context of the Law of Armed Conflict (LOAC).

The analysis revealed a regulatory divergence between NATO's technological imperative and the UN's normative postulate, resulting in a normative gap-the absence of a legally binding, operationally viable definition of MHC. This deficiency leads to the risk of a „responsibility gap” and the erosion of IHL principles like distinction and proportionality. Nationally, the 2039 AI Strategy, while strategically sound, lacks the executive legal acts to operationalize MHC requirements.

The study concludes that the effective implementation of AI is critically dependent on an urgent shift from strategic declarations to binding, national executive regulations and the establishment of a legally binding MHC definition to ensure full compliance with IHL.

Keywords

artificial intelligence, lethal autonomous weapons systems, law of armed conflict, AI legal regulations, NATO AI strategy

Streszczenie

Celem badań jest szczegółowa analiza i porównanie ram prawnych oraz regulacyjnych, zarówno na poziomie międzynarodowym jak i krajowym, dotyczących wdrażania sztucznej inteligencji (AI) w systemach uzbrojenia, dążąc do ustalenia, w jaki sposób istniejące regulacje równoważą dążenie do przewagi technologicznej z wymogami prawnymi Międzynarodowego Prawa Humanitarnego (MPH), kontroli oraz odpowiedzialności. Zastosowano metodę analizy i porównania dokumentów źródłowych (traktatów, rezolucji, strategii, raportów), krytycznie oceniając stopień szczegółowości regulacji dotyczących Istotnej Ludzkiej Kontroli (MHC), atrybucji odpowiedzialności oraz wymogów wyjaśnialności AI w kontekście Prawa Konfliktów Zbrojnych.

Analiza wykazała rozbieżność regulacyjną między imperatywem technologicznym NATO a postulatem normatywnym ONZ, kluczowym wynikiem czego jest luka normatywna – brak prawnie wiążącej, operacyjnie wdrażalnej definicji MHC. Sytuacja ta prowadzi do ryzyka pustej odpowiedzialności oraz erozji zasad rozróżnienia i proporcjonalności MPH. W kontekście krajowym, Strategia Sztucznej Inteligencji do roku 2039, mimo słuszności strategicznej, nie posiada aktów wykonawczych operacjonalizujących wymogi MHC.

W konkluzji, stwierdzono, że skuteczność implementacji AI jest ściśle uwarunkowana pilnym przejściem od deklaracji strategicznych do wiążących regulacji wykonawczych na poziomie krajowym oraz ustanowieniem prawnie wiążącej definicji MHC, by zapewnić pełną zgodność systemów z MPH.

Słowa kluczowe

sztuczna inteligencja, autonomiczne systemy uzbrojenia, prawo konfliktów zbrojnych, regulacje prawne AI, Strategia AI NATO

Wprowadzenie

Sztuczna inteligencja (AI) rewolucjonizuje charakter operacji wojskowych, wprowadzając radykalne zmiany w procesach decyzyjnych, rozpoznaniu i użyciu siły. Jej wdrażanie w systemach militarnych, zwłaszcza wperspektywie autonomicznych systemów uzbrojenia (LAWS¹), stawia przed społecznością międzynarodową i państwami narodowymi fundamentalne wyzwania w zakresie prawa międzynarodowego, odpowiedzialności i etyki. Centralnym problemem

¹ Lethal Autonomous Weapons Systems.

badawczym niniejszego artykułu jest identyfikacja i analiza ram prawnych oraz regulacyjnych na poziomie międzynarodowym i krajowym, które mają za zadanie określić szybkie tempo innowacji technologicznej w sektorze obronnym.

Wprowadzenie technologii sztucznej inteligencji do działań zbrojnych jest napędzane nie tylko imperatywem militarnym, ale i gwałtownie rosnącym rynkiem. Raporty branżowe, takie jak analiza Artificial Intelligence In Military Market (2025 – 2030)², potwierdzają rosnące tempo i skalę inwestycji. Globalne wydatki na militarną AI przechodzą obecnie fazę intensywnej ekspansji geometrycznej, a według prognoz, rynek ten odnotowuje szybkie, dwucyfrowe roczne wskaźniki wzrostu³. Ten trend jest stymulowany przez strategiczną rywalizację mocarstw i ich dążenie do przewagi decyzyjnej. Ta dynamika rynkowa wymusza na państwach, w tym Polsce, przyspieszenie prac nad implementacją AI w Siłach Zbrojnych RP, obejmując szerokie spektrum zastosowań – od zautomatyzowanej analizy danych z rozpoznania (ISR⁴), przez optymalizację logistyki, aż po autonomiczne systemy bojowe. Taki wzrost komercyjnej i strategicznej presji na wdrożenie AI bezpośrednio przekłada się na pilną potrzebę ustanowienia klarownych granic prawnych i mechanizmów kontrolnych.

Obecna sytuacja w sektorze obronnym charakteryzuje się znaczącym rozdziwieniem między tempem innowacji technologicznej a powolnością procesów legislacyjnych. O ile strategiczne dokumenty międzynarodowe i krajowe wyrażają poparcie dla rozwoju AI, o tyle nie dostarczają wystarczająco szczegółowych i prawnie wiążących rozwiązań dla kluczowych dylematów. Niewystarczające regulacje prawne i brak spójnych ram odpowiedzialności dotyczących wykorzystania algorytmów sztucznej inteligencji w systemach bojowych stanowią kluczową barierę prawną, generując ryzyko niezgodności z Międzynarodowym Prawem Humanitarnym (MPH) oraz erozję zasady „Istotnej Ludzkiej Kontroli” (MHC⁵) nad krytycznymi funkcjami bojowymi.

Celem niniejszych badań jest dokonanie szczegółowej analizy i porównania kluczowych dokumentów prawnych i strategicznych, na poziomie międzynarodowym i krajowym, które ustanawiają zasady wykorzystania AI w działaniach zbrojnych, uwzględniając dynamiczny wzrost sektora militarnego AI. Dążono do ustalenia, w jakim zakresie istniejące regulacje międzynarodowe oraz krajowe skutecznie równoważą dążenie do innowacji wojskowej w zakresie

² Grand View Research, Artificial Intelligence in Military Market (2025 – 2030) Size, Share & Trends Analysis Report By Offering (Hardware, Software, Services), By Application (Warfare Platform, Logistics & Transportation), By Technology, By Platform, By Installation, By Region, And Segment Forecasts, Report ID GVR-4-68040-021-3, Grand View Research [online], San Francisco, 2025, <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-military-market-report> [dostęp: 10.11.2025].

³ Ibidem.

⁴ K. Matela, Wybrane aspekty systemów wywiadu, obserwacji i rozpoznania (ISR), „Wiedza Obronna” 2021, t. 276, nr 3, s. 239-252, <https://doi.org/10.34752/2021-I276>.

⁵ H.M. Roff, R. Moyes, Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016, <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf> [dostęp: 10.11.2025].

AI z wymogami kontroli, odpowiedzialności i przestrzegania Międzynarodowego Prawa Humanitarnego w obliczu delegowania części decyzji bojowych maszynom.

W świetle powyższego, główny problem badawczy artykułu można sformułować w postaci następującego pytania: W jaki sposób istniejące ramy prawne i regulacyjne na poziomie międzynarodowym i krajowym skutecznie równoważą dążenie do innowacji wojskowej w zakresie AI z wymogami kontroli, odpowiedzialności i przestrzegania Międzynarodowego Prawa Humanitarnego, oraz jakie luki normatywne wynikają z nieskoordynowanego tempa rozwoju technologicznego i legislacyjnego?

W wyniku analizy wstępnych założeń badawczych sformułowano główną hipotezę badawczą w następującej postaci: Zastosowanie sztucznej inteligencji w operacjach wojskowych, choć wzmacnia bezpieczeństwo narodowe poprzez przewagę technologiczną, jest ściśle ograniczone przez istniejące ramy regulacyjne, a skuteczność jej implementacji w systemie obronnym państwa jest uwarunkowana pilnym ustanowieniem spójnych, krajowych mechanizmów kontroli i odpowiedzialności, które harmonizują z wymogiem Istotnej Ludzkiej Kontroli wynikającym z zobowiązań sojuszniczych i MPH.

Przegląd literatury

Niniejszy przegląd literatury ma na celu identyfikację i uzasadnienie wyboru kluczowych dokumentów i źródeł normatywnych, stanowiących fundament analizy ram prawnych wykorzystania sztucznej inteligencji w działaniach zbrojnych. Przyjęte podejście badawcze wymaga koncentracji nie tylko na tradycyjnej literaturze naukowej, ale przede wszystkim na oficjalnych dokumentach strategicznych, rezolucjach międzynarodowych i politykach obronnych, które bezpośrednio kształtują regulacje dotyczące militarnych zastosowań sztucznej inteligencji. Podstawę analizy stanowią normy Międzynarodowego Prawa Humanitarnego, a w szczególności zapisy takie jak Klauzula Martensa i Artykuł 36 Protokołu i Dodatkowego. MPH, choć nie wymienia wprost AI, narzuca na państwa obowiązek prawnej oceny każdej nowej broni, co czyni jego zasady (rozdzielenie, proporcjonalność) pierwszym i nieodzownym punktem odniesienia dla wszystkich dalszych regulacji.

Główny korpus analizowanej literatury stanowią dokumenty wykraczające poza ogólne ramy MPH, a koncentrujące się na wyzwaniach związanych z ludzką kontrolą, odpowiedzialnością i rozliczalnością w dobie autonomii wojskowej. W związku z tym, wybrano kluczowe akty regulacyjne z poziomu globalnego i sojuszniczego. Na poziomie ONZ jest to Rezolucja Zgromadzenia Ogólnego nr 78/241⁶, która stanowi wyraz globalnej woli politycznej dążącej do ustanowienia

⁶ Organizacja Narodów Zjednoczonych, Resolution 78/241, Lethal autonomous weapons systems, Zgromadzenie Ogólne, 22 grudnia 2023 r., <https://docs->

wiążących norm dotyczących LAWS. Uzupełnieniem tego jest Deklaracja Polityczna w sprawie Odpowiedzialnego Wykorzystania Wojskowego Sztucznej Inteligencji i Autonomii⁷, jako kluczowy, choć niewiążący prawnie, akt wyznaczający standardy etyczne i operacyjne dla państw, w tym Polski.

Kolejnym filarem literatury są dokumenty instytucjonalne, które bezpośrednio wpływają na polską doktrynę obronną. Wybrano Strategię Sztucznej Inteligencji NATO⁸ jako nadrzędny dokument sojuszniczy, definiujący zasady etyczne (legalność, odpowiedzialność, przewidywalność) dla wdrażania AI przez członków Paktu. Na poziomie krajowym, analizie poddano Resortową strategię sztucznej inteligencji do roku 2039 Ministerstwa Obrony Narodowej⁹, która jest kluczowym, choć strategicznym, dokumentem określającym kierunki rozwoju polskiej armii w tym zakresie. Analiza ta umożliwia zidentyfikowanie luki regulacyjnej między ogólnymi aspiracjami sojuszu a konkretnymi krajowymi planami wdrożeniowymi w Polsce.

Wreszcie, przegląd literatury uwzględnia również akty pośrednio regulujące bezpieczne i legalne wykorzystanie AI. W tym kontekście, analizowane są porozumienia międzynarodowe, takie jak Memorandum of Understanding on Cyber Security of Information Technology Cooperation¹⁰ podpisane przez polski rząd, które wzmacnia ramy prawne w zakresie cyberbezpieczeństwa i wiarygodności danych. Ponieważ militarna AI jest głęboko zależna od jakości i bezpieczeństwa przetwarzanych informacji, te akty regulacyjne stanowią istotny element literatury, pozwalający ocenić, w jakim stopniu prawne ramy ochrony danych wspierają legalność i niezawodność systemów autonomicznych.

Metodologia

W procesie badawczym zastosowano metody charakterystyczne dla analizy regulacyjnej w naukach o bezpieczeństwie i prawie międzynarodowym, co pozwoliło na kompleksową ocenę prawnych ram wykorzystania AI w działaniach zbrojnych.

library.unoda.org/General_Assembly_First_Committee_-Seventy-Ninth_session_(2024)/78-241-US-EN.pdf [dostęp: 10.11.2025].

⁷ U.S. Department of State, Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, 2023.

⁸ NATO, Summary of the NATO AI Strategy, 2021, https://www.nato.int/cps/en/natohq/official_texts_187617.htm [dostęp: 10.11.2025]; NATO, Summary of NATO's revised AI Strategy, 2024, https://www.nato.int/cps/en/natohq/official_texts_227237.htm [dostęp: 10.11.2025].

⁹ Ministerstwo Obrony Narodowej, Resortowa strategia sztucznej inteligencji do roku 2039, Warszawa 2024.

¹⁰ U.S. Department of Homeland Security, DHS, Polish Ministry of Digital Affairs Sign Memorandum of Understanding [online], 3.10.2024, <https://www.dhs.gov/archive/news/2024/10/03/dhs-polish-ministry-digital-affairs-sign-memorandum-understanding-promoting> [dostęp: 10.11.2025].

Podstawową metodą była analiza dokumentów źródłowych, skupiająca się na identyfikacji norm prawnych, standardów etycznych i zobowiązań strategicznych zawartych w kluczowych dokumentach międzynarodowych i krajowych. Krytycznej ocenie poddano treść, zakres obowiązywania oraz stopień szczegółowości regulacji, w szczególności w obszarach:

- kontroli człowieka nad użyciem siły;
- atrybucji odpowiedzialności za naruszenia Prawa Konfliktów Zbrojnych (PKZ¹¹);
- wymogów transparentności i wyjaśnialności algorytmów militarnych (Explainable AI – XAI¹²).

Kolejnym etapem było zastosowanie metody porównawczej. Metoda ta została użyta do zestawienia i kontrastowania podejść regulacyjnych na poziomie międzynarodowym (sojuszniczym i globalnym) z podejściem narodowym (polskim). Porównanie koncentrowało się na różnicach w akcentach – na przykład na stopniu nacisku na innowacje w stosunku do kontroli prawnej i etycznej.

Zastosowano również metodę analizy i syntezy w celu rozłożenia kompleksowego problemu na czynniki składowe (np. odpowiedzialność operacyjna, prawna, etyczna) i ich ponowne zestawienie. Analiza umożliwiła precyzyjne wydzielenie kluczowych mechanizmów kontrolnych w każdym dokumencie, natomiast synteza pozwoliła na ujęcie ram prawnych jako integralnego elementu systemu bezpieczeństwa, który musi zapewnić ciągłość odpowiedzialności w łańcuchu dowodzenia w środowisku zautomatyzowanym.

Wnioskowanie logiczne, w tym indukcja i dedukcja, posłużyło do sformułowania ostatecznych wniosków i weryfikacji przyjętej hipotezy badawczej w świetle teorii bezpieczeństwa i prawa międzynarodowego.

Wyniki i dyskusja

Analiza porównawcza kluczowych dokumentów międzynarodowych i krajowych potwierdza, że strategiczne wykorzystanie sztucznej inteligencji w działaniach zbrojnych jest ściśle uwarunkowane stworzeniem prawnie ugruntowanych ram kontroli. Wyniki badań wskazują na istnienie rozbieżności regulacyjnej między imperatywem utrzymania przewagi technologicznej, silnie promowanym przez Strategię Sztucznej Inteligencji NATO, a dążeniem do ustanowienia rygorystycznych norm kontroli, widocznym w inicjatywach globalnych, takich jak Rezolucja Zgromadzenia Ogólnego ONZ nr 78/241 i Deklaracja Polityczna w sprawie Odpowiedzialnego Wykorzystania

¹¹ W literaturze przedmiotu terminy Prawo Konfliktów Zbrojnych (PKZ) oraz Międzynarodowe Prawo Humanitarne (MPH) są często używane zamiennie i odnoszą się do tego samego zbioru norm prawnych regulujących prowadzenie działań wojennych i ochronę ofiar. W niniejszej pracy terminy te będą używane naprzemiennie jako synonimy.

¹² N.G. Wood, Explainable AI in the military domain, „Ethics Inf Technol” 2024, t. 26, nr 29, <https://doi.org/10.1007/s10676-024-09762-w>.

Wojskowego Sztucznej Inteligencji i Autonomii. Ta fundamentalna rozbieżność generuje niestabilność prawną dla państw, które muszą równocześnie spełniać wymogi sojusznicze i przestrzegać zasad Międzynarodowego Prawa Humanitarnego, którego zasady (jak Klauzula Martensa i Artykuł 36 Protokołu i Dodatkowego) stanowią punkt wyjścia dla oceny każdej nowej broni.

Kluczowym wynikiem jest stwierdzenie luki regulacyjnej w kontekście Istotnej Ludzkiej Kontroli. Choć eksperci prawa, tacy jak Heather M. Roff i Richard Moyes w raporcie *Meaningful Human Control, Artificial Intelligence and Autonomous Weapons*, argumentują za pilnym ustanowieniem prawnie wiążących norm MHC w celu utrzymania indywidualnej odpowiedzialności karnej (trudnej do atrybucji w przypadku autonomicznych decyzji algorytmów), żadne z analizowanych porozumień ani strategii nie dostarcza jednolitej, operacyjnie wdrażalnej definicji tego, jaki poziom interwencji ludzkiej jest prawnie wymagany. Interpretacja wyników wskazuje, że ta luka prowadzi do ryzyka „pustej odpowiedzialności” (*responsibility gap*) w świetle Prawa Konfliktów Zbrojnych. Jednocześnie, Raport Sekretarza Generalnego ONZ identyfikuje, że państwa członkowskie zgadzają się co do kluczowych elementów MHC, które powinny obejmować zdolność ludzi do sprawowania nadzoru, zachowania osądu wymaganego przez MPH oraz możliwości ograniczania typów zadań i celów¹³. Mimo tego szerokiego konsensusu co do elementów składowych, nie zdefiniowano jednoznacznej i wiążącej definicji samego pojęcia MHC, co jest kluczowe dla legalnego i bezpiecznego wykorzystywania systemów autonomicznych w działaniach zbrojnych.

W kontekście polskim, Resortowa Strategia Sztucznej Inteligencji do roku 2039, choć słusznie nakierowana na rozwój zdolności, ma charakter strategiczny, a nie regulacyjny. Jej zgodność z MPH i NATO jest deklaratywna, lecz brakuje szczegółowych aktów wykonawczych przekładających ogólne zasady MHC na krajowe zasady użycia siły (ROE¹⁴). Rosnące, szybkie tempo globalnych inwestycji w militarną AI generuje presję operacyjną, która może prowadzić do wdrażania systemów przed wypełnieniem tej luki prawnej.

Porównanie Polskich i Zagranicznych Regulacji

Strategia Sztucznej Inteligencji NATO z 2021 roku ustanowiła kluczowe ramy sojusznicze, koncentrując się na odpowiedzialnej adaptacji nowych technologii w celu utrzymania przewagi technologicznej Sojuszu. Dokument ten wprowadził sześć nadrzędnych zasad etycznych wdrażania AI: legalność, odpowiedzialność i rozliczalność, wyjaśnialność i identyfikowalność,

¹³ United Nations, Report of the Secretary-General on Lethal Autonomous Weapons Systems, A/79/88, 2024, s. 6-7, <https://documents.un.org/doc/undoc/gen/n24/154/32/pdf/n2415432.pdf> [dostęp: 10.11.2025].

¹⁴ Rules of Engagement.

niezawodność, zarządzalność i łagodzenie uprzedzeń (bias)¹⁵. Zasady te mają zapewnić, że zastosowanie AI będzie zgodne z prawem międzynarodowym i wartościami demokratycznymi. Zrewidowana Strategia Sztucznej Inteligencji NATO z 2024 roku wzmocniła te zasady, koncentrując się na operacjonalizacji AI, zwiększając nacisk na interoperacyjność i bezpieczeństwo łańcucha dostaw. Potwierdziła konieczność zachowania ludzkiego nadzoru nad systemami o kluczowym znaczeniu bojowym, unikając jednak formułowania prawnie wiążącej definicji istotnej kontroli, co pozostawia państwom członkowskim elastyczność w implementacji krajowej.

Na forum globalnym dominują postulaty zmierzające do ustanowienia wiążących norm prawnych. Rezolucja Zgromadzenia Ogólnego ONZ nr 78/241 w sprawie LAWS jest wyrazem międzynarodowej woli politycznej do pilnego opracowania instrumentu regulacyjnego (np. traktatu) w celu rozwiązania wyzwań prawnych i etycznych związanych z systemami uzbrojenia podejmującymi krytyczne decyzje bez udziału człowieka. Dodatkowo, Raport Sekretarza Generalnego ONZ z 2024 roku, analizując stanowiska państw członkowskich, wskazał na istnienie szerokiego konsensusu co do konieczności zachowania zdolności do sprawowania nadzoru. Kontrola ta musi obejmować¹⁶:

- wystarczające informacje, w tym dotyczące zdolności systemu uzbrojenia oraz kontekstu operacyjnego, umożliwiające zapewnienie zgodności z prawem międzynarodowym;
- zdolność do wykonywania oceny i stosowania własnego osądu w zakresie wymaganym przez Międzynarodowe Prawo Humanitarne;
- zdolność do ograniczania rodzajów zadań oraz celów, których wykonanie system może realizować;
- zdolność do nałożenia ograniczeń co do czasu trwania, zasięgu geograficznego oraz skali użycia;
- zdolność do zmiany lub modyfikacji celu operacyjnego bądź misji systemu;
- zdolność do przerwania działania lub dezaktywacji systemu.

Stany Zjednoczone jako jeden z liderów technologicznych, zainicjowały Deklarację Polityczną w sprawie Odpowiedzialnego Wykorzystania Wojskowego AI i Autonomii, która obecnie (stan na rok 2025) została podpisana przez ponad 50 państw. Dokument ten ma charakter politycznie wiążący, lecz niewiążący prawnie, koncentrując się na najlepszych praktykach i budowaniu zaufania. Deklaracja wyznacza standardy dla transparentności, zarządzania ryzykiem i konieczności ludzkiego osądu w krytycznych decyzjach operacyjnych. Wymaga od sygnatariuszy, w tym Polski, aby zapewniły odpowiedzialność i rozliczalność za użycie systemów AI/autonomicznych. To podejście kładzie nacisk na

¹⁵ NATO, Summary of the NATO AI Strategy, 2021, https://www.nato.int/cps/en/natohq/official_texts_187617.htm [dostęp: 10.11.2025].

¹⁶ United Nations, Report of the Secretary-General on Lethal Autonomous Weapons Systems, A/79/88, 2024, s. 6-7, <https://documents.un.org/doc/undoc/gen/n24/154/32/pdf/n2415432.pdf> [dostęp: 10.11.2025].

samoregulację i dobrowolne przyjęcie standardów, zamiast na narzucone z góry, sztywne regulacje prawno-traktatowe.

Krajowym dokumentem strategicznym, wyznaczającym kierunki wdrażania AI w Siłach Zbrojnych RP, jest Resortowa Strategia Sztucznej Inteligencji do roku 2039 Ministerstwa Obrony Narodowej. Jest to akt o charakterze planistyczno-rozwojowym, a nie regulacyjnym, skupiający się na rozwoju zdolności i integracji AI z procesami dowodzenia, rozpoznania i wsparcia operacyjnego. Wśród celów operacyjnych Strategia wymienia m.in. automatyzację analizy danych wywiadowczych oraz wsparcie dla dowódców w podejmowaniu decyzji, co bezpośrednio przekłada się na zastosowanie AI w kluczowych obszarach działań zbrojnych. W celu realizacji tych założeń przewidziano m.in. utworzenie Centrum Implementacji Sztucznych Inteligencji (CISI)¹⁷. Polityczny mandat dla tej transformacji wynika z Strategii Bezpieczeństwa Narodowego RP z 2020 roku, która wskazuje na konieczność transformacji technologicznej Sił Zbrojnych RP i modernizacji systemów dowodzenia, które w praktyce opierają się na zaawansowanych technologiach, w tym na AI¹⁸.

Prawne luki i wyzwania w kontekście MPH

Implementacja sztucznej inteligencji i autonomicznych systemów uzbrojenia w operacjach wojskowych stawia przed Międzynarodowym Prawem Humanitarnym wyzwania o charakterze fundamentalnym. MPH, choć nie zawiera bezpośrednich zapisów regulujących AI, jest uniwersalnie stosowalne do każdej nowej broni, a jego zasady stanowią bazę dla oceny systemów LAWS. Zasada rozróżnienia jest tu kamieniem węgielnym, gdyż wymaga, by strony konfliktu odróżniały ludność cywilną od kombatanów i obiekty cywilne od celów wojskowych. Artykuł 48 Protokołu i Dodatkowego stanowi, że strony konfliktu mają kierować swoje operacje wyłącznie przeciwko celom wojskowym¹⁹. Idąc tym tropem, zasada ostrożności, skodyfikowana w Artykule 57 Protokołu i Dodatkowego, nakłada na strony obowiązek podjęcia wszelkich możliwych do wykonania środków w celu weryfikacji celów, wyboru metod minimalizujących straty cywilne oraz powstrzymania się od ataków, gdy spodziewane szkody cywilne są nieproporcjonalne do korzyści wojskowej²⁰. Ponieważ wojskowe rozwiązania AI nie gwarantują stuprocentowej skuteczności namierzania celów, zasady te są bezpośrednio zagrożone. Dodatkowe zasady istotne dla regulacji autonomicznego uzbrojenia obejmują zakaz broni powodującej niepotrzebne

¹⁷ Ministerstwo Obrony Narodowej, Resortowa strategia sztucznej inteligencji do roku 2039, Warszawa 2024, s. 22.

¹⁸ Rzeczpospolita Polska, Strategia Bezpieczeństwa Narodowego Rzeczypospolitej Polskiej, Warszawa 2020, s. 7-8.

¹⁹ Protokół dodatkowy i do Konwencji genewskich z 12 sierpnia 1949 r. dotyczący ochrony ofiar międzynarodowych konfliktów zbrojnych, sporządzony w Genewie dnia 8 czerwca 1977r., art. 48.

²⁰ Ibidem, art. 57.

cierpienia, wymóg prawnego przeglądu nowej broni na podstawie Artykułu 36 Protokołu i Dodatkowego²¹ oraz ochronę wynikającą z Klauzuli Martensa²².

Najbardziej krytyczną luką prawną jest problem atrybucji odpowiedzialności karnej za naruszenia Prawa Konfliktów Zbrojnych popełnione przez systemy AI. PKZ opiera się na zasadzie indywidualnej odpowiedzialności karnej (zawartej m.in. w Statucie Rzymskim Międzynarodowego Trybunału Karnego oraz w Konwencjach Genewskich), która wymaga dowodu winy umyślnej (*mens rea*) lub rażącego niedbalstwa ze strony człowieka (dowódcy, operatora). W sytuacji, gdy algorytm podejmie nieprzewidzianą decyzję skutkującą atakiem naruszającym PKZ, ustalenie winnego w łańcuchu dowodzenia staje się niezwykle trudne. Jak podkreślają eksperci, ta niepewność prowadzi do ryzyka pustej odpowiedzialności, gdzie odpowiedzialność prawna ulega rozmyciu, podważając skuteczność MPH.

Podstawową przeszkodą w regulacji technicznej jest brak jednolitej, prawnie wiążącej definicji Istotnej Ludzkiej Kontroli. Mimo osiągnięcia konsensusu państw członkowskich ONZ co do elementów kontroli (np. możliwość dezaktywacji, zachowanie osądu), nie ma porozumienia co do poziomu i natury tej kontroli. Zgodnie z analizą Lieber Institute, brak jasnych standardów MHC uniemożliwia państwom rzetelną ocenę, czy ich systemy autonomiczne spełniają wymogi zasady rozróżnienia i proporcjonalności w ataku, co jest centralne dla MPH²³. To jest kluczowy problem normatywny, który pozwala na zbyt luźną interpretację MHC przez państwa dążące do przewagi technologicznej.

Dodatkowym wyzwaniem jest wyjaśnialność i niezawodność samych algorytmów, co bezpośrednio wpływa na ich legalność. Zaawansowane systemy AI często działają w sposób nieprzezroczysty (tzw. *black box*), co uniemożliwia prawnikom pełne prześledzenie procesu decyzyjnego, który doprowadził do użycia siły. Jak wynika z analizy opublikowanej na łamach *European Journal of Law and Political Science*, ta bariera transparentności jest niezgodna z wymogami MPH dotyczącymi badania naruszeń, ponieważ nie pozwala na rzetelną weryfikację czy zasady prawa były przestrzegane²⁴.

Ponadto, systemy AI są podatne na uprzedzenia, odzwierciedlające historyczne lub społeczne tendencje zawarte w danych szkoleniowych. Jeśli algorytm został przeszkolony na danych, które nadmiernie identyfikują pewne obiekty lub grupy jako zagrożenie, może to skutkować automatycznym

²¹ Ibidem, art. 36.

²² IV Konwencja haska dotycząca praw i zwyczajów wojny lądowej z dnia 18 października 1907 r., preambuła (tzw. „klauzula Martensa”), (Dz.U. 1927, nr 21, poz. 161).

²³ M. Zamani, How Meaningful Is ‘Meaningful Human Control’ in LAWS Regulation?, *Articles of War – Lieber Studies*, „Lieber Institute at West Point” [online], 26.03.2025, <https://lieber.westpoint.edu/how-meaningful-is-meaningful-human-control-laws-regulation> [dostęp: 10.11.2025].

²⁴ B. Hamad, The Impact of Digital Technology on International Humanitarian Law: Ethical and Legal Implications of Autonomous Weapons Systems, *„European Journal of Law and Political Science”* 2025, t. 4, nr 4, s. 1-14, <https://doi.org/10.24018/ejpolitics.2025.4.4.182>.

naruszeniem zasady rozróżnienia i neutralności działania, co stanowi bezpośrednie zagrożenie prawne. Wyzwanie to pokazuje, jak wady technologiczne bezpośrednio przekładają się na zagrożenia zgodności z prawem, czyniąc systemy AI potencjalnie niezgodnymi z podstawowymi normami MPH.

Wnioski i rekomendacje

Przeprowadzona analiza i porównanie kluczowych dokumentów prawnych oraz strategicznych potwierdza, że zastosowanie sztucznej inteligencji w operacjach wojskowych, choć strategicznie niezbędne, jest ściśle ograniczone przez istniejące ramy prawne, a skuteczność jej implementacji jest uwarunkowana pilnym ustanowieniem spójnych mechanizmów kontroli. Głównym ustaleniem jest rozbieżność regulacyjna między sojuszniczym imperatywem technologicznym promowanym przez NATO a globalnym postulatem normatywnym ONZ. Ta fundamentalna niezgodność prowadzi do kluczowej luki normatywnej: brak jest prawnie wiążącej, operacyjnie wdrażalnej definicji tego, jaki poziom interwencji człowieka jest prawnie wymagany. Bez jasnych standardów MHC, tradycyjne zasady Międzynarodowego Prawa Humanitarnego, takie jak zasada rozróżnienia i zasada ostrożności, ulegają rozmyciu, co rodzi ryzyko pustej odpowiedzialności.

W kontekście krajowym, Resortowa Strategia Sztucznej Inteligencji do roku 2039, mimo słusznych celów operacyjnych (w tym utworzenia CISI) i deklarowanej zgodności z prawem międzynarodowym, ma charakter planistyczny, a nie regulacyjny. Oznacza to, że ambicje modernizacyjne nie są poparte szczegółowymi aktami wykonawczymi i krajowymi doktrynami operacjonalizującymi wymogi kontroli człowieka i atrybucji odpowiedzialności w łańcuchu dowodzenia. Ta nieobecność krajowych norm stanowi barierę prawną dla bezpiecznego i zgodnego z MPH wdrożenia AI. Dodatkowo, wyzwania technologiczne związane z wyjaśnialnością i uprzedzeniami algorytmicznymi stanowią bezpośrednie zagrożenie prawne, ponieważ utrudniają badanie naruszeń MPH i mogą prowadzić do automatycznego łamania zasady rozróżnienia.

Na podstawie dokonanej analizy, w celu osiągnięcia pełnej zgodności wdrażanych systemów AI z prawem międzynarodowym i sojuszniczymi zobowiązaniami, sformułowano następujące rekomendacje dla decydentów. Najwyższym priorytetem jest pilne ustanowienie i wdrożenie prawnie wiążącej definicji Istotnej Ludzkiej Kontroli, precyzującej minimalny poziom nadzoru ludzkiego nad krytycznymi funkcjami bojowymi. Równolegle, należy opracować doktryny wykonawcze i zasady użycia siły, które w jasny sposób przypiszą odpowiedzialność prawną (w tym karną) w przypadku nieprzewidzianych działań algorytmów.

W obszarze technicznym, konieczne jest ustanowienie prawnie wiążących standardów weryfikacji i certyfikacji dla algorytmów AI przed ich wdrożeniem. Wymogi te muszą obejmować audytowalność procesów decyzyjnych oraz testowanie systemów pod kątem neutralizacji uprzedzeń algorytmicznych, co jest warunkiem wstępnym zgodności z zasadami rozróżnienia i ostrożności. Ponadto, Polska powinna aktywnie uczestniczyć w pracach ONZ i forach CCW²⁵ w celu wypracowania wspólnego, prawnie wiążącego instrumentu międzynarodowego, który ujednolici ramy regulacyjne, minimalizując ryzyko prawne i etyczne w regionie.

Bibliografia

- Grand View Research, Artificial Intelligence in Military Market (2025 – 2030) Size, Share & Trends Analysis Report By Offering (Hardware, Software, Services), By Application (Warfare Platform, Logistics & Transportation), By Technology, By Platform, By Installation, By Region, And Segment Forecasts, Report ID GVR-4-68040-021-3, Grand View Research [online], San Francisco, 2025, <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-military-market-report>.
- Hamad B., The Impact of Digital Technology on International Humanitarian Law: Ethical and Legal Implications of Autonomous Weapons Systems, „European Journal of Law and Political Science” 2025, t. 4, nr 4, <https://doi.org/10.24018/ejpolitics.2025.4.4.182>.
- IV Konwencja haska dotycząca praw i zwyczajów wojny lądowej z dnia 18 października 1907 r. (Dz.U. 1927, nr 21, poz. 161).
- Matela K., Wybrane aspekty systemów wywiadu, obserwacji i rozpoznania (ISR), „Wiedza Obronna” 2021, t. 276, nr 3, <https://doi.org/10.34752/2021-l276>.
- Ministerstwo Obrony Narodowej, Resortowa strategia sztucznej inteligencji do roku 2039, Warszawa 2024.
- NATO Artificial Intelligence Strategy, 2021.
- NATO's revised Artificial Intelligence Strategy, 2024.
- Organizacja Narodów Zjednoczonych, Resolution 78/241, Lethal autonomous weapons systems, Zgromadzenie Ogólne, 22 grudnia 2023 r., [https://docs-library.unoda.org/General_Assembly_First_Committee_-_Seventy-Ninth_session_\(2024\)/78-241-US-EN.pdf](https://docs-library.unoda.org/General_Assembly_First_Committee_-_Seventy-Ninth_session_(2024)/78-241-US-EN.pdf).
- Pilarski G., Zawadzki T., Zdolności do prowadzenia pełnego spektrum działań militarnych w cyberprzestrzeni, „Wiedza Obronna” 2025, t. 292, nr 3, <https://doi.org/10.34752/2025-3-8>.
- Protokół dodatkowy i do Konwencji genewskich z 12 sierpnia 1949 r. dotyczący ochrony ofiar międzynarodowych konfliktów zbrojnych, sporządzony w Genewie dnia 8 czerwca 1977 r. (Dz.U. 1992, nr 41, poz. 175).
- Roff H. M., Moyes R., Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. Briefing paper prepared for the Informal Meeting of Experts on Lethal

²⁵ Konwencja o zakazie lub ograniczeniu użycia pewnych broni konwencjonalnych, które mogą być uważane za powodujące nadmierne cierpienia lub mające niekontrolowane skutki.

- Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016, <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>.
- Rzeczpospolita Polska, Strategia Bezpieczeństwa Narodowego Rzeczypospolitej Polskiej, Warszawa 2020.
- U.S. Department of Homeland Security, DHS, Polish Ministry of Digital Affairs Sign Memorandum of Understanding [online], 3.10.2024, <https://www.dhs.gov/archive/news/2024/10/03/dhs-polish-ministry-digital-affairs-sign-memorandum-understanding-promoting>.
- U.S. Department of State, Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, 2023.
- United Nations, Report of the Secretary-General on Lethal Autonomous Weapons Systems, A/79/88, 2024, <https://documents.un.org/doc/undoc/gen/n24/154/32/pdf/n2415432.pdf>.
- Wood N.G., Explainable AI in the military domain, „Ethics Inf Technol” 2024, t. 26, nr 29, <https://doi.org/10.1007/s10676-024-09762-w>.
- Zamani M., How Meaningful Is ‘Meaningful Human Control’ in LAWS Regulation?, Articles of War – Lieber Studies, Lieber Institute at West Point [online], 26.03.2025, <https://lieber.westpoint.edu/how-meaningful-is-meaningful-human-control-laws-regulation>.