Social Dissertations 2026, Volume 20
eISSN 2657-9332

SECURITY THREATS AND RISKS IN EDGE AI

ZAGROŻENIA I RYZYKA BEZPIECZEŃSTWA W EDGE AI

Wiktor Sędkowski[1,2 A-F]

[1]Department of Cybersecurity, Opole University of Technology, Poland
[1]Katedra Cyberbezpieczeństwa, Politechnika Opolska, Polska
[2]TECH, Nokia Solutions and Networks, Poland
[2]TECH, Nokia Solutions and Networks, Polska

Sędkowski, W. (2026). Security threats and risks in Edge AI / Zagrożenia i ryzyka bezpieczeństwa w Edge AI. Social Dissertations / Rozprawy Społeczne, 20(1), 97-107. https://doi.org/10.29316/rs/220428

Authors' contribution / Wkład autorów:
A. Study design / Zaplanowanie badań
B. Data collection / Zebranie danych
C. Data analysis / Dane – analiza i statystyki
D. Data interpretation / Interpretacja danych
E. Preparation of manuscript / Przygotowanie artykułu
F. Literature analysis / Wyszukiwanie i analiza literatury
G. Funds collection / Zebranie funduszy

Tables / Tabele: 3
Figures / Ryciny: 2
References / Literatura: 21
Submitted / Otrzymano: 2025-12-10
Accepted / Zaakceptowano: 2026-04-08

Abstract: The aim of this article is to conduct a comprehensive analysis of security threats affecting end devices in Edge AI environments, with particular emphasis on their specific vulnerabilities resulting from architectural and technical limitations.
Materials and methods: The study was based on a review of the literature and the latest reports on Edge AI security. The STRIDE methodology was used for the systematic identification and classification of threats, analyzing all layers of the distributed AI systems architecture (edge, fog, cloud).
Results: A wide spectrum of hardware, software, and operational threats was identified, including unique risks related to physical access to devices, manipulation of AI models, and data leakage.
Conclusions: Recommendations were formulated regarding protection mechanisms covering authentication, cryptography, monitoring, and redundancy. The need to continuously adapt security strategies to evolving attack techniques was emphasized as a prerequisite for ensuring the reliability of Edge AI systems.

Keywords: Edge AI, cybersecurity, threats, artificial intelligence

Streszczenie: Celem artykułu jest przeprowadzenie kompleksowej analizy zagrożeń bezpieczeństwa dotyczących urządzeń końcowych w środowiskach Edge AI, ze szczególnym uwzględnieniem ich specyficznych podatności wynikających z architektury i ograniczeń technicznych.
Materiał i metody: Badanie oparto na przeglądzie literatury oraz najnowszych raportów dotyczących bezpieczeństwa Edge AI. Do systematycznej identyfikacji i klasyfikacji zagrożeń zastosowano metodykę STRIDE, analizując wszystkie warstwy architektury rozproszonych systemów AI (edge, fog, cloud).
Wyniki: Zidentyfikowano szerokie spektrum zagrożeń sprzętowych, programowych i operacyjnych, w tym unikalne ryzyka związane z fizycznym dostępem do urządzeń, manipulacją modelami AI i wyciekiem danych.
Wnioski: Sformułowano zalecenia dotyczące mechanizmów ochrony obejmujących uwierzytelnianie, kryptografię, monitoring i redundancję. Podkreślono konieczność ciągłego dostosowywania strategii bezpieczeństwa do ewoluujących technik ataków jako warunku zapewnienia niezawodności systemów Edge AI.

Słowa kluczowe: Edge AI, cyberbezpieczeństwo, zagrożenia, sztuczna inteligencja

Adres korespondencyjny: Wiktor Sędkowski, Katedra Cyberbezpieczeństwa, Politechnika Opolska, ul. Prószkowska 76, 45-758 Opole, Polska; email: w.sedkowski@po.edu.pl ORCID: 0000-0002-4543-0499

Era sprzętu AI

Edge AI, czyli sztuczna inteligencja brzegowa, polega na uruchamianiu algorytmów AI bezpośrednio na urządzeniach końcowych użytkownika. Głównie mowa tu o smartfonach, urządzeniach ubierlanych (ang. wearables), czujnikach IoT, robotach czy samochodach. Kod uruchamiany jest na nich zamiast w scentralizowanych centrach danych lub chmurze. Takie lokalne przetwarzanie umożliwia natychmiastową analizę danych, skraca czas reakcji oraz pozwala zachować wyższy poziom prywatności i bezpieczeństwa, ponieważ wrażliwe informacje nie muszą opuszczać miejsca ich powstania (Shafee i in., 2025).

Aby wspierać możliwość wykorzystywania AI na urządzeniach końcowych, producenci implementują dedykowane akceleratory AI. Najczęściej spotykanymi z nich są jednostki przetwarzania neuronowego (NPU), jednostki przetwarzania tensorów (TPU), procesory graficzne (GPU), programowalne macierze bramek (FPGA), specjalizowane układy scalone (ASIC), mikrokontrolery czy wyspecjalizowane moduły do zadań analitycznych i rozpoznawania wzorców, np. w kamerach monitoringu, systemach automatyzacji domu, urządzeniach biomedycznych czy autonomicznych pojazdach (Singh, 2023). Takie rozwiązania zapewniają użytkownikom natychmiastową reakcję urządzenia na polecenia głosowe, gesty, analizę obrazu oraz lokalne sterowanie bez konieczności przesyłania danych do zdalnych serwerów.

Tabela 1. Porównanie akceleratorów AI

Typ jednostki	Wydajność obliczeniowa dla AI (względem CPU)	Elastyczność programowania	Efektywność energetyczna	Typowe zastosowania AI
CPU (Standardowy procesor)	Bazowa	Wysoka	Niska	Sterowanie, zadania ogólne
GPU (Procesor graficzny)	~10-30x szybsza	Średnia (równoległe programowanie)	Średnia do wysokiej	Trenowanie modeli, przetwarzanie dużych danych, wizja komputerowa
TPU (jednostka przetwarzania tensorów)	~50 x szybsza od CPU w TensorFlow	Niska (specjalizowany do TensorFlow)	Średnia	Duże modele uczenia głębokiego, inferencja na dużą skalę
NPU (jednostka przetwarzania neuronowego)	~10-20x szybsza od CPU	Średnia	Wysoka	Urządzenia mobilne, rozpoznawanie wzorców
FPGA (Programowalna macierz bramek)	~5-15x szybsza	Wysoka (reprogramowalna)	Wysoka	Edge AI, niskie opóźnienia, zadania specjalistyczne
ASIC (Specjalizowany układ scalony)	Najwyższa, nawet 50x+ od CPU	Niska (dedykowany sprzęt)	Najwyższa	Produkcja masowa, dedykowane akceleratory AI, IoT, autonomiczne pojazdy

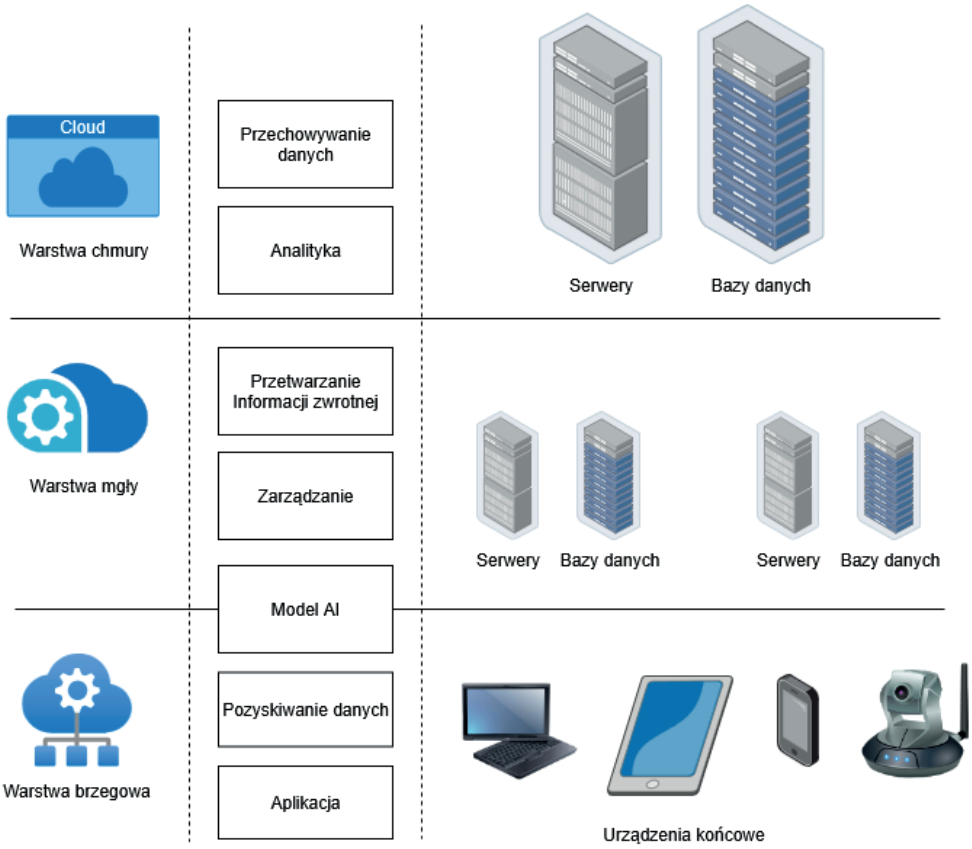
Źródło: opracowanie własne na podstawie: Hesse, 2021; Bieglow 2024.

Szybki rozwój Edge AI niesie jednak ze sobą istotne ryzyka bezpieczeństwa: podatność urządzeń na ataki fizyczne, manipulacje firmware, wycieki danych, ingerencje w modele AI czy trudności z zapewnieniem aktualności zabezpieczeń na wielu rozproszonych urządzeniach (Shafee i in., 2025; Wingarz i in., 2024). Szeroka gama zagrożeń wymaga od naukowców i praktyków opracowania dedykowanych modeli zagrożeń oraz wdrożenia mechanizmów ochrony zgodnych ze specyfiką zdecentralizowanych systemów AI.

Przegląd architektury Edge AI i typowych zastosowań

Architektura Edge Artificial Intelligence opiera się na rozproszeniu mocy obliczeniowej w bliskim sąsiedztwie urządzeń generujących dane, w przeciwieństwie do tradycyjnego modelu opartego na scentralizowanych chmurach obliczeniowych. W praktyce oznacza to, że analiza, uczenie maszynowe oraz inferencja są wykonywane na lokalnych urządzeniach końcowych lub w ich bezpośrednim otoczeniu, co znacząco redukuje opóźnienia komunikacyjne, zmniejsza obciążenie sieci oraz podnosi poziom prywatności danych osobowych (Shi i in., 2016; Shafee i in., 2025).

Architektura ta zwykle składa się z kilku warstw. Warstwa brzegowa (edge layer) obejmuje urządzenia końcowe wyposażone w wyspecjalizowane akceleratory realizujące szybkie przetwarzanie danych, rozpoznawanie wzorców czy klasyfikację bez konieczności komunikacji z chmurą. Nad warstwą brzegową występuje warstwa mgły (fog layer), czyli lokalne węzły obliczeniowe (np. serwery w pobliżu urządzeń), które mogą agregować, filtrują oraz analizują dane na poziomie pośrednim, przed ich przesłaniem do centralnej chmury. Ostatecznie za koordynacje i przetwarzanie dużych zbiorów danych klasycznie przechowywane oraz analizowane są w warstwie chmurowej (cloud layer).



Rycina 1. Architektura Edge AI

Źródło: opracowanie własne na podstawie: Shafee i in., 2025; Wingarz i in., 2024.

W literaturze naukowej i technicznej można znaleźć różne warianty architektury Edge AI, które różnią się w zależności od potrzeb aplikacji, lokalizacji infrastruktury czy wymagań dotyczących opóźnień i bezpieczeństwa. Najczęściej spotykany jest opisywany powyżej model składający

się z trzech warstw: warstwy chmurowej (cloud), warstwy mgły (fog) oraz warstwy brzegowej (Edge). Innym wariantem jest podejście oparte na wielu małych, lokalnych centrach danych (mikrosieciach), które współpracują z urządzeniami Edge i oferują funkcje uruchamiania i aktualizacji modeli AI dynamicznie, co zwiększa odporność na awarie i ataki (Jones, 2025). Kolejnym z wariantów stosowany jest w rozwiązaniach, w których większy nacisk kładzie się na przetwarzanie i uczenie lokalne w warstwie mgły (fog), traktując ją jako kluczowy punkt sterowania danymi i regulacji lokalnych zdarzeń, co może zmniejszyć potrzebę przesyłania danych do centralnej chmury (Yi i in., 2015).

Jak łatwo zauważyć Edge Computing i Edge AI to pojęcia blisko powiązane. Oba odzwierciedlają jednak różne aspekty przetwarzania danych na urządzeniach brzegowych. Edge Computing odnosi się do ogólnego przetwarzania danych blisko ich źródła (na lokalnych serwerach, bramach sieciowych czy urządzeniach zbierających dane). Jego celem jest zmniejszenie opóźnień i odciążenie łączy do chmury poprzez lokalne filtrowanie i wstępną analizę danych (Shi, 2016). Natomiast Edge AI koncentruje się na uruchamianiu zaawansowanych algorytmów sztucznej inteligencji bezpośrednio na urządzeniach końcowych, takich jak kamery, sensory, roboty czy smartfony, umożliwiając podejmowanie natychmiastowych decyzji i działanie w czasie rzeczywistym, bez konieczności przesyłania danych do chmury.

Metodyka badań

Niniejsze badanie zostało ukierunkowane przez dwa pytania badawcze. Jakie zagrożenia bezpieczeństwa są charakterystyczne dla urządzeń brzegowych w środowiskach Edge AI i w jakim stopniu metodologia STRIDE pozwala na systematyczną identyfikację i klasyfikację zagrożeń we wszystkich warstwach architektury Edge AI. Dodatkowo praca miała za zadanie zidentyfikować mechanizmy ochrony są adekwatne do specyficznych ograniczeń sprzętowych i operacyjnych urządzeń kocowych. Na podstawie przeglądu literatury postawiono następującą hipotezę roboczą: środowisko Edge AI generuje unikalny profil ryzyka, wynikający z fizycznej dostępności urządzeń, ograniczonych zasobów obliczeniowych oraz rozproszonej architektury, który nie jest w pełni adresowany przez standardowe podejścia do bezpieczeństwa systemów scentralizowanych. Weryfikacja tej hipotezy została przeprowadzona przez systematyczne zastosowanie metodologii STRIDE.

Badanie przeprowadzono w oparciu o następujące dwa główne etapy. W pierwszym dokonano analizy publikacji naukowych, raportów technicznych oraz dokumentacji dotyczącej bezpieczeństwa Edge AI, ze szczególnym uwzględnieniem prac dotyczących ataków sprzętowych, manipulacji modeli AI oraz podatności w akceleratorach AI, prace wymienione są w sekcji bibliografii. Priorytetowo wybrane zostały publikacje z lat 2023-2025, uwzględniając najnowsze zagrożenia i techniki ochrony w dynamicznie rozwijającej się dziedzinie Edge AI. Poszukując świeżych informacji na temat rozwijającej się technologii część publikacji pozyskano z bazy arXiv oferującej dostęp do artykułów naukowych w formie preprintów. Zostały też podjęte kroki, aby wybrać prace publikowane w renomowanych czasopismach (n.p. IEEE), raz dokumentację wiodących organizacji (Microsoft, OWASP, MITRE). Literatura obejmuje trzy istotne dla pracy obszary: fundamenty architektoniczne Edge Computing, metodyki modelowania zagrożeń oraz niebezpieczeństwa występujące w środowiskach AI.

Po zapoznaniu się z literaturą oraz ze szczegółami technicznymi architektury Edge AI w drugim etapie pracy zastosowana została metodyka modelowania zagrożeń STRIDE. Należy zaznaczyć, że praca ma charakter teoretyczno-analityczny i nie obejmuje praktycznej weryfikacji

zidentyfikowanych zagrożeń w środowisku testowym ani kwantyfikacji ryzyka. Ponadto, dynamika związana z rozwojem technologii AI oraz nowo pojawiające się wektory ataków mogą wymagać aktualizacji zidentyfikowanych zagrożeń.

Modelowanie zagrożeń dla Edge AI

Skuteczne zarządzanie bezpieczeństwem w środowiskach Edge wymaga precyzyjnej identyfikacji i klasyfikacji zagrożeń, co zgodnie z najlepszymi praktykami realizuje się za pomocą sprawdzonych metodyk modelowania zagrożeń (threat modeling). Wśród dostępnych podejść do analizy zagrożeń w cyberbezpieczeństwie, metodyka STRIDE wyróżnia się wszechstronnością, uniwersalnością i skutecznością, dzięki czemu jest najczęściej rekomendowana do zastosowań w systemach rozproszonych, w tym Edge AI (Shostack, 2014). STRIDE to akronim obejmujący sześć kategorii zagrożeń: Spoofing (podszywanie się), Tampering (manipulacje), Repudiation (zaprzeczenie),Information Disclosure (ujawnienie informacji), Denial of Service (odmowa usługi), Elevation of Privilege (podniesienie uprawnień). Metodyka ta umożliwia systematyczne sprawdzenie każdego elementu architektury systemu pod kątem tych sześciu kategorii, wspierając identyfikację potencjalnych słabości zarówno na poziomie urządzeń brzegowych, lokalnych węzłów (fog) jak i warstwy chmurowej (Marshall i in., 2025).

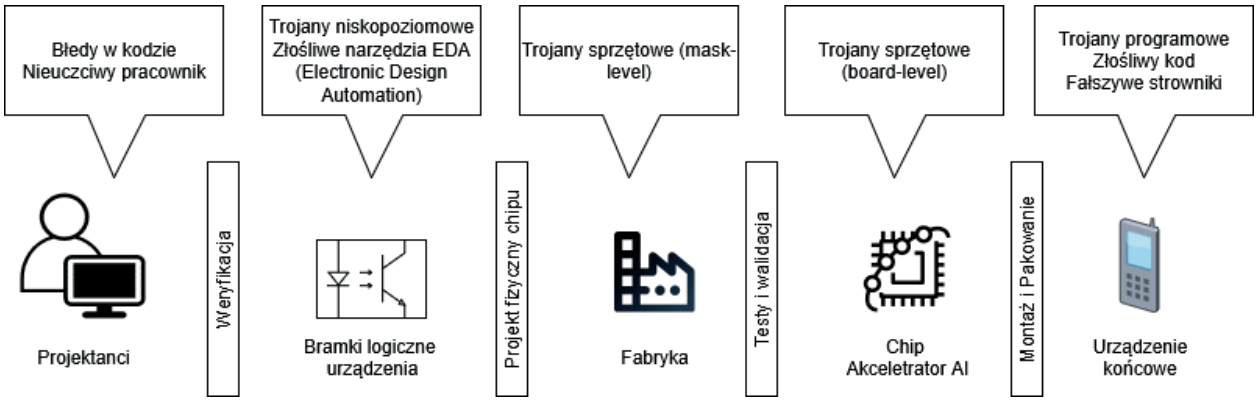
STRIDE jako metodologia jest tylko jednym z elementów szerszego spektrum technik i narzędzi, które mogą wspierać proces modelowania zagrożeń w złożonych, rozproszonych systemach AI działających na brzegu sieci. W kontekście Edge AI konieczne jest uwzględnienie specyfiki tej architektury, obejmującej wiele heterogenicznych urządzeń, różnorodność protokołów komunikacyjnych i wymogi niskiego opóźnienia. Z tego powodu modelowanie zagrożeń można też przeprowadzić przy użyciu technik lepiej dopasowanych do danego systemu. Jedną z nich jest podejście Asset-Centric Threat Modeling, które skupia się na identyfikacji i ochronie kluczowych zasobów. W kontekście Edge AI może to być model AI, dane sensoryczne, czy specyficzny układ sprzętowy, jaki występuje na urządzeniach brzegowych. Model Asset-Centric pomaga skoncentrować wysiłki analityczne i obronne na tych elementach, które mają kluczowe znaczenie dla funkcjonowania systemu i prywatności użytkowników (von der Assen i in., 2023). Innym podejściem jest użycie adaptowanego z modelu cyberataku opisanego przez Lockheed Martin, modelu Kill Chain. Pozwala on na śledzenie i rozpoznanie etapów ataku na urządzenia Edge AI: od rozpoznania, przez uzyskanie dostępu, wykonanie złośliwych działań, po utrzymanie kontroli nad urządzeniem i ewentualną eskalację. Ten model pomaga też projektować mechanizmy detekcji i przerywania ataków na różnych fazach, co w architekturze rozproszonej Edge AI jest szczególnie ważne (Kazimierczak, 2024). Z kolei tworzenie drzew ataku (Attack Tree) pozwala na wizualne przedstawienie potencjalnych ścieżek ataku na system, pokazując różnorodne techniki i sposoby, które atakujący mogą wykorzystać, by osiągnąć określony cel. Dla systemów Edge AI, gdzie zagrożenia mogą wynikać zarówno z fizycznego dostępu do urządzenia, jak i z aspektów sieciowych, drzewo ataku umożliwia kompletne i szczegółowe mapowanie możliwych punktów awarii lub włamania. Istnieje też stosunkowo nowy framework zaproponowany specjalnie dla systemów wieloagentowych AI i Edge AI. MAESTRO, o którym mowa, skupia się na modelowaniu dynamicznych interakcji pomiędzy agentami AI, urządzeniami Edge oraz środowiskiem, analizując potencjalne zagrożenia wynikające z autonomicznych decyzji, współpracy agentów i naruszeń zasad bezpieczeństwa na poziomie systemu (Zambare, 2025).

Ponieważ w niniejszej pracy poruszana jest szeroka gama zagrożeń dla niesprecyzowanego systemu Edge AI, użyta została metodologia STRIDE, ze względu na jej wszechstronność oraz uniwersalność.

Identyfikacja i klasyfikacja zagrożeń dla urządzeń Edge AI

W środowisku Edge AI występuje szereg różnorodnych zagrożeń, które można podzielić na kilka podstawowych typów odpowiadających specyfice sprzętu, oprogramowania oraz sposobów komunikacji i przetwarzania danych ale także wpasowujących się w kategorie metodologii STRIDE. Zagrożenia związane z podszywaniem się (Spoofing) dotyczą sytuacji, gdy złośliwy podmiot lub urządzenie fałszywie przedstawia się jako zaufany komponent sieci lub użytkownik, aby uzyskać nieautoryzowany dostęp do danych bądź funkcji systemu AI. W kontekście Edge AI może to oznaczać podszywanie się pod sensory lub urządzenia końcowe, co skutkuje wprowadzaniem fałszywych danych. Takie manipulacje mogą prowadzić do błędnych decyzji AI lub umożliwić dalsze ataki na system.

Zagrożenia polegające na manipulacji (Tampering) są w środowisku Edge AI szczególnie niebezpieczne, gdyż atakujący może modyfikować firmware urządzeń, wstrzykiwać złośliwe fragmenty kodu do modeli AI lub ingerować w fizyczne komponenty sprzętowe. Modyfikacje takie często pozostają niewykryte i mogą prowadzić do poważnych konsekwencji, takich jak uszkodzenie integralności modeli, ujawnienie danych lub utrata kontroli nad urządzeniem. Tego typu zagrożenia dokładnie opisano w pracy „Challenges in Detecting and Mitigating Hardware Trojans in ML Accelerators" (Gubbi i in., 2023). W cytowanym opracowaniu autorzy przedstawiają własne podejścia do identyfikacji zagrożeń, oceny ryzyka oraz implementacji środków zaradczych, z naciskiem na ich zastosowanie w procesie wytarzania urządzeń końcowych Edge AI, słusznie zauważając że zagrożenia mogą pojawić się w całym cyklu życia tego typu produktów.



Rycina 2. Zagrożenia w procesie wytwarzania urządzeń końcowych Edge AI
Źródło: opracowanie własne.

Zagrożenia związane z zaprzeczeniem wykonania operacji (Repudiation) dotyczą braku lub niedostatecznego monitorowania oraz logowania czynności wykonywanych przez urządzenia Edge lub użytkowników. W efekcie po incydencie bezpieczeństwa trudno jest ustalić faktyczne zdarzenia, co pozwala atakującym na ukrycie śladów swoich działań lub zaprzeczenie ich popełnienia. W Edge AI ze względu na rozproszenie infrastruktury i heterogeniczność systemów problem ten

jest jednym z trudniejszych do rozwiązania. Zagrożenia wynikające z ujawnienia informacji (Information Disclosure) obejmują nieautoryzowany dostęp do danych przetwarzanych lub przechowywanych na urządzeniach Edge AI. Są to np.: poufne dane osobowe, wyniki analiz AI czy dane konfiguracyjne. Ze względu na fizyczną dostępność tych urządzeń oraz nierzadko ograniczoną ochronę kanałów komunikacyjnych, ryzyko wycieku jest bardzo wysokie. Przykładem są ataki side-channel, które wykorzystują analizę poboru mocy lub emisji elektromagnetycznej do wykradania informacji. Odrębnym problemem jest możliwość ujawnienia kodu i struktury modelu AI. Jest to problem obarczony bardzo dużym ryzkiem w środowiskach Edge AI, gdzie modele są instalowane bezpośrednio na urządzeniach końcowych. Modele AI, często chronione jako własność intelektualna, zawierają cenne informacje, takie jak unikalne architektury sieci neuronowych, zoptymalizowane parametry i dane treningowe, które są efektem kosztownych badań i rozwoju. W przypadku instalacji na urządzeniach fizycznie dostępnych dla użytkowników, istnieje ryzyko przeprowadzenia inżynierii wstecznej (reversing), która umożliwia odtworzenie kodu, struktury modelu lub odzyskanie danych, co prowadzi do kradzieży własności intelektualnej oraz potencjalnego nieautoryzowanego powielania. Proces inżynierii wstecznej może obejmować analizę binarną, dekompilację lub ekstrakcję wag i parametrów z pamięci urządzenia. Stanowi on też potencjalne naruszenie prywatności, gdy model może zawierać dane pochodzące od innych użytkowników. W literaturze zwraca się uwagę na potrzebę stosowania technik ochrony, takich jak szyfrowanie modeli, zabezpieczenia sprzętowe (Secure Enclaves, Trusted Execution Environments) oraz mechanizmy zaciemniania kodu i detekcji (model watermarking, obfuscation), które utrudniają lub uniemożliwiają efektywne przeprowadzenie odwróconej inżynierii (Zhao i in., 2025). Zagrożenia odmowy usługi (Denial of Service) to ataki mające na celu zaburzenie funkcjonowania urządzenia Edge, przeciążenie jego zasobów obliczeniowych lub blokadę dostępu do niego. W przypadku Edge AI takie działania mogą uniemożliwić realizację czasowo krytycznych zadań AI, np. w systemach autonomicznych czy medycznych. Ataki DDoS na lokalne sieci, przeciążanie procesorów AI lub blokowanie interfejsów aktualizacji to typowe przykłady (Wingarz i in., 2024). Zagrożenia polegające na eskalacji uprawnień (Elevation of Privilege) pojawiają się, gdy atakujący zyskuje uprawnienia wyższe niż przewidziane, np. administratora systemu, co pozwala mu na pełną kontrolę nad urządzeniem i modelami AI. Taka sytuacja może wynikać z luk w systemie operacyjnym, błędów w konfiguracji lub zaniedbań w zabezpieczeniach aplikacji na urządzeniach Edge.

Tabela 2. Wybrane zagrożenia dla systemów Edge AI

Kategoria STRIDE	Zagrożenie	Warstwa		
		Edge	Fog	Cloud
Spoofing	Podszywanie się pod legalne urządzenia Edge AI	X	X	
	Fałszywe uwierzytelnianie użytkownika (np. Backdoor w modelu biometrii)	X	X	
	Fałszywe sygnały sensoryczne (sensor spoofing)	X		
	Podszywanie się pod usługę aktualizacji firmware	X	X	X
	Atak typu man-in-the-middle na łączu sensor-edge	X	X	
Tampering	Modyfikacje firmware'u urządzeń Edge, wprowadzanie backdoorów	X		
	Fizyczna ingerencja w układy scalone przez porty debugu JTAG	X		
	Próby sabotażu modeli AI (model poisoning)	X	X	
	Manipulacje baz danych i plików konfiguracyjnych i prodokołów komunikacyjnych	X	X	

Kategoria STRIDE	Zagrożenie	Warstwa		
		Edge	Fog	Cloud
Repudiation	Brak spójnego i bezpiecznego logowania operacji	X	X	X
	Usuwanie lub modyfikacja logów przez atakującego	X	X	X
	Brak kryptograficznego potwierdzania autentyczności operacji	X	X	X
	Możliwość zaprzeczania wykonania operacji	X	X	X
	Niewystarczający nadzór i redundancja w systemach monitoringu	X	X	X
Information Disclosure	Przechwycenie niezabezpieczonych danych na łączu sensor-edge	X	X	
	Ujawnienie poufnych danych użytkownika	X	X	X
	Ujawnienie szczegółów modelu	X		
	Wycieki przez ataki side-channel (analiza poboru mocy, sygnałów elektromagnetycznych)	X		
	Nieodpowiednie zarządzanie kluczami szyfrowania	X	X	
	Podsłuchiwanie interfejsów komunikacyjnych (Wi-Fi, Bluetooth)	X	X	
Denial of Service	Ataki DDoS na urządzenia Edge	X	X	
	Wyczerpanie zasobów (CPU, pamięć, energia)	X		
	Ataki na sieć lokalną uniemożliwiające dostępność danych	X	X	
	Ataki logiczne powodujące restart lub awarię oprogramowania	X	X	
	Blokowanie krytycznych aktualizacji firmware	X	X	X
Elevation of Privilege	Luka w systemie umożliwiająca dostęp administratora	X	X	
	Nieautoryzowana instalacja oprogramowania z wyższymi uprawnieniami	X	X	
	Użycie exploitów systemowych do przejęcia kontroli nad urządzeniem Edge AI	X		

Źródło: opracowanie własne.

Analiza zagrożeń w środowisku Edge AI, oparta na metodologii STRIDE i zilustrowana w tabeli 2, obejmuje zróżnicowane kategorie ataków z podziałem na warstwy Edge, Fog i Cloud, uwzględniając specyfikę sprzętową, komunikacyjną oraz programową, co umożliwia identyfikację precyzyjnych punktów podatności w całym stosie architektonicznym. Wyniki analizy mają bezpośrednie odzwierciedlenie w realnych implementacja Edge AI – na przykład w pojazdach autonomicznych, gdzie ataki spoofingowe na LiDAR lub manipulacje modelami powodują błędne decyzje nawigacyjne, lub w urządzeniach medycznych IoT, gdzie ataki typu DoS i side-channel zagrażają bezpieczeństwu pacjentów. W przemyśle naftowym i gazowym przypadki cyberataków na rozproszoną infrastrukturę prowadziły do przestojów i strat (Beerman, 2023), podkreślając potrzebę wdrożenia wielowarstwowych kontroli dostępu i monitoringu, co dotyczy również zastosowań rozwiązań Edge-AI w tej branży.

Wnioski i kierunki dalszych badań

Analiza zagrożeń w środowiskach Edge AI z uwagi na rosnącą liczbę zastosowań tej technologii w kluczowych sektorach, takich jak medycyna, transport czy przemysł jest niezwykle ważna. Wdrożenie rozproszonych urządzeń inteligentnych blisko źródeł danych pozwala na błyskawiczne podejmowanie decyzji, ale jednocześnie stawia nowe, złożone wyzwania związane z bezpieczeństwem.

Wnioski płynące z analizy metodą STRIDE pokazują, że ochrona tych systemów musi być wielowymiarowa i dostosowana do specyfiki warstwy sprzętowej, komunikacyjnej oraz programowej. Zagrożenia związane z podszywaniem się (Spoofing) wskazują na konieczność wdrożenia mechanizmów uwierzytelniania tożsamości urządzeń (jak chociażby implementacje mTLS) i użytkowników na wszystkich poziomach architektury Edge AI. Ponieważ urządzenia funkcjonują często w otwartym, rozproszonym środowisku z wieloma punktami dostępu, ich fałszywe identyfikowanie może doprowadzić do poważnych naruszeń integralności danych oraz działania systemów AI. Wprowadzenie zaawansowanych metod kryptograficznych i certyfikacji sprzętu wzmocni ochronę przed takimi atakami, ale musi iść w parze z ciągłym monitoringiem i reagowaniem na anomalie. Manipulacje (Tampering) dotyczące firmware'u, oprogramowania i modeli AI to zagrożenia, które mogą mieć długotrwałe konsekwencje dla całego systemu. Nieautoryzowane ingerencje mogą powodować błędne analizy czy nawet przejęcie kontroli nad urządzeniami. Z uwagi na fizyczny dostęp do urządzeń brzegowych, zabezpieczenia muszą obejmować obronę na poziomie sprzętowym, jak i na poziomie oprogramowania – w tym mechanizmy wykrywające nieautoryzowane zmiany kodu lub próby modyfikacji danych treningowych. Zaprzeczenie wykonania operacji (Repudiation) stanowi poważne wyzwanie dla audytu i odpowiedzialności w systemach Edge AI. Złożoność i rozproszenie infrastruktury często utrudniają spójne i niezawodne logowanie zdarzeń. Skuteczne systemy bezpieczeństwa powinny zapewniać integralność i dostępność logów, wykorzystując kryptograficzne sygnatury zdarzeń, tak aby uniemożliwić manipulację lub ich usunięcie po ataku. Transparentność jest niezbędna do zachowania zaufania zarówno użytkowników, jak i administratorów systemu. W zakresie ujawnienia informacji (Information Disclosure) uwagę należy zwrócić na zabezpieczenie danych osobowych i wrażliwych, które są przetwarzane lokalnie na urządzeniach Edge. Z powodu specyfiki rozproszonych systemów i potencjalnego fizycznego dostępu, konieczne jest stosowanie wielowarstwowych metod ochrony, w tym silnego szyfrowania transmisji danych, izolacji środowiska wykonawczego oraz zabezpieczeń fizycznych. Ataki typu side-channel, które wykorzystują poboczne sygnały wymuszają potrzebę projektowania układów sprzętowych oraz oprogramowania z myślą o odporności na takie metody wycieku informacji. Ataki odmowy usługi (Denial of Service) mogą sparaliżować funkcjonowanie krytycznych komponentów Edge AI, szczególnie w systemach wymagających ciągłej dostępności i niskich opóźnień, np. inteligentnych pojazdach czy systemach medycznych. Ostatecznie zagrożenia wynikające z podniesienia uprawnień (Elevation of Privilege) mogą umożliwić całkowite przejęcie kontroli nad urządzeniami, co w konsekwencji prowadzi do masowych naruszeń bezpieczeństwa i prywatności. Zabezpieczenia muszą obejmować ścisłą kontrolę dostępu, audyt uprawnień oraz regularne aktualizacje systemów operacyjnych i aplikacyjnych.

Tabela 3. Proponowane techniki obronne dla wybranych zagrożeń

Kategoria STRIDE	Kluczowe zagrożenie	Działania mitygujące
Spoofing	Fałszywe urządzenia	mTLS z certyfikatami sprzętowymi (TPM), poświadczanie tożsamości urządzeń, behawioralne odciski palców
Tampering	Modyfikacja firmware'u	Trusted Boot + Secure Boot. Monitorowanie integralności w czasie rzeczywistym (IMA)
	Model poisoning	Podpisywanie cyfrowe modeli oraz zbiorów danych, hashowanie
Repudiation	Brak logów manipulacja danych	Kryptografia, syslog z sygnaturami HMAC

Kategoria STRIDE	Kluczowe zagrożenie	Działania mitygujące
Information Disclosure	Ujawnienie struktury modelu AI	Zaufane środowiska wykonawcze (TEE), kryptografia, znakowanie wodne i zaciemnianie
Denial of Service	Wyczerpanie zasobów	Web Application Firewall, limity zapytań
Elevation of Privilege	Nadużywanie interfejsów fizycznych	Uszczelnianie portów (wyłączenie JTAG po wdrożeniu)

Źródło: opracowanie własne.

Teoretyczne rozważania przedstawione w niniejszym opracowaniu powinny być wsparte przez szerokie badania praktyczne, prowadzone w zbliżonych do rzeczywistych środowiskach Edge AI. Istniejące platformy eksperymentalne i zbiory danych, takie jak Edge-IIoTset (Ferrag 2022) obejmujący zróżnicowane urządzenia, protokoły oraz konfiguracje chmura-edge, służą dziś głównie do oceny skuteczności algorytmów detekcji intruzów w trybie scentralizowanym. Koncentrują się przede wszystkim na klasycznych atakach sieciowych i integralności ruchu, a nie na pełnym spektrum zagrożeń specyficznych dla Edge AI, takich jak manipulacje modelami, ataki typu side-channel czy eskalacja uprawnień. Równolegle pojawiają się testbedy i ramy eksperymentalne pozwalające symulować i oceniać bezpieczeństwo rozproszonych architektur IoT/Edge (Ghadiri 2023), co pokazuje, że tworzenie powtarzalnych, dobrze udokumentowanych scenariuszy ataków znacząco ułatwia porównywanie metod obrony oraz analizę kompromisu między bezpieczeństwem a wydajnością. W kontekście Edge AI brakuje jednak znormalizowanych benchmarków obejmujących jednocześnie ataki na warstwę sprzętową (np. fizyczną manipulację), modyfikacje firmware'u oraz modyfikacje modeli i danych treningowych. Dalsze prace powinny zatem obejmować projektowanie środowisk testowych oraz zestandaryzowanych scenariuszy eksperymentalnych, które pozwolą ocenić skuteczność zabezpieczeń.

Jak wynika z powyższej analizy, skuteczne zabezpieczenie urządzeń Edge AI wymaga kompleksowego podejścia opartego na wielowarstwowych mechanizmach ochronnych, które weźmie pod uwagę unikalne charakterystyki i ograniczenia tych systemów. Modelowanie zagrożeń według STRIDE dostarcza uniwersalnej ramy do identyfikacji ryzyk i planowania strategii obronnych. Niezbędne jest ciągłe monitorowanie nowych typów ataków i adaptowanie zabezpieczeń do zmieniającego się krajobrazu technologicznego.

Literatura:

1. Beerman, J., Berent, D., Falter, Z., Bhunia, S. (2023). *A Review of Colonial Pipeline Ransomware Attack*. 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), Bangalore, 8-15. https://doi.org/10.1109/ccgridw59191.2023.00017
2. Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., Janicke, H. (2022). Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. *IEEE Access, 10*, 40281-40306. https://doi.org/10.1109/ACCESS.2022.3165809
3. Ghadiri, R., ElHajj, M. (2023). *Security and Performance Analysis of Edge Computing in IoT*. IEEE International Conference on Communication, Networks and Satellite, COMNETSAT 2023, Malang, 542-548. https://doi.org/10.1109/COMNETSAT59769.2023.10420709
4. Gubbi, K. I., Kaur, I., Hashem, A., Sai Manoj, P. D., Homayoun, H., Sasan, A., Salehi, S. (2023). *Securing AI Hardware: Challenges in Detecting and Mitigating Hardware Trojans in ML Accelerators*. IEEE 66th International Midwest Symposium on Circuits and Systems (MWSCAS), Tempe, 821-825. https://doi.org/10.1109/MWSCAS57524.2023.10406065

5. Jones, N. F. (2025). Decentralized Edge-AI Strategies for Micro-Datacenter Optimization and Resource-Conscious Query Execution. *International Journal of Information Technology Research and Development (IJITRD), 6*(3), 19-24.
6. Kazimierczak, M., Habib, N., Chan, J. H., Thanapattheerakul, T. (2024). Impact of AI on the Cyber Kill Chain: A Systematic Review. *Heliyon, 10*(24), e40699. https://doi.org/10.1016/j.heliyon.2024.e40699
7. Li, G., Hari, S. K. S., Sullivan, M., Tsai, T., Pattabiraman, K., Emer, J., Keckler, S. W. (2017). *Understanding error propagation in deep learning neural network (DNN) accelerators and applications*. SC '17: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, 1-12. https://doi.org/10.1145/3126908.3126964
8. Marshall, A., Parikh, J., Kiciman, E., Shankar, R., Kumar, S. (2025). *Threat Modeling AI/ML Systems and Dependencies*. Microsoft. Pobrane z: https://learn.microsoft.com/en-us/security/engineering/threat-modeling-aiml (data dostępu 10.02.2025).
9. Mukherjee, R., Chakraborty, R. S. (2022). Novel hardware trojan attack on activation parameters of FPGA-based DNN accelerators. *IEEE Embedded Systems Letters, 14*(3), 131-134. https://doi.org/10.1109/LES.2022.3159541
10. Shafee, A., Hasan, S. R., Awaad, T. A. (2025). Privacy and security vulnerabilities in edge intelligence: An analysis and countermeasures. *Computers and Electrical Engineering, 123*, 110146. https://doi.org/10.1016/j.compeleceng.2025.110146
11. Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L. (2016). Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*, *3*(5), 637-646. https://doi.org/10.1109/JIOT.2016.2579198
12. Shostack, A. (2014). *Threat Modeling: Designing for Security*. Wiley.
13. Singh, R., Gill, S. S. (2023). Edge AI: A survey. *Internet of Things and Cyber-Physical Systems, 3*. https://doi.org/10.1016/j.iotcps.2023.02.004
14. Strickland, E. (2024). *15 Graphs That Explain the State of AI in 2024: The AI Index Tracks the Gener-ative AI Boom, Model Costs, and Responsible AI Use*. IEEE Spectrum. Pobrane z: https://spectrum. ieee.org/ai-index-2024 (data dostępu 10.02.2025).
15. Sung, J., Han, S. (2024). Use of edge resources for DNN model maintenance in 5G IoT networks. *Cluster Computing, 27*(4), 5093-5105. https://doi.org/10.1007/s10586-023-04236-y
16. Tuli, S., Mirhakimi, F., Pallewatta, S., Zawad, S., Casale, G., Javadi, B., Yan, F., Buyya, R., Jennings, N. R. (2023). AI augmented Edge and Fog computing: Trends and challenges. *Journal of Network and Computer Applications, 216*, 103648. https://doi.org/10.1016/j.jnca.2023.103648
17. von der Assen, J., Sharif, J., Feng, Ch., Bovet, G., Stiller, B. (2024). Asset-driven Threat Modeling for AI-based Systems. *arXiv preprint*. https://arxiv.org/html/2403.06512v1
18. Wingarz, S., Lauscher, A., Edinger, J., Kaaser, D., Schulte, S., Fischer, M. (2024). SoK: Towards Security and Safety of Edge AI. *arXiv preprint*. https://arxiv.org/html/2410.05349v1
19. Yi, S., Li, C., Li, Q. (2015). *A Survey of Fog Computing*. Mobidata '15: Proceedings of the 2015 Workshop on Mobile Big Data Pages, 37-42. https://doi.org/10.1145/2757384.2757397
20. Zambare, P., Thanikella, V. N., Liu, Y. (2025). Securing agentic ai: Threat modeling and risk analysis for network monitoring agentic ai system. *arXiv preprint*. https://arxiv.org/html/2508.10043v1
21. Zhao, K., Li, L., Ding, K., Gong, N. Z., Zhao, Y., Dong, Y. (2025). A systematic survey of model extraction attacks and defenses: State-of-the-art and perspectives. *arXiv preprint*. https://arxiv.org/html/2508.15031v2