dr hab. inż. Mikołaj Leszczuk
Instytut Telekomunikacji
Wydział Informatyki, Elektroniki i Telekomunikacji
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
al. Mickiewicza 30
30-059 Kraków
mikolaj.leszczuk@agh.edu.pl

Kraków, dn. 9 lutego 2024 r.

RECENZJA ROZPRAWY DOKTORSKIEJ
mgr inż. Katarzyny Filus
„Wyjaśnialność i bezpieczeństwo systemów inteligentnych”
Promotor: dr hab. inż. Joanna Domańska, prof. IITiS PAN
Instytut Informatyki Teoretycznej i Stosowanej
Polskiej Akademii Nauk
Dziedzina: nauki inżynieryjno-techniczne
Dyscyplina: informatyka techniczna i telekomunikacja

Aktualność i znaczenie rozprawy doktorskiej

Aktualność i znaczenie rozprawy doktorskiej Katarzyny Filus pt. „Wyjaśnialność i bezpieczeństwo systemów inteligentnych” są istotne, gdyż dotyczą kluczowych aspektów rozwoju sztucznej inteligencji (ang. *Artificial Intelligence*, AI) w dziedzinie informatyki technicznej i telekomunikacji. Autorka skupia się na bezpieczeństwie i wyjaśnialności systemów AI, oferując nowe metody i rozwiązania do poprawy tych aspektów. Rozprawa bada tradycyjne zagrożenia informatyczne oraz specyficzne dla AI, koncentrując się także na wyjaśnialności i interpretowalności modeli uczenia maszynowego. Praca ta jest odpowiedzią na aktualne wyzwania i standardy w dziedzinie AI.

Rozprawa Katarzyny Filus stanowi ważny wkład w rozwój i zrozumienie systemów inteligentnych. Praca ta kładzie nacisk na istotne zagadnienia w kontekście rosnącego zapotrzebowania na przejrzyste i bezpieczne technologie AI. Autorka podejmuje trudne i znaczące kwestie związane z interpretowalnością i wiarygodnością algorytmów AI, które są niezbędne do budowania zaufania i akceptacji wśród użytkowników i społeczeństwa. Proponowane przez nią metody i badania są odpowiedzią na kluczowe problemy w dziedzinie bezpieczeństwa informatycznego i wyjaśnialności AI, co ma znaczący wpływ na praktyczne zastosowania tych technologii. Rozprawa ta, przez połączenie teorii z praktycznymi rozwiązaniami, pokazuje, jak ważne jest zrozumienie i rozwiązywanie problemów związanych z wyjaśnialnością i bezpieczeństwem systemów inteligentnych, nie tylko dla postępu naukowego, ale także dla ich skutecznego i odpowiedzialnego wdrażania w różnych dziedzinach życia.

Problem naukowy wyjaśnialności i bezpieczeństwa systemów AI

Problem naukowy podejmowany przez Katarzynę Filus dotyczy wyjaśnialności i bezpieczeństwa systemów inteligentnych. Autorka skupia się na kluczowych wyzwaniach związanych z bezpieczeństwem systemów opartych na algorytmach AI oraz ich

wyjaśnialnością. Proponowane metody mają na celu poprawę tych aspektów. Autorka analizuje różne aspekty bezpieczeństwa systemów inteligentnych, w tym tradycyjne zagrożenia informacyjne oraz zagrożenia bezpośrednio związane z algorytmami AI. W pracy badane są także nowe podejścia do wyjaśniania działania algorytmów uczenia głębokiego, mające na celu zwiększenie ich interpretowalności i zrozumienia przez użytkowników. Rozprawa wnosi istotny wkład w dziedzinie cyber-bezpieczeństwa, koncentrując się na analizie podatności i ataków sieciowych, a także na projektowaniu i testowaniu nowych metod obrony. Autorka wprowadza innowacyjne podejścia do wykrywania podatności oraz metody testowania algorytmów AI, uwzględniając zarówno aspekty techniczne, jak i etyczne. Praca podkreśla związek między bezpieczeństwem a wyjaśnialnością systemów AI, proponując rozwiązania, które mają na celu zwiększenie transparentności i zaufania do technologii AI. Problem został sformułowany w sposób trafny, odzwierciedlając aktualne wyzwania i potrzeby w dziedzinie AI.

Sformułowanie hipotezy

Hipoteza pracy Katarzyny Filus głęboko penetruje problematykę bezpieczeństwa i wyjaśnialności w dziedzinie AI, dwóch kluczowych wyzwań współczesnej AI. Praca ta nie tylko zgłębia aktualne problemy niskiego zaufania społecznego do AI i ograniczonego stosowania systemów inteligentnych w krytycznych obszarach, ale również proponuje nowatorskie metody mające na celu poprawę obu tych aspektów. Metody te, takie jak interpretacja działania głębokich sieci splotowych czy inteligentny system lokalizacji użytkowników, są intuicyjne w użyciu i proste w interpretacji, co przyczynia się do zwiększenia wyjaśnialności działania systemów AI.

W kontekście globalnych trendów w AI na ostatnie lata, szczególnie istotne staje się dążenie do tworzenia bezpiecznych i etycznych systemów AI, z naciskiem na przejrzystość i ochronę prywatności. Wyzwania te wymagają nowych regulacji, jak te przykładowo zapowiadane przez Departament Zdrowia i Opieki Społecznej USA (ang. *United States Department of Health and Human Services*, HHS), nakładające na twórców AI obowiązek zapewnienia przejrzystości i bezpieczeństwa algorytmów stosowanych w ochronie zdrowia. Dodatkowo, zauważa się rosnące zapotrzebowanie na rozwój AI w obszarach takich jak cyber-bezpieczeństwo, z użyciem zaawansowanych algorytmów uczenia maszynowego do wykrywania i reagowania na zagrożenia.

Katarzyna Filus, adresując te kwestie w swojej pracy, wpisuje się w aktualne dyskusje dotyczące AI i przyczynia się do rozwoju metod poprawiających bezpieczeństwo i wyjaśnialność systemów inteligentnych, co jest zgodne z globalnymi dążeniami i potrzebami współczesnego świata.

Rozwiązania problemów bezpieczeństwa i wyjaśnialności

Autorka rozprawy doktorskiej skupiła się na poprawie bezpieczeństwa i wyjaśnialności systemów inteligentnych. W pracy doktorskiej zaproponowano i zastosowano różnorodne metody, które skutecznie adresują te zagadnienia. Metody te obejmują nowe podejścia do inicjalizacji i trenowania losowych sieci neuronowych, analizę podatności oprogramowania, a także rozwój metod wizualizacji i wyjaśniania działania głębokich sieci splotowych. Badania potwierdziły skuteczność tych metod, co wskazuje na to, że Autorka trafnie sformułowała i rozwiązała postawione problemy. Praca ta stanowi istotny wkład w dziedzinę bezpieczeństwa i wyjaśnialności systemów inteligentnych, otwierając nowe perspektywy dla dalszych badań i praktycznego zastosowania AI. Szczególnie godne uwagi są opracowane metody wizualizacji

aktywacji sieci neuronowych, które znacząco przyczyniają się do lepszego zrozumienia i interpretacji działania modeli AI, co jest kluczowe dla budowania zaufania do tych technologii.

Innowacyjny wkład w rozwój bezpieczeństwa i wyjaśnialności

Praca Katarzyny Filus wnosi znaczący wkład w dziedzinę bezpieczeństwa i wyjaśnialności systemów inteligentnych poprzez rozwój i wdrożenie nowatorskich metod. Szczególnie istotne jest zastosowanie modyfikacji Losowych Sieci Neuronowych (ang. *Random Neural Networks*, RNN) do skutecznego wykrywania zagrożeń bezpieczeństwa oraz innowacyjne podejście do wyjaśniania działania głębokich sieci splotowych przez zaawansowane metody wizualizacji. Autorka przyczynia się do lepszego zrozumienia i testowania systemów AI, co ma kluczowe znaczenie dla ich praktycznego zastosowania w krytycznych obszarach, takich jak cyber-bezpieczeństwo i wizja komputerowa. Praca wyznacza nowe ścieżki w badaniach nad wyjaśnialnością algorytmów, oferując jednocześnie praktyczne rozwiązania do testowania i poprawy bezpieczeństwa systemów opartych na AI.

Oryginalny dorobek Autorki i jego znaczenie

Oryginalny dorobek naukowy Katarzyny Filus, prezentowany w jej rozprawie doktorskiej, obejmuje kilka kluczowych osiągnięć w dziedzinie AI, skupiając się na bezpieczeństwie i wytłumaczalności systemów AI. W jej dorobku szczególnie wyróżniają się następujące punkty:

1. Rozwój nowatorskich metod usprawniających bezpieczeństwo i wytłumaczalność systemów opartych na algorytmach AI. Autorka skupiła się na łączeniu teoretycznych podstaw z praktycznymi zastosowaniami, proponując rozwiązania, które mogą być zastosowane w rzeczywistych systemach.
2. Współautorstwo w publikacjach naukowych, prezentujących wyniki jej badań. Katarzyna Filus miała istotny wkład w wiele publikacji:
    a. “NetSat: Network Saturation Adversarial Attack” (BigData, 2023, 70 punktów)
    b. “Recycling of Generic ImageNet-trained Models for Smart-city Applications” (DSAA, 2023, 140 punktów)
    c. “Global Entropy Pooling Layer for Convolutional Neural Networks” (Neurocomputing, 2023, 140 punktów)
    d. “Software Vulnerabilities in TensorFlow-Based Deep Learning Applications” (Computers and Security, 2023, 140 punktów)
    e. “Visual examination of relations between known classes for deep neural network classifiers” (BigData, 2023, 70 punktów)
    f. “Robust category recognition based on deep templates for educational mobile applications” (BigData, 2023, 70 punktów)
    g. “HierAct: a Hierarchical Model for Human Activity Recognition in Game-Like Educational Applications” (BigData, 2023, 70 punktów)
    h. “Real-time testing of vision-based systems for AGVs with ArUco markers” (BigData, 2022, 70 punktów)
    i. “NAM: What Does a Neural Network See?” (IJCNN, 2022, 70 punktów)
    j. “Cost-Effective Filtering of Unreliable Proximity Detection Results Based on BLE RSSI and IMU Readings Using Smartphones” (Scientific Reports, 2022, 140 punktów)
    k. “Building a real- time testing platform for unmanned ground vehicles with UDP Bridge” (Sensors, 2022, 100 punktów)

l. “Finding the best hardware configuration for 2D SLAM in indoor environments via simulation based on Google Cartographer” (Scientific Reports, 2022, 140 punktów)
m. “Adaptive Hurst-Sensitive Active Queue Management” (Entropy, 2022, 100 punktów)
n. “Approximation Models for the Evaluation of TCP/AQM Networks” (Bulletin of the Polish Academy of Sciences: Technical Sciences, 2022, 100 punktów)
o. “SDK4ED: One-click platform for Energy-aware, Maintainable and Dependable Applications” (DATE, 2022, 70 punktów)
p. “Efficient Feature Selection for Static Analysis Vulnerability Prediction” (Sensors, 2021, 100 punktów)
q. “LIDAR Point Cloud generation for SLAM algorithm evaluation” (Sensors, 2021, 100 punktów)
r. “Supervised learning of Neural Networks for Active Queue Management on the Internet” (Sensors, 2021, 100 punktów)
s. “Random Neural Network for Lightweight Attack Detection in the IoT” (MASCOTS 2021, 70 punktów)
t. “The Random Neural Network as a Bonding Model for Software Vulnerability Prediction” (MASCOTS 2020, 70 punktów)
u. “Long-Range Dependent Traffic Classification with Convolutional Neural Networks Based on Hurst Exponent Analysis” (Entropy, 2020, 100 punktów)
v. “The self-similar markovian sources” (Applied Sciences, 2020, 100 punktów)

3. Prace Katarzyny Filus przyczyniają się do rozwoju narzędzi i technik, które mają praktyczne zastosowanie w zakresie bezpieczeństwa oraz interpretowalności systemów AI, co jest kluczowe dla zaufania i akceptacji tych technologii przez społeczeństwo.

Podsumowując, dorobek Katarzyny Filus w jej rozprawie doktorskiej jest znaczący z punktu widzenia rozwoju bezpiecznych i wytłumaczalnych systemów AI, co ma duże znaczenie zarówno teoretyczne, jak i praktyczne w kontekście coraz szerszego wdrażania AI w różnych aspektach życia codziennego i przemysłu.

Znaczenie poznawcze i praktyczne wkładu w dyscyplinę

Wkład Katarzyny Filus w dyscyplinę informatyki technicznej i telekomunikacji (a w szczególności w obszar AI) jest znaczący zarówno z poznawczego, jak i praktycznego punktu widzenia. Autorka skoncentrowała się na poprawie bezpieczeństwa i wyjaśnialności systemów inteligentnych, proponując nowe metody wykrywania zagrożeń bezpieczeństwa i podatności, jak również rozwiązania umożliwiające lepsze zrozumienie działania algorytmów uczenia głębokiego. Znaczenie poznawcze pracy wynika z głębokiej analizy aktualnych wyzwań w dziedzinie bezpieczeństwa i wyjaśnialności AI, a praktyczne znaczenie wiąże się z opracowaniem skutecznych narzędzi i metod, które mogą być stosowane w rzeczywistych zastosowaniach AI. Praca ta w istotny sposób przyczynia się do rozwoju bezpieczniejszych i bardziej transparentnych systemów AI, oferując zarówno teoretyczne podstawy, jak i praktyczne implementacje, które mogą znaleźć zastosowanie w różnorodnych aplikacjach. Jej wyniki mogą być wykorzystane do dalszych badań i rozwoju w dziedzinie AI, zwiększając zaufanie społeczne i otwierając nowe możliwości w wykorzystaniu technologii inteligentnych w krytycznych i odpowiedzialnych zastosowaniach.

Poziom wiedzy technicznej w rozprawie doktorskiej

Rozprawa doktorska Katarzyny Filus wykracza daleko poza podstawową wiedzę techniczną, prezentując innowacyjne podejście do problematyki bezpieczeństwa i wyjaśnialności w obszarze AI. Autorka nie tylko szczegółowo analizuje znane zagrożenia w popularnych bibliotekach AI, jak TensorFlow, ale również rozwija nowatorskie metody inicjalizacji i trenowania RNN, adresując kluczowe wyzwania w dziedzinie. Jej praca wyraźnie pokazuje głębokie zrozumienie zarówno teoretycznych, jak i praktycznych aspektów bezpieczeństwa systemów inteligentnych, jednocześnie skutecznie rozwijając metody ich wyjaśnialności. To podejście stanowi znaczący wkład w rozwój inteligentnych systemów, które są nie tylko skuteczne, ale także transparentne i bezpieczne dla użytkowników.

Ograniczenia i wyzwania w rozprawie doktorskiej

Słabe strony rozprawy Katarzyny Filus obejmują ograniczenia w zakresie zastosowania proponowanych metod w różnorodnych kontekstach rzeczywistych, możliwe wyzwania związane ze skalowaniem metod do bardzo dużych zbiorów danych oraz potencjalne trudności w interpretacji i zastosowaniu wypracowanych metod przez osoby niebędące ekspertami w dziedzinie AI. Ponadto, praca ma pewne ograniczenia wynikające z zakresu badań, dostępnych zasobów i wybranych metodologii badawczych. Praca, choć proponuje nowatorskie metody w zakresie bezpieczeństwa i wyjaśnialności algorytmów AI, mogłaby skorzystać na bardziej szczegółowym przedstawieniu przypadków ich praktycznego zastosowania w realnych środowiskach. Ponadto, w kontekście interdyscyplinarności, pracy brakuje głębszego omówienia współpracy między różnymi dyscyplinami naukowymi i praktycznymi, co mogłoby wzbogacić zrozumienie i aplikację proponowanych metod. Praca mogłaby również skupić się na dalszym rozwoju metod testowania i oceny skuteczności algorytmów, szczególnie w kontekście różnorodnych danych i zastosowań.

Ocena rozprawy w kontekście spełniania wymagań naukowych

Rozprawa doktorska Katarzyny Filus wyraźnie spełnia wymagania naukowe, zwyczajowo stawiane rozprawom doktorskim. Praca ta odzwierciedla dogłębną wiedzę i rozumienie zagadnień związanych z bezpieczeństwem i wyjaśnialnością systemów inteligentnych, co jest kluczowe w obecnych czasach rozwoju technologii AI. Autorka skutecznie łączy teoretyczne aspekty z praktycznymi aplikacjami, demonstrując znaczące umiejętności badawcze. Rozprawa zawiera nowatorskie metody i praktyczne obserwacje, które przyczyniają się do postępu w dziedzinie AI, skupiając się na poprawie bezpieczeństwa i wyjaśnialności systemów AI. W pracy zastosowano innowacyjne podejścia, takie jak użycie RNN do wykrywania zagrożeń bezpieczeństwa i metody wizualizacji dla wyjaśniania działania głębokich sieci splotowych. Te elementy wskazują na wysoki poziom innowacyjności i istotny wkład w dziedzinie informatyki technicznej i telekomunikacji, szczególnie w zakresie technologii systemów inteligentnych.

Końcowe wnioski recenzji (konkluzja)

Pomimo pewnych wad i słabych stron rozprawy, oświadczam, że:

1. Cel rozprawy został generalnie osiągnięty. Analiza wyników eksperymentów badawczych, przeprowadzonych podczas badań do niniejszej rozprawy, które zostały opublikowane w czasopismach naukowym oraz zaprezentowane na konferencjach naukowych, potwierdzają słuszność postawionych w rozprawie tez.

2. Osiągnięcia doktorantki zostały ujęte w rozprawie, która jest oryginalną odpowiedzią na problem naukowy z dziedziny nauk technicznych, z dyscypliny informatyki technicznej i telekomunikacji.
3. Tematyka rozprawy jest wyraźnie znana doktorantce, o czym świadczy dobór materiału i jego analiza. W bazach publikacji (a także w wykazie literatury zawartym w treści rozprawy) znajdują się artykuły ściśle związane z tematem rozprawy, których współautorką jest Autorka rozprawy, wskazujący, że doktorantka ma wcześniejszy dorobek naukowy na tym polu.
4. Rozprawa dodatkowo pokazuje zdolność doktorantki do samodzielnego prowadzenia badań.

Ustawa z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dziennik Ustaw z 2003 r. numer 65, pozycja 595, art. 13, ust. 1) stanowi:

Rozprawa doktorska, przygotowywana pod opieką promotora, powinna stanowić oryginalne rozwiązanie problemu naukowego lub artystycznego oraz wykazywać ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej lub artystycznej, a także umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej.

Na podstawie punktów 1, 2, 3 i 4 podsumowania niniejszej recenzji stwierdzam, że przedstawiona przez Panią mgr inż. Katarzynę Filus rozprawa doktorska spełnia wymagania warunki określone w art. 13 ust. 1 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. nr 65, poz. 595 z późniejszymi zmianami) i wnoszę o przyjęcie w/w rozprawy doktorskiej i jej dopuszczenie do publicznej obrony.

Dodatkowo, chciałbym podkreślić wyjątkowe osiągnięcia mgr inż. Katarzyny Filus w zakresie publikacji naukowych, które znacząco przyczyniają się do podkreślenia naukowej wartości jej pracy doktorskiej. Współautorstwo w licznych artykułach, opublikowanych w recenzowanych międzynarodowych czasopismach naukowych, które znajdują się w prestiżowych wykazach i są doceniane przez naukową społeczność za wysoki poziom merytoryczny, podkreśla znaczenie i aktualność przeprowadzonych przez nią badań.

Szczególnie godne uwagi są publikacje takie jak:

- **“NetSat: Network Saturation Adversarial Attack” oraz “Recycling of Generic ImageNet-trained Models for Smart-city Applications”, prezentujące innowacyjne podejścia do wykorzystania i zabezpieczenia systemów inteligentnych w aplikacjach miejskich,**
- **“Global Entropy Pooling Layer for Convolutional Neural Networks” i “Software Vulnerabilities in TensorFlow-Based Deep Learning Applications”, które skupiają się na kluczowych aspektach bezpieczeństwa i wydajności algorytmów AI,**
- **oraz prace takie jak “Visual examination of relations between known classes for deep neural network classifiers” czy “HierAct: a Hierarchical Model for Human Activity Recognition in Game-Like Educational Applications”, oferujące nowe spojrzenie na interpretowalność i zastosowania AI w edukacji i innych dziedzinach.**

Te i inne publikacje, wymienione w recenzji, nie tylko demonstrują głębokie zrozumienie i umiejętność adresowania aktualnych problemów naukowych przez mgr inż. Katarzynę Filus, ale także jej zdolność do przekraczania tradycyjnych granic dyscyplin, wprowadzając innowacje, które mają potencjał do szerokiego zastosowania w przyszłości.

Biorąc pod uwagę wyjątkowy wkład naukowy mgr inż. Katarzyny Filus, jej aktywny udział w życiu naukowym, jak również znaczące osiągnięcia publikacyjne, z przekonaniem

rekomenduję wyróżnienie jej rozprawy doktorskiej. Jestem przekonany, że takie wyróżnienie będzie nie tylko uhonorowaniem jej dotychczasowego wkładu w rozwój nauk technicznych, ale także zachętą do dalszego prowadzenia przełomowych badań w dziedzinie informatyki technicznej i telekomunikacji.
Praca doktorska mgr inż. Katarzyny Filus, z jej bogatym dorobkiem publikacyjnym, stanowi wzór doskonałości naukowej i praktycznej, odzwierciedlający najwyższe standardy badawcze. Tym samym, bez żadnych zastrzeżeń, rekomenduję jej rozprawę do wyróżnienia, podkreślając jej wybitny wkład w rozwój i zrozumienie wyjaśnialności i bezpieczeństwa systemów inteligentnych.

Podpisano przez/ Signed by:
MIKOŁAJ
LESZCZUK
Data/ Date: 12.02.2024 19:47
mSzafir

dr hab. inż. Mikołaj Leszczuk