AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

DZIEDZINA: Nauki inżynieryjno-techniczne

DYSCYPLINA: Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne

Autoreferat Rozprawy Doktorskiej

Elektroniczne Systemy Zabezpieczenia Technicznego jako źródło danych w procesie automatycznej obrony przed cyberatakami oraz wykrywania sprawcy przestępstwa

Autor: Jan Kapusta

Promotor rozprawy: Prof. dr hab. inż. Jerzy Baranowski
Promotor pomocniczy: Dr inż. Waldemar Bauer

Praca wykonana: AGH, Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

Kraków, 2026

1. Wstęp i uzasadnienie podjęcia tematu

Przedmiotem niniejszej rozprawy jest analiza i modelowanie cyber-fizycznych systemów ochrony (ang. *Cyber-Physical Protection Systems*, CPPS) z wykorzystaniem metod sztucznej inteligencji, probabilistycznego modelowania degradacji oraz automatyzacji analizy danych bezpieczeństwa. Punktem wyjścia jest założenie, że elektroniczne systemy zabezpieczenia technicznego (ESZT) nie powinny być traktowane wyłącznie jako pasywne elementy infrastruktury ochronnej, lecz jako istotne źródło danych, które po odpowiednim przetworzeniu może wspierać zarówno ocenę skuteczności ochrony, jak i planowanie działań zwiększających odporność obiektu na zagrożenia fizyczne i cybernetyczne.

Motywacja badawcza wynika z obserwacji, że współczesna infrastruktura krytyczna funkcjonuje w środowisku rosnącej złożoności zagrożeń. Klasyczne systemy ochrony fizycznej są silnie powiązane z technologiami teleinformatycznymi (ICT), co oznacza, że ich skuteczność zależy nie tylko od parametrów fizycznych barier, detekcji i reakcji, lecz także od podatności cybernetycznych, jakości integracji systemowej oraz zmian zachodzących w czasie eksploatacji. Tradycyjne podejścia do oceny skuteczności systemów ochrony fizycznej nie nadążają za tymi zmianami, co może prowadzić do przeszacowania poziomu bezpieczeństwa i do decyzji utrzymaniowych lub inwestycyjnych nieadekwatnych do rzeczywistego ryzyka.

W pracy zidentyfikowano pięć zasadniczych luk badawczych: brak uwzględnienia cyberzagrożeń w klasycznej metodologii EASI (Estimated Adversary Sequence Interruption), brak modeli degradacji elementów specyficznych dla ESZT, brak standaryzacji i automatyzacji interpretacji symboli technicznych na planach, ograniczoną automatyzację modelowania ścieżek ataku oraz fragmentaryczność dotychczasowych rozwiązań. Rozprawa stanowi odpowiedź na te deficyty w postaci zintegrowanej, trzywarstwowej metodologii obejmującej warstwę percepcyjną, decyzyjną i raportującą.

2. Teza pracy

Teza rozprawy została sformułowana następująco:

> *Możliwe jest opracowanie zintegrowanego, probabilistycznego modelu oceny skuteczności Cyber-Fizycznych Systemów Ochrony (CPPS), który jednocześnie uwzględnia degradację komponentów technicznych w czasie, zagrożenia cybernetyczne oraz operacyjny kontekst funkcjonowania systemu.*

Teza ma charakter zarówno poznawczy, jak i aplikacyjny. W wymiarze poznawczym dotyczy możliwości spójnego połączenia modeli ryzyka fizycznego, cybernetycznego, degradacyjnego i kontekstowego. W wymiarze aplikacyjnym zakłada stworzenie rozwiązania pozwalającego przejść od ręcznej, statycznej oceny bezpieczeństwa do procedury zautomatyzowanej, skalowalnej i opartej na danych.

3. Struktura pracy

Rozprawa ma układ problemowo-metodyczny. Rozdział 2 zawiera przegląd metod oceny systemów bezpieczeństwa fizycznego oraz identyfikuje kluczowe deficyty, ze szczególnym uwzględnieniem braku uwzględniania cyberzagrożeń i degradacji. Rozdział 3 omawia zastosowanie algorytmów komputerowej analizy obrazów (YOLO, COG) do automatycznej detekcji elementów infrastruktury budynków na planach 2D. Rozdział 4 przedstawia metodologię generowania grafów ścieżek ataku oraz ich heurystyczną ocenę z wykorzystaniem algorytmu A* i autorskiej heurystyki *stealth*. Rozdział 5 opisuje zastosowanie dużych modeli językowych (LLM) w syntezie danych bezpieczeństwa oraz automatycznym generowaniu raportów i rekomendacji. Rozdział 6 przedstawia kompleksową implementację narzędzia AI-SecPA oraz przykłady jego zastosowania. Rozdział 7 podsumowuje wnioski, analizuje ograniczenia i wskazuje kierunki dalszego rozwoju.

4. Metody badawcze i metodologia

Rozprawa ma charakter interdyscyplinarny i integruje metody typowe dla inżynierii bezpieczeństwa, analizy ryzyka, sztucznej inteligencji, modelowania probabilistycznego, teorii grafów oraz przetwarzania języka naturalnego. W ujęciu metodologicznym wyróżniono trzy komplementarne warstwy rozwiązania.

Warstwa percepcyjna: detekcja i interpretacja elementów architektonicznych

Jej zadaniem jest automatyczna analiza planów technicznych i identyfikacja elementów architektonicznych oraz elementów bezpieczeństwa. Wykorzystano metody detekcji obiektów oparte na rodzinie YOLOv8, aby ostatecznie użyć jako punktu startowego wag modelu YOLOv8m, pretrenowanego na zbiorze COCO (yolov8m.pt), który następnie został doszkolony na własnych zbiorach treningowych. W zakresie detekcji elementów budowlanych zbiór treningowy zawierał obrazy uwzględniające elementy architektoniczne (symbole drzwi, ścian okien itp., podkłady budowlane oraz aranżacje obiektów budowlanych). Zbiór wykorzystany do treningu modelu służącego do detekcji symboli systemów bezpieczeństwa z wykorzystaniem zaproponowanej koncepcji COG (Contextual Object Grouping) składał się z obrazów zawierających schematy techniczne wraz z legendami opisującymi użyte w nich symbole. Celem zaproponowanego podejścia COG jest nie tylko wykrywanie obiektów (symboli ESZT), lecz także ich interpretacja w kontekście lokalnym i funkcjonalnym. Model YOLOv8m uczy się zarówno klas atomowych (“symbol”, “label”), jak i klas kontekstowych (“row_leg”, “legend”), co umożliwia automatyczne parowanie symboli z ich etykietami w oparciu o strukturę legendy rysunku, niezależnie od stosowanego standardu rysunkowego.

Wyniki walidacyjne: model architektoniczny osiągnął mAP@0.5 = 96,8% przy 74 FPR (klatek na sekundę), zaś model COG uzyskał mAP@0.5 = 99% oraz mAP@0.50-0.95 = 81%, przy 100% poprawności parowań symbol–etykieta w testach. Zbiór danych treningowych i

walidacyjnych dla COG obejmował łącznie 1573 instancje zdefiniowanych klas (1167 w zestawie treningowym i 406 w zestawie walidacyjnym). Zdefiniowane klasy w zakresie COG są następujące:

- Legend (legenda jako całość)
- Symbol (symbol wewnątrz legendy)
- Label (etykieta)
- Row_Leg (wiersz legendy zawierający symbol i etykietę)
- L_title (tytuł legendy)
- Column_S (kolumna legendy zawierająca symbole)
- Column_L (kolumna legendy zawierająca etykiety)

W zakresie detekcji elementów architektonicznych zbiór danych obejmował 452 plany architektoniczne (304 treningowe, 148 walidacyjnych). Zdefinowane w tym zakresie klasy elementów architektonicznych są następujące:

- SingleDoor (drzwi z jednym skrzydłem)
- DLDoor (Drzwi z dwoma skrzydłami)
- Window (okno)
- IntWall (ściana wewnętrzna)
- ExtWall (ściana zewnętrzna)
- Stairs (schody)
- SlidingDoor (drzwi przesuwne)

Warstwa decyzyjna: modelowanie ścieżek ataku i ocena skuteczności

Warstwa decyzyjna koncentruje się na modelowaniu ścieżek ataku oraz ocenie skuteczności ochrony. Podstawą jest krytyczna analiza i rozszerzenie metodologii EASI (zob. sekcja 5).

Warstwa raportująca: LLM i synteza wyników

Warstwa raportująca wykorzystuje duże modele językowe (LLM) wraz z biblioteką Outlines do konsolidacji danych, wspomagania interpretacji wyników oraz generowania ustrukturyzowanych rekomendacji. Modele językowe nie zastępują modeli analitycznych, lecz pełnią funkcję warstwy syntezy wspierającej użytkownika w interpretacji wyników oceny bezpieczeństwa. Biblioteka Outlines implementuje tzw. *structured generation* czyli mechanizm wymuszający określone formaty wyjściowe bezpośrednio na poziomie procesu generowania tekstu przez LLM, co eliminuje ryzyko niestrukturyzowanych lub nieprzewidywalnych odpowiedzi.

5. Kluczowe elementy metodologiczne i wzory

Klasyczna metodologia EASI

Metodologia EASI (ang. *Estimated Adversary Sequence Interruption*), wprowadzona przez Bennetta w 1977 r., definiuje prawdopodobieństwo przerwania ataku $P(I)$ jako:

$$P(I) \;=\; PD_1 \cdot PC_1 \cdot P(R|A_1) \;+\; \sum_{i=2}^{n} P(R|A_i) \cdot PC_i \cdot PD_i \cdot \prod_{j=1}^{i-1}\bigl(1 - PD_j\bigr) \qquad (1)$$

gdzie:

- PD_i — prawdopodobieństwo detekcji podczas i-tej akcji przeciwnika,
- PC_i — prawdopodobieństwo skutecznej komunikacji z siłami reagowania,
- $P(R|A_i)$ — warunkowe prawdopodobieństwo skutecznej odpowiedzi sił reagowania.

Metodologia opiera się na trzech założeniach: (1) przeciwnik działa wyłącznie fizycznie, (2) skuteczność komponentów jest stała w czasie, (3) analiza wymaga ręcznego modelowania ścieżek ataku. Założenia te, uzasadnione w latach 70., stanowią dziś główne ograniczenia metodologii.

EASI rozszerzone o cyberodporność

W odpowiedzi na lukę związaną z brakiem uwzględnienia cyberzagrożeń do klasycznej formuły EASI wprowadzono współczynnik cyberodporności CR_i dla każdego cyfrowego komponentu systemu. Modyfikacja polega na skorygowaniu prawdopodobieństwa detekcji:

$$PD_i^{\text{cyber}} \;=\; PD_i \cdot CR_i \qquad (2)$$

gdzie:

- PD_i^{cyber} — skorygowane prawdopodobieństwo detekcji z uwzględnieniem cyberzagrożeń,
- PD_i — nominalne prawdopodobieństwo detekcji przy założeniu braku cyberataków,
- $CR_i \in [0, 1]$ — współczynnik cyberodporności i-tego komponentu.

Wartość $CR_i = 1$ odpowiada pełnej odporności na cyberzagrożenia, $CR_i = 0$ oznacza całkowite przejęcie komponentu przez atakującego. W analizowanym studium przypadku przyjęto trzy poziomy klasyfikacji: wysoki ($CR \approx 1{,}0$), umiarkowany ($CR \approx 0{,}5$) oraz niski ($CR \approx 0{,}1$). Wyniki wskazują na istotne przeszacowanie w klasycznym ujęciu: ta sama ścieżka ataku, oceniana klasyczną metodą EASI, uzyskała $P(I) = 89{,}88\%$; po uwzględnieniu współczynników cyberodporności wartość ta spadła do 1,63%.

EASI z modelem degradacji

Klasyczna metodologia pomija stopniowe obniżanie się skuteczności komponentów ESZT w czasie eksploatacji. W niniejszej pracy zaproponowano probabilistyczny model degradacji oparty na pięciu czynnikach kumulatywnych. Efektywność systemu w miesiącu i definiuje się jako:

$$SE_i \;=\; \max\bigl(100 - X_i,\; 0\bigr) \tag{3}$$

gdzie skumulowany wpływ degradacyjny X_i oblicza się jako:

$$X_i \;=\; \sum_{j=1}^{i} \bigl(AG_j + KG_j + TI_j + VI_j - IE_j\bigr) \tag{4}$$

Poszczególne składniki, modelowane odpowiednimi rozkładami probabilistycznymi, oznaczają:

- $AG_j \sim \mathcal{N}(0{,}5;\; 0{,}15^2)$ — starzenie się komponentów (ang. *Aging*); odzwierciedla średni spadek sprawności o ok. 50 pp w dekadzie,
- $KG_j \;\sim\; \mathcal{N}(0{,}1;\;\; 0{,}03^2)$ — wzrost wiedzy i umiejętności przeciwnika (ang. *Knowledge Growth*); ok. 10% rocznego przyrostu skuteczności ataków,
- $TI_j = \mu \cdot \log(j) + \varepsilon$, $\mu = 0{,}7$, $\varepsilon \sim \mathcal{N}(0;\; 0{,}21^2)$ — postęp technologiczny narzędzi ataku (ang. *Technology Improvement*); dynamika malejących przyrostów,
- $VI_j \sim \text{Bernoulli}(p = 0{,}2)$ z jednorazowym spadkiem o 10 pp — wpływ nowych podatności (ang. *Vulnerability Introduction*); zdarzenia rzadkie, skokowe, avg. 1 na 5 miesięcy,
- $IE_j \;\sim\; \mathcal{N}(0{,}5;\;\; 0{,}15^2)$ — efekty kwartalnych działań konserwacyjnych (ang. *Improvement Effects*); redukuje skumulowany spadek o średnio 50 pp.

Zastosowanie funkcji $\max(\cdot, 0)$ zapewnia, że efektywność systemu nie przyjmuje wartości ujemnych. Integracja modelu degradacji z formułą EASI polega na modyfikacji prawdopodobieństwa detekcji każdego elementu:

$$PD_i^{\text{deg}} \;=\; PD_i \cdot \frac{SE_i}{100} \tag{5}$$

gdzie PD_i^{deg} to prawdopodobieństwo detekcji z uwzględnieniem degradacji, PD_i to nominalne prawdopodobieństwo detekcji nowego elementu, a SE_i to skuteczność elementu po i miesiącach eksploatacji. Symulacje Monte Carlo ($N = 10\,000$ przebiegów) wykazały, że w rozważanym scenariuszu mediana czasu do przekroczenia progu 50% efektywności CPPS wynosi 75–80 miesięcy (ok. 6,5–6,7 roku).

Zintegrowany model EASI-Δ

Naturalna konsekwencja odrębnych analiz trzech perspektyw jest ich łączne ujęcie w spójnej ocenie końcowej. Model EASI-Δ definiuje zintegrowane prawdopodobieństwo

przerwania ataku jako **średnią arytmetyczną** trzech cząstkowych ocen EASI: klasycznej, rozszerzonej o cyberodporność oraz rozszerzonej o degradację:

$$P(I)^{\Delta} \;=\; \frac{P(I)^{\text{klasyczna}} \;+\; P(I)^{\text{cyber}} \;+\; P(I)^{\text{deg}}}{3} \tag{6}$$

gdzie:

- $P(I)^{\text{klasyczna}}$ — prawdopodobieństwo przerwania ataku wyliczone klasyczną formułą EASI (1), z nominalnymi wartościami PD_i,
- $P(I)^{\text{cyber}}$ — prawdopodobieństwo przerwania ataku przy zastąpieniu PD_i przez $PD_i^{\text{cyber}} = PD_i \cdot CR_i$ zgodnie z równaniem (2),
- $P(I)^{\text{deg}}$ — prawdopodobieństwo przerwania ataku przy zastąpieniu PD_i przez $PD_i^{\text{deg}} = PD_i \cdot SE_i/100$ zgodnie z równaniem (5).

Podejście oparte na średniej arytmetycznej odzwierciedla przekonanie, że żadna z trzech perspektyw nie powinna jednostronnie dominować końcowej oceny. Model EASI-Δ zachowuje kompatybilność z aparatem obliczeniowym EASI, przy czym każdy z trzech składników jest wyznaczany niezależnie, a ich uśrednienie daje ocenę konserwatywną wobec podejścia klasycznego, a zarazem bardziej ostrożną interpretacyjnie niż samo ujęcie cybernetyczne. Dla analizowanego studium przypadku ($\overline{CR} = 0{,}5$; $\overline{SE} = 70\%$):

$$P(I)^{\Delta} \;=\; \frac{89{,}88\% + 1{,}63\% + 76{,}19\%}{3} \;\approx\; 55{,}9\%$$

Wynik ten (ok. 34 punkty procentowe poniżej oceny klasycznej) stanowi znacznie bardziej realistyczną miarę skuteczności systemu eksploatowanego od kilku lat w środowisku z istotnymi zagrożeniami cybernetycznymi. Zestawienie wyników czterech wariantów oceny dla tej samej ścieżki ataku przedstawia poniższa tabela:

Table 1: Porównanie przykładowych wyników oceny skuteczności $P(I)$ dla tej samej ścieżki ataku przy czterech wariantach metodologicznych.

Wariant metodologiczny	$\boldsymbol{P(I)}$
Klasyczna EASI	89,88%
EASI z uwzględnieniem cyberzagrożeń	1,63%
EASI z uwzględnieniem degradacji	76,19%
EASI-Δ — średnia arytmetyczna trzech ocen	$\approx$ 55,9%

Różnica ponad 88 punktów procentowych między oceną klasyczną a oceną uwzględniającą cyberzagrożenia jednoznacznie wskazuje na konieczność uwzględnienia tego aspektu w każdej rzetelnej ocenie skuteczności CPPS.

Heurystyka *stealth* dla algorytmu A*

Do heurystycznej eksploracji grafów ścieżek ataku zastosowano algorytm A* z oryginalną funkcją heurystyczną *stealth*, opracowaną dla potrzeb modelowania zachowania przeciwnika

dążącego do minimalizacji ryzyka wykrycia:

$$h_{\text{stealth}}(n,\, t) \;=\; w_1 \cdot D_{\text{Manhattan}}(n,\, t) \cdot S_{\text{scale}} \;+\; w_2 \cdot \frac{1}{DD_n} \;+\; w_3 \cdot BT_n \;+\; w_4 \cdot DR_n \quad (7)$$

gdzie:

- $D_{\text{Manhattan}}(n,t) \cdot S_{\text{scale}}$ — podstawowa orientacja przestrzenna: odległość Manhattan od węzła n do celu t przeskalowana do wymiarów planu; człon odpowiedzialny za directionality przeszukiwania,
- $1/DD_n$ — odwrotność opóźnienia detekcji (*Detection Delay*) węzła n; im większe opóźnienie detekcji, tym niższy koszt heurystyczny węzła, co preferuje trasy trudno wykrywalne,
- BT_n — czas pokonania bariery (*Barrier Time*) węzła n; penalizuje węzły wymagające długiego czasu przełamywania barier,
- DR_n — inherentne ryzyko detekcji (*Detection Risk*) węzła n; bezpośrednia penalizacja węzłów o wysokim prawdopodobieństwie wykrycia,
- $w_1, w_2, w_3, w_4 \;\geq\; 0$ — współczynniki wagowe umożliwiające modelowanie różnych profili przeciwnika.

Kalibracja wag pozwala symulować odmienne scenariusze: profil *stealth professional* (wysokie w_2 i w_4) odzwierciedla priorytety doświadczonego adwersarza unikającego detekcji; profil *time-critical attack* (dominujące w_3) reprezentuje priorytet szybkości przy ataku pod presją czasu; profil *amateur intruder* (zbalansowane wagi) oddaje ograniczone umiejętności sprawcy. Studium przypadku na rzeczywistym planie obiektu wykazało identyfikację 115 racjonalnych ścieżek ataku, z kosztem optymalnej ścieżki wynoszącym 271,42 oraz kosztem najgorszej ścieżki 9264,50. Wartość kosztu nie jest wyrażona w jednostkach fizycznych, lecz stanowi bezwymiarową miarę numeryczną wynikającą z agregacji składników funkcji heurystycznej *Stealth* (długości ścieżki, opóźnienia detekcji, czasu pokonania barier oraz ryzyka wykrycia). Im wyższa wartość kosztu, tym ścieżka jest trudniejsza do skutecznego przebycia i mniej atrakcyjna z perspektywy adwersarza.

6. Implementacja narzędzia AI-SecPA

Wszystkie opisane elementy metodologiczne zostały zintegrowane w spójnym narzędziu programistycznym AI-SecPA (ang. *AI – Security Paths Analyzer*). Narzędzie prowadzi użytkownika przez dziesięcioetapowy, ustrukturyzowany cykl analityczny, podzielony na trzy główne części.

Część I – Przygotowanie danych i kontekstu (etapy I–II): Etap I polega na wczytaniu planu architektonicznego obiektu *bez* naniesionego rozmieszczenia systemów bezpieczeństwa. Jest to podejście zgodne z metodologią zarządzania ryzykiem, w której ryzyko inherentne (bez środków kontrolnych) jest wyznaczane przed oceną ryzyka

rezydualnego. Etap II obejmuje przeprowadzenie ankiety bezpieczeństwa (tzw. Security Survey), czyli systematyczne zbieranie danych kontekstowych niewidocznych na planach: specyfikacji technicznych urządzeń i ich wieku (dane wejściowe do modelu degradacji), parametrów grup interwencyjnych, współczynników cyberodporności poszczególnych komponentów, a także danych o kontekście geopolitycznym i organizacyjnym.

Część II – Detekcja i analiza obrazu (etapy III–VI): Etap III to automatyczna detekcja elementów architektonicznych (ściany, drzwi, okna, korytarze) z użyciem YOLOv8 oraz post-processingu obejmującego eliminację duplikatów (NMS) i geometryczną segmentację ścian na logiczne odcinki. Etap IV umożliwia manualną korektę błędów detekcji przez eksperta. Etap V dotyczy identyfikacji elementów ESZT na planie bezpieczeństwa z użyciem modelu COG. Etap VI obejmuje segmentację pomieszczeń i weryfikację graficznej ciągłości struktury. Miejsca nieciągłości są sygnalizowane wizualnie, a ich usunięcie jest warunkiem poprawnej analizy grafowej.

Część III – Modelowanie grafowe, ocena i raportowanie (etapy VII–X): Etap VII to automatyczne generowanie skierowanego grafu $G = (V, E)$ ze wszystkich względnie racjonalnych ścieżek ataku na podstawie stworzonej topologii obiektu, z przypisaniem węzłom i krawędziom parametrów bezpieczeństwa (prawdopodobieństwa detekcji, czasy barier, współczynniki cyberodporności, wiek komponentów). Etap VIII to heurystyczna eksploracja grafu algorytmem A* z heurystyką *stealth* oraz walidacja i filtracja ścieżek. System analizuje setki scenariuszy ataku w czasie niedostępnym w podejściu manualnym. Etap IX to ocena EASI-Δ dla zidentyfikowanych ścieżek z automatyczną identyfikacją ścieżek krytycznych i węzłów kluczowych. Etap X to generowanie raportu bezpieczeństwa z użyciem LLM sterowanego przez bibliotekę Outlines, obejmującego syntezę wyników ilościowych, analizę ryzyka na poziomach inherentnym i rezydualnym oraz priorytetyzowane rekomendacje dla zarządcy obiektu.

Narzędzie zaprojektowano z myślą o możliwości uruchomienia lokalnie, na komputerze wyposażonym w akcelerator GPU, co eliminuje konieczność posiadania infrastruktury sieciowej.

7. Wyniki i wkład w stan wiedzy

Rozprawa przedstawia szereg elementów o charakterze metodologicznym i technicznym. W zakresie modelowania degradacji ESZT zaproponowano probabilistyczne ujęcie procesów degradacji specyficznych dla CPPS, uwzględniające równocześnie czynniki techniczne (starzenie komponentów elektronicznych), środowiskowe (ewolucja zagrożeń i narzędzi ataku) oraz operacyjne (efekty działań konserwacyjnych). Rozszerzenie uznanej metodologii EASI o wskaźniki degradacji i cyberodporności stanowi krok w kierunku urealnionego modelu z uwzględnieniem długoterminowej oceny skuteczności systemów ochronnych. Podejście to różni się w tym zakresie od statycznych, nominalnych szacunków dominujących w dotychczasowej praktyce.

Omówiona w rozprawie integracja metodologii COG z heurystyczną analizą ścieżek ataku tworzy spójny łańcuch analityczny: od automatycznej interpretacji diagramów

technicznych, przez budowę grafów obiektów chronionych, po generowanie i ocenę scenariuszy zagrożeń. Zastosowanie dużych modeli językowych do automatycznej syntezy i raportowania wyników otwiera nowe możliwości w zakresie komunikacji między warstwą techniczną systemu, a użytkownikiem końcowym, który zazwyczaj, jako specjalista w zarządzaniu ryzykiem i bezpieczeństwem, niekoniecznie jest biegły w metodach numerycznych.

Koncepcja klasy kontekstowej COG jako pośredniego poziomu ontologicznego pomiędzy atomową detekcją obiektów, a interpretacją semantyczną stanowi zmianę podejścia w analizie diagramów technicznych. Podejście to pozwala modelowi bezpośrednio „postrzegać" kontekst (strukturę legendy, relacje przestrzenne symboli) zamiast wnioskować o nim w osobnym etapie post-processingu.

Zaproponowane metody i narzędzia mają charakter interdyscyplinarny. Łączą one w sobie dorobek inżynierii niezawodności, cyberbezpieczeństwa, sztucznej inteligencji, teorii grafów oraz zarządzania bezpieczeństwem w spójną metodologię, której abstrakcyjne ujęcie CPPS jako systemu o określonych właściwościach probabilistycznych umożliwia stosowanie jej w różnych sektorach infrastruktury krytycznej (od energetyki i transportu po sektor finansowy).

Główne elementy wkładu naukowego pracy można ująć następująco:

- **Probabilistyczny model degradacji ESZT** — nowatorskie ujęcie pięciu klas czynników degradacyjnych (AG, KG, TI, VI, IE) w kontekście CPPS, zweryfikowane symulacjami Monte Carlo ($N = 10\,000$) w horyzoncie 120 miesięcy.
- **Rozszerzenie EASI o cyberodporność i degradację (EASI-Δ)** — modyfikacja uznanej metodologii poprzez wprowadzenie współczynnika CR_i oraz wskaźnika SE_i, a następnie uśrednienie trzech cząstkowych ocen $P(I)$ w jedną miarę zintegrowaną.
- **Metodologia COG** — koncepcja kontekstowej interpretacji symboli technicznych w dokumentacji bezpieczeństwa z użyciem hierarchicznych klas ontologicznych, osiągająca mAP@0.5 $\approx 99\%$ i 100% poprawności parowań symbol–etykieta.

Powyższy wkład posłużył jako podstawa do stworzenia i implementacji kompletnego narzędzia oceny CPPS łączącego detekcję YOLOv8, modelowanie grafowe z autorską heurystyką *Stealth*, ocenę EASI-Δ oraz generowanie raportów przez LLM sterowany biblioteką Outlines. Narzędzie, to łączy w sobie również trzy perspektywy: degradację techniczną, wpływ zagrożeń cybernetycznych oraz operacyjny kontekst funkcjonowania systemu.

8. Ograniczenia i kierunki dalszych badań

W rozprawie zidentyfikowano kilka istotnych ograniczeń proponowanego podejścia. W odniesieniu do warstwy percepcyjnej wskazuje się, że modele trenowano na ograniczonych zbiorach danych, a różnice w jakości skanów, standardach rysunkowych i skali mogą

wpływać na skuteczność detekcji; model wykazuje ponadto wrażliwość na rotację symboli ESZT względem orientacji bazowej w legendzie. W odniesieniu do warstwy grafowej zauważa się, że heurystyki nie obejmują wszystkich aspektów rzeczywistych scenariuszy ataku, w szczególności zachowań nieracjonalnych ani pełnej dynamiki organizacyjnej obiektu. Ograniczeniem modelu degradacyjnego pozostaje konieczność kalibracji na danych operacyjnych z rzeczywistych instalacji CPPS.

Wśród kierunków dalszych badań wskazano: rozszerzenie zbiorów danych treningowych o plany z różnych regionów geograficznych i standardów projektowych; rozwój algorytmów niezmienności względem transformacji geometrycznych (rotacja, skalowanie) w module COG; walidację modelu degradacyjnego na danych operacyjnych; rozszerzenie o zaawansowaną sensorykę IoT umożliwiającą ciągłe monitorowanie efektywności; dalszą integrację z podejściami Digital Twin.

9. Literatura cytowana w referacie

1. Bennett, H. A., *EASI Approach to Physical Security Evaluation*, Sandia Laboratories, SAND-76-0500, 1977.
2. Garcia, M. L., *The Design and Evaluation of Physical Protection Systems*, 2nd ed., Butterworth-Heinemann, 2008.
3. Clem, J., Atkins, W. D., Urias, V., *Investigation of Cyber-Enabled Physical Attack Scenarios*, Sandia National Laboratories, 2015.
4. Porter, S., *Physical Protection Systems and the Cyber Security Component*, Lawrence Livermore National Laboratory, 2015.
5. Depoy, J., Phelan, J., Sholander, P., Smith, B., Varnado, G., Wyss, G., "Risk assessment for physical and cyber attacks on critical infrastructures," in *MILCOM 2005 – IEEE Military Communications Conference*, 2005.
6. Snell, M., Rivers, J., "The treatment of blended attacks in nuclear security effectiveness assessments," Sandia National Laboratories, 2015.
7. Guo, Y., Yan, A., Wang, J., "Cyber security risk analysis of physical protection systems of nuclear power plants and research on the cyber security test platform using digital twin technology," in *International Conference on Power System Technology (POWERCON)*, 2021.
8. Bowen, Z., Ming, Y., Hidekazu, Y., Hongxing, L., "Evaluation of physical protection systems using an integrated platform for analysis and design," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(11), 2945–2955, 2017.
9. Andiwijayakusuma, D., Setiadipura, T., Purqon, A., Su'ud, Z., "The development of EASI-based multi-path analysis code for nuclear security system with variability extension," *Nuclear Engineering and Technology*, 54(10), 3604–3613, 2022.
10. Kapusta, J., Bauer, W., Baranowski, J., "Evaluation of the effectiveness of physical protection systems with consideration of its cyber-resilience," in *27th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 457–461, 2023.
11. Kapusta, J., Bauer, W., Baranowski, J., "Contextual Object Grouping (COG): A Specialized Framework for Dynamic Symbol Interpretation in Technical Security Diagrams," *Algorithms*, 18(10), 642, 2025. doi:10.3390/a18100642.
12. Kapusta, J., Bauer, W., Baranowski, J., "Ocena skuteczności systemów ochrony cyberfizycznej z uwzględnieniem degradacji elektronicznych systemów zabezpieczeń," *Pomiary Automatyka Robotyka*, 29(3), 2025. doi:10.14313/PAR_257/29.
13. Narale, S., Verma, A., Anand, S., "Structure and degradation of aluminum electrolytic capacitors," in *National Power Electronics Conference (NPEC)*, pp. 1–6, 2019.

14. Yu, C., Jiang, L., Yuan, J., "Study of performance degradations in DC–DC converter due to hot carrier stress by simulation," *Microelectronics Reliability*, 46(9–11), 1840–1843, 2006.

15. Alam, M. K., Khan, F. H., "Reliability analysis and performance degradation of a boost converter," in *IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 5592–5597, 2013.

16. Huang, H., Boyer, A., Dhia, S. B., "Analysis and modeling of passive device degradation for a long-term electromagnetic emission study of a DC-DC converter," *Microelectronics Reliability*, 55(9–10), 2061–2066, 2015.

17. Redmon, J., Divvala, S., Girshick, R., Farhadi, A., "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

18. Jocher, G., Chaurasia, A., Qiu, J., *YOLO by Ultralytics*, 2023.

19. Krizhevsky, A., Sutskever, I., Hinton, G. E., "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

20. He, K., Zhang, X., Ren, S., Sun, J., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

21. Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., Zhou, M., "LayoutLM: Pre-training of text and layout for document image understanding," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1192–1200, 2020.

22. Huang, Y., Lv, T., Cui, L., Lu, Y., Wei, F., "LayoutLMv3: Pre-training for document AI with unified text and image masking," *arXiv preprint* arXiv:2204.08387, 2022.

Publikacje autora rozprawy

1. **Contextual Object Grouping (COG): A Specialized Framework for Dynamic Symbol Interpretation in Technical Security Diagrams**
   Kapusta, J.; Bauer, W.; Baranowski, J.
   Algorithms, 18(10), 642, 2025.
   DOI: 10.3390/a18100642
   Główny autor — wkład: 80%

2. **Ocena skuteczności systemów ochrony cyberfizycznej z uwzględnieniem degradacji elektronicznych systemów zabezpieczeń**
   Kapusta, J.; Bauer, W.; Baranowski, J.
   Pomiary Automatyka Robotyka, R. 29, Nr 3/2025, s. 29–37.
   DOI: 10.14313/PAR_257/29
   Główny autor — wkład: 80%

3. **Detection of building infrastructure elements in layouts using YOLO as a foundation for graph-based adversarial attack path modeling**
   Kapusta, J.; Bauer, W.; Baranowski, J.
   W: *MMAR 2025 — 29th International Conference on Methods and Models in Automation and Robotics*, 26–29 sierpnia 2025, Międzyzdroje, Polska. Piscataway: IEEE, 2025, s. 357–361.
   Główny autor i prezenter — wkład: 90%

4. **Use of machine learning methods in assessing physical security with cybersecurity elements**
   Kapusta, J.; Bauer, W.; Baranowski, J.
   W: *XLVI SPETO — Konferencja z Podstaw Elektrotechniki i Teorii Obwodów*, Gliwice–Wisła, 21–23 maja 2025. Gliwice: Politechnika Śląska, 2025, s. 29–30.
   Główny autor i prezenter — wkład: 90%

5. **A stochastic approach to modeling long-term degradation of the electronic security systems as a crucial element of evaluation of Physical Protection Systems (PPS)**
   Kapusta, J.; Bauer, W.; Baranowski, J.
   W: *IECON 2024 — 50th Annual Conference of the IEEE Industrial Electronics Society*, Chicago, USA, 3–6 listopada 2024. Piscataway: IEEE, 2024, s. 1–7.
   Główny autor — wkład: 90%

6. **Evaluation of the effectiveness of physical protection systems with consideration of its cyber-resilience**
   Kapusta, J.; Bauer, W.; Baranowski, J.
   W: *MMAR 2023 — 27th International Conference on Methods and Models in Automation and Robotics*, 22–25 sierpnia 2023, Międzyzdroje, Polska. Piscataway: IEEE, 2023, s. 457–461.
   Główny autor i prezenter — wkład: 90%

7. **Analiza danych i optymalizacja w Przemyśle 4.0**
   Baranowski, J.; Kucharska, E.; Bauer, W.; Grobler-Dębska, K.; Kashpruk, N.; Kraszewska, M.; Mularczyk, R.; Piskor-Ignatowicz, C.; Dudek, A.; Dworak, D.; **Kapusta, J.**; Kawa, K.
   W: *Nauka – Technika – Technologia: Seria Wydawnicza AGH*, T. 5. Kraków: Wydawnictwa AGH, 2022, s. 43–52.
   Współautor — wkład: 10%