

Gliwice, dn.26.03.2025 r.

Prof. dr hab. inż. Marcin Gorawski
Katedra Informatyki Stosowanej
Wydział Automatyki, Elektroniki i Informatyki
Politechnika Śląska
ul. Akademicka 16,
44-100 Gliwice,
Marcin.Gorawski@polsl.pl

Recenzja rozprawy doktorskiej
mgr inż. Marcin Krystek
z tytułowanej:
Data Mesh for Smart Cities

Podstawą opracowania recenzji jest pismo Prodziekana Wydziału Informatyki i Telekomunikacji Politechniki Poznańskiej, dr hab. inż. Szymona Wilka, Prof. PP (DIIIT-63-27/2024) z dnia 27.11.2024r. w związku z uchwałą Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Poznańskiej z dnia 26.11.2024r. Rozprawa liczy sobie 97 stron i jest przygotowana w języku angielskim. Uzyskałem wgląd do dokumentacji rozprawy od dnia 26.12.2024 r.

1. Dane o doktorancie:

1) Studia

Absolwent studiów magisterskich w Politechnice Poznańskiej, w latach od [REDAKTOWANE] r. W latach [REDAKTOWANE] rozpoczął studia doktoranckie w Szkole Doktorskiej Politechniki Poznańskiej.

2) Rodzaj doktoratu

Przedstawiona do recenzji rozprawa doktorska mgr inż. Marcina Krystka pt. *Data Mesh for Smart Cities* dokumentuje przebieg realizacji doktoratu wdrożeniowego. Motywacją była Strategia Rozwoju Miasta Poznania 2030, uchwała RMP z dnia 11.05.2010 r. Po aktualizacji i ogłoszeniu "Strategii 2020 +" uchwałą Nr XLI/708/VII/2017 z dnia 24 stycznia 2017 r. 1.2.16. - wpisanie efektywnego wdrażania w Poznaniu koncepcji i rozwiązań z zakresu Smart City do priorytetowych zagadnień i kluczowych zadań.

3) Promotor i opiekun merytoryczny

Rozprawa ta wykonana została pod kierunkiem promotora dr hab. inż. Mikołaja Morzego, prof. PP Wydziału Informatyki i Telekomunikacji Politechniki Poznańskiej oraz opiekuna merytorycznego dr inż. Cezarego Mazurka z Poznańskiego Centrum Superkomputerowo-Sieciowego (PCSS) afiliowanego przy Instytucie Chemii Bioorganicznej Polskiej Akademii Nauk (ICHB PAN).

4) Miejsce pracy z podaniem zajmowanych stanowisk lub pełnionych funkcji

Mgr inż. Marcin Krystka pracuje od [REDAKTOWANE] at w PCSS przy ICHB PAN, od początku na stanowisku [REDAKTOWANE] gdzie tematyka pt. *Data Mesh for Smart Cities* była realizowana we współpracy z UM Poznań nad cyfryzacją zasobów miejskich i ich udostępnianiem szerokiej publiczności.

<p>2</p>	<p>Cel rozprawy</p> <p>Celem pracy doktorskiej mgr inż. Marcina Krystka pt.: Data Mesh for Smart Cities było stworzenie unikalnej platformy otwartych danych miejskich w oparciu o adaptacyjną analizę badawczo – eksperymentalną koncepcji Smart City 4.0 (SC4.0), która dostarcza w praktyce gotowe usługi i funkcjonalne aplikacje na potrzeby mieszkańców i urzędników. Ostatecznie doktorant mgr inż. Marcin Krystek jest współautorem projektu, budowy i wdrożenia Platformy otwartych danych dla inteligentnych miast (Open Data platform for Smart Cities (ODSC)) uszczegółowioną dla Poznania (ODSC [Poznań]). ODSC [Poznań] znajdujemy pod linkiem Poznań-Otwarte Dane. Odbiorcami ODSC [Poznań] są: a) Mieszkańcy i grupy Mieszkańców (np. zorganizowanych w Rady Osiedli, Stowarzyszenia, NGO...), b) Twórcy stron internetowych, aplikacji mobilnych, nowych e-usług, c) Przedsiębiorcy sektora MŚP. Istotą ODSC[Poznań] jest zagwarantowanie długofalowego wsparcia dla zrównoważonego rozwoju Poznania dzięki publicznemu udostępnieniu i wykorzystaniu zasobów treści i danych miejskich, tworzonych i gromadzonych za publiczne pieniądze.</p> <p>ODSC[Poznań] udostępnia (www.poznan.pl/opendata/service-purpose):</p> <ul style="list-style-type: none"> • Zasoby www.poznan.pl (treści i usługi); • Zasoby bip.poznan.pl (treści); • Wybrane zasoby Systemu Informacji Przestrzennej (SIK); • Zorientowane zasoby do ponownego użycia; • Web Services oraz API e-usługi. <p>Doktorant był współautorem procesu projektowania i prototypowania wybranych elementów ODSC[Poznań] w postaci gotowych usług i aplikacji tworzonych wg. koncepcji wypracowaną w części naukowej.</p> <p>Ocena celu i tezy rozprawy: Celem rozprawy było przedstawienie nowego podejścia do rozwiązań ODSC[Poznań]. Cel ten mógłby zostać przedstawiony bardziej precyzyjnie. W dysertacji, na stronie 10 (1.4 Thesis), postawiona została teza badawczo-wdrożeniowa. Prezentowane rozwiązania ODSC mające na celu uzasadnienie dokonań badawczo-eksperymentalnych i potwierdzenie postawionej tezy, przedstawiono w sposób dość rozwlekły, ale w sumie czytelny.</p>
<p>3</p>	<p>Ocena merytoryczna rozprawy</p> <p>Niniejszą rozprawę zaliczam do klasy prac badawczo-konstrukcyjno-eksperymentalnych. Nie ulega wątpliwości, że Doktorant osiągnął wysoki, ale nie kompletny stopień opanowania teorii inżynierii danych, a zwłaszcza wiedzy o badaniu, projektowaniu i prototypowaniu platform otwartych danych. Dysertacja jest skonstruowana poprawnie, zawiera wstęp wraz z niepełnym omówieniem stanu wiedzy, ale nie dość wyraźnie sformułowany jest cel oraz nazbyt obszernie, mało precyzyjnie postawioną tezę. Z drugiej strony, pewnym utrudnieniem w czytaniu rozprawy była momentami nieprecyzyjność sformułowań technicznych i nie wprowadzeniu na początku pracy podstawowych pojęć i definicji. Stąd przedstawiam poniższe kompendium wiedzy dla uzyskania większej przejrzystości i czytelności tematyki pt:</p> <p>Problem badawczy - jego charakter i znaczenie</p> <p>Generalnie tematyka pt.: Data Mesh for Smart Cities jest rozumiana bardzo szeroko, ma charakter problemu ogólnego. Stąd przyjmuję pierwsze przybliżenie do oceny tematyki w ramach uszczegółowienia, w zagadnieniu pt.: Data Mesh for Smart Cities (DaMeSC) w kontekście przedstawia próby podejścia do modelu Inteligentnego miasta czwartej generacji (SC4.0) poprzez wdrożenie Platformy otwartych danych dla inteligentnych miast (ang. Open Data platform for Smart Cities (ODSC)). Wtedy, w takim doktoracie wdrożeniowym, wyróżniamy tematykę</p>

badawczą, tu opisaną jako wielokontekstowa analiza badawcza koncepcji DaMe i uzasadnienie możliwości jej efektywnej adaptacji spełnienia wymagań platformy otwartych danych miejskich. Tematykę wdrożeniową w tym doktoracie opisać można jako wdrożenie prototypu systemu zarządzania danymi miejskimi wykorzystującego wielkie modele językowe.

Dokonanie oceny niniejszej rozprawy umożliwiają uzyskane odpowiedzi na 2 zasadnicze pytania: 1) Czy problem badawczy ma charakter naukowy? oraz 2) Czy problem badawczy ma znaczenie praktyczne? wymaga perspektywicznego ramowego opisu koncepcji DaMe. Wobec globalnych wyzwań przeludnienia, zagrożeń klimatycznych i szybszego wyczerpywania się zasobów Zgromadzenie Ogólne Organizacji Narodów Zjednoczonych w 2015 roku oficjalnie ustanowiło 17 Celów Zrównoważonego Rozwoju (ang. Sustainable Development Goals – SDG), z perspektywą realizacji do 2030 r. Komisja Europejska i Eurostat analizuje przebieg prac nad dążeniem do realizacji 17 SDGs tj.: eliminacja ubóstwa i głodu, zapewnienie dobrego zdrowia i dobrobytu oraz dobrej jakości edukacji i uczenia, równość płci, dostępna czysta woda i warunki sanitarne, dostępna i czysta energia, wzrost gospodarczy zatrudnienie i godna praca, odporność na skutki katastrof, społeczeństwa pokojowe, zrównoważona konsumpcja i produkcja, przemysł i infrastruktura, zrównoważona industrializacja i innowacyjność, silne instytucje, sprawiedliwość, partnerstwa celowe. Wielowymiarowy model SC4.0 jest mocno zorientowaną koncepcją na zrealizowanie SDGs przy dobrze zdefiniowanej aktywnej relacji z interesariuszami tj. podmiotami publicznymi i prywatnymi w zakresie większej efektywności, optymalności i nowych metod finansowania.

Uważam, że wielowymiarowy model SC4.0 należy budować, obejmując obszary 3 wymiarów zarządzania IT:

- 1) Zarządzanie strategiami SC4.0;
- 2) Strategie zarządzania Data Mesh;
- 3) Zarządzanie otwartymi danymi w SC4.0, (ang. Open Data Management in SC4.0 (ODMSC4.0))

1. Zarządzanie strategiami SC4.0 to zarządzanie i rozwój ukierunkowanych systemów transakcyjnych i analitycznych m.in systemy tj.: inteligentnego transportu, CCTV, klimatyczne. W tej fazie rozwoju SC (1.0 – 3.0) większość tych systemów powinna być zoptymalizowana pod kątem wydajności i efektywności z ograniczeniami narzuconymi przez klasyczne architektury Hurtowni Danych (ang. Data Warehouses) i Jezior Danych (ang. Data Lake). Wadami tych klasycznych architektur, pomimo ich elastyczności i skalowalności, jest trudne zarządzanie wielkimi zbiorami danych, a złożone potoki danych strumieniowych i słabo specyfikowane zbiory danych są wyzwaniem dla utrzymania optymalnego przetwarzania z zachowaniem ich wysokiej jakości, pochodzenia i zaufania. Także złożoność projektowania architektury jezior danych prowadziła czasami do stopniowego pogarszania ich użyteczności i skuteczności. Z badań wynika, że generalnie centralizacja zarządzania danymi utrudnia efektywną integrację i wykorzystanie różnych źródeł danych powodując opóźnienia w dostosowywaniu się do nowych źródeł danych i potrzeb konsumentów. Strategie zarządzania architekturami tych systemów nie zapewniają efektywnej transformacji cyfrowej w kierunku budowy platform danych (Data Platforms (DP)); I dalej, nawet najnowsza generacja architektury DP, tzw. multimodalna architektura chmury (ang. Multimodal Cloud Architecture (MCA)) zapewniając przetwarzanie zarówno wsadowe, jak i strumieniowe oraz transformację danych w czasie rzeczywistym nie jest wydajna w zarządzaniu infrastrukturami wielkiej skali. MCA nie zapewnia integralności i dostępności danych w różnych przypadkach użycia. Odpowiedzią na brak wymaganej efektywności jest pojawienie się nowej koncepcji budowy DP w oparciu o **Architekturę Integracji Danych** (ang. Data Integration Architectures (DIA)). Według trójpoziomowej taksonomii DIA w warstwie wirtualnej wyróżniamy architektury Data

Mesh (DaMe). Data Mesh jest innowacyjnym i strategicznym podejściem DIA do zarządzania wielkimi zbiorami danych na platformach DP.

2. Strategie zarządzania Data Mesh zsynchronizują działanie w rozproszeniu i decentralizacji, pozwalając na wysoce wydajne i efektywne przetwarzanie danych. Architektura DaMe opiera się o 4 zasady: **Własności Domeny** (ang. Domain Ownership (DoOw)), **Dane jako Produkt** (ang. Data as a Product (DaPr)), **Samoobsługowa Platforma Danych** (ang. Self-Serve Data Platform (SePI)) i **Federacyjny Nadzór Obliczeniowy** (Federated Computational Governance (FeGo)). Łącznie zasady DoOw, DaPr, SePI i FeGo ustanawiają inteligentną i skalowalną architekturę DaMe, która efektywnie udostępnia dane całej organizacji/korporacji. **Zasada DoOw** w architekturze DaMe obejmuje przeniesienie odpowiedzialności za zarządzanie danymi analitycznymi (jako źródło lub odbiorca) na domeny biznesowe, dynamicznie skalowalne i spójne dla operacji biznesowych, technologii i analizy danych. Technicznie DoOw nie zakłada klasycznych interfejsów operacyjnych API tj REST, GraphQL lub gRPC, ale integruje się wraz z nimi, poprzez nowe interfejsy analityczne, dodając kolejną warstwę odpowiedzialności za domeny do analiz danych szczegółowych. **Zasada DaPr** w architekturze DaMe określa dane dziedzinowe powinny być traktowane jako produkt, nazywany kwantem danych, działa niezależnie i obejmuje elementy strukturalne służące do udostępniania danych na podstawie umów prawnych. Technicznie DaPr zapewnia, że każdy produkt danych posiada hermetyczną, wysoką samowystarczalną spójność funkcjonalną, integrując wewnętrzny kod transformacji, dane, metadane i infrastrukturę; co zwiększa ich użyteczność, wykrywalność i możliwości bezpiecznego udostępniania. Produkty danych za pośrednictwem portów danych I/O, umożliwiając ich przepływ i wykorzystanie danych w różnych domenach, do konkretnych potrzeb, za pomocą metod synchronicznych, tj. zapytania API, lub asynchronicznych, tj. subskrypcje zdarzeń. **Zasada SePI** w architekturze DaMe zapewnia efektywne zarządzanie danymi przy zdecentralizowaniu DoOw, usprawnia cykl życia DaPr i automatyzuje procesy bezpieczeństwa i zgodności w sieci połączonych zasobów danych. Technicznie SePI w architekturze DaMe jest skalownie ustrukturyzowana w 3 płaszczyznach tj.: a) infrastruktury danych (ang. Data Infrastructure Plane) zarządza podstawowymi elementami zasobów infrastruktury, b) doświadczenia produktu danych (ang. Data Product Experience Plane) upraszcza opracowywanie i wykorzystania produktów danych, c) doświadczenia siatki (ang. Mesh Experience Plane) zapewnia spójność funkcjonalność i zarządzanie w całej DaMe. **Zasada FeGo** w architekturze DaMe wprowadza ustrukturyzowane i zdecentralizowane zarządzanie interoperacyjnością danych, zgodnych z przepisami podejmowaniu decyzji i rozliczalności, maksymalizując wartości zagregowanych danych. Technicznie FeGo egzekwuje zasady w lokalnym kontekście każdego DaPr, unikając w ten sposób scentralizowanych punktów awarii lub wąskie gardła operacyjne(kontrola dostępu i szyfrowanie) oraz zapewnia ustandaryzowane interfejsy API i zarządzanie wskaźnikami docelowego poziomu usług (SLO).

3. Zarządzanie otwartymi danymi w SC4.0 (ODMSC4.0)

Zaimplementowanie pradygmatu DaMe do realiów organizacyjnych administracji miejskiej wymusza przyjęcie szeregu założeń i strategii zarządzania w projektowaniu koncepcji ODMSC4.0. Przydatne do analizy wstępnej ODMSC4.0 powinny być rozważane w kontekście zarówno **nieefektywności wdrożeń**, jak i **inżynierii danych**. Istniejąca literatura naukowa i publicystyka gminna na temat budowy i funkcjonowania ODMSC4.0 koncentruje się na osiągnięciu nadrzędnego celu - SDGs. Unijna dokumentacja techniczna i specyfikacje czy inicjatywy, tj.: "European data space for smart communities" skupiają się na narzędziach i normalizacji (tj. FIWARE foundation). Ryzyka związane z **nieefektywnością wdrożeń ODMSC4.0** finansowanych ze środków publicznych (inwestycje komunalne) muszą być

zminimalizowane, aby uniknąć oskarżeń o niegospodarność. Rosnąca liczba systemów IT oraz obszarów/punktów pomiarowych (np. IoT) generują dane (np. surowe) o zachodzących procesach, w dowolnym momencie (np. w czasie rzeczywistym), znając ich kontekst (np. status i zmienne związane) wymaga nowych modeli przetwarzania danych. **Inżynieria danych** pozwala zbudować warstwy infrastruktury ODMSC4.0 m.in. w zakresie: a) wyszukiwania danych, weryfikowanie i ich katalogowanie, b) ekstrakcji danych (ETL), niezawodnego przepływu potoków danych, oraz przechowywanie ustrukturyzowanych i nieustrukturyzowanych danych, c) jakości danych (czyszczenie, wykrywanie anomalii, uzupełnianie, i przygotowanie danych). Stąd wskazane technologie Big Data tj. Apache Hadoop oraz Apache Spark. Apache Hadoop klasy fault - tolerance to otwarta platforma programistyczna (Java) do rozproszonego dwuetapowego przetwarzania (MapReduce) wielkich zbiorów danych np. Apache Hadoop 3.4.1 (HDFS). Apache Spark to klastrowa platforma wieloetapowego przetwarzania Big Data z większością obliczeń wprost w PaO oraz interfejsy API (Scala, Python, Java i R), co daje różnice wydajnościowe sięgające nawet 100-krotności na niekorzyść platformy Hadoop (algorytmy iteracyjne lub DM i IoT). ODMSC4.0 odpowiada za skuteczność, skalowalność i bezpieczeństwo DaMe w kontekście: masy/ilości danych, różnorodności i heterogeniczności danych, jakości danych, szybkości transferów danych oraz użyteczności i zmienności danych tzw. cyklu życia danych z generowaniem alertów o występujących wyjątkach i anomaliach. Infrastruktura ODMSC4.0 w architekturze **platformy otwartych danych SC4.0** spełnia kluczowe zadanie zachowania spójnych i aktualnych zbiorów danych i ich opisów przeznaczonych dla pomyślnego wdrożenia zarówno usług i aplikacji dla obywateli, jak i ich dostępu dla grup specjalnych (tj. badaczy, deweloperów) w celach innowacyjnych zastosowań. Z perspektywy zbiorów danych otwartych ważna jest ich charakterystyka opisana przez metadane, struktury danych otwartych i danych kontekstowych, oraz potencjalne użycie. Z drugiej strony, w praktyce często brakuje szczegółowej dokumentacji zbiorów danych, co utrudnia przygotowania odpowiednich ich charakterystyk i metadanych. Brak kluczowych metadanych może być uzupełniony automatycznie lub półautomatycznie poprzez wykorzystanie **dużych modeli językowych** (ang. Large language models (LLM)). Wdrażanie LLM w ODMSC4.0 może zapewnić skuteczną strategię analizy danych strukturalnych. Stąd kolejne uściślenie tematyki rozprawy pt.: **Data Mesh for Smart Cities** dotyczy przedstawienia próby modelowania i prototypowania architektury **Inteligentnego miasta czwartej generacji (SC4.0)** poprzez wdrożenie **Platformy otwartych danych dla inteligentnych miast** (ang. Open Data platform for Smart Cities (ODSC)), uszczegółowioną dla Poznania (**ODSC [Poznań]**).

Teraz można przejść do udzielenia odpowiedzi na wcześniejsze pytanie:

2.1. Czy w recenzowanej rozprawie doktorskiej rozważany problem ma charakter naukowy?, w zakresie sprecyzowanej tematyki badawczej jako zagadnienie pt.:

Wielokontekstowa analiza badawcza koncepcji DaMe i uzasadnienie możliwości jej efektywnej adaptacji spełnienia wymagań ODSC [Poznań].

Pozytywną odpowiedź daje częściowo rozdział 4 rozprawy pt.: **Smart City data landscape using Data Mesh**. Doktorant zauważa, że trudność projektowania ODSC polega na dostosowaniu koncepcji DaMe, która wywodzi się z otoczenia korporacyjnego, do realiów organizacyjnych administracji miejskiej. To wymaga przyjęcia szeregu założeń i przemyślanej strategii wdrożenia koncepcji ODSC[Poznań]. W rozdziale 4 przedstawiono aktualny stan wiedzy o ODSC, podsumowano doświadczenia zdobyte podczas wdrożenia ODSC[Poznań] z perspektywy socjologicznej i technicznej. Wyniki te opublikowane w artykule autorstwa: M. Krystek, C. Mazurek, M. Morzy, and J. Pukacki. *Introducing data mesh paradigm for smart city platforms design*. *Proceedings of the 56th Hawaii International Conference*

on System Sciences, pp.6885, 01 2023. Tu aspekt naukowy występuje w analizie zastosowania koncepcji DaMe w realizacji miejskich DP, w celu zminimalizowania ryzyka nieefektywnych wdrożeń, skupiając się bardziej na procesach przetwarzania danych niż na samych technologiach. Dokonano głębokiej analizy literaturowej w zakresie ODSC (rozdz. 4.1). W kolejnym podrozdziale 4.2 (Smart City Data Platform pillars) omówiono ODSC w aspekcie technologii, użyteczności, zrównoważonego rozwoju i zarządzania. Aspekt technologiczny ODSC rozumiany jako zdolność do budowy integrującej, skalowalnej i adaptacyjnej ODSC szybkiego przetwarzania różnych typów danych. Aspekt użyteczności podkreśla zdolność dostosowania się ODSC do różnych i nietypowych scenariuszy danych i przyszłych potrzeb użytkowników oraz zarządzania dynamicznym zakresem zainteresowanych stron. Aspekt zrównoważonego rozwoju dotyczy zdolności ODSC do obsługi nowych typów danych, dostarczania innowacyjnych produktów, dopasowania się do różnych scenariuszy i nowych interesariuszy. Aspekt zarządzania obejmuje wybór między scentralizowanym a zdecentralizowanym zarządzaniem, ma kluczowe znaczenie dla utrzymania jakości danych i dostosowania do potrzeb wszystkich użytkowników). Doktorant w podrozdziale 4.3 (Basic tenets of DataMesh in the context of Smart City (DaMeSC) przedstawił podstawowe założenia DaMeSC na 3 poziomach: socjologiczny, technologiczny i infrastruktury. Kolejno przedstawiono wnioski z wdrożenia DaMeSC wymiarach: Partnerstwo, Małe kroki, Zrozumienie, Wartość danych, Promocja i upowszechnianie, Scenariusz jazdy, Zarządzanie zadaniami. Oceny jednak wymaga rozdział 4.5 (Technical aspects). Projektowanie ODSC[Poznań] ma uwzględniać określone przez miasto wymagania biznesowe i niezbędne funkcjonalności. Proces wdrożenia został podzielony na etapy, zgodnie z priorytetami Poznania. Środowiska docelowe wdrożenia ODSC[Poznań] wykorzystuje zasoby obliczeniowe PCSS w modelu PaaS (ang. Platform as a Service), przy jasnym przekazaniu obowiązków jej utrzymania i zapewnienia ciągłości infrastruktury dedykowanym administratorom. Zasoby obliczeniowe, bazodanowe i sieciowe są traktowane jako usługi, które mogą być swobodnie wykorzystywane w ramach podzielonej puli, zmniejszając potrzebę angażowania administracji do minimum. Stąd wybór Samoobsługowej Platformy Danych, (ang Self-Serve Data Platform (SSDP)) sformułowana w modelu DaMe. Natomiast infrastruktura w modelu PaaS pozwala optymalizować zasoby poprzez alokację odpowiedniej ich liczby, w sposób iteracyjny i rozszerzony w procesie projektowania ODSC. SSDP zapewnić ma automatyzację wszystkich procesów wytwarzania oprogramowania wg zasady "domen danych" oraz usług dostępnych w ramach ODSC. Do tego pomocny jest proces CI/CD (ang. Continuous Integration, Delivery, and Deployment) z szablonami konfiguracji, które ponownie użyte w wielu aplikacjach do osiągnięcia pełnej elastyczności w realizacji potrzeb "domen danych", przy jednoczesnym zminimalizowaniu liczby godzin pracy. Kolejnym dobrym rozwiązaniem jest ODSC oparte o ustandaryzowaną metodykę MLOps (ang. Machine Learning Operations) bazującą na automatyzacji procesów uczenia maszynowego (ML) wraz z operacyjnym wdrożeniem i utrzymaniem produkcji DevOps (ang. Develop Operations). Doktorant uzasadnia także wybór koncepcji jednokrotnego logowania SSO (ang. Single Sign-On) w ODSC jako usługę odpowiedzialną za uwierzytelnianie oraz autoryzację dostępu użytkowników do poszczególnych usług (oparte o standardy OpenID Connect oraz OAuth2). To podejście obniża koszty wdrożenia ODSC przez ustandaryzowane procedury, uproszczone zarządzanie aplikacjami i ich kodami oraz przechowywanie plików.

Teraz odpowiedzmy na kolejne pytanie:

2.2 Czy w recenzowanej rozprawie doktorskiej rozważany problem ma charakter naukowy?, w zakresie sprecyzowanej tematyki badawczej jako zagadnienie pt.:

Analiza badawczo- eksperymentalna użycia dużych modeli językowych w podnoszeniu efektywności i użyteczności platformy otwartych danych miejskich Poznania (ODSC [Poznań]).

Wdrażanie dużych modeli językowych (LLM) w ODMSC4.0 może zapewnić strategię automatycznego rozpoznawania struktury i właściwości danych. LLM są bardzo skuteczne w analizowaniu danych strukturalnych (obsługiwane przez konwencjonalne techniki statystyczne i uczenie maszynowe). Pozytywną odpowiedzią na pytanie 2.2 zawiera rozdział 5 rozprawy pt.: **Large Language Models in the service of Open Data platform for Smart City**. W kontekście ODSC[Poznań] jednym z najważniejszych celów LLM jest automatyzacja generowania metadanych związanych z samymi produktami danych. Ewaluacja generatywnej sztucznej inteligencji (ang. Generative AI) przekształca analizę danych i tworzenie treści przez różnych liderów w tej dziedzinie tj.: OpenAI GPT (GPT-2, GPT-3, GPT-4), LaMDA, OPT-175B, Hugging Face's BLOOM, Jurassic-1 Jumbo..., najnowsze to Gemini 2, DeepMind i LLaMA 3). LLM przewyższają zdolności "ludzkiego" zrozumienia semantyki zbioru danych w zadaniach niestandardowych, identyfikują wzorce i generują przewidywania zakorzenione w logice i spójności tematycznej. Zdolność LLM do radzenia sobie z różnymi tekstami jest dobrze udokumentowana, natomiast mniej zbadana jest jej zdolność w posługiwaniu się danymi tabelarycznymi, powszechnych w różnych scenariuszach DPSC. Doktorant ocenia zrozumienie przez LLM tabelarycznych struktur danych i ich potencjału dla dużych zbiorów danych. Badał także zdolności LLM do dostarczania agregacji danych, polecań intuicyjnych nazw funkcji i sugestii odpowiednich zestawów danych dla określonych zapytań lub potrzeb biznesowych. Sekcje 5.3, 5.4, 5.5 przedstawiają eksperymenty jaką LLM mają zdolność do obsługi określone scenariuszy użytkowników ODSC[Poznań];

Sekcja 5.3 Eksperyment 1: Zidentyfikuj i oznacz pojęcia

Eksperyment 1 pozwala ocenić, czy LLM może dokładnie identyfikować i oznaczać pojęcia reprezentowane w bazie danych z małej próbki surowych danych, bez wskazówek o znaczeniu atrybutów, ich semantyki i relacji. Użyto pseudonimizację dla maskowania nazw atrybutów z czteroznakowymi kodami Base64 wygenerowanymi przez narzędzie OpenSSL, aby zmniejszyć prawdopodobieństwo halucynacji w modelu. W eksperymencie, GPT-4-turbo został zaprezentowany ze zbiorem danych składającym się z 50 losowo wybranych krotek z tabeli, wraz z minimalistycznym opisem, który nadaje podstawowy kontekst tabeli (użyto interfejsu API OpenAI, przy użyciu biblioteki programowania langchain). LLM został przetestowany na czterech zestawach danych z ODSC[Poznań]: a) Groby - informacje o grobach znajdujących się na cmentarzu miejskim, b) Punkty adresowe - punkty adresowe, ich lokalizacje, kategorie i przynależność do różnych obszarów funkcjonalnych miasta, c) Stacje rowerowe - lokalizacja i aktualny stan zajętości stacji rowerów miejskich, d) MPKStops - informacje o przystankach komunikacji miejskiej, ich lokalizacji, kategoriach, typach pojazdów i możliwych przesiadkach. Dla każdego zestawu danych utworzono zapytanie zawierające wymagania formatu wyjściowego i zawartości. Przykładowe zapytanie: *Przeanalizuj następujący dokument JSON, w którym każda funkcja reprezentuje opis pojedynczego punktu adresowego w mieście Poznań, Polska. Analizuj wartości i relacje między wszystkimi atrybutami. Szczegółowe wyniki dla poszczególnych tabel przedstawiono w Tabeli 5.2 (Groby), Tabeli 5.3 (Przystanki MPK), Tabeli 5.4 (Stacje Rowerowe) oraz Tabeli 5.5 (Punkty Adresowe).*

Sekcja 5.4 Eksperyment 2: generowanie opisów podobnych do ludzkich.

Doktorant przetestował zdolności modelu LLM do wygenerowania informacyjnego opisu zbioru danych zawierający podsumowanie, znaczenia poszczególnych atrybutów, określone obszary domen oraz potencjalne pola użycia w przyszłych analizach. Ekspert stworzył opisy 18 zestawów danych, które zostały wygenerowane

przez/lub dla służb miejskich w Poznaniu (tab.5.6) oraz zbiór 15 zapytań sparametryzowanych, aby umożliwić dynamiczne odwoływanie się do odpowiednich zestawów danych. Odpowiedzi na zapytania były jednocześnie przechowywane w pliku tekstowym w celu walidacji oraz w wektorowej bazie danych do wykorzystania w kolejnym eksperymencie. W tej sekcji dołączono wiele odpowiedzi LLM na wybrane zapytania dla zestawu danych Stacje rowerowe.

Sekcja 5.5. Eksperyment 3: wsparcie w podejmowaniu decyzji. W eksperymencie 2 wygenerowano szczegółowe opisy 18 zestawów danych z indywidualnymi atrybutami i ogólną charakterystykę. Tu celem było sprawdzenie, czy na podstawie wygenerowanych informacji LLM może wesprzeć analityka w doborze najlepszych zbiorów danych i nadaje się do wykorzystania w danym scenariuszu badawczym. Stworzono 10 zapytań systemowych, z których każde reprezentowało potencjalne zagadnienie badawcze. Każde zapytanie zostało podzielone na trzy sekcje: a) określenie roli użytkownika, b) szczegółowy opis zadania analitycznego, c) zakres informacji źródłowych wykorzystanych do udzielenia odpowiedzi, oraz formatu samej odpowiedzi. Przykładowe zapytanie: *Jako statystyk chciałbym przeanalizować częstotliwość występowania imion i nazwisk mieszkańców miasta. Interesuje mnie identyfikowanie i śledzenie ogólnych trendów w długim okresie czasu oraz tego, jak imiona i nazwy użytkowników funkcjonują w przestrzeni publicznej. Zaproponuj najistotniejsze zestawy danych z kontekstu, które mogą wesprzeć to zadanie analityczne. Wymień nazwy zestawów danych i uzasadnij wybór, wyjaśniając, dlaczego określone zestawy danych są odpowiednie dla mojego przypadku użycia.*

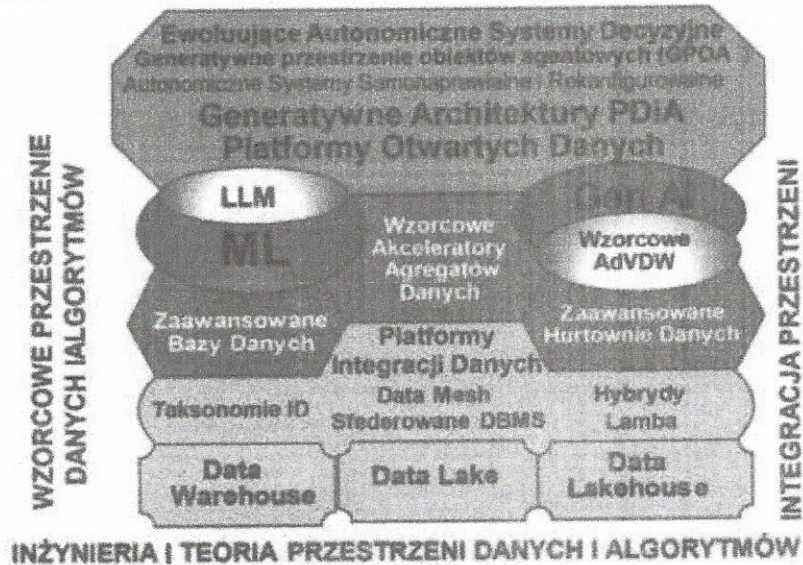
Odpowiedzi na to zapytanie znajdujemy w sekcji 5.5. Doktorant przedstawił także wyniki z eksperymentu w sekcji 5.6 oraz w artykule: M. Krystek, M. Basinski, M. Morzy, C.Mazurek: **Managing Data Platforms for Smart Cities Using Large Language Models**. ISD 2024.

4 **Wkład naukowo-wdrożeniowy Doktoranta opisywany w rozprawie**

Wkład naukowo – wdrożeniowy doktoranta Marcina Krystka należy postrzegać jako znaczący udział we współpracy Zespołu trzech Instytucji tj.: Politechniki Poznańskiej, Poznańskiego Centrum Superkomputerowo-Sieciowego (PCSS) afiliowanego przy ICHB PAN oraz Urzędu Miasta Poznania. Stopień zaangażowania i szczegółowy wkład naukowo – wdrożeniowy doktoranta przy tak złożonym projekcie najlepiej może określić jego promotor dr hab. inż. Mikołaj Morzy, prof. PP Wydziału Informatyki i Telekomunikacji Politechniki Poznańskiej oraz opiekun merytoryczny dr inż. Cezary Mazurek z PCSS'u. Recenzent tego doktoratu wdrożeniowego ocenia przedstawione wyniki w rozprawie traktując je jako znaczący wkład doktoranta w projekcie ODSC [Poznań]. Motywacją badań była konieczność dostarczenia miastu Poznań środowiska rozwojowego do zrealizowania obowiązku dostępu do otwartych danych gminnych zgodnie z oficjalną strategią inteligentnego miasta Poznania. Pozwoliło to na zdefiniowanie głównych wyzwań budowy platformy danych, która odpowiada na potrzeby UM Poznań, gdzie urzędnicy miejscy przetwarzają i wyciągają wnioski z danych, którymi dysponują. Doktorant opracował fundamentalne założenia, które stały się podstawą dostosowania koncepcji DaMe do uwarunkowań ODSC i zdefiniowania ODSC [Poznań] o charakterze zarówno technologicznym, jak i społeczno-kulturowym. W recenzji niniejszej w pkt. 2 pt.: **Cel rozprawy** i w pkt.3 pt.: **Problem badawczy - jego charakter i znaczenie** odpowiedziałem pozytywnie na pytania:

2.1. Czy w recenzowanej rozprawie doktorskiej rozważany problem ma charakter naukowy?, w zakresie sprecyzowanej tematyki badawczej jako zagadnienie pt.:

	<p style="text-align: center;">Wielokontekstowa analiza badawcza koncepcji DaMe i uzasadnienie możliwości jej efektywnej adaptacji spełnienia wymagań ODSC [Poznań].</p> <p>2.2 Czy w recenzowanej rozprawie doktorskiej rozważany problem ma charakter naukowy?, w zakresie sprecyzowanej tematyki badawczej jako zagadnienie pt.: Analiza badawczo-eksperymentalna użycia dużych modeli językowych w podnoszeniu efektywności i użyteczności platformy otwartych danych miejskich Poznania (ODSC [Poznań]).</p> <p>Podsumowując ww. tematyka recenzowanej rozprawy ma średni charakter naukowy.</p> <p>Wydzieloną ważną tematyką wdrożeniową doktoranta w rozprawie opisać można inaczej jako wdrożenie prototypu systemu zarządzania danymi miejskimi wykorzystującego wielkie modele językowe.</p> <p>Celem tego badania doktoranta było przyspieszenie rozwoju aplikacji ODSC[Poznań], generalnie usprawniające życie w inteligentnych miastach. Wstępne wyniki sugerują, że LLM mogą służyć jako dostępne, opłacalne narzędzia programistyczne w celu uproszczenia interakcji ze złożonością ODSC[Poznań]. Praktyczne stosowanie LLM wymaga opracowania nowego interfejsu konwersacyjnego dla ODSC[Poznań], koncentrując się na testowaniu jego użyteczności i wydajności z użytkownikami końcowymi w rzeczywistych scenariuszach. Zobaczmy ww. problematykę szerzej jako Nową koncepcję Inteligentnych Miast Przyszłości (Smart Cities of the Future SC(F)).</p>
5	<p>Architektura Inteligentnych Miast Przyszłości</p> <p>Nowa koncepcja Inteligentnych Miast Przyszłości SC(F)), w praktyce platforma (Open Data platform for Smart Cities of the Future (ODSC(F)))</p> <p>Recenzent uważa, że w kontekście ODSC(F) wymagane jest opracowanie nowej</p> <p>4) Strategii Zarządzanie Przestrzenią Danych i Algorytmów w SC(F) (ang. Data and Algorithm Space Management in SC(F) (DASM SC(F)))</p> <p>Analityka danych (nadmiarowo ang. Data Sciences) bazuje na osiągnięciach inżynierii danych, wykorzystując dane do predykcji, decyzji i obserwowalności. Klasyczną analitykę danych tworzą dwie warstwy and ID tj.: a) analizy, metryki, segmenty, zagregowane danych i funkcje zorientowane oraz b) generatywna sztuczna inteligencja i głębokie uczenie. W wykładzie pod linkiem <u>Wykład Profesorski - prof. dr hab. inż Marcin Gorawski</u> przedstawiono nową strategię zarządzania Przestrzenią Danych i Algorytmów (PDIA).</p> <p>Poniżej prezentuję ramowo rozwiązania dla DASM SC(F).</p>



6 Grupa pytań rozstrzygających

Niejasności i zbyt ogólne opisy tematyki w niniejszej rozprawie doktorskiej powinny być wyjaśnione przez Doktoranta w 3 płaszczyznach **DASM SC(F)** jak niżej.

6.1. Platformy Integracji Danych (PID)

Obok różnych platform integracji danych wybrano architekturę Data Mesh (DeMe), gdzie dane są definiowane jako produkt (**Dane jako Produkt** (ang. Data as a Product (**DaPr**)) zarządzany przez dedykowany Zespół, odpowiedzialny zarówno za jego dostępność i jakość danych, jak i bezpieczeństwo, w środowisku rozproszonym SC. Stąd koncepcja „produktu danych” ma kluczowe znaczenie dla architektury DaMeSC4.0.

W rozprawie nie zdefiniowano pojęcia „produktu danych”, stąd pytania:

- A. Jakie konkretne przykłady DaPr występują w SC4.0?
- B. W jaki sposób te produkty danych są projektowane i wdrażane DaMeSC4.0?

W rozprawie podkreśla się znaczenie i konieczność uwzględnienia społeczno-technicznych i organizacyjnych wyzwań w platformach danych (DP).

- C. W jaki sposób architektura DaMe szczegółowo odpowiada na te wyzwania w kontekście DaMeSC4.0?
- D. Jak łącznie zasady DoOw, DaPr, SePI i FeGo ustanawiają inteligentną i skalowalną architekturę DaMeSC4.0?
- E. Jak zasady DaMe prowadzą do lepszych rezultatów w porównaniu z tradycyjnymi podejściami w środowisku SC4.0?

Doktorant koncentrując się na technologicznych i organizacyjnych aspektach wdrażania DaMeSC, wskazuje na możliwe liczne implikacje społeczne i etyczne.

- F. Jakie są potencjalne ryzyka i wyzwania związane z zachowaniem prywatności danych, bezpieczeństwem i zapobieganiem dyskryminacji mieszkańców?
- G. W jaki sposób można zapewnić odpowiedzialne i etyczne wykorzystanie danych SC4.0?

W rozprawie brak jest opisu rozwiązań PID dla sieci brzegowych określonych przez wielość urządzeń, realizujących np. monitorowanie miejsc parkingowych z binarnym odczytem informacji (miejsce zajęte/wolne), do zaawansowanych urządzeń

pomiarowych z możliwościami wykonywania akcji (np. sterowanie oświetleniem miejskim). Problemem w sieciach IoT (ang. Internet of Things) w DaMeSC jest rozbudowa urządzeń w lokalnych sieciach Mesh w narzędzie zapewniające dynamicznie zarządzalne, automatycznie konfigurowalne środowisko komunikacji bezprzewodowej. W celu rozwiązania tego problemu, można zastosować rozwiązania znane z Semantic Web i zaadoptować je w środowisku IoT.

H. Jak zaprojektować dynamiczne środowisko komunikacji bezprzewodowej dla infrastruktury IoT dla SC4.0 miasta Poznania?

I. Jak opracować semantyczny moduł rozszerzający opis urządzeń IoT na potrzeby DaMeSC4.0 ?

Na poziomie zarządzania SC4.0 można postawić pytania:

J. Jak zarządzać strategiami SC4.0 na przykładzie Systemu Inteligentnego Transportu?

K. Jak zaprojektować optymalne zarządzanie strategiami DaMeSC4.0 ?

6.2. Platformy Otwartych Danych

W recenzowanej pracy doktorskiej przedstawiono ogólnie wyzwania związane z prototypowaniem i wdrażaniem architektury ODSC[Poznań] (link: [Poznań-Otwarte Dane](#)) wykazując dużą złożoność organizacyjną i tematyczną ODMSC4.0.

Stąd pytania:

L. Jak zoptymalizować zarządzanie otwartymi danymi w SC4.0 - ODMSC4.0 ?

M. Jakie kluczowe kroki mogą podjąć miasta, aby wspierać kulturę opartą na otwartych danych i współuczestniczyć w ODMSC4.0 różnych departamentów UM i interesariuszami?

N. W jaki sposób można poprawić dostępność danych, jakość i świeżość danych oraz proces wspomaganie podejmowania decyzji w Poznaniu bazując na ODSC[Poznań]?

Podkreślam, że przedstawiane w rozprawie wdrożenie ODSC[Poznań] ma charakter ramowo - opisowy. Przygotowano prototyp ODSC[Poznań] zarządzany przez ODMSC4.0 ze wsparciem środowiskiem wielkich modeli językowych LLMs. Przejście z procesu prototypowania na pełne wdrożenie ODSC[Poznań] należy koniecznie wyznaczyć metryki dla takiego wdrożenia. Zasadne są odpowiedzi na pytania:

O. Kiedy nastąpi zwrot z inwestycji, wg wskaźnika ROI (ang. Return on Investment) z ODSC[Poznań]?

P. Jakie są kluczowe wskaźniki efektywności wg wskaźników KPI (ang. Key Performance Indicators) dla ODSC[Poznań]?

6.3. Generatywne Platformy Inteligentnych Miast Przyszłości

W rozprawie doktorskiej przedstawiono interesujące eksperymenty i ich analizę z użyciem wielkich modeli językowych (LLMs). Jednak, zważywszy na przedstawioną w niniejszej recenzji, w punkcie 5, **Nową koncepcją Inteligentnych Miast Przyszłości SC(F)**, w pierwszej fazie rozwoju ODSC[Poznań] można postawić pytanie:

Q. Jak należy pogłębić badania nad LLM-ami dla ODSC[Poznań] w kontekście agentowej sztucznej inteligencji?

R. Jakie korzyści lub wyzwania mogą przynieść lokalne modele językowe i potencjalna synergia z LLMs?

Obok różnych architektur i platform integracji danych dla wybranej koncepcji ODSC(4.0) i ODSC[Poznań] jest niewystarczające zarządzanie otwartymi danymi ODMSC4.0. W nowej SC(F) **Generatywne Platformy Inteligentnych Miast Przyszłości** będą tworzyć Strategię Zarządzania Przestrzenią Danych i Algorytmów

w SC(F) (ang. Data and Algorithm Space Management in SC(F) (DASMSC(F)).
Kolejne pytanie:

- S. Jak Doktorant postrzega możliwość rozszerzenia swoich badań nad użyciem DASMSC(F) do projektowania ODSC(F)[Poznań]?

7 Dane o dorobku naukowym Doktoranta

Poniżej tabelarycznie zebrano dane o dorobku naukowym Doktoranta.

Rodzaj publikacji	Liczba
Artykuły w czasopiśmie	2
Monografie naukowe	1
Rozdziały w monografiach	1
Publikacje konferencyjne	7
Pozostałe publikacje	1
Razem publikacji	12

Publikacje wyróżnione	Liczba	Pozycja poniżej
Artykuły w czasopiśmie z IF	2	[1] , [2]
Publikacje konferencyjne (liczone 140 pkt. MNiSzW)	2	[3] , [4]
Publikacje konferencyjne (pierwszy współautor)	3	[3],[4],[5],[6]

- 1) Andrés García-Silva, José Manuel Gómez-Pérez, Raúl Palma, Marcin Krystek, Simone Mantovani, Federica Foglini, Valentina Grande, Francesco De Leo, Stefano Salvi, Elisa Trasatti, Vito Romaniello, Mirko Albani, Cristiano Silvagni, Rosemarie Leone, Fulvio Marelli, Sergio Albani, Michele Lazzarini, Hazel J. Napier, Ilkay Altintas: **Enabling FAIR research in Earth Science through research objects**. Future Gener. Comput. Syst. 98: 550-564 (2019).
- 2) Sławomir Bak, Marcin Krystek, Krzysztof Kurowski, Ariel Oleksiak, Wojciech Piatek, Jan Węglarz: **GSSIM - A tool for distributed computing experiments**. Sci. Program.19(4): 231-251 (2011)
- 3) Marcin Krystek, Mikołaj Basinski, Mikołaj Morzy, Cezary Mazurek: **Managing Data Platforms for Smart Cities Using Large Language Models**. ISD 2024
- 4) Marcin Krystek, Cezary Mazurek, Mikołaj Morzy, Juliusz Pukacki: **Introducing Data Mesh Paradigm for Smart City Platforms Design**. HICSS 2023: 6885-6892
- 5) M. Krystek, C. Mazurek, M. Morzy, and J. Pukacki. **Introducing data mesh paradigm for smart city platforms design**. Proceedings of the 56th Hawaii International Conference on System Sciences, page 6885, 01 2023. URL <https://hdl.handle.net/10125/103468>. (cited on pages 13 and 33)
- 6) Marcin Krystek, Krzysztof Kurowski, Ariel Oleksiak, Krzysztof Rządca: **Comparison Of Centralized And Decentralized Scheduling Algorithms Using GSSIM Simulation Environment**. CoreGRID Integration Workshop 2008: 185-19

Liczba publikacji	Wartość
Liczba publikacji	Wartość
Według bazy DBLP	11
Według bazy Web of Science	8
Według bazy Scopus	9
Liczba cytowań ogółem	Wartość
Według bazy Web of Science	40
Według bazy Scopus	58
Indeks Hirsch	Wartość
Według bazy Web of Science	3
Według bazy Scopus	4

Ocena dorobku naukowego Doktoranta

Cytowane w rozprawie artykuły zostały napisane starannie, precyzyjnym językiem, w związku tym nie wnoszę do nich uwag. Godne uwagi są dwa artykuły [1] i [2] w czasopiśmie z IF (Impact Factor), dwie publikacje konferencyjne [3] i [4] z punktacją 140 pkt. MNIŠzW oraz cztery publikacje konferencyjne [3],[4],[5],[6] z Doktorantem jako z pierwszym współautorem. W omówionych w dysertacji publikacjach zakładam, że głównym pomysłodawcą kluczowych rozwiązań i metod jest mgr. inż. Marcin Krystek. Szkoda, że nie dołączono, potwierdzających powyższe, oświadczeń autorów publikacji i umieszczonych jako załączniki rozprawy – co byłoby pożądaną, dobrą praktyką. Osiągnięcia naukowe mgr inż. Marcina Krystka mieszczą się w zakresie wnioskowanej dziedziny nauki inżynierjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja. Zebrane dane o dorobku naukowym doktoranta są wystarczające do wystawienia oceny pozytywnej.

8 Ocena rozprawy doktorskiej

Przedstawiona rozprawa została przygotowana w języku angielskim, jako praca doktorska o charakterze wdrożeniowym. Rozprawa składa się z siedmiu rozdziałów, bibliografii, spisu tabel. Ogólna objętość rozprawy to 97 strony o spisie treści jak niżej.

Contents „Data Mesh for Smart Cities	4.2.4 Governance .39
Acknowledgments vii	4.3 Basic tenets of Data Mesh in the context of Smart City . 40
Abstract ix	4.3.1 Sociological level . 40
Streszczenie xi	4.3.2 Technology level . 42
1 Introduction 1	4.3.3 Infrastructure level . 43
1.1 Institutional synergy 1	4.4 Lessons learned from implementing Data Mesh in a Smart City . 44
1.1.1 About PSNC . 1	4.4.1 Partnership . . 45
1.1.2 About Poznań . 2	4.4.2 Small steps . 45
1.1.3 Role of partners . 5	4.4.3 Understanding . 46
1.2 Data context . 6	4.4.4 Data value . 47
1.2.1 Poznań Smart City infrastructures and systems .6	4.4.5 Promotion and dissemination. 47
1.2.2 Poznań Open Data . 8	4.4.6 Driving scenario . 47
1.3 City data in the era of generative AI revolution . 9	4.4.7 Task management . 48
1.4 Thesis . 10	4.5 Technical aspects .48
1.5 Research framework . 10	5 Large Language Models in the service of Open Data platform for Smart City 53
2 Smart City 13	5.1 Motivation . 53
2.1 Definition of Smart City . 13	5.2 Large Language Models in structured data analysis . 55
2.2 Poznań Smart City strategy . 14	5.3 Experiment 1: identify and label concepts . 56
2.3 Poznań Smart City data assets .6	5.4 Experiment 2: generate human-like descriptions . 57
2.4 Challenges in implementing Open Data platform for Smart City . 18	5.5 Experiment 3: support in decision making . 64
3 Data Mesh 21	5.6 Experiment conclusions . 65

3.1 Data platform architectures . 21	6 ODSC Deployment 69
3.2 Design pattern weaknesses . 22	6.1 Deployment approach in Poznań . 69
3.3 Data Mesh principles . 24	6.1.1 System layers . 69
3.4 Data Mesh approach for data platform design . 27	6.1.2 Data management . 71
3.4.1 Data domains and data sharing . 28	6.2 Poznań Open Data portal . 72
3.4.2 Data product . 29	6.2.1 Data product catalog . 72
3.4.3 Self-served data platform . 30	6.2.2 Conversational interface for open data . 75
3.4.4 Federated computational governance . 32	6.3 Data analytics . 77
4 Smart City data landscape using Data Mesh . 33	6.4 ODSC Concierge . 77
4.1 Urban data platforms in literature . 33	6.4.1 Toolset . 79
4.2 Smart City Data Platform pillars . 35	6.4.2 Dataset reverse engineering . 81
4.2.1 Technology . 35	6.4.3 Data characterisation . 82
4.2.2 Usability . 37	7 Conclusion . 87
4.2.3 Sustainability . 38	List of Tables . 89
4.2.3 Sustainability . 38	Bibliography 93 - 97

Ogólny stan wiedzy kandydata w zakresie Informatyki na podstawie oceny wybranych rozdziałów rozprawy

Rozdział 3: Data Mesh. Doktorant powołuje się na pracę: Zhamak Dehghani. Data Mesh Delivering Data-Driven Value at Scale. OREILLY, 2022, gdzie pojęcie DaMe opisano szczegółowo. Opisano elementy usługi DaMe. Ich dostosowanie do warunków i specyficznych wymagań środowiska miejskiego będą stanowić podstawę, które będziemy rozwijać ideę DPSC. W sekcji 3.1 podsumowano poprzednie podejścia w celu projektowania DPSC. Przedstawiono ich problemy i związane z nimi słabości w sekcji 3.2. W sekcji 3.3 przedstawiono definicję siatki danych i najważniejsze aspekty tej architektury przedstawiono w sekcji 3.4

Ocena rozdziału 3: Poprawny opis (str. 21–32) ale niejednoznaczny kontekst, pobieżny wobec uwag niniejszej recenzji w punkcie 3 pt. **Problem badawczy - jego charakter i znaczenie.** Rozdział potwierdza średni stan wiedzy kandydata w zakresie Informatyki

Rozdział 4: Smart City data landscape using Data Mesh. Rozdział przedstawia zakres wyzwań i zagrożeń związanych z projektowaniem ODSC. W sekcji 4.1 przedstawiono aktualny stan wiedzy w tym zakresie. Sekcja 4.2 opisuje fundamenty ODSC. Założenia dotyczące wdrożenia DaMe w kontekście Smart City przedstawiono w sekcji 4.3. Podsumowano doświadczenia z wdrożenia w Poznaniu zarówno z perspektywy socjologicznej (sek. 4.4), jak i z perspektywy technicznej (sek. 4.5).

Ocena rozdziału 4: Ocena i analizę tego rozdziału (str. 33–52) przeprowadzono w niniejszej recenzji w punkcie 3 pt. **Problem badawczy - jego charakter i znaczenie** jako odpowiedź na pytanie pt.: **Wielokontekstowa analiza badawcza koncepcji DaMe i uzasadnienie możliwości jej efektywnej adaptacji spełnienia wymagań ODSC [Poznań].** Doktorant nie uwzględnił badań nad Architekturę Integracji Danych w budowie DPSC. Rozdział 4 potwierdza wysoki stan wiedzy kandydata w tym zakresie Informatyki.

Rozdział 5: Large Language Models in the service of Open Data platform for Smart City. W sekcji 5.1 uzasadniono automatyzację powtarzalnych zadań DP dużej skali. Sekcja 5.2 zawiera krótki przegląd ogólnej zdolności LLM, a sekcje 5.3, 5.4, 5.5 przedstawiają szczegółowo eksperymenty sprawdzające, czy LLM obsługują określone scenariusze użytkowników ODSC. Końcowe wyniki podsumowano w punkcie 5.6, wcześniej opublikowane w M. Krystek, C. Mazurek, M. Morzy, and M. Basiński: Managing Data Platforms for Smart Cities Using Large Language Models. Proc. of the 32th International Conf. on Information Systems Development, 08 2024.

Ocena rozdziału 5. Ocena i analizę tego rozdziału ((str.53-67)) przeprowadzono w niniejszej recenzji w punkcie 3 pt. **Problem badawczy - jego charakter i znaczenie** jako odpowiedź na pytanie pt.: **Analiza badawczo- eksperymentalna użycia dużych**

modeli językowych w podnoszeniu efektywności i użyteczności platformy otwartych danych miejskich Poznania (ODSC [Poznań]) jest pozytywna.

Rozdział 6: ODSC Deployment. W sekcji 6.1 przedstawiono koncepcję wdrożenia i kluczowe technologie wdrożenie ODSC [Poznań]. W sekcji 6.2 przedstawiono I fazę wdrożenia skupioną na zbiorach danych katalogowanych. Oprogramowanie udostępniania i analizy danych to sekcja 6.3. Ostatnia sekcja 6.4 poświęcona jest prototypom narzędzi stworzonych do automatyzacji procesów generowania wartościowych opisów dla zbiorów danych. Przeprowadzone eksperymenty wykazały, że LLM potrafi skutecznie pomagać użytkownikom w realizacji powtarzalnych zadań i procesów ODSC [Poznań].

Ocena rozdziału 6. Wysoko oceniam ten rozdział (str. 69 – 85) za jakość i wiedzę praktyczną i eksperymentalną doktoranta w dyscyplinie Informatyka techniczna i telekomunikacja.

Podsumowując ocenę rozprawy doktorskiej:

- 1.) Doktorant wykazał się odpowiednią wiedzę badawczo-eksperymentalną w dyscyplinie Informatyka techniczna i telekomunikacja.
- 2.) Doktorant uzyskał wartościowe doświadczenie oparte o eksperymenty z LLM w ODSC [Poznań].
- 3.) Wskazuje duże praktyczne znaczenie wdrożonych rozwiązań ODSC [Poznań] i stwarzające potencjalne możliwości ich dalszego użycia.
- 4.) ponadto:
 - o stwierdzenia zawarte w rozprawie są godne zaufania,
 - o dobra jakość prototypowego oprogramowania ODSC [Poznań].
 - o rozprawa zawiera prawie kompletną bibliografią,
 - o rozprawa napisana poprawnym i zrozumiałym językiem angielskim z zachowaniem staranności.

Uwagi krytyczne

1. Doktorant uważa, że możliwości modeli generatywnych wydają się obecnie prawie nie ograniczone, a badania naukowe nad nimi fascynujące. Wg mnie przyszłością sztucznej inteligencji są modele Generatywnych Przestrzeni Danych i Algorytmów (modele GPDIA). Modele GPDIA uczą się, jak dane, algorytmy i architektury PDiA są dystrybuowane, poszukują architektur specjalizowanych, schematów i wzorców, a następnie generują nowe architektury IT korzystając z tej wiedzy. Wstępną wiedzę na temat modeli GPDIA przedstawiłem w punkcie 5 niniejszej recenzji.
2. Doktorant nie wykazał posiadania patentów, wniosków racjonalizatorskich, czy też zgłoszonych wniosków patentowych.
3. Brak oświadczeń współautorów.
4. Brak spisu nagród i wyróżnień autora.

Oceniając rozprawę pragnę podkreślić, iż została ona wykonana na wysokim poziomie wdrożeniowym o dużym znaczeniu praktycznym i jest wartościowa z punktu widzenia pogłębienia wiedzy w dyscyplinie Informatyka Techniczna i Telekomunikacja. Wnosi ona także oryginalny wkład eksperymentalny LLM i potwierdza wysokie kwalifikacje Doktoranta, które znajdują także potwierdzenie na drodze uzyskania zadawalających odpowiedzi na pytania zawarte w niniejszej recenzji, w punkcie 6.1.w grupie pytań rozstrzygających.

8

Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami) ¹ moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak X)

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Ponadto, biorąc pod uwagę powyższe analizy rekomenduję wyróżnienie rozprawy doktorskiej.

M. Górowska
Podpis

M. Górowska

(podpis Recenzenta)

¹ <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>