Prof. dr hab. inż. Bogusław Cyganek
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Informatyki, Elektroniki i Telekomunikacji
Instytut Elektroniki
cyganek@agh.edu.pl

Kraków, 23.03.2026

Recenzja rozprawy doktorskiej
Pana magistra inżyniera Jana Michała Dubińskiego
zatytułowanej:
Reliable and Safe Generative Models

Wstęp

Recenzja rozprawy doktorskiej Pana magistra inżyniera Jana Dubińskiego pt. „*Reliable and Safe Generative Models*[1]", która powstała w roku 2025 na Politechnice Warszawskiej. Promotorem pracy jest Pan prof. dr hab. inż. Przemysław Rokita. Przygotowanie recenzji zostało wykonane na zlecenie Przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja, Pana prof. dra hab. inż. Jarosława Arabasa. Rozprawa doktorska została napisana po angielsku. Recenzja została przygotowana w formie odpowiedzi na pytania dotyczące rozprawy doktorskiej Pana mgra inż. Jana Dubińskiego.

1. Problem badawczy oraz jego znaczenie

Jaki jest najważniejszy problem rozważany w rozprawie?

Rozprawa doktorska Pana mgra inż. Jana Dubińskiego wpisuje się w nowoczesną dziedzinę współczesnej nauki związanej z rozwojem sztucznej inteligencji (ang. *artificial intelligence* – AI) i dotyczy bardzo istotnych tzw. modeli generatywnych. Są to systemy AI zdolne do tworzenia nowych treści, takich jak teksty czy obrazy, a jednocześnie spełniających ściśle określone założenia, takie jak statystyczna zgodność z danymi użytymi do ich trenowania. Modele tego typu zyskały wielkie zainteresowanie właśnie ze względu na możliwości twórcze, a nie tylko analityczne, takie jak dla przykładu klasyfikacje, czy detekcje obiektów w obrazach. Dzięki modelom generatywnym możliwe jest więc tworzenie nowych utworów muzycznych, tekstów, obrazów wideo, czy też symulacji zjawisk fizycznych – te ostatnie stanowią jeden z głównych obszarów zainteresowania Doktoranta. Możliwości generatywnego AI otwierają olbrzymie i wcześniej nie spotykane możliwości stosunkowo łatwego i szybkiego w realizacji tworzenia danych multimedialnych. Oczywiście, i jak w przypadku większości nowoczesnych technologii, oprócz niewątpliwych walorów pozytywnych, istnieją też liczne zagrożenia, zarówno dla użytkowników, jak i samych twórców modeli oraz właścicieli danych, za pomocą których modele te są uczone, nie zawsze za zgodą, czy też nawet wiedzą tych ostatnich. Również tego typu

[1] Niezawodne i bezpieczne modele generatywne

zagadnieniom szeroko rozumianej ochronnych wartości intelektualnej w modelach uczenia głębokiego poświęcona jest znaczna część rozprawy doktorskiej Pana mgra inż. Jana Dubińskiego.

Czy praca ma charakter naukowy?

Nie ulega wątpliwości, że praca doktorska Pana mgra inż. Jana Dubińskiego ma charakter naukowy. Tak jak już wspomniano, dotyczy ona głównie najnowszych metod tzw. generatywnej AI, a w szczególności opracowaniu takich modeli, które w najlepszy znany sposób mogą służyć do symulacji eksperymentów z zakresu fizyki wysokich energii, takich jakie są przeprowadzane np. w Wielkim Zderzaczu Hadronów CERN. W rozprawie podjęta została również tematyka problemów związanych z zapewnieniem wartości intelektualnej oraz jej ochrony w tego typu modelach generatywnych. Są to zagadnienia nie tylko będące jednym z pierwszoplanowych przedmiotów współczesnej nauki, ale również dotyczą jej wielu różnych dyscyplin, takich jak informatyka i fizyka.

Czy praca ma znaczenie praktyczne?

Oprócz niewątpliwych walorów czysto naukowych, praca doktorska Pana mgra inż. Jana Dubińskiego ma niebagatelne znaczenie praktyczne. Opracowane przez Doktoranta metody oraz algorytmy miały na celu rozwiązanie istotnych zagadnień praktycznych, takich jak m.in. efektywna symulacja eksperymentów fizyki cząstek wysokich energii. Również zagadnienia dotyczące szeroko rozumianego bezpieczeństwa oraz prywatności danych w kontekście AI mają olbrzymie znaczenie praktyczne, które niewątpliwie będzie tylko rosnąć wraz z dynamicznym rozwojem AI.

2. Wkład Autora

Jaki jest najważniejszy wkład Autora rozprawy?

Najważniejsze osiągnięcia naukowe przedstawione w rozprawie Pana magistra inżyniera Jana Dubińskiego można podzielić na dwa główne zagadnienia dotyczące modeli generatywnych:

A. *Metody rozszerzające możliwości generatywnych metod AI*.

B. *Metody ochrony wartości intelektualnej w modelach uczenia głębokiego*.

Każda z powyższych grup została dodatkowo podzielona na bardziej szczegółowe metody dotyczące rozwiązania specyficznych problemów naukowych, które zostaną przedstawione w poniższym zestawieniu.

A.1. *Metoda selektywnego zwiększania różnorodności próbek generowanych przez GAN*.

Metoda selektywnego zwiększania różnorodności próbek generowanych przez GAN (ang. *generative adversarial networks*) została opracowana głównie do umożliwienia symulacji procesów zachodzących podczas eksperymentów fizyki wysokich energii w Wielkim Zderzaczu Hadronów (LHC) w ośrodku CERN. Dotychczas stosowane metody symulacji detektorów cząstek wykorzystywały metody typu Monte Carlo, które mimo iż o uznanej renomie, to jednak wymagają niebagatelnych nakładów obliczeniowych. W tym kontekście, generatywne metody AI mogą stanowić ich istotną alternatywę. Tym niemniej, wyzwaniem jest tutaj niezawodność tego typu rozwiązań, które muszą uwzględniać zmienne warunkowe, które istotnie wpływają na odpowiedzi detektora padających cząstek. Jak zauważył Doktorant,

problemem jest tu moduł generatora, który wykazuje skłonność do ignorowania szumów wejściowych i koncentruje się wyłącznie na zmiennych warunkujących, generując ograniczony zakres wyników. W rezultacie uniemożliwia to modelowi uchwycenie pełnej zmienności danych zderzeniowych, co zmniejsza jego użyteczność w zadaniach symulacyjnych. Dzięki szczegółowej analizie niedoskonałości istniejących metod generatywnych, które mają tendencję do operowania tak jakby rozkład zmiennych warunkowych był jednorodny, Pan mgr inż. Jan Dubiński wraz ze współautorami zdołał zaproponować oryginalną modyfikację funkcji kosztu, która pozwala obejść te ograniczenia. Jest to metoda nazwana Selective Diversity GAN (SDI-GAN), która umożliwia regularyzację generatora, która z kolei wymusza różnorodność generowanych próbek w sposób zależny od samych danych treningowych. Metoda ta umożliwia więc skalowanie efektu różnorodności generowanych próbek zgodnie z wariancją zmiennych warunkujących, co w efekcie wpływa na generowanie próbek bardziej różnorodnych i realistycznych, a jednocześnie bez utraty dokładności w stosunku do rozkładów danych treningowych. Pozytywne efekty takiego podejścia zostały zaprezentowane w realistycznym przypadku symulacji tzw. kalorymetru zerowego w eksperymencie ALICE.

Metoda ta została opublikowana w roku 2022 na konferencji International Conference on Neural Information Processing (ICONIP), w publikacji pt. „*Selectively increasing the diversity of GAN-generated samples*”, w której Pan mgr inż. Jan Dubiński jest pierwszym z autorów.

A.2. *Metoda symulacji detektora szybkich cząstek z wykorzystaniem mieszanki ekspertów generatywnych – ExpertSim*.

Jest to kontynuacja i udoskonalenie metody zaprezentowanej w poprzednim zadaniu (A.1). Okazało się bowiem, że mimo iż zaprezentowana tam metoda wykazuje świetne rezultaty, to są one ciągle niewystarczające do precyzyjnej symulacji procesów zachodzących pod wpływem multimodalnych danych fizyki wysokich energii. Tak złożone oddziaływania cząstek w detektorach prowadzą do rozkładów, które trudno aproksymować wyłącznie za pomocą tylko jednego modelu generatywnego. To co Pan mgr inż. Jan Dubiński zaproponował wraz z zespołem, to zastąpienie jednego takiego modelu zespołem w postaci złożenia współpracujących modeli. W tym podejściu, zamiast trenować pojedynczy generator obejmujący całą przestrzeń odpowiedzi cząstek, trenuje się kilka generatorów równocześnie, tyle że każdy z nich koncentruje się na innej części rozkładu charakterystycznego dla eksperymentów fizyki wysokich energii. Oryginalnym rozwiązaniem zaproponowanym przez Pana mgra inż. Jana Dubińskiego ze współautorami jest specjalny trenowany moduł rozdzielająco-kontrolny – tzw. ruter – który łączy dane wejściowe z najbardziej odpowiednim dla nich ekspertem. W rezultacie system ten w znacznie lepszy sposób może uchwycić różnorodność odpowiedzi detektorów. Z kolei, każdy z indywidulanych ekspertów w tym systemie, to już wcześniej opracowany SDI-GAN (A.1). Jak zostało to pokazane za pomocą eksperymentów, system ten osiąga wyższą dokładność w odtwarzaniu rozkładów eksperymentalnych, zapewniając jednocześnie znaczne przyspieszenie w porównaniu z tradycyjnymi symulacjami typu Monte Carlo.

Metoda ta została opublikowana na prestiżowej konferencji European Conference on Artificial Intelligence ECAI 2025 w publikacji pt. „*ExpertSim: Fast Particle Detector Simulation Using Mixture-of-Generative-Experts*”, w której Pan mgr inż. Jan Dubiński jest drugim z autorów.

B.1. *Metoda aktywnej ochrony przed kradzieżą enkoderów – Bucks for Buckets* (*B4B*).

Jest to pierwsza z serii metod ochrony wartości intelektualnej systemów powstałych w wyniku kosztownego trenowania modeli AI. Jednym z rosnących zagrożeń jest kradzież modelu – możliwe jest to np. poprzez odpowiednią interakcję z systemem za pośrednictwem dostępnego interfejsu API, za pomocą którego próbuje się zbudować substytut modelu – rodzaj modelu-modelu – który ściśle imituje zachowanie modelu oryginalnego. Pan mgr inż. Jan Dubiński wraz z zespołem zauważyli, że istniejące sposoby obrony przed tego typu kradzieżami są w dużej mierze pasywne, koncentrując się na ograniczaniu dostępu do zapytań lub dodawaniu szumu. Jednakże zawodzą one w konfrontacji z pomysłowymi hakerami, którzy potrafią dynamicznie przystosować i zaadaptować swoje strategie ataku. Dzięki tej obserwacji Pan mgr inż. Jan Dubiński wraz z zespołem zaproponowali metodę Bucks for Buckets (B4B). Jej główna idea to monitorowanie zapytań użytkowników w celu określenia stopnia eksploracji/pokrycia przestrzeni tzw. osadzeń (ang. *embeddings*) danego modelu. Sama idea jest tu dość prosta – jeśli ten stopień „przeszukiwania" jest zbyt rozległy, to może to świadczyć właśnie o próbie wyłudzenia wewnętrznej struktury modelu w celu jego podrobienia. Po zidentyfikowaniu tego typu podejrzanego zachowania, w zaproponowanej metodzie nakładane są dodatkowe koszty adaptacyjne, które penalizują próby ekstrakcji, w rezultacie uniemożliwiając stworzenie repliki modelu. Dodatkowo, aby przeciwdziałać atakom z wykorzystaniem wielu kont, Pan mgr inż. Jan Dubiński zaproponował transformacje przestrzeni reprezentacji dla każdego użytkownika, co uniemożliwia koordynację między kontami, ale jednocześnie jest niezauważalne dla uczciwego użytkownika. Prawidłowe działanie metody zostało wykazane eksperymentalnie z wykorzystaniem takich zbiorów jak FashionMNIST, SVHN, STL10 oraz CIFAR10, jak również enkoderów typu SimSiam oraz DINO.
Metoda ta została opublikowana na prestiżowej konferencji: Conference on Neural Information Processing Systems (NeurIPS) 2023, pt. „*Bucks for Buckets (B4B): Active Defenses Against Stealing Encoders*". Pan mgr inż. Jan Dubiński jest pierwszym z autorów.

B.2. *Metoda ochrony przed atakami wglądu w dane w wielkich modelach dyfuzyjnych*.

Jest to kolejna z metod dotyczących ochrony modeli AI, poszerzona jednak o ochronę danych na podstawie których są one trenowane. Główny problem to stwierdzenie, czy zbiór danych został wykorzystany w treningu modelu, co może się wiązać właśnie z naruszeniem praw autorskich. Jednakże nie jest to zadanie ani łatwe, ani proste w realizacji. Głównym problemem jest tu przeważnie olbrzymia liczba danych używanych do trenowania modeli. W tym przypadku stwierdzenie, czy dana próbka rzeczywiście została użyta w procesie trenowania i w istocie „należy" ona do modelu (jest zapisana w jego wagach) jest bardzo trudne. W tym kontekście Pan mgr inż. Jan Dubiński wraz z zespołem analizowali ataki oparte na ocenie przynależności danych do modeli dyfuzyjnych. Pierwsze osiągnięcie to krytyczne spojrzenie na dotychczasowe metody ewaluacji tego typu metod i zaproponowanie nowych sposobów ewaluacji, które są dostosowane do skali nowoczesnych modeli generatywnych. Kluczowe jest tutaj opracowanie specjalnego zbioru danych do oceny metod wykrycia ataków tego typu. Dzięki opracowaniu tej platformy badawczej Pan mgr inż. Jan Dubiński wraz z zespołem przeprowadzili liczne eksperymenty w celu oceny metod wnioskowania. Wyniki te sugerują, że ataki tego typu są mniej skuteczne niż było to sugerowane w pracach innych autorów.

Metoda została opublikowana na konferencji IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2023, w pracy pt. „*Towards More Realistic Membership Inference Attacks on Large Diffusion Models*", w której Pan mgr inż. Jan Dubiński jest pierwszym z autorów.

B.3. *Metoda identyfikacji łamania prawa własności danych w modelach dyfuzyjnych*.

Mimo iż wcześniej opracowane metody detekcji ataków zostały zweryfikowane w wybranych zadaniach, Pan mgr inż. Jan Dubiński wraz z zespołem zauważyli, że nie dostarczają one wiarygodnych wyników w przypadku systemów o większej skali. W tych przypadkach, nie jest jasne w jaki sposób można zweryfikować czy czyjeś zbiory danych zostały użyte, czy też nie, do treningu danego modelu generatywnego. W tego typu zadaniach, ze względu na olbrzymie ilości danych użytych do trenowania bardziej naturalne – jak zauważył Pan mgr inż. Jan Dubiński – jest pytanie o obecność całych zbiorów danych, a nie pojedynczych próbek. Przy tych założeniach Doktorant zaproponował więc nową metodę identyfikacji danych chronionych prawami autorskimi – nazwaną Copyrighted Data Identification (CDI) – która łączy wartości przynależności do modelu nie z jednej, lecz z wielu próbek, a następnie dokonuje ich analizy za pomocą testów statystycznych. CDI została następnie zweryfikowana eksperymentalnie, pokazując że osiąga ona wysoką niezawodność w wykrywaniu zbiorów danych chronionych prawem autorskim, znacznie lepszą od metod już istniejących. Istotne jest tutaj też to, że możliwe jest to bez dostępu do samego modelu, co czyni ją szczególnie atrakcyjną w praktyce, gdzie modele generatywne są często dostępne za pośrednictwem interfejsów API typu black-box.

Metoda ta została opublikowana na prestiżowej konferencji: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2025, w pracy pt. „*CDI: Copyrighted Data Identification in Diffusion Models*". W publikacji tej Pan mgr inż. Jan Dubiński jest pierwszym z autorów.

B.4. *Metoda ochrony przed atakami na prywatność modeli auto-regresyjnych*.

Jest to kolejna z opracowanych przez Pana mgra inż. Jana Dubińskiego oraz współpracowników metod dotyczących wiarygodnego wykrywania użycia danych, często chronionych prawami autorskimi, do trenowania dużych modeli, nie tylko dyfuzyjnych, ale i nowszych tzw. modeli autoregresyjnych. Te ostatnie nie dokonują procesu odszumiania w celu wykrycia ukrytych danych, ale generuje próbki w sposób sekwencyjny na podstawie tzw. tokenów. Proces ten, przebiegający w następujących po sobie krokach, ułatwia jednak zapamiętywanie danych treningowych. Pan mgr inż. Jan Dubiński wraz z zespołem badali zagrożenia w modelach autoregresyjnych. Dzięki określeniu procesu zapamiętywania przez te modele danych treningowych, zdołali opracować nowe metody, które dokonują ekstrakcji treści wprost z wyjść modeli autoregresyjnych. Badacze zaobserwowali tutaj, że modele autoregresyjne są szczególnie podatne na swoisty wyciek prywatności, często przekraczający podatność modeli dyfuzyjnych. W rezultacie, opracowane metody, które dokonują agregacji wartości przynależności (ang. *membership inference*) pomiędzy wieloma próbkami danych, umożliwiają skuteczną detekcję nieautoryzowanego użycia danych objętych prawami autorskimi. Właściwości metody zostały zbadane eksperymentalnie, które potwierdziły podejrzenia że modele oparte na autoregresji są znacznie bardziej narażone na nadużycie prywatności niż ma to miejsce w przypadku modeli dyfuzyjnych.

Opisana metoda została opublikowana na prestiżowej konferencji: International Conference on Machine Learning (ICML) 2025, pt. „*Privacy Attacks on Image AutoRegressive Models*". Jan Dubiński jest tu drugim autorem.

Wyżej wymienione oryginalne osiągnięcia naukowe Pana mgra inż. Jana Dubińskiego są wysoce nowatorskie i dotyczą najnowszych i często jeszcze dobrze nie zbadanych zagadnień generatywnych modeli AI. Świadczy to o głębokiej wiedzy oraz pomysłowości, jak również dojrzałości naukowej Doktoranta. Opracowanie przez Niego oraz współpracujący zespół badawczy metody dotyczą kluczowych zagadnień naukowych współczesnej informatyki. Metody te zostały opublikowane na wiodących konferencjach naukowych dotyczących AI. Zostały one również dogłębnie zweryfikowane eksperymentalnie. Często też został dołączony kod, który umożliwia weryfikację wyników, jak również dalsze rozszerzenie proponowanych metod przez inne zespoły.

3. Poprawność naukowa

Czy stwierdzenia zawarte w rozprawie są godne zaufania? Czy uzasadnienia są poprawne? Wskaż zauważone słabości i błędy. Wskaż także te aspekty dotyczące poprawności, które są najbardziej wartościowe.

Pan mgr inż. Jan Dubiński podjął uzasadnioną naukowo metodykę prowadzenia badań bazującą na wnikliwej analizie problemów oraz ich rozwiązań opublikowanych przez inne zespoły badawcze, a następnie na analizie ich ograniczeń. Bazując na tej wiedzy Doktorant opracował rozwiązania nowsze i konkurencyjne do już istniejących. Wszystkie w ten sposób opracowane metody, które głównie dotyczyły generatywnych modeli AI, zostały przez Pana mgra inż. Jana Dubińskiego dogłębnie zweryfikowane eksperymentalnie. Podejście takie oceniam jako jak najbardziej poprawne z naukowego punktu widzenia.

W rozprawie Pana mgra inż. Jana Dubińskiego natknąłem się jednak na pewne stwierdzenia oraz zagadnienia, które wymagają albo doprecyzowania, albo też dalszego rozszerzenia.

1. Doktorant wraz z zespołem poświęcili wiele uwagi opracowaniu metod detekcji prób kradzieży, czy też podrobienia wytrenowanego modelu generatywnego, jak również nieautoryzowanego użycia danych do trenowania modeli tego typu. Tymczasem, i jak jest to wielokrotnie raportowane w licznych pracach naukowych, a nawet doniesieniach medialnych, zagrożenie istnieje również ze strony samych generatywnych modeli AI, a być może jest jeszcze bardziej istotne ze społecznego punktu wiedzenia. Chodzi tu głównie o możliwość łatwego i szybkiego tworzenia – generowania – fałszywych treści, które jednak do złudzenia przypominają, czy też udają prawdziwe. Są to tzw. *fake news*, takie jak zmontowane obrazy, czy też treści multimedialne mające właśnie na celu wprowadzenie w błąd osoby, do których są adresowane. Jednakże Autor w ogóle nie wspomina o tym fakcie, mimo iż sama rozprawa jest bardzo obszerna; zawiera też i to wielokrotnie przytaczane przeglądy prac naukowych w tej dziedzinie. Warto więc aby Doktorant przybliżył te właściwości sieci generatywnych, a korzystając ze swojej wiedzy na ten temat, przybliżył też możliwości opracowywania metod identyfikacji tego typu *fake news*, czy też innych falsyfikacji danych.
2. Jak już wielokrotnie podkreśliłem przy ocenie osiągnięć Doktoranta, wszystkie one zostały opracowane w ścisłej współpracy z dość licznym zespołem badawczym. Zaprezentowane w rozprawie metody zostały opublikowane w pracach wielo-autorskich, w których Doktorant

jest pierwszym, a w kilku przypadkach – drugim, z autorów. Rozprawa zawiera również opis oryginalnych osiągnięć Pana mgra inż. Jana Dubińskiego. Tym niemniej, nieco brakuje mi jaśniejszego określenia, które dokładnie metody zostały oryginalnie *wymyślone* przez Doktoranta, a które przez innych członków zespołu.

3. Str 41, wzór (4.3) – nie jest jasne czym X_c różni się od $\mathcal{X}_c$ (tzn. pisanego czcionką kaligrafowaną).
4. Str. 45, rozdz. 4.4.3 – warto by właściwości metody sprawdzić na jakiś innych danych, na których inni badacze sprawdzają jakość GAN, np. MS-GAN itd.
5. Str. 46, rys. 4.4.4 („diverse results") –mimo wszystko są to rezultaty dość różne, przynajmniej wizualnie, od tych "real data" (kształt, położenie itd.), więc chyba trudno tu jedną metodę np. SDI-GAN specjalnie wyróżniać?
6. Str. 47, rozdz. 4.5 – w ewaluacji eksperymentalnej tej metody brakuje mi tzw. *ablation study*; brakuje analizy choćby wpływu parametru λ ze wzoru (4.5).
7. Str. 53, rozdz. 5.2.2 – Autor nie sięga do źródeł metod Mixture-of-Experts, ograniczając się tylko do ery deep learningu. Skoro już pisać rozdział tego typu, to warto by jednak sięgnąć do prac będących u podstaw tego typu metod.
8. Str. 56,– wzór (5.1) chyba jest to samo co (4.4) ale – dla utrudnienia – nieco inaczej zapisane?
9. Str. 57 – wzór (5.5) dobrze by było porównać z (4.5); co tu nowego, czemu?
10. Str. 57 – nie jest jasne jak router jest trenowany; supervised, unsupervised, czy też inaczej? Nie jest też jasne jak jest wyznaczany „most suitable expert" – jeśli był uczony w trybie nadzorowanym, to jak były oznaczane dane?
11. Str. 58 – wzór (5.8), jeśli system był trenowany w trybie end-to-end to już różnych parametrów λ jest kilkanaście, czego jednak Autor nie przedyskutował.
12. Str. 61, rys. 5.5.2 – co to jest „true" zaprezentowane na wykresach? Trudno też coś powiedzieć o samej metodzie ExpertSim z samej obserwacji tych wykresów.
13. Str. 62 – parametry λ są dobierane w procesie, które Doktorant opisał jako „*tuning efforts*", ale może dałoby się i te parametry wytrenować, skoro metoda i tak jest end-to-end? Z drugiej strony może te λ routera można by wpisać "wmnożyć" w wagi modelu i tak trenować?
14. Str. 95 – lepszego wyjaśnienia wymaga odległość FID, dlaczego została użyta ta właśnie?
15. Str. 101, tabela 7.7.1 – jest pewien problem z tymi eksperymentami, bo Autorzy porównują je tylko sami ze sobą; Czy nie ma innych badań w tej dziedzinie, do których można by się porównać?
16. Str.112 – wyjaśnienia wymaga stwierdzenie „*observing that training data representations exhibit a markedly different distribution from test data*", choćby w kontekście trenowania modeli z wykorzystaniem tzw. multi-fold. Czy w tym przypadku, gdy dane wielokrotnie losowo dzielimy na fold treningowy oraz testowy, to rzeczywiście mamy do czynienia z tak różnymi dystrybucjami?
17. Str. 155 i kolejne – Bardzo obszerna bibliografia, która liczy 267 pozycje. Jednakże wiele z nich niekompletnych, np. [8] [24] [50] itd. , bądź też w ogóle bezużytecznych [70] [185].

Należy tutaj jednoznacznie stwierdzić, że powyższe pytania i sugestie poszerzenia dodatkowych zagadnień mają wyłącznie charakter dyskusyjny i w żadnym stopniu nie umniejszają istotnego wkładu merytorycznego przedstawionego w rozprawie Pana mgra inż. Jana Dubińskiego.

4. Wiedza Kandydata

Które z rozdziałów rozprawy omawiają istniejący stan wiedzy i dzięki temu potwierdzają ogólny stan wiedzy kandydata w zakresie informatyki? Jakie obszary tych dyscyplin zostały omówione w tych rozdziałach/sekcjach? Jaka jest opinia recenzenta o bibliografii? Prosimy o podanie innych argumentów za lub przeciw, że kandydat posiada ogólną wiedzę w dyscyplinie ITT.

Rozprawa doktorska Pana mgra inż. Jana Dubińskiego to dzieło bardzo obszerne jak na prace tego typu – liczy aż 250 stron. Składa się ona z 10 rozdziałów głównych, spisu publikacji oraz suplementów. Najważniejsze jednak jest to, że zaprezentowany w rozprawie materiał naukowy jest bardzo obszerny i merytorycznie bogaty. Zaprezentowane metody zostały już opublikowane w 6 publikacjach na czołowych i wysoce punktowanych konferencjach ML/AI, takich jak NeurIPS 2023, ECAI 2025, czy też CVPR 2025 (dokładny spis tych publikacji znajduje się na str. 152 rozprawy). W 4 z tych 6 publikacji Pan mgr inż. Jan Dubiński jest pierwszym autorem. Natomiast w 2 pozostałych, Doktorant jest drugim z autorów. Mimo iż są one wieloautorskie, to najwyższy światowy poziom tych publikacji świadczy też o niekwestionowanej wiedzy Pana mgra inż. Jana Dubińskiego w dyscyplinie informatyka techniczna i telekomunikacja.

Pełniejsza lista istotnych publikacji naukowych, których jednym z autorów jest Pan mgr inż. Jan Dubiński jest bardzo imponująca, uwzględniwszy młody wiek Doktoranta. Baza Web of Science przytacza 11 publikacji tego typu, które są cytowane 26 razy, a tzw. indeks Hirscha Pan mgr inż. Jan Dubiński wynosi 3. Są to wysokie wartości, ale szczególnie istotne są tutaj wysoko punktowane konferencje.

Wszystko powyższe świadczy o ponadprzeciętnej wiedzy Doktoranta Pana magistra inżyniera Jana Dubińskiego w dyscyplinie informatyka techniczna i telekomunikacja, a w szczególności w dziedzinie metod uczenia maszynowego i sztucznej inteligencji.

5. Podsumowanie

Rozprawa doktorska Pana magistra inżyniera Jana Dubińskiego prezentuje ogólną i głęboką wiedzę Kandydata w dyscyplinie informatyka techniczna i telekomunikacja, a w szczególności w zakresie nowoczesnych metod sztucznej inteligencji. Świadczy też o umiejętność samodzielnego prowadzenia badań naukowych na najwyższym światowym poziomie, jak również o dobrym opanowaniu warsztatu badawczego. Zaprezentowane w rozprawie doktorskiej Pana magistra inżyniera Jana Dubińskiego metody i algorytmy są oryginalne i dotyczą rozwiązania istotnych problemów naukowych dotyczących generatywnej sztucznej inteligencji. Opracowane metody stanowią oryginalny i własny wkład Doktoranta Pana magistra inżyniera Jana Dubińskiego w rozwój dyscypliny naukowej informatyka techniczna i telekomunikacja.

Recenzowaną pracę oceniam jako z spełniającą ze znacznym nadmiarem formalne oraz zwyczajowe wymagania stawiane rozprawom doktorskim. Wnioskuję o jej przyjęcie oraz o dopuszczenie Pana magistra inżyniera Jana Dubińskiego do publicznej obrony.

Zważywszy również na bardzo wysoki poziom naukowy opracowanych metod oraz bardzo cenne publikacje, w których Pan mgr inż. Jan Dubiński jest pierwszym z autorów, **wnioskuję o wyróżnienie rozprawy**.