

dr hab. inż. Tomasz Kajdanowicz, prof. uczelni
Katedra Sztucznej Inteligencji
Wydział Informatyki i Telekomunikacji
Politechnika Wrocławska

Wrocław, 09.06.2026

Recenzja rozprawy doktorskiej

dla Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Uniwersytetu
Jagiellońskiego

Tytuł rozprawy: *Bridging the Generalization Gap: A Methodological Approach to Building and Benchmarking NLP Models*

Autor rozprawy: Michał Pietruszka

Promotor: dr hab. inż. Paweł Morawiecki, prof. IPI PAN

Promotor pomocniczy: dr hab. Przemysław Spurek, prof. UJ

Dyscyplina naukowa: informatyka techniczna i telekomunikacja

Charakter rozprawy: zbiór powiązanych tematycznie publikacji naukowych

1 Analiza strony merytorycznej rozprawy

1.1 Obszar problemowy i temat rozprawy

Rozprawę doktorską stanowi zbiór dziewięciu powiązanych tematycznie publikacji naukowych, z których osiem zostało opublikowanych w prestiżowych konferencjach i czasopismach, a jedna, wg stanu opracowania autoreferatu, znajdowała się w procesie recenzji, a na dzień sporządzenia niniejszej recenzji rozprawy została opublikowana w materiałach konferencyjnych. Rozprawa doktorska Michała Pietruszki koncentruje się na problemie luki generalizacji (generalization gap) w przetwarzaniu języka naturalnego (NLP), czyli rozbieżności między osiągnięciami modeli w warunkach akademickich a ich praktyczną użytecznością w zastosowaniach przemysłowych. Doktorant formułuje tezę, że przezwyciężenie tej luki wymaga systematycznego, wieloaspektowego podejścia technicznego, obejmującego kolejno: ewaluację, innowację architektoniczną, optymalizację praktyczną, holistyczną aplikację oraz integrację zaawansowanych umiejętności.

Temat rozprawy jest naukowo istotny i aktualny. Problem generalizacji modeli NLP stanowi jedno z centralnych wyzwań współczesnej sztucznej inteligencji, a jego praktyczne implikacje dotyczą licznych sektorów przemysłu, od finansów i prawa po medycynę i logistykę. Doktorant identyfikuje pięć wzajemnie powiązanych wyzwań technicznych: (1) wyzwanie ewaluacyjne i nieadekwatność istniejących benchmarków, (2) wyzwanie architektoniczne z ograniczoną zdolnością generalizacji modeli, (3) wyzwanie związane z kosztami obliczeniowymi, (4) wyzwanie złożoności danych, w szczególności z danymi multimodalnymi, oraz (5) wyzwanie integracji i transferu wiedzy w łączeniu wielu kompetencji w rozwiązywaniu problemów. Doktorant nie traktuje tych wyzwań jako izolowanych, lecz identyfikuje sprzężenia zwrotne (feedback loops) między nimi, argumentując, że brak postępu w jednym obszarze hamuje rozwój w pozostałych. Praca lokuje się na styku badań nad architekturami sieci neuronowych, rozumieniem dokumentów,

optymalizacją modeli językowych oraz ewaluacją systemów AI, co nadaje jej charakter interdyscyplinarny w obrębie informatyki.

1.2 Organizacja pracy i przegląd literatury

Opis zbioru powiązanych tematycznie publikacji naukowych jest zorganizowana w dziewięciu rozdziałach odpowiadających zaproponowanej mapie metodologicznej. Rozdział P1 stanowi wprowadzenie, w którym Doktorant formułuje cele badawcze, tezę i hipotezę oraz dokonuje przeglądu istniejących modeli cyklu życia technologii, argumentując za potrzebą stworzenia dedykowanego podejścia dla NLP. Rozdział P2 przedstawia mapę metodologiczną składającą się z pięciu faz, identyfikuje sprzężenia zwrotne między nimi i mapuje poszczególne kontrybucje dysertacji na te fazy. Rozdział P3 zawiera notację i tło techniczne obejmujące Transformatory, przetwarzanie dokumentów, ekstrakcję informacji kluczowych (KIE), wizualne QA dokumentów i feature engineering, co jest nadmiarową i niewiele wnoszącą zawartością o naturze wiedzy ogólnej.

Rozdziały P4-P8 stanowią merytoryczny rdzeń pracy, każdy odpowiadający jednej fazie mapy metodologicznej i prezentujący konkretne kontrybucje badawcze w postaci opublikowanych artykułów. Rozdział P9 zawiera podsumowanie, refleksję nad ograniczeniami oraz kierunki przyszłych badań. Struktura ta jest przejrzysta i spójna, a każdy rozdział kończy się sekcją wiążącą wyniki z szerszą narracją metodologiczną. Należy podkreślić, że przegląd literatury jest prowadzony w sposób rozproszony, a kontekst literaturowy jest związany z poszczególnymi publikacjami, co jest uzasadnione formatem zbioru, choć szerszy, przekrojowy przegląd stanu badań w osobnym rozdziale mógłby wzmocnić spójność teoretyczną rozprawy.

1.3 Ocena wyników oraz stopnia ich oryginalności

W rozprawie doktorskiej Michała Pietruszki zaprezentowano szereg oryginalnych osiągnięć badawczych o znaczącym wpływie na rozwój dziedziny przetwarzania języka naturalnego i rozumienia dokumentów. Poniżej omawiam poszczególne publikacje wchodzące w skład zbioru, oceniając ich oryginalność i wkład Doktoranta.

P4.1 From Dataset Recycling to Multi-Property Extraction and Beyond (CoNLL 2020, CORE: A, MNiSW: 140)

Praca Dwojak, Pietruszka, Borchmann, Chłędowski, Graliński prezentuje przekształcenie zbioru danych WikiReading w celu umożliwienia rzetelnej ewaluacji generalizacji w zadaniach ekstrakcji informacji kluczowych (KIE). Doktorant ze współautorami zidentyfikowali krytyczny problem w oryginalnym zbiorze WikiReading, gdzie te same artykuły pojawiały się zarówno w zbiorze treningowym, jak i testowym, co umożliwiało modelom zapamiętywanie zamiast rzeczywistej generalizacji. Zaproponowano redefinicję zadania z ekstrakcji pojedynczych właściwości (Single Property Extraction) na ekstrakcję wielu właściwości jednocześnie (Multiple Property Extraction, MPE) oraz fundamentalną restrukturyzację podziału danych. Dodatkowo usunięto wycieki danych i niekompletne anotacje. Oryginalność tej pracy polega na trafnym zdiagnozowaniu wady metodologicznej powszechnie używanego benchmarku i zaproponowaniu rozwiązania, które wymusza na modelach prawdziwe rozumienie semantyczne. Wkład Doktoranta obejmuje, jak deklaruje, przede wszystkim redefinicję zbioru danych, co stanowi wkład do artykułu, ale pozostaje niejasne w jakim konkretnie zakresie.

P5.1 Going Full-TILT Boogie on Document Understanding with Text-Image-Layout Transformer (ICDAR 2021, CORE: A, MNIŚW: 140)

Praca Powalski, Borchmann, Jurkiewicz, Dwojak, Pietruszka, Pałka przedstawia multimodalną architekturę TILT opartą na modelu T5, integrującą informacje tekstowe, wizualne i przestrzenne (layout) dokumentów. Model wprowadza względne cechy przestrzenne uwzględniające odległości zarówno poziome, jak i pionowe między słowami, a także kontekstualizowane embeddingi obrazów generowane przez sieć konwolucyjną opartą na U-Net. TILT jest trenowany end-to-end bez potrzeby dodatkowych sygnałów treningowych, a jego skuteczność została potwierdzona zwycięstwem w dwóch międzynarodowych konkursach. Wkład Doktoranta obejmował koordynację współpracy z zespołem LAMBERT w celu wypracowania architektonicznego podejścia do KIE z rozumieniem layoutu, a także pozyskiwanie, transformację i ewaluację zbiorów danych do nadzorowanego multimodalnego QA, co znacząco poprawiło proces treningu. Osiągnięcie merytoryczne stojące za prezentowanym rezultatem jest bezpośrednio zauważalne, ale nie do uściślenia pozostaje wkład Doktoranta w konkretach związanych uzyskaniem rezultatów opisanych w publikacji.

P5.2 STable Table Generation Framework for Encoder-Decoder Models (EACL 2024, CORE: A, MNIŚW: 140)

Praca Pietruszka, Turski, Borchmann, Dwojak, Pałka, Szyndler, Jurkiewicz, Garncarek adresuje problem generowania tabel z tekstu w modelach encoder-decoder. Kluczową innowacją jest rezygnacja z ustalonej kolejności generowania komórek (np. od lewej do prawej, od góry do dołu) na rzecz dynamicznej kolejności opartej na wynikach pewności modelu (confidence scores). Doktorant zaproponował trening permutacyjny, gdzie model jest eksponowany na liczne losowe sekwencje generowania komórek, co wymusza zdolność do wypełniania dowolnej komórki na podstawie dowolnej częściowo wypełnionej tabeli. Podczas dekodowania model wypełnia komórki o najwyższym wyniku pewności, co stabilizuje proces generowania i zapobiega kaskadowym błędom. STable przewyższył konwencjonalne metody na pięciu z ośmiu zbiorów danych. Wkład Doktoranta w tym rezultacie obejmuje główną ideę algorytmu (zarówno treningu, jak i inferencji), zmiany architektoniczne, implementację i badania ablacyjne, ale nie jest do końca klarowny z uwagi na wskazany w publikacji równy wkład trzech pierwszych autorów i współautorstwie dalszych pięciu.

P6.1 Sparsifying Transformer Models with Trainable Representation Pooling oraz Successive Halving Top-k Operator (ACL 2022, CORE: A*, MNIŚW: 200 oraz AAAI 2021 student abstract, CORE: A* MNIŚW: 0)

Prace Pietruszka, Borchmann, Graliński/Garncarek prezentują operator Soft Top-k i architekturę DeepPyramidion, które adresują kwadratową złożoność obliczeniową mechanizmu uwagi w Transformerach. Operator Soft Top-k stanowi różniczkowalny mechanizm selekcji, który stopniowo redukuje długość sekwencji po każdej warstwie kodowania. Formalnie, każdej reprezentacji tokenu przypisywany jest wynik ważności przez liniową funkcję skorującą, a następnie nisko- i wysoko-skorowane reprezentacje są parowane i łączone za pomocą softmaxu (wysoko-skorowana reprezentacja dominuje, ale informacja z nisko-skorowanej jest zachowywana w grafie obliczeniowym). Proces ten jest stosowany iteracyjnie, zmniejszając o połowę liczbę reprezentacji aż do uzyskania k najistotniejszych. Architektura DeepPyramidion używając ten mechanizm uzyskała do znaczną redukcję złożoności obliczeniowej na zadaniach streszczania długich dokumentów (arXiv, PubMed) bez utraty jakości, osiągając wyniki porównywalne lub lepsze od modeli state-of-the-art, które były znacznie większe i inicjalizowane z pretrenowanych modeli RoBERTa lub BART. Wkład Doktoranta polegał na zaprojektowaniu algorytmu, integracji z modelem DeepPyramidion,

przeprowadzeniu wszystkich eksperymentów i badań ablacyjnych, implementacji połowy modeli bazowych, przygotowaniu zbiorów danych oraz zarządzaniu projektem badawczym.

P7.1 DUE – End-to-End Document Understanding Benchmark (NeurIPS 2021, CORE: A*, MNiSW: 200)

Praca Borchmann, Pietruszka, Stanislawek, Jurkiewicz, Turski, Szyndler, Graliński proponuje zunifikowany benchmark do rozumienia dokumentów, integrujący sześć kluczowych obszarów: ekstrakcję informacji kluczowych, klasyfikację, analizę layoutu, odpowiadanie na pytania, QA nad wykresami oraz QA nad tabelami. Główną innowacją jest ujednolicenie tych zadań w jednym logicznym zadaniu, gdzie wejściem jest obraz dokumentu, a tekstem wyjściem. Dla każdego zbioru danych przeprowadzono reformulację zadań, manualną anotację lub walidację, zapewniono zestawy danych testowych, zmierzono wydajność ludzką. Benchmark eliminuje problem fragmentacji ewaluacji w rozumieniu dokumentów. Wkład Doktoranta obejmuje współtworzenie koncepcji benchmarku i metodologii ewaluacji, identyfikację odpowiednich zbiorów danych, przygotowanie większości z nich, współudział w implementacji modeli bazowych oraz organizację kodu i zbiorów danych. Do potwierdzenia jest znaczenie udziału Doktoranta w prezentowanym rezultacie w świetle oznaczenia równego wkładu trzech pierwszych autorów i współautorstwie czterech kolejnych.

P7.2 ICDAR 2023 Competition on Document UnderstanDing of Everything (DUDE) (ICCV 2023, CORE: A*, MNiSW: 200; ICDAR 2023, CORE: A, MNiSW: 140)

Prace Van Landeghem, Tito, (Borchmann), Pietruszka et al. stanowią rozszerzenie idei DUE na pełnoskalowe rozumienie wizualnych dokumentów. DUDE zawiera ponad 5 000 wielostronicowych dokumentów i ponad 41 000 pytań z różnorodnych dziedzin (medycyna, prawo, technika, finanse). Pytania wymagają zaawansowanych operacji wielokrokowego rozumowania i wykraczają poza ekstrakcję tekstu i obejmują zapytania o właściwości layoutu, liczenie elementów, nawigację między stronami, a także pytania bez odpowiedzi. DUDE wprowadza nowe typy odpowiedzi (ekstrakcyjne, abstrakcyjne, listowe) i podzbiory diagnostyczne kategoryzujące pytania według złożoności. Benchmark był prezentowany jako konkurs na ICDAR 2023. Wkład Doktoranta obejmował manualne zbieranie dokumentów, ulepszenie implementacji rodziny modeli T5, eksperymenty poprawiające modele bazowe oraz uzyskanie wyników state-of-the-art., co dla potwierdzenia istotności w dorobku Doktoranta wymaga precyzyjnego uściślenia.

P8.1 Can Models Help Us Create Better Models? Evaluating LLMs as Data Scientists (pierwotnie arXiv preprint, w recenzji – obecnie EACL 2026, CORE: A MNiSW: 140)

Praca Pietruszka, Borchmann, Jędrasz, Morawiecki proponuje FeatEng, benchmark oparty na zautomatyzowanej ekstrakcji cech jako zadaniu testującym integrację umiejętności. Benchmark obejmuje 103 zbiory danych i wymaga od modeli połączenia interpretacji danych, wiedzy domenowej, kreatywnego generowania hipotez i praktycznej implementacji kodu w celu poprawy wydajności modelu docelowego. Kluczowym wynikiem jest wykazanie silnej korelacji między zaproponowaną metryką a rankingiem Chatbot Arena Elo, co sugeruje, że benchmark dobrze aproksymuje praktyczną użyteczność modeli bez kosztownej ewaluacji ludzkiej. Wkład Doktoranta jest znaczący i obejmuje koncepcję benchmarku, pozyskanie i ujednolicenie zbiorów danych, przygotowanie modeli bazowych, zarządzanie metadanymi, rozwój kodu do udostępniania i ewaluacji, kompilację wyników, korelację z rankingiem Chatbot Arena Elo oraz stworzenie ontologii typów cech.

1.4 Ocena wkładu naukowego i samodzielności Doktoranta

Publikacje wchodzące w skład rozprawy zostały opublikowane w najwyżej cenionych konferencjach w dziedzinie AI i NLP. Jest to dorobek nadzwyczaj silny, znacząco przewyższający typowe wymagania stawiane rozprawom doktorskim w dyscyplinie informatyki. Łączna punktacja ministerialna publikacji z wyłączeniem preprintu (P8.1) oraz pracy AAAI 2021 (0 pkt.) wynosi 1 300 punktów, co stanowi wynik wybitny.

Wszystkie publikacje są wieloautorskie, co jest normą w informatyce, szczególnie w obszarze uczenia maszynowego. W rozprawie Doktorant każdorazowo opisuje swój wkład w poszczególne publikacje. Analiza tych opisów pozwala na następującą ocenę samodzielności: w pracach P5.2 (STable), P6.1 (DeepPyramidion/Soft Top-k) i P8.1 (FeatEng) Doktorant pełnił rolę wiodącą lub dominującą i był autorem głównych idei algorytmicznych, przeprowadzał eksperymenty i zarządzał zakresem badań. W pracach P7.1 (DUE) i P7.2 (DUDE) był jednym z głównych współtwórców koncepcji i metodologii. W pracach P4.1 (WikiReading Recycled) i P5.1 (TILT) jego wkład miał charakter istotny, ale bardziej ukierunkowany – odpowiednio redefinicja zbioru danych oraz pozyskiwanie i ewaluacja danych treningowych.

Całościowy wkład Doktoranta w rozwój dziedziny jest wielowymiarowy i obejmuje: (1) nowe narzędzia ewaluacyjne (WikiReading Recycled, DUE, DUDE, FeatEng), które stały się standardami referencyjnymi w społeczności badawczej, (2) nowe architektury (TILT, STable, DeepPyramidion) demonstrujące, że innowacja architekuralna i efektywność obliczeniowa nie muszą stać w sprzeczności, (3) techniki optymalizacji (Soft Top-k) redukujące złożoność uwagi do subliniowej, oraz (4) weryfikację praktyczną, a komercjalizacja TILT i STable potwierdza, że proponowane rozwiązania mają realną wartość przemysłową.

1.5 Zagadnienia dyskusyjne

Zawarte poniżej uwagi i pytania nie mają na celu krytykowania prezentowanej rozprawy, a stanowią zaproszenie do dyskusji i dalszej refleksji nad przedstawionymi w pracy zagadnieniami.

1. Doktorant prezentuje pięciofazową mapę metodologiczną jako główny wkład konceptualny rozprawy. Choć narracja jest przekonująca, a poszczególne kontrybucje zasadniczo adresują zidentyfikowane wyzwania, mapa ta ma wydźwięk *ex post*. Poszczególne publikacje powstawały w latach 2020-2026, a przedstawiona ramowa narracja metodologiczna wygląda na zbudowaną retrospektywnie. Czy Doktorant mógłby odnieść się do tego, w jakim stopniu mapa była apriorycznym planem badawczym, a w jakim retrospektywną racjonalizacją uzyskanych wyników?

2. Identyfikacja pięciu feedback loops (Benchmark-Model Development Loop, Complexity-Optimization Trade-off Loop, itd.) stanowi interesującą konceptualizację. Jednakże ich empiryczna weryfikacja w rozprawie jest ograniczona. Czy Doktorant dysponuje konkretnymi dowodami ilościowymi na to, że przerwanie któregośkolwiek z tych cykli istotnie przyspieszyło postęp w kolejnej fazie?

3. Generalizowalność mapy metodologicznej. Doktorant sam zauważa w ograniczeniach (P9.3), że mapa może nie być uniwersalnie stosowalna do wszystkich poddziedzin NLP. Jak Doktorant postrzega stosowalność zaproponowanej mapy do dziedzin pokrewnych, np. wizji komputerowej, uczenia ze wzmocnieniem, uczenia reprezentacji, gdzie wyzwania generalizacji przyjmują odmienne formy?

4. TILT i ewolucja modeli. Architektura TILT powstała w 2021 roku, przed eksplozją dużych modeli multimodalnych (GPT, Gemini). Jak Doktorant ocenia aktualność TILT w kontekście

współczesnych modeli bazowych (foundation models), które również integrują tekst, obraz i layout? Czy podejście TILT zachowuje przewagi np. w warunkach ograniczonych zasobów obliczeniowych?

5. W pracy nad STable trening permutacyjny wymaga ekspozycji na wiele losowych kolejności generowania komórek. Jaka jest złożoność obliczeniowa tego podejścia względem standardowego treningu? Jak STable skaluje się na tabelach o dużych rozmiarach (setki wierszy/kolumn)?

6. Od czasu publikacji DeepPyramidion pojawiło się wiele alternatywnych podejść do efektywnej uwagi (Flash Attention, Mamba, RWKV, Linear Attention). Jak Doktorant pozycjonuje Soft Top-k operator względem tych nowszych rozwiązań?

2 Analiza strony formalnej rozprawy

2.1 Ocena układu pracy i redakcji manuskryptu

Rozprawa doktorska napisana jest w języku angielskim i pod względem edytorskim jest opracowana z zachowaniem stylu. Zawiera streszczenie (Abstract), spis treści, rozdział wprowadzający, z notacją i tłem technicznym, pięć rozdziałów merytorycznych odpowiadających poszczególnym publikacjom, rozdział podsumowujący oraz biografię naukową. Praca liczy 45 stron tekstu ramowego (bez przedruków artykułów), co jest objętością odpowiednią dla rozprawy opartej na zbiorze publikacji, gdzie tekst ramowy pełni funkcję integrującą, łącząc poszczególne artykuły w spójną narrację badawczą.

Terminologia używana w pracy jest właściwa dla analizowanego obszaru badań. Doktorant posługuje się aparatem pojęciowym z zakresu uczenia maszynowego, przetwarzania języka naturalnego i rozumienia dokumentów. Notacja jest wprowadzona w rozdziale P3 w sposób dość zdawczy i stosowana nie do końca konsekwentnie (pomiędzy publikacjami). Bibliografia do części ramowej obejmuje 71 pozycji, reprezentujących aktualne jak również podstawowe prace z dziedziny. Układ pracy jest logiczny i wspierający główną argumentację. Warto jednak odnotować, że tekst ramowy mógłby zyskać na wartości, gdyby zawierał bardziej rozbudowaną sekcję dyskusyjną porównującą zaproponowaną mapę metodologiczną z rzeczywistym przebiegiem badań (np. czas kontrybucji vs. logiczna sekwencja faz). Sama mapa jest dość specyficzna i rozwija ciekawy wątek organizowania metodycznego w zakresie tematu pracy, ale właśnie przez to nie jest miarodajnie usytuowana w kursie dzieła, w szczególności w naukach badających metodę naukową. Jest to jednakże uwaga drugorzędna, w znikomym stopniu wpływająca na odbiór zasadniczej części pracy, czyli publikacje. Podsumowując, tekst ramowy wprowadzający do publikacji mógłby być przygotowany staranniej.

2.2 Uwagi szczegółowe

Większość treści rozprawy jest poprawna pod względem językowym, stylistycznym i merytorycznym. W trakcie czytania zwróciłem uwagę na następujące drobne kwestie: w bibliografii występują zdublowane pozycje – niektóre publikacje są cytowane dwukrotnie pod różnymi numerami co wymagałoby korekty edytorskiej. Ponadto, opis cykli życia technologii w rozdziale P1.3 jest obszerny, ale jego związek z konkretnymi wynikami badawczymi rozprawy mógłby być bardziej ścisły.

3 Komentarz do oceny rozprawy

Ewaluacja rozprawy doktorskiej składającej się ze zbioru wieloautorskich publikacji stanowi wyzwanie dla recenzenta, które w przypadku doktoratu P. Michała Pietruszki jest szczególne. Po pierwsze, Doktorant niekonsekwentnie i w sposób nie do końca przejrzysty deklaruje swój wkład w każdą z publikacji, choć stara się opisać go w specjalnych ramkach 'Publication' zamieszczanych na końcu każdego podrozdziału. Po drugie, w przypadku kluczowych kontrybucji (STable, DeepPyramidion, FeatEng) rola Doktoranta jako głównego twórcy wymaga jasnego doprecyzowania i przekonującego udokumentowania. Niemniej, spójność tematyczna zbioru, od ewaluacji, przez architekturę i optymalizację, po aplikację, świadczy o planowym i konsekwentnym programie badawczym, a nie o przypadkowym zbiorze okazjonalnych współprac, co działa pozytywnie na odbiór cyklu.

Na skomentowanie zasługuje fakt, że Doktorant łączy perspektywę badawczą z praktyczną, a jego kariera obejmuje zarówno badania akademickie (Uniwersytet Jagielloński), jak i przemysłowe. Ten podwójny kontekst nadaje rozprawie autentyczność, szczególnie w zakresie argumentacji o luce generalizacji między akademią a przemysłem. Domniemuję, że Doktorant pisze o problemie, który sam doświadczał zawodowo.

4 Ocena spełnienia wymogów ustawowych

Rozprawa prezentuje szeroką wiedzę teoretyczną Autora w dyscyplinie informatyki, ze szczególnym uwzględnieniem uczenia maszynowego, przetwarzania języka naturalnego, rozumienia dokumentów i ewaluacji systemów AI. Doktorant demonstruje swobodę w poruszaniu się między teorią (np. formalizm matematyczny operatora Soft Top-k), projektowaniem architektonicznym (TILT, STable, DeepPyramidion), metodologią ewaluacyjną (WikiReading Recycled, DUE, DUDE, FeatEng) i zastosowaniami przemysłowymi.

Praca wykazuje umiejętność samodzielnego prowadzenia pracy naukowej. Doktorant jest współautorem i autorem oryginalnych idei algorytmicznych, projektantem eksperymentów, autorem implementacji i osobą zarządzającą zakresem projektów badawczych. Dla pełnego poparcia tego stwierdzenia, w nawiązaniu do wcześniej przedstawionych uwag, wartościowym byłoby przedstawienie w zdokumentowanej formie uzupełnienia potwierdzającego wkład w poszczególne artykuły.

Rozprawa stanowi oryginalne rozwiązanie problemów naukowych w kilku aspektach. Pojawiają się nowe podejścia architektoniczne (TILT, STable, DeepPyramidion), nowe techniki optymalizacji (Soft Top-k), nowe narzędzia ewaluacyjne (WikiReading Recycled, DUE, DUDE, FeatEng) oraz nowa perspektywa metodologiczna na problem luki generalizacji w NLP. Praktyczna weryfikacja tych rozwiązań, jak komercjalizacja TILT i STable, patenty, przejęcie firmy wraz z IP z zakresu doktoratu dodatkowo potwierdza wartość aplikacyjną wyników.

5 Konkluzja

Przedstawiona praca doktorska, autorstwa mgr. inż. Michała Pietruszki, spełnia kryteria określone przez Ustawę z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668). Praca doktorska ukazuje głęboką znajomość dziedziny przetwarzania języka naturalnego, rozumienia dokumentów i uczenia maszynowego, i pozwala wywnioskować o wysokich umiejętnościach samodzielnego prowadzenia pracy naukowej przez Doktoranta.

Rozprawa koncentruje się na systematycznym adresowaniu luki generalizacji w NLP poprzez spójną mapę metodologiczną, obejmującą ewaluację, innowację architektralną, optymalizację, holistyczną aplikację i integrację zaawansowanych umiejętności. Wkład naukowy, choć jest wielowymiarowy i obejmuje zarówno nowe architektury i algorytmy o potwierdzonej wartości, jak i benchmarki, które stały się standardami referencyjnymi w społeczności badawczej, wymaga potwierdzenia pod kątem samodzielności wykonanych zakresów prac i wkładu. Dorobek publikacyjny, obejmujący prace w ACL, NeurIPS, ICCV, ICDAR, EACL, AACL i CoNLL, jest wybitny pod względem jakości konferencji.

Wszelkie uwagi krytyczne, które pojawiły się w recenzji, mają charakter dyskusyjny i nie wpływają na moją jednoznaczną i pozytywną ocenę rozprawy, zarówno pod kątem oryginalności, jak i wartości merytorycznej. Określenie celów badawczych, proponowane innowacyjne metody ich realizacji, przeprowadzenie badań na najwyższym poziomie i kompetencje w formułowaniu wniosków dowodzą, że Autor jest przygotowany do dalszej działalności naukowej.

Bazując na tych przesłankach wnioskuję, by mgr inż. Michał Pietruszka był dopuszczony do kolejnych etapów postępowania.

Tomasz
Kajdanowicz

Digitally signed by
Tomasz Kajdanowicz
Date: 2026.06.09
22:20:13 +02'00'