Wrocław, 30.09.2025,

z uzupełnieniami z dnia 6.10.2025 dotyczących podstaw prawnych

Prof. dr hab. Mirosław Kutyłowski

NASK -Państwowy Instytut Badawczy

Recenzja dorobku naukowego

przedstawionego we wniosku habilitacyjnym dra Marka Pawlickiego

„Zaawansowane metody uczenia maszynowego oraz wyjaśnialnej sztucznej inteligencji (xAI) dla wykrywania ataków sieciowych"

Niniejsza opinia została przygotowana na podstawie

- decyzji Rady Doskonałości Naukowej z dnia 14.3.25, DRKN.Z2.400.1.2025.
- U C HWA Ł Y NR 9/2025 Rady Naukowej Dyscypliny informatyka techniczna i telekomunikacja Politechniki Bydgoskiej im. Jana i Jędrzeja Śniadeckich z dnia 20 marca 2025 r. powołania Komisji habilitacyjnej w postępowaniu w sprawie nadania stopnia doktora habilitowanego w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja, wszczętego na wniosek dr. inż. Marka Pawlickiego
- Zawiadomienie nr4 (NCS.521.1.2025) o wyznaczeniu na Recenzenta Komisji habilitacyjnej w postępowaniu w sprawie nadania stopnia doktora habilitowanego

w powiązaniu z pismem 4/NCS.521.1.2025 skierowanym do mnie przez Przewodniczącego Rady Naukowej Dyscypliny informatyka techniczna i telekomunikacja, Politechniki Bydgoskiej.

Niniejsza opinia została opracowana na podstawie art. 221.8 z ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2024 r. poz. 1571 z późn. zm.).

Po przeanalizowaniu przedstawionych we wniosku publikacji stwierdzam, że brak jest przesłanek do stwierdzenia, że kandydat „posiada w dorobku osiągnięcia naukowe [...] stanowiące znaczny wkład w rozwój dyscypliny", to jest nie spełnia warunku art. 219 ust. 1 z ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2024 r. poz. 1571 z późn. zm.),

Tym samym, zmuszony jestem

negatywnie

zaopiniować omawiany wniosek.

Retrospektywnie, biorąc pod uwagę kilkadziesiąt wniosków habilitacyjnych, z jakimi miałem do czynienia bądź jako recenzent bądź jako członek Centralnej Komisji, wniosek ten oceniam jako jeden z najmniej uzasadnionych pod względem merytorycznym.

UZASADNIENIE

Kwestie ogólne

(1) Większość wyników przedstawionych we wniosku ma charakter eksperymentalny i jest przyczynkiem do technik zaproponowanych przez innych autorów, niejako sprawdzając lub aplikując je w konkretnym przypadku. W pozostałych przypadkach (np. praca [C8]), zaproponowana technika jest w niewielkim stopniu odkrywcza i sprawia wrażenie pomysłu ad hoc. Pod względem ciężaru merytorycznego, prace mogłyby stanowić podstawę nadania stopnia doktora, trudno jednak mówić o „znacznym wkładzie w rozwój dyscypliny". Nie negując potrzeby wykonywania takich żmudnych i czasami nieefektownych lub wręcz chybionych prac, „znaczny wkład" powinien być moim zdaniem rozumiany jak wkład jakościowy a nie jako wkład ilościowy. O koncentracji Kandydata na ilości świadczy choćby liczba opublikowanych prac (29 na liście DBLP w 2024 roku, co daje średnio ok 2 tygodnie na prace badawcze i sformułowanie publikacji).

(2) Prezentowane prace cechują się niskim poziomem warsztatu naukowego, który obejmuje nie tylko poszukiwanie wyników, ale także ich dokumentowanie w rygorystyczny sposób. Zwraca uwagę brak stosowania zasady Brzytwy Occama: prace zawierają często treści mało istotne lub poboczne w stosunku do prezentowanych wyników. Na przykład, obszerne są rozdziały related work (łatwe do skomponowania obecnie przy pomocy narzędzi AI), jednak bez syntezy istotnych faktów z dotychczasowego stanu wiedzy. Co więcej, przywoływane pojęcia czy techniki z literatury opisywane są w sposób graniczący z obfuskacją. Techniczna i matematyczna precyzja zastępowana jest zbiorem komentarzy do pojęć i ich roli.

Wiele fragmentów tekstów sprawia wrażenie wygenerowania przez narzędzia AI. Przy wysokiej poprawności stylistycznej i stosowaniu niekiedy niestandardowego słownictwa utrudniającego zrozumienie dla osoby, dla której angielski nie jest językiem macierzystym, niejasny jest ogólny przekaz merytoryczny. Podczas gdy każde zdanie jest w pewnym sensie prawdziwe, nie układają się one w moim odbiorze w logiczny ciąg wyjaśnień skierowanych do Czytelnika. Nie jest stosowana zasada maksymalizacji treści przekazu przy minimalizacji użytych środków.

O ile zjawiska obniżania jakości edycyjnej są niestety często spotykane w literaturze naukowej wskutek postępującej dewaluacji procesów peer review, jednak w przypadku pretendowania do „znacznego wkładu w rozwój dyscypliny" są nieco dyskwalifikujące.

(3) Przedstawione prace w niewielkim stopniu dotyczą wykrywania ataków sieciowych mimo iż jest to zadeklarowano w tytule wniosku. Związki z cyberbezpieczeństwem są luźne – prezentowane wyniki mogą odnosić się niemal do każdego pola zastosowania AI. Jest to

zestaw prac raczej ilustrowanych zastosowaniami AI do wykrywania ataków niż związanych ze specyfiką AI dla systemów IDS (Intrusion Detection Systems). Trenowanie i eksperymenty są często wykonywane na zbiorach danych związanych z cyberbezpieczeństwem, ale nie jest to regułą, a wyniki nie wydają się szczególnie specyficzne.
Habilitacja w zakresie badania systemów IDS powinna, moim zdaniem, być oparta o jakieś rozpoznawalne narzędzie do ochrony systemów lub wykazujące aktywnie nieszczelność ochrony. Jeśli miałaby być oparta o rezultaty teoretyczne, to powinna być oparta o reguły jakości nauk ścisłych, w tym na przykład dokumentowania wyników statystycznych czy przedstawiania przekonywujących przesłanek dotyczących proponowanych technik.

Szczegółowy przegląd publikacji przedstawionych we wniosku

[C1] Intrusion detection approach based on optimised artificial neural network, Michał Choraś, Marek Pawlicki, Neurocomputing 452 (2021) 705–715

Wkładem niniejszej pracy jest przetestowanie wielu konfiguracji topologii sieci neuronowych w celu wykrywania i klasyfikacji włamań. Testy przeprowadzono na publicznie dostępnych zbiorach danych. Praca zawiera dyskusję na temat uzyskanych wyników oraz porównuje konsekwencje różnych wyborów topologii i hiperparametrów.

Uzyskanie wyników z pracy mogło wymagać znacznego wysiłku obliczeniowego, jednak praca ma charakter czysto rutynowy. Z punktu widzenia poziomu nowości technicznej wkład wydaje się minimalny. Znaczna część artykułu to ogólne wprowadzenie wskazujące na znaczenie wykrywania włamań oraz podstawowe wprowadzenie do sieci neuronowych.

[C2] A survey on neural networks for (cyber-) security and (cyber-) security of neural networks, Marek Pawlicki, Rafał Kozik, Michał Choraś, Neurocomputing 500 (2022) 1075–1087

Zgodnie z tytułem i treścią, jest to artykuł przeglądowy. Jako taki nie należy do kategorii oryginalnych osiągnięć badawczych stanowiących podstawę do nadania stopnia dr hab.

Chociaż pisanie przeglądów i inne działania związane z „systematyzacją wiedzy" są ważne dla ekosystemu badawczego, zgodnie z obowiązującym prawem polskim działania w tym obszarze nie wspierają ani nie dyskwalifikują przyznania stopnia doktora habilitowanego.

Czasami „przeglądy" wykraczają daleko poza zebranie dotychczasowych osiągnięć. Niestety, nie dostrzegłem żadnego wkładu tego rodzaju w omawianym artykule. Sprawia on wrażenie pracy czysto rutynowej.

[C3] How to Boost Machine Learning Network Intrusion Detection Performance with Encoding Schemes, Marek Pawlicki, Aleksandra Pawlicka, Rafał Kozik, Michał Choraś, CISIM 2023, LNCS 14164, pp. 283–297, 2023.

Praca ta jest kontynuacją kierunku wyznaczonego przez pracę "Performance Evaluation of Different Feature Encoding Schemes on Cybersecurity Logs" Erica Jacksona i Rajeeva Agrawala (cytowanej w

{C3} jako [22]). Wyniki powyższej pracy wykazują znaczenie kodowania danych tekstowych w ramach kategorii. Praca [C3] powtarza to samo badanie na innym zbiorze danych (dotyczącym wykrywania włamań). Nic dziwnego, że również w tym przypadku sposób kodowania danych ma wpływ na testowane klasyfikatory.

Zaletą tej pracy jest zwrócenie uwagi na to, że każdy zbiór danych, ze względu na swoje szczególne własności, wymagać może specyficznego kodowania Z drugiej strony jest to czysto rutynowa praca, daleka od uznania jej za znaczący wkład w badania naukowe.

Z redakcyjnego punktu widzenia, warto zauważyć, że artykuł Jacksona i Agrawala stoi na dużo wyższym poziomie edytorskim. Można się o tym przekonać choćby porównując definicje „kodowania" . Jackson i Agrawal czynią to w dość rygorystyczny sposób, podczas gdy wprowadzenie w [C3] jest niejasne i wieloznaczne.

[C4] **Towards Deployment Shift Inhibition Through Transfer Learning in Network Intrusion Detection**, Marek Pawicki, Rafał Kozik, Michał Choraś, ARES 2022

W artykule rozważono ponowne szkolenie klasyfikatora (wyszkolonego dla pewnego zbioru danych dotyczących wykrywania włamań do sieci) przy użyciu niewielkiej ilości nowych próbek. Podejście to stanowi odpowiedź na zmieniający się charakter włamań, podczas gdy systemy IDS muszą ewoluować, gdy dostępna jest ograniczona liczba próbek, niewystarczająca do stworzenia klasyfikatora od podstaw. Tak więc podejście to, zwane „transfer learning", stanowi próbę utrzymania skuteczności wykrywania w miarę ewolucji ataków.

Autorzy przeprowadzają eksperymenty na dostępnych zbiorach danych, oceniając skuteczność transfer learning w kontekście systemów IDS. Autorzy konkludują, podejście to jest skuteczne. Pojawiają się twierdzenia, że „eksperymenty sugerują, iż zaledwie 100 próbek z każdej klasy wystarczy, aby komponent osiągnął pełną wydajność w nowej sytuacji". Jest to daleko idący wniosek: wyniki mówią jedynie, że w przeszłości tempo ewolucji ataków włamań było na tyle niskie, że można było je zaadaptować poprzez transferlearning. Być może będzie to prawdą również w przyszłości. Z drugiej strony transfer learning może pomóc atakującemu w dostosowaniu się do strategii obrony.

Artykuł [C4] zawiera cenne spostrzeżenia na temat natury ataków włamań z punktu widzenia narzędzi AI. Jednak autorzy wydają się wyolbrzymiać swój wkład, stwierdzając, że *"This paper's significant contribution is presenting the results of an innovative experiment aimed at minimising the deployment shift in machine-learning-powered intrusion detection systems, by means of employing the techniques of transfer learning."* Moim zdaniem transfer learning jest jednym z pierwszych pomysłow, jaki może mieć w tej sytuacji właściciel system IDS w obliczu pojawiających się nowych scenariuszy ataku.

[C5] **Improving Siamese Neural Networks with Border Extraction Sampling for the use in Real-Time Network Intrusion Detection**, Marek Pawicki, Rafał Kozik, Michał Choraś, 2023 International Joint Conference on Neural Networks (IJCNN)

Jest to kontynuacja pracy "Leveraging Siamese networks for one-shot intrusion detection model" Hindy'ego i innych. Istotną zmianą jest strategia ponownego szkolenia sieci Siamese w przypadku pojawienia się nowych danych oraz ograniczenie par do tych dostarczonych przez k-NN.

Pomysł wydaje się dość obiecujący. Wyniki eksperymentów potwierdzają te twierdzenia. Wypada jednak zaznaczyć, że w artykule brakuje precyzji w kluczowej części dotyczącej jasnego określenia algorytmu wyboru par. Obowiązkiem autora jest jednoznaczne przedstawienie wszystkich szczegółów metody, tak aby wyniki mogły zostać odtworzone przez osobę trzecią.

[C6] Towards Quality Measures for xAI algorithms: Explanation Stability, Marek Pawlicki, DSAA'2023

W artykule zbadano stabilność charakterystyk SHAP klasyfikatora RandomForest przy różnych modelach zaburzeń danych. Autor sformułował trzy hipotezy i zweryfikował je za pomocą eksperymentów. Wyniki zostały szeroko udokumentowane. Oczywiście zaburzenia danych powinny wpływać na SHAP (w przeciwnym razie wyjaśnienia SHAP byłyby bez znaczenia dla modelu). Motywacją autora jest zbadanie stabilności SHAN – gwałtowne zmiany wynikające z niewielkich zmian w zbiorze danych świadczyłyby o niestabilności SHAP. Moja intuicja podpowiada mi, że w przypadku „normalnych" zbiorów danych (cokolwiek to oznacza) SHAP powinien być stosunkowo stabilny, ponieważ upraszcza model.

Negatywnym aspektem jest to, że artykuł jest napisany w stylu tekstu generowanego przez sztuczną inteligencję. Na przykład podrozdział opisujący SHAP jest poprawny w tym sensie, że każde zdanie można zinterpretować jako prawdziwe, jednak dla czytelnika jest on bezużyteczny jako wprowadzenie do opisywanej techniki. Dla porównania, w cytowanej pracy [7] SHAP jest zdefiniowany w elegancki sposób matematyczny, przy mniej więcej takiej samej długości tekście.

[C7] Explainability versus Security: The Unintended Consequences of xAI in Cybersecurity, Marek Pawlicki, Aleksandra Pawlicka, Rafał Kozik, Michał Choraś, SecTL'24

Artykuł dotyczy kwestii tzw. Counterfactual Explanations (CE) jako narzędzi wspomagających projektowanie ataków. Wkładem artykułu jest zestaw eksperymentów generujących CE i ZOO dla przypadków True Positives uzyskanych przez Random Forest wytrenowany do wykrywania złośliwego oprogramowania. Same eksperymenty mają charakter rutynowy.

Jest dość oczywiste, że CE mogą być wykorzystywane nie tylko do wyjaśniania decyzji podejmowanych przez klasyfikator, ale są również dobrym źródłem informacji dla atakującego wskazujących jak uniknąć wykrycia ataku. Zrozumienie klasyfikatora jest bowiem zawsze pomocne w oszukaniu go. Wyniki eksperymentów przedstawione przez autorów nie dostarczają żadnych istotnych nowych informacji na temat tego problemu.

Wnioski sformułowane w artykule mówią: *"Thus, in the future, we plan to design and develop methods to counter the very effect described in this paper. We will focus on tailoring secure xAI solutions for use in critical infrastructure sectors, such as cybersecurity, healthcare, fake news detection and law enforcement applications, where security is paramount ."* Chociaż cel ten jest

atrakcyjny, składanie jakichkolwiek obietnic jest ryzykowne i przekracza granice rzetelności naukowej – nie powinniśmy, choćby w ukryty sposób, obiecywać rzeczy niewykonalnych.

[C8] ARIA, HaRIA, and GeRIA: Novel Metrics for Pre-Model Interpretability, Marek Pawlicki, IEEE Access *10.1109/ACCESS.2024.3454084*

W artykule zaproponowano metodę analizy zbioru danych mającą na celu ilościowe określenie roli poszczególnych cech. Proponowana analiza opiera się na obliczeniu wzajemnej informacji i statystyki F. Po zastosowaniu maksymalnego skalowania bezwzględnego jako wynik końcowy obliczana jest średnia. Średnia ta może być arytmetyczna, harmoniczna lub geometryczna, co autor nazwał odpowiednio ARIA, HaRIA i GeRIA.

Główna część artykułu stanowi prezentacja wyników obliczeń uzyskanych dla wybranych zbiorów danych oraz porównanie wyników ARIA, HaRIA i GeRIA z metrykami opartymi na modelach. Nie ma jednoznacznych wniosków, a wyniki różnią się w zależności od zbioru danych. Dlatego nie jestem przekonany co do wartości proponowanych metryk. W statystyce dość łatwo jest zaproponować nowe metryki i przedstawić obszerne wyniki obliczeń, ale kluczem jest znalezienie statystyk, które prawidłowo odzwierciedlają właściwości zbioru danych. Na podstawie przedstawionych wyników nie widzę ogólnej silnej korelacji między proponowanymi metrykami dla zbioru danych a metrykami dla wyszkolonych modeli. Gwarantuje to, że podczas szkolenia „coś się dzieje" i prosta analiza danych przed trenowaniem nie pozwala ujawnić właściwości wynikowego modelu.

Artykuł napisany jest w sposób graniczący z obfuskacją. Jako przykład podać można podsekcję "Maximum Absolute scaling" w sekcji II. Materials and Methods.

> *is a scaling technique based on the maximum absolute value observed in the set. This method ensures that the maximal absolute value of each feature is adjusted to 1.0, without shifting or centering the data. This preserves inherent sparsity. MaxAbs Scaling is a fundamental preprocessing step in ML, bringing numerical data to a common scale. For positive values, MaxAbs Scaler brings the data to the [0,1] range, which enhances the comparability across different variables [32].*

Jako część sekcji „Materiały i metody" powinno to pomóc czytelnikowi zrozumieć pojęcia użyte w dalszej części artykułu. Jednak znacznie lepiej byłoby po prostu powiedzieć, że każda wartość cechy x jest przekształcana na x/mx, gdzie mx jest maksymalną wartością bezwzględną tej cechy w zbiorze danych. Niestety, inne definicje są sformułowane w ten sam sposób: techniczna i matematyczna precyzja została zastąpiona niejasnymi opisami, komentarzami i odniesieniami do literatury.

[C9] The survey on the dual nature of xAI challenges in intrusion detection and their potential for AI innovation, Marek Pawlicki, Aleksandra Pawlicka, Rafał Kozik, Michał Choraś, Artificial Intelligence Review (2024) 57:330

Podobnie jak artykuł [C2], jest to artykuł przeglądowy. Istnieją więc formalne powody, które mogą dyskwalifikować tę pracę, niezależnie od jej zalet. Przyjmując bardziej liberalne podejście, można argumentować, że analizując opinie innych autorów, można uzyskać nową wiedzę, której nie dostrzega się, mając jedynie lokalny ogląd tematu. Niestety nie jestem przekonany, że tak jest w przypadku omawianego artykułu:

- Chociaż autorzy twierdzą, że skupiają się na wykrywaniu włamań, omówione wyzwania są dość uniwersalne i stosunkowo oczywiste. Czytelnik mógłby oczekiwać głębszego skupienia się na wykrywaniu włamań, a nie tylko kilku uwag, takich jak wspomnienie o wymaganym czasie dostarczenia wyjaśnień. W związku z tym treść artykułu w ograniczonym stopniu odpowiada tytułowi.
- Wyodrębnienie wyzwań wymienionych w wybranych artykułach jest zadaniem, które wydaje się wykonalne dla wykwalifikowanego promt engineer. W tym przypadku indywidualnym wkładem autora nie jest wynik wyszukiwania i podsumowanie (które można dość dobrze skomponować za pomocą sztucznej inteligencji), ale sztuka zadawania pytań i przeglądania szkiców. Zadawanie właściwych pytań można uznać za cenny wkład, jeśli pytanie lub odpowiedź można uznać za odkrycie naukowe. W przypadku omawianego artykułu trudno zaryzykować że dokonano jakiegoś odkrycia
- Omawiając wyzwania związane z xAI, autorzy nie wykraczają poza listę życzeń (czasami raczej nierealistycznych), podczas gdy rolą nauk technicznych jest nie tylko konstruowanie rozwiązań, ale także określanie, co jest niewykonalne. Jeśli, dla przykładu, osiągnięcie wszystkich celów RODO w przypadku prostych baz danych jest prawie niemożliwe, to jak możemy oczekiwać sukcesu w przypadku sztucznej inteligencji? Uważam, że autorzy powinni szczerze przyznać, że w niektórych obszarach xAI nie uda się znaleźć ogólnego, satysfakcjonującego rozwiązania, a nie w pośrednio sugerować, że jest to osiągalne.
- „Dual nature" wyzwań związanych z xAI polega na tym, że, jak rozumiem z tekstu, samo wyzwanie jest okazją do rozwiązania problemu. Czy istnieje jednak jakieś wyzwanie, które nie stwarza okazji do prób znalezienia rozwiązania? Awansowanie „dual nature" do tytułu artykułu wydaje się czysto marketingowym zabiegiem.

Artykuł nie jest badaniem wyzwań związanych z xAI, ale badaniem tego, jakie wyzwania związane z xAI są dostrzegane przez autorów publikacji. Byłoby to interesującym wkładem w przypadku wykrycia powszechnego nieporozumienia lub zaniedbania ważnych kwestii. Jednak w artykule [C9] nie ma takich ustaleń o charakterze strategicznym co do kierunku badań.

Pozostałe uwagi

1. Mimo iż większość prac ma charakter eksperymentalny, autorzy przedstawionych prac nie umożliwiają niezależnej weryfikacji wyników poprzez podanie linków do repozytoriów oprogramowania (same analizowane zbiory danych pochodzą z publicznie dostępnych źródeł). W ostatnim okresie, tego typu materiał dokumentujący wyniki badań staje się coraz powszechniejszą walką z plagą publikowania sfałszowanych lub fikcyjnych wyników badań. Brak (choćby hipotetycznej) możliwości sprawdzenia wyników oznacza brak rękojmi dla prawdziwości wyników. W naukach ścisłych, takich jak matematyka, odpowiadałoby publikowaniu twierdzeń bez możliwych do sprawdzenia dowodów.
2. Współczynniki bibliometryczne Kandydata należy traktować z sposób uwzględniający specyfikę dziedziny sztucznej inteligencji i liczby publikowanych prac w tym obszarze. Przyjęty algorytm (indeks cytowań) powoduje zawyżenie rankingu czasopism z obszarów bioinformatyki, sztucznej inteligencji itp wskutek praktyk publikacyjnych zbliżonych do zwyczajów np. w chemii, gdzie podstawą publikacji może być wynik eksperymentu a nie rozwiązanie określonego naukowego (np. zbudowanie efektywniejszego niż dotychczasowe algorytmu).

Uwaga szczegółowa: zauważyłem przypisanie 140 pkt pracy [C7], podczas gdy ACM Workshop on Secure and Trustworthy Deep Learning Systems nie występuje ani na wykazie ministerialnym ani nie był ewaluowany przez CORE. Podstawą skojarzenia go z punktacją był zapewne fakt kolokacji workshopu z konferencją ASIA CCS (zasłużenie widniejącej na liście CORE z kategorią A i liście ministerialnej z 140 punktami).

3. Przedstawione deklaracje współautorów nie odnoszą się do „odkryć naukowych" (które są podstawą oceny wniosku) a jedynie do klasyfikacji CRediT. Nie pozwala to na rozgraniczenie (tam gdzie rola współautorów nie jest marginalna w obszarze wkładu naukowego) udziału współautorów. Bardziej celowe byłoby wskazanie, która idea lub rozwiązanie należy przypisać Kandydatowi i w jakim stopniu (gdy odkrycie jest wspólne). Udział w obszarach takich jak „recenzja" czy „pozyskanie funduszy" nie są relewantne dla oceny wniosku habilitacyjnego. Użycie taksonomii CRediT zaciemnia obraz sytuacji i pozwala uniknąć odpowiedzi na istotne aspekty oceny wniosku.

Mirosław Kutyłowski

Mirosław Kutyłowski