Prof. Marek Klonowski
Katedra Sztucznej Inteligencji
Politechnika Wrocławska

Wrocław

Recenzja w postępowaniu habilitacyjnym Pana dra Marka Pawlickiego

Niniejsza recenzja dotyczy wniosku Pana dra Marka Pawlickiego o nadanie stopnia naukowego doktora habilitowanego w dyscyplinie informatyka techniczna i telekomunikacja. Wniosek ten jest oceniany na gruncie ustawy *Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r. (Dz. U. z 2024 r. poz. 1571).

Warunkiem koniecznym nadania stopnia doktora habilitowanego jest spełnienie trzech warunków. Pierwszy z nich to posiadanie stopnia doktora. Habilitant uzyskał ten stopień na Uniwersytecie Technologiczno-Przyrodniczym w sierpniu 2020 roku na podstawie dysertacji pt. *Zastosowanie metod uczenia maszynowego do wykrywania ataków sieciowych*.

Drugi warunek, czyli wykazanie się aktywnością naukową realizowaną w więcej niż jednej instytucji naukowej, również został w oczywisty sposób spełniony. Dr Marek Pawlicki odbył trzymiesięczny staż na uniwersytecie w Neapolu, który zaowocował publikacjami dobrej rangi. Współpracował także z zespołami badawczymi z Politechniki Wrocławskiej i Politechniki Warszawskiej, co przyniosło kilka wspólnych publikacji. Aktywność naukowa przejawia się również w realizacji licznych projektów z udziałem różnych podmiotów. Mimo silnego i stałego związania z jednostką macierzystą, Kandydat był bez wątpienia aktywny także w innych ośrodkach.

Ocena osiągnięcia naukowego

Zasadniczym elementem recenzji jest ocena osiągnięcia naukowego, czyli ocena spełnienie trzeciego warunku. W tym przypadku jest to cykl dziewięciu prac oznaczonych [C01]–[C09], opublikowanych w latach 2021–2024. Osiągnięcie zostało zatytułowane:

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

Pierwsze spostrzeżenie dotyczy podobieństwa tematycznego tego cyklu do rozprawy doktorskiej Habilitanta. Widać zasadniczo tylko dwie różnice. Pierwsza polega na tym, że osiągnięcie habilitacyjne ma dotyczyć „zaawansowanych” metod, jednak poziom zaawansowania przedstawionych rozwiązań wydaje się podobny do tych analizowanych w doktoracie. Drugi wyróżnik to odwołanie do „wyjaśnialnej sztucznej inteligencji (xAI)”, której znaczenie w przedstawionym cyklu wydaje się jednak marginalne. Można więc stwierdzić, że tematyka doktoratu i osiągnięcia habilitacyjnego jest niemal tożsama — zarówno pod względem merytorycznym, jak i personalnym. Sześć spośród dziewięciu publikacji współtworzone jest z oboma promotorami doktoratu, a siódma z jednym z nich. Tylko dwie prace mają charakter samodzielny.

Taki dobór cyklu nie przekonuje o samodzielności Habilitanta, a tematyka osiągnięcia habilitacyjnego wydaje się jedynie rozwinięciem doktoratu. Jest to jednak sytuacja, wydaje się, dopuszczalna w świetle obecnych przepisów.

Problemem jest niejasny udział Habilitanta w powstaniu prac wieloautorskich. Z dostarczonych deklaracji (też według konwencji CRediT) trudno jednoznacznie ustalić faktyczny wkład poszczególnych autorów w przypadku części prac.

Wszystkie prace wchodzące w skład cyklu zostały opublikowane w obiegu międzynarodowym. Bez wątpliwości można stwierdzić, że przedstawione osiągnięcie habilitacyjne, jak i cały dorobek dra Marka Pawlickiego, trafnie przypisane są do dyscypliny informatyka techniczna i telekomunikacja.

Dwie prace [C01] i [C02] ukazały się w czasopiśmie *Neurocomputing* – periodyku o solidnej renomie, choć umiarkowanie selektywnym. Obydwie prace są dobrze cytowane. Praca [C08] została opublikowana w *IEEE Access*, czasopiśmie bardzo rozpoznawalnym, lecz mało selektywnym i zdecydowanie spoza czołówki w sensie jakości publikowanych prac. Praca [C09] ukazała się w *Artificial Intelligence Review*. Pozostałe publikacje pochodzą z konferencji o randze CORE B/C lub wydarzeniu afiliowanym przy konferencji ESORICS. Ranga wydawnictw, w których ukazało się osiągnięcie habilitacyjne, jest więc przeciętna dla dyscypliny.

Kwestia spójności tematycznej cyklu pozostaje dyskusyjna. Argumentacja Habilitanta na rzecz tej spójności (Autoreferat, rys. 1, s. 10) nie jest przekonująca. Kluczowym niedostatkiem osiągnięcia jest także niewielka liczba oryginalnych, rozwiniętych i należycie opisanych koncepcji.

Analiza poszczególnych prac wchodzących w skład osiągnięcia naukowego

Praca [C01] (*Neurocomputing, 2021*) została opublikowana przez Habilitanta wspólnie z jednym z promotorów doktoratu. Dotyczy optymalizacji różnych modeli uczenia maszynowego (ML) w kontekście systemów wykrywania intruzów (IDS). Zasadniczą wartością tej publikacji jest pokazanie, że systemy IDS oparte na rozwiązaniach głębokiego uczenia (DL) często nie działają optymalnie, a odpowiednia optymalizacja hiperparametrów może znacznie poprawić ich skuteczność. Należy jednak zauważyć, że przedstawione działania optymalizacyjne mają charakter standardowego przeszukiwania przestrzeni parametrów. Choć praca jest pożyteczna i zwraca uwagę na istotny problem inżynierski, trudno w niej dostrzec nowe, oryginalne pomysły. Zaproponowane podejście jest ogólne i nie odnosi się w sposób szczególny do specyfiki systemów IDS. Praca, stosunkowo krótka (10 stron), zawiera liczne ogólne opisy – np. dotyczące działania sieci neuronowych, co wydaje się mocno nadmiarowe w specjalistycznym artykule.

Praca [C02] (*Neurocomputing*) została napisana przez Habilitanta wspólnie z oboma promotorami doktoratu. Większość treści stanowi omówienie powszechnie znanych problemów, takich jak niezbalansowanie danych czy dobór funkcji aktywacji sieci. Oryginalnych elementów można doszukać się w rozdziale 4.2, poświęconym XAI. Ta część jednak jest krótka i sprawia wrażenie niedopracowanej, co obniża wartość naukową całej publikacji.

Praca [C03] (konferencja CORE C) również została napisana wspólnie z oboma promotorami. Dotyczy wpływu sposobu kodowania danych na działanie systemów IDS opartych o ML. W rzeczywistości jednak rozważania te nie mają szczególnego związku z IDS – dane z tej dziedziny wykorzystano raczej jako przykład do ogólnych analiz wpływu kodowania.

Praca [C04] (konferencja ARES) jest kolejną publikacją współautorską Habilitanta i jego promotorów. Dotyczy zastosowania standardowych technik uczenia transferowego do analizy danych IDS. Nie widać w niej nowatorskiego charakteru – jest to raczej weryfikacja, że uczenie transferowe działa (w określonym zakresie) dla konkretnych modeli i zbiorów danych. Publikacja ta jest poprawna warsztatowo, ale bez oryginalnych rozwiązań.

Praca [C05] została napisana wspólnie z oboma promotorami i dotyczy poprawy jakości trenowania sieci syjamskich. Autorzy proponują, aby proces uczenia koncentrował się na danych z okolic granicy decyzyjnej. Choć pomysł jest interesujący, sam fakt, że dane z tego obszaru są najbardziej informatywne, jest bardzo intuicyjny. Brakuje analizy teoretycznej, a liczba wyników eksperymentalnych jest niewystarczająca. Praca jest krótka (6 stron), a znaczna część tekstu to opis powszechnie znanych architektur. Mimo pewnej pomysłowości, publikacja sprawia wrażenie niedopracowanej i trudno uznać ją za istotną część osiągnięcia habilitacyjnego.

Praca [C06] jest samodzielna i dotyczy analizy stabilności funkcji wyjaśnialności. Autor ogranicza się do popularnej techniki SHAP oraz modelu lasów losowych. W Autoreferacie Habilitant stwierdza, że *„mała perturbacja nie powinna powodować drastycznych zmian w wyjaśnialności”* (s. 26), co stanowi punkt wyjścia dalszej analizy. Choć Autor uzasadnia potrzebę badania stabilności wyjaśnień, argumentacja nie jest w pełni przekonująca, zwłaszcza w kontekście perturbacji opartych na permutacji atrybutów. Brakuje tu wprowadzenia formalizmu, a w części eksperymentalnej pominięto szczegółową analizę zmian w przedziale 20–50% permutacji, mimo że to właśnie tam zachodzą najciekawsze efekty. Praca ma pewne walory i elementy oryginalności, jednak również sprawia wrażenie niedopracowanej.

Praca [C07] została opublikowana na mało znanej konferencji, a współautorami są obaj promotorzy Habilitanta. Opisuje ona rozszerzoną metodę generowania wyjaśnień kontrfaktycznych dla ataków sieciowych, opartą na zmodyfikowanym algorytmie DiCE. Publikacja zawiera ciekawe uwagi na temat relacji między bezpieczeństwem a wydajnością, jednak temat ten nie został pogłębiony ani odpowiednio zbadany.

Praca [C08] (samodzielna, *IEEE Access*) dotyczy oceny znaczenia danych przed ich użyciem w modelu. Koncepcja jest interesująca, jednak metodologia nie została jasno przedstawiona.Brakuje porównania z istniejącymi metodami oceny ważności danych (np. przegląd w pracy: https://arxiv.org/pdf/2212.04612). Ograniczenie się do niewielkich zbiorów danych i brak pogłębionej interpretacji wyników znacząco obniżają wartość naukową tej publikacji.

Praca [C09] (*Artificial Intelligence Review*) została napisana wspólnie z oboma promotorami doktoratu. Dotyczy wyzwań i kierunków rozwoju AI w kontekście IDS. Sposób wytypowania omawianych zagadnień został opisany szczegółowo, jednak motywacja stojąca za nim pozostaje niejasna. W efekcie powstała lista 25 „wyzwań” lub „kierunków”, które są omówione w sposób bardzo skrótowy. Wskazane hasła mają różny poziom abstrakcji, często się częściowo pokrywają a w całej pracy brakuje precyzji. Na przykład punkt „Privacy and Security” obejmuje liczne inne problemy i częściowo nakłada się na „Adversarial use of xAI”. Wrażenie potęguje bałagan

pojęciowy i brak odniesień do definicji. Autorzy przytaczają m.in. stwierdzenie: *„xAI models must explicitly consider privacy concerns throughout their lifecycle”* (s. 16), lecz nie uzasadniają go. Podobnie zdanie: *„Limiting the search to English-language papers was practical for the researchers, it also made the study more accessible to a wider international audience”* (s. 6) domaga się uzasadnienia. Praca, choć zawiera ciekawe spostrzeżenia, nie powinna w obecnej formie stanowić części osiągnięcia habilitacyjnego.

Podsumowanie części analitycznej

Podsumowując analizę poszczególnych prac, należy stwierdzić, że mimo pewnych interesujących elementów i poprawności metodologicznej, cały cykl ma braki. Udział Habilitanta w pracach wieloautorskich jest częściowo niejasny, oryginalność części prac jest niska, a poziom opracowania wielu publikacji powierzchowny. Brakuje głębszych analiz teoretycznych i spójności tematycznej.

Pozostały dorobek

Pan dr Marek Pawlicki ma w swoim dorobku bardzo dużą liczbę publikacji – ponad 120. Są to w większości prace wieloautorskie, a na podstawie dostarczonej dokumentacji trudno jednoznacznie ustalić, jaki był indywidualny wkład Habilitanta. Publikacje reprezentują zróżnicowany poziom i obejmują tematykę koncentrującą się na pograniczu bezpieczeństwa oraz sztucznej inteligencji. Nie odnotowano prac w najbardziej prestiżowych czasopismach lub na konferencjach najwyższej rangi, choć w dorobku znajdują się publikacje w solidnych, rozpoznawalnych wydawnictwach. Całość ujawnionego dorobku ukazała się w wydawnictwach o obiegu międzynarodowym.

Pan dr Marek Pawlicki wymienia również kilka prac jako „kolejne osiągnięcia”, w szczególności prace oznaczone jako [DA01] oraz [DA02]. Wkład Habilitanta w te prace nie został jasno określony, brak także deklaracji w konwencji CRediT. Same prace dotyczą problematyki przykładów zwodniczych i prezentują rozwiązania oparte na standardowych podejściach (m.in. odszumianie).

Ocena wniosku i autoreferatu

Autoreferat niestety sprawia złe wrażenie. Zawiera informacje zbędne, a jednocześnie pomija kluczowe elementy. Styl niekiedy jest potoczny i nieprzystający do tekstu naukowego (np. fragment: *„(...) które potrafią określić coś Y na podstawie informacji X*”, s. 12). Część sformułowań jest niezrozumiała lub nieprecyzyjna. Trudne do interpretacji są np. akapity zaczynające się od słowa „Charakterystyka” (s. 12) czy zdania typu: *„Badacze często korzystają z zestawów danych referencyjnych (benchmarków), które zawierają cechy niemożliwe do uzyskania w czasie rzeczywistym”* (s. 12). Podobnie niejasne pozostaje: *„Z tego samego powodu modele trenowane na różnej dystrybucji danych osiągają gorsze wyniki, gdy są testowane na danych pochodzących z pokrewnej, ale innej dziedziny”* (s. 16). Wiele wypowiedzi wymagałoby doprecyzowania lub uzasadnienia, np. *„Poza dziedziną AI, dobre wyjaśnienia są zazwyczaj stabilne” (s. 26)* – nie wiadomo, co Autor rozumie przez „dobre wyjaśnienia” ani jak definiuje „stabilność”. Podobnie zdanie: *„Zrozumienie stabilności wyjaśnień w różnych scenariuszach danych jest zatem niezbędne dla wykorzystania w praktyce”* (s. 26) nie zostało poparte argumentacją i wydaje się wątpliwe z punktu widzenia doświadczenia empirycznego. Wiele sformułowań jest nieprecyzyjnych (*„zaokrąglona ocena znaczenia cech”*, s. 35) lub zbyt ogólnych (np. opis znaczenia XAI, s. 33).

Całość autoreferatu sprawia wrażenie chaotycznej i niejasnej prezentacji. Brakuje pogłębionej analizy literatury, choć trzeba przyznać, że liczba potencjalnych pozycji bibliograficznych jest gigantyczna, co utrudnia pełny przegląd.

W autoreferacie pojawiają się również nieścisłości dotyczące klasyfikacji konferencji. Warsztaty przy konferencji *ESORICS* nie są traktowane jako CORE A i – w odróżnieniu od głównej konferencji – charakteryzują się znacznie mniejszą selektywnością.

Pozostałe przejawy aktywności zawodowej

Wszystkie inne przejawy aktywności zawodowej Habilitanta należy ocenić pozytywnie, a niektóre nawet wysoko. Pan dr Marek Pawlicki posiada liczne publikacje we współautorstwie z badaczami z różnych ośrodków naukowych. Jego wskaźniki bibliometryczne, jak na obecny etap kariery, są bardzo dobre i przekraczają typowe wymagania stawiane kandydatom do stopnia doktora habilitowanego.

Dobre wrażenie robi także zaangażowanie w działalność recenzencką oraz udział w radach programowych konferencji. Choć nie były to wydarzenia najwyższej rangi, świadczą o aktywności i rozpoznawalności Habilitanta w środowisku naukowym. Pozytywnie oceniam również działalność dydaktyczną, współpracę ze studentami oraz widoczną działalność popularyzatorską.

Na uwagę zasługuje duża aktywność w zakresie transferu technologii do przemysłu – szczególnie w obszarze bezpieczeństwa teleinformatycznego, który jest zbieżny z tematyką osiągnięcia habilitacyjnego. W dorobku Habilitanta znajdują się również projekty o innej tematyce. Wysoko oceniam fakt uczestnictwa w licznych projektach, w tym międzynarodowych, z których czterema Pan dr Pawlicki kierował.

Podsumowanie

Całościowa ocena wniosku jest trudna. Zastrzeżenia dotyczące dorobku – zwłaszcza osiągnięcia habilitacyjnego – są liczne i istotne. Nie ma dostatecznie jasnego opisu indywidualnego wkładu Habilitanta w poszczególne prace; wiele koncepcji jest niedopracowanych, a część prezentowanych treści – powierzchowna lub oczywista. Brakuje formalizacji pojęć, pogłębionej analizy teoretycznej i solidnych badań eksperymentalnych. Cykl publikacji sprawia wrażenie zbioru prac jedynie luźno powiązanych tematyką IDS poprzez wspólne dane, a nie spójnego osiągnięcia naukowego.

Z drugiej strony, w pracach nie stwierdzono istotnych błędów metodologicznych, a niektóre prace zawierają pewne elementy oryginalne. Należy także docenić zaangażowanie dra Marka Pawlickiego w projekty badawczo-rozwojowe oraz wysokie wskaźniki bibliometryczne.

Uważam, że Habilitant zasługuje na **dopuszczenie do kolokwium habilitacyjnego**. **Na obecnym etapie pozytywnie opiniuję wniosek habilitacyjny, mimo wyraźnie krytycznej oceny jego zasadniczych elementów.** Zaznaczam jednak, że rozważam ocenę negatywną wniosku Habilitanta podczas posiedzenia komisji, o ile mój odbiór nie ulegnie zmianie po zapoznaniu się z pozostałymi recenzjami lub w wyniku przebiegu kolokwium habilitacyjnego.