

Dr hab. inż. Krzysztof Pietruszewicz, prof. ZUT
Katedra Automatyki i Robotyki
Wydział Elektryczny
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Al. Piastów 17, 70-310 Szczecin

Szczecin, dnia 30 marca 2025 r.

Recenzja rozprawy doktorskiej mgr. inż. Marcina Krystka

Tytuł rozprawy: **Data Mesh for Smart Cities**
Autor rozprawy: **mgr inż. Marcin Krystek**
Promotor rozprawy: **dr hab. inż. Mikołaj Morzy, Prof. PP**
Opiekun merytoryczny: **dr inż. Cezary Mazurek**
Dziedzina: **nauki inżynieryjno-techniczne**
Dyscyplina: **informatyka techniczna i telekomunikacja**

1. Wstęp i podstawa prawna sporządzenia recenzji

Recenzowana przeze mnie rozprawa doktorska porusza istotne, zarówno z naukowego jak i przede wszystkim praktycznego a w tym wdrożeniowego punktu widzenia, zagadnienie opracowania innowacyjnej architektury Data Mesh w kontekście inteligentnych miast (Smart Cities), oferując nowe podejście do zarządzania złożonymi systemami danych miejskich. Kluczowym jest tutaj opracowanie i implementacja platformy otwartych danych dla Smart Cities (ODSC) w Poznaniu, mającej na celu ułatwienie dostępu do danych miejskich i ich ponowne wykorzystanie. Istotnym elementem pracy jest również badanie możliwości wykorzystania dużych modeli językowych sztucznej inteligencji LLM (ang. *Large Language Models*) do automatyzacji zadań związanych z przygotowaniem i interpretacją danych na tej platformie.

Recenzja została przygotowana w odpowiedzi na uchwałę Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Poznańskiej z dnia 26 listopada 2024 r. w sprawie powołania recenzentów mgr. inż. Marcina Krystka pt. „Data Mesh for Smart Cities”, przekazaną przez Prodziekana Wydziału Informatyki i Telekomunikacji dr hab. inż. Szymona Wilka, prof. PP, pismem z dnia 27 listopada 2024 r. (**pismo DIiT-63-27/2024**).

Promotorem rozprawy doktorskiej jest **dr hab. inż. Mikołaj Morzy, Prof. PP**, natomiast opiekunem merytorycznym jest **dr inż. Cezary Mazurek**.

2. Ocena formalna, charakterystyka pracy. Poprawność

Recenzowana praca doktorska obejmuje łącznie 97 stron, numerowanych od pierwszej strony rozdziału „Introduction” do ostatniej strony bibliografii. Strony poprzedzające rozdział pierwszy (od strony tytułowej do spisu treści) są numerowane rzymskimi cyframi. Praca zawiera dedykację, po której następują podziękowania (str. VII). Główna część pracy jest podzielona na 7 rozdziałów: 1. Introduction (strony 1-12); 2. Smart City (strony 13-20); 3. Data Mesh (strony 21-32); 4. Smart City data landscape using Data Mesh (strony 33-52); 5. Large Language Models in

the service of Open Data platform for Smart City (strony 53-68); 6. ODSC Deployment (strony 69-86) oraz 7. Conclusion (strony 87-88). Bibliografia (Bibliography) znajduje się na stronach 93-97 i zawiera 53 pozycje literatury. W bibliografii znajdują się dwie pozycje, w których doktorant (pozycje 12, 13) jest współautorem:

1. **M. Krystek**, C. Mazurek, M. Morzy, 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> ;
2. **M. Krystek**, C. Mazurek, M. Morzy, M. Basiński. *Managing Data Platforms for Smart Cities Using Large Language Models*. Proceedings of the 32th International Conference on Information Systems Development, 08 2024

Obie cytowane prace mają ścisły związek z recenzowaną rozprawą.

Bibliografia zawiera, poza artykułami i materiałami konferencyjnymi, również książki, dokumenty i raporty instytucjonalne, materiały nt. inicjatyw europejskich jak również odwołania do stron internetowych (strony instytucji i organizacji, artykuły on-line, portal danych).

Praca napisana została w języku angielskim. Zawiera również spis rysunków (str. 89-90) oraz tabel (91-92).

Na uwagę zasługuje fakt realizacji pracy przez pracownika Poznańskiego Centrum Superkomputerowo-Sieciowego, znaczącego w kraju i za granicą ośrodka badawczo-rozwojowego oraz lidera technologicznego w obszarach związanych z dyscypliną naukową rozprawy. Współpraca PCSS z Miastem Poznań w zakresie Smart City nadaje rozprawie praktyczny kontekst i potencjalnie zwiększa jej wiarygodność. W pracy przyjęto określoną metodologię badawczą: opisano problem badawczy, ograniczenia, przeprowadzono eksperymenty mające na celu weryfikację możliwości wykorzystania LLM. Opisy eksperymentów są charakterystyczne dla prac naukowych dyscypliny naukowej informatyka techniczna i telekomunikacja.

Pewnym zaskoczeniem była dla mnie informacja o wykorzystaniu dużych modeli językowych (LLM) również w kontekście weryfikacji poprawności językowej, spójności oraz stylistyki tekstu rozprawy. Autor wskazuje: „Duże modele językowe zostały wykorzystane do weryfikacji poprawności językowej, zachowania spójności tekstu i poprawy jakości stylistycznej tekstu.” W mojej ocenie, choć intencja ta może być zrozumiała z perspektywy eksploracji możliwości nowoczesnych narzędzi, warto rozważyć, czy zastosowanie takich rozwiązań w obszarze redakcyjnym nie osłabia przekazu dotyczącego samodzielności warsztatu badawczego. Uważam, że merytoryczna wartość rozprawy jest wystarczająco wysoka, aby obejść się bez takich zabiegów – tym bardziej, że doskonalenie formy językowej stanowi integralny element rozwoju kompetencji naukowych doktoranta.

Praca przygotowana jest bardzo starannie, dowody prezentowane są w sposób bardzo wyrazisty, eksperymenty opisane precyzyjnie. Trudno jest ocenić jakość prototypowego oprogramowania, wytworzonego w ramach doktoratu – wierzę, że opinie użytkowników

końcowych są gromadzone na bieżąco i tym samym doświadczenia użytkowników optymalizowane.

Poniżej przedstawiam pokrótce zawartość poszczególnych rozdziałów pracy.

Rozdział 1. Wprowadzenie

W rozdziale tym przedstawiono kontekst pracy doktorskiej, która koncentruje się na zastosowaniu architektury Data Mesh w inteligentnych miastach, w ramach programu doktoratu wdrożeniowego łączącego badania akademickie z praktyką biznesową. Opisano synergii instytucjonalną pomiędzy Poznańskim Centrum Superkomputerowo-Sieciowym (PCSS) a miastem Poznań, ich role w inicjatywach Smart City oraz kontekst danych miejskich, w tym inicjatywy Open Data. Wprowadzono również koncepcję platformy Otwartych Danych dla Inteligentnych Miast (ODSC), która jest głównym przedmiotem badań, oraz przedstawiono tezę i ramy badawcze pracy.

Rozdział 2. Smart City

Rozdział ten poświęcony jest koncepcji Smart City, rozpoczynając od analizy różnych definicji i poszukiwania wspólnych elementów. Następnie przedstawiono strategię Smart City przyjętą przez miasto Poznań, opartą na zasadach zrównoważonego rozwoju, innowacji technologicznych, rozwoju gospodarczego, integracji społecznej i cyfryzacji, z silnym naciskiem na wykorzystanie danych. Na koniec omówiono wyzwania związane z implementacją platformy Open Data dla Smart City w specyficznym kontekście organizacyjnym i technicznym Poznania.

Rozdział 3. Data Mesh

W tym rozdziale wprowadzono koncepcję architektury Data Mesh jako alternatywnego podejścia do budowy platform danych, które ma na celu przezwycięzenie ograniczenia tradycyjnego podejścia. Omówiono słabości wzorców projektowych platform danych, które często nie uwzględniają wyzwań społeczno-technicznych i organizacyjnych. Następnie przedstawiono podstawowe zasady koncepcji Data Mesh, a także opisano podejście do projektowania platform danych.

Rozdział 4. Smart City data landscape using Data Mesh

Rozdział ten analizuje dane inteligentnego miasta z perspektywy Data Mesh, podkreślając, że wdrożenie platformy danych w takim środowisku wiąże się z szerokim spektrum wyzwań, głównie o charakterze kulturowym i organizacyjnym. Przedstawiono przebieg implementacji koncepcji ODSC w Poznaniu, wraz z przyjętymi założeniami i zdobytymi doświadczeniami, wskazując na różnice między środowiskiem korporacyjnym a administracją miejską. Omówiono podstawowe założenia Data Mesh w kontekście Smart City na poziomie socjologicznym, technologicznym i infrastrukturalnym.

Rozdział 5. Large Language Models in the service of Open Data platform for Smart City

Rozdział 5 zawiera szczegółowe omówienie wyników przeprowadzonych eksperymentów, poparte tabelami prezentującymi metryki skuteczności modeli LLM w różnych zadaniach. Z praktycznego punktu widzenia najbardziej wartościowe wydają się sekcje 5.3, 5.4 oraz 5.5, opisujące przeprowadzone eksperymenty. Eksperymenty te badały różne aspekty wykorzystania LLM w kontekście platformy otwartych danych dla inteligentnego miasta, od automatycznego opisywania danych, przez generowanie zrozumiałych dla człowieka wyjaśnień, po wsparcie w

procesie wyszukiwania i doboru danych do konkretnych analiz. W sekcji 5.6 dokonano analizy danych empirycznych wynikających z przeprowadzonych eksperymentów.

Rozdział 6. ODSC Deployment

Praca opisuje wdrożenie prototypowych rozwiązań, takich jak portal Otwartych Danych Miasta Poznania (sekcja 6.2) z jego różnymi funkcjonalnościami (katalog zasobów, wyszukiwanie, wizualizacje, interfejs konwersacyjny) oraz narzędzia ODSC Concierge do automatyzacji generowania metadanych (sekcja 6.4). Przedstawienie w rozprawie działających prototypów i ich konkretnych funkcji demonstruje praktyczną realizację koncepcji i weryfikuje jej wykonalność. Fakt, że portal Otwarte Dane (<https://www.poznan.pl/opendata>) został wdrożony produkcyjnie, dodatkowo wzmacnia wartość recenzowanej rozprawy.

Za najważniejsze osiągnięcie rozprawy doktorskiej uważam rozdziały: 5 oraz 6. Zawierają one bowiem wyniki eksperymentów (5) potwierdzających prezentowanych w rozdziałach 3 i 4 koncepcji oraz wyniki wdrożenia (6). Uważam, że tematyka poruszana w pracy jest bardzo aktualna i ma olbrzymie znaczenie praktyczne. Proponowane rozwiązanie wpisuje się w trendy technologiczne w zakresie nowoczesnego projektowania złożonych systemów informatycznych.

3. Problem badawczy i jego znaczenie

Tezy pracy skupiają się wokół trzech najważniejszych problemów badawczych:

1. Zbadanie i adaptacja architektury nazwanej Data Mesh do zarządzania złożonymi danymi w kontekście Inteligentnego Miasta. Praca koncentruje się na zrozumieniu, jak zasady Data Mesh mogą być zastosowane w środowisku miejskim w celu poprawy zarządzania i wykorzystania danych;
2. Zaprojektowanie i wdrożenie koncepcji platformy Otwartych Danych dla Inteligentnych Miast (ODSC) opartej na adaptowanej tutaj architekturze Data Mesh. Celem pracy jest stworzenie nowego modelu platformy danych, która ułatwi udostępnianie i ponowne wykorzystanie danych miejskich, z uwzględnieniem specyficznych potrzeb i wyzwań środowisk miejskich, takich jak Poznań;
3. Wykorzystanie dużych modeli językowych sztucznej inteligencji (LLM) do automatyzacji i ulepszenia procesów związanych z otwartymi danymi na wdrożonej platformie ODSC. Autor rozprawy bada, w jaki sposób LLM mogą wspierać użytkowników platformy, na przykład w odkrywaniu, interpretowaniu i opisywaniu zbiorów danych, co w dłuższej perspektywie przyczynić się ma do zwiększenia ich użyteczności i dostępności.

Najważniejszym problemem rozważanym w rozprawie jest możliwość wdrożenia architektury Data Mesh w kontekście inteligentnych miast (Smart Cities). Rozprawa bada, w jaki sposób to innowacyjne podejście do zarządzania i wykorzystywania złożonych systemów danych może sprostać wyzwaniom społeczno-technicznym i organizacyjnym, które często są pomijane przez tradycyjne podejścia do projektowania platform danych.

Koncepcja „Data Mesh”, opisywana w pracy, przedstawiana jest jako zdecentralizowane podejście do udostępniania, uzyskiwania dostępu i zarządzania danymi analitycznymi

w złożonych środowiskach o dużej skali, wewnątrz lub między organizacjami. Koncepcja ta ma na celu poprawę relacji między danymi domenowymi i wyeliminowanie problemów spowalniających istniejące procesy oraz ograniczających zdolność do adaptacji.

Rozprawa koncentruje się na specyficznych potrzebach i wyzwaniach miasta Poznań, jak również na ocenie praktycznych skutków wykorzystania nowoczesnych technologii. W pracy wskazuje się duże modele językowe (LLM) jako szansę na polepszenie funkcjonalności platform danych.

Głównym celem rozprawy, jak wskazuje autor, jest zbadanie i zaproponowanie koncepcji platformy danych dla inteligentnych miast (Smart Cities) opartej na architekturze Data Mesh. Celem tym jest stworzenie innowacyjnego podejścia do zarządzania i wykorzystywania złożonych systemów danych w środowisku miejskim. Autor stawia sobie za cel praktyczny opracowanie i wdrożenie innowacyjnej platformy danych dla inteligentnego miasta, łączącej zasady Data Mesh z możliwościami dużych modeli językowych (LLM), dla efektywnego zarządzania danymi miejskimi, udostępniania ich oraz wspierania zrównoważonego rozwoju Poznania.

W mojej ocenie postawiony cel rozprawy został osiągnięty. Praca, pomimo swojego wysoce aplikacyjnego charakteru posiada liczne walory naukowe.

Praca podejmuje konkretny problem badawczy, jakim jest efektywne zarządzanie danymi w kontekście Smart Cities. Autor proponuje nowatorskie rozwiązanie w postaci platformy ODSC opartej na architekturze Data Mesh.

Doktorant nie tylko opisuje istniejące rozwiązania i problemy, ale także prowadzi eksperymenty mające na celu weryfikację hipotez dotyczących zastosowania nowoczesnych technologii (LLM) do poprawy funkcjonalności platform danych. Rozprawa krytycznie analizuje dotychczasowe podejścia do budowy platform danych dla miast i na tej podstawie formułuje własną koncepcję (ODSC). Celem nowego podejścia jest przewyższenie ograniczeń i lepsze wspieranie zrównoważonego rozwoju miast. Tym samym, praca dąży do pogłębienia wiedzy w dziedzinie miejskich platform danych i ułatwienia ich skutecznego wdrażania w praktyce.

Uważam, że zdefiniowane cele pracy są istotne i spełniają wymagania stawiane rozprawom doktorskim. Według mnie, wszystkie cele pracy zostały w rozprawie osiągnięte.

4. Wkład autora w rozwój dyscypliny

Najważniejszym wkładem autora w rozprawie jest koncepcja Platformy Otwartych Danych dla Inteligentnych Miast (ODSC) oraz jej wdrożenie w kontekście Poznania. Autor zaadaptował zasady architektury Data Mesh do specyfiki środowiska miejskiego, promując zdecentralizowane zarządzanie danymi i traktowanie ich jako produktów danych przez poszczególne jednostki administracji.

Innowacyjnym aspektem pracy jest wykorzystanie technologii LLM do ulepszenia funkcjonalności ODSC. Przeprowadzone eksperymenty wykazały potencjał LLM w identyfikacji pojęć, generowaniu opisów danych oraz wsparciu decyzji dotyczących wyboru zbiorów danych.

W wyniku tych badań powstały prototypy narzędzi (ODSC Concierge) automatyzujące generowanie metadanych.

Pierwszym etapem wdrożenia ODSC jest portal Otwartych Danych Miasta Poznania, służący jako katalog produktów danych. Rozprawa podkreśla, że sukces platformy zależy nie tylko od technologii, ale także od zmian organizacyjnych i kulturowych w administracji miejskiej. Wkład autora łączy nowatorską koncepcję zarządzania danymi, zastosowanie AI oraz praktyczną implementację, uwzględniającą wyzwania organizacyjne.

Moja opinia najważniejszych dokonań autora rozprawy jest zbieżna z deklarowanym przez doktoranta. Rozprawa wnosi istotny wkład w dyscyplinę „informatyka techniczna i telekomunikacja” poprzez zaproponowanie nowatorskiej architektury platformy danych dla inteligentnych miast (ODSC), innowacyjne zastosowanie LLM do zarządzania danymi, udokumentowanie praktycznego procesu wdrożenia zidentyfikowanymi wyzwaniami i wnioskami, oraz integrację nowoczesnych technologii w spójne i funkcjonalne rozwiązanie. Te dokonania poszerzają wiedzę i praktykę w zakresie projektowania i implementacji zaawansowanych systemów informatycznych i platform danych w kontekście miejskim.

Stwierdzam, że doktorant dysponuje wymaganym do prowadzenia badań naukowych zasobem wiedzy oraz umiejętnościami, jak również kompetencjami z zakresu dyscypliny naukowej **informatyka techniczna i telekomunikacja**.

5. Wiedza kandydata w obszarze rozprawy

Kilka rozdziałów i sekcji w rozprawie omawia istniejący stan wiedzy, tym samym potwierdzając ogólny stan wiedzy kandydata w zakresie dyscypliny naukowej „informatyka techniczna i telekomunikacja”:

1. Sekcja 2.1 przedstawia definicje oraz ogólną koncepcję Smart City;
2. Sekcja 2.4 omawia wyzwania implementacji otwartych platform danych dla potrzeb Inteligentnych Miast;
3. Rozdział 3 jako całość, opisujący koncepcję i zasady podejścia Data Mesh;
4. Sekcja 4.1 prezentuje przegląd istniejących rozwiązań i podejść do budowy miejskich platform danych;
5. W sekcji 4.3 autor rozprawy łączy omówioną wcześniej koncepcję Data Mesh ze specyfiką Smart City;
6. Sekcja 5.2 wprowadza czytelnika w problematykę wykorzystania LLM w analizie danych strukturalnych;
7. W sekcji 6.1.1 opisano warstwy proponowanego modelu architektury wdrażanej w ramach badań platformy (patrz rysunek 6.1); doktorant wykazał się znajomością standardowych wzorców architektonicznych w zakresie wdrażania technologii.

Podsumowując, rozdziały 2, 3, 4 i 5, a także wybrane sekcje rozdziału 6 omawiają istniejący stan wiedzy w różnych obszarach informatyki, takich jak:

1. architektury systemów informatycznych,
2. zagadnienia zarządzanie danymi,

3. problematykę baz danych,
4. inżynieria oprogramowania,
5. sztuczna inteligencja oraz
6. zastosowania informatyki w kontekście miejskim.

Poprzez odniesienia do literatury przedmiotu i dyskusję istniejących rozwiązań, autor rozprawy potwierdza swoje solidne fundamenty w zakresie wiedzy informatycznej.

Zakres przygotowanej bibliografii jest dość szeroki – cytowane pozycje literatury odnoszą się do kluczowych koncepcji poruszanych w rozprawie. Autor uwzględnił również dwie współautorskie publikacje, co należy uznać za istotne z punktu widzenia realizacji celów pracy. Warto jednak zauważyć, że analiza literatury mogłaby zostać pogłębiona, zwłaszcza w obszarze architektur danych. Biorąc pod uwagę, że rozprawa dotyczy również zagadnień związanych ze sztuczną inteligencją, wskazane byłoby odniesienie się do fundamentalnych koncepcji leżących u podstaw wykorzystywanej technologii LLM – mogłoby to stanowić cenne uzupełnienie dla czytelnika zainteresowanego tym aspektem pracy.

W literaturze rozprawy zabrakło w mojej ocenie prac poruszających zagadnienie metodyk wytwarzania oprogramowania i ogólniej zarządzania projektami IT w tak złożonych problemach jak ten, który rozwiązuje w rozprawie doktorant. Powyższe uwagi nie umniejszają mojej wysokiej oceny oraz swobody, z jaką autor porusza się po źródłach literaturowych związanych z pracą.

6. Podsumowanie. Wnioski końcowe

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) moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

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

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?

Zdecydowanie
TAK

Raczej TAK

Trudno
powiedzieć

Raczej NIE

Zdecydowanie
NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

Zdecydowanie
TAK

Raczej TAK

Trudno
powiedzieć

Raczej NIE

Zdecydowanie
NIE

Po zapoznaniu się z pracą doktorską pt. „Data Mesh for Smart Cities” stwierdzam, że:

przedstawiona rozprawa spełnia

właściwe ustawowe wymagania stawiane rozprawom doktorskim oraz mieści się w dyscyplinie informatyka techniczna i telekomunikacja i w związku z powyższym

wniosuję o przyjęcie rozprawy doktorskiej oraz dopuszczenie

mgr. inż. Marcina Krystka do publicznej obrony w wyżej wspomnianej dyscyplinie.

Prętnica.

Podpis