

Warszawa, 6 maja 2024 r.

Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie
 Wydział Matematyczno-Przyrodniczy. Szkoła Nauk Ścisłych
 ul. Wóycickiego 1/3, 01-938 Warszawa



**ZESTAWIENIE PROPONOWANYCH TEMATÓW PRAC MAGISTERSKICH
 NA ROK AKADEMICKI 2024/25 NA KIERUNKU INFORMATYKA**

opiekun pracy	proponowany temat pracy	uwagi, zakres pracy, literatura
dr inż. Jakub Gąsior	Analiza i symulacja wybranych protokołów routingu w sieciach sensorowych	
dr inż. Jakub Gąsior	Opracowanie i implementacja wybranego algorytmu eksploracji danych z wykorzystaniem technologii CUDA	
dr inż. Jakub Gąsior	Lasy losowe - ocena jakości prognostycznej na podstawie analizy sentymentu tweetów	
dr inż. Jakub Gąsior	Uczenie głębokie w problemie wykrywania fałszywych wiadomości (fake news detection)	
dr inż. Robert Kłopotek	Budowa metod generatywnych terenów lub poziomów w grach (GANs lub GPT)	Sprawdzanie możliwości generowania terenu/mapy lub poziomu gry przy użyciu sieci GANs lub GPT. Weryfikacja grywalności lub zgodności z rzeczywistym terenem. Literatura: 1. J. Kowalski, R. Miernik, P. Pytlik, M. Pawlikowski, K. Piecuch and J. Sękowski, "Strategic Features and Terrain Generation for Balanced Heroes of Might and Magic III Maps," 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, Netherlands, 2018, pp. 1-8, doi: 10.1109/CIG.2018.8490430. 2. Spick, Ryan John and Walker, James Alfred (Accepted: 2019) Realistic and Textured Terrain Generation using GANs. In: The 16th ACM SIGGRAPH European Conference on Visual Media Production (CVMP). ACM . https://eprints.whiterose.ac.uk/153088/1/Real_world_Textured_terrain_generation_using_GANs_1.pdf 3. Sudhakaran, S., Gonz'alez-Duque, M., Glanois, C., Freiburger, M., Najarro, E., & Risi, S. (2023). MarioGPT: Open-Ended Text2Level Generation through Large Language Models. ArXiv, abs/2302.05981. https://arxiv.org/pdf/2302.05981.pdf
dr inż. Robert Kłopotek	Badanie jakości modeli generatywnych dla dialogów w budowie NPC	Sprawdzanie możliwości tworzenia postaci niegrywanych NPC w grze RPG lub podobnej. Wykorzystanie modeli z rodziny GPT: GPT-2, GPT-J, GPT-JT lub GPT-Neo. Jakość odpowiedzi NPC zostanie oceniona za pomocą popularnych miar, takich jak ROUGE, METEOR czy Semantic Answer Similarity. Literatura: 1. Judith van Stegeren and Jakub Myśliwiec. 2021. Fine-tuning GPT-2 on annotated RPG quests for NPC dialogue generation. In Proceedings of the 16th International Conference on the Foundations of Digital Games (FDG '21). Association for Computing Machinery, New York, NY, USA, Article 2, 1–8. https://doi.org/10.1145/3472538.3472595 2. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv, abs/1910.01108.

opiekun pracy	proponowany temat pracy	uwagi, zakres pracy, literatura
dr inż. Robert Kłopotek	Analiza podejść advanced RAG w zadaniu odpowiedzi na pytania do pliku PDF	<p>Analiza podejść typu "advanced RAG" (Retriever-Augmented Generation) w zadaniu odpowiedzi na pytania z plików PDF koncentruje się na ulepszeniach i modyfikacjach modelu RAG, który łączy techniki wyszukiwania informacji (retrieval) z generacją tekstu, aby efektywnie odpowiadać na postawione pytania bazując na zawartości dokumentu PDF. RAG dzieli proces na dwa główne etapy: odzyskiwanie odpowiednich fragmentów tekstu (retriever - część bazy wiedzy) i generowanie odpowiedzi na ich podstawie (generator - LLM). W kontekście plików PDF, dodatkowym wyzwaniem jest konwersja dokumentów do formatu, który można efektywnie przeszukiwać i przetwarzać przez model.</p> <p>Celem pracy jest stworzenie platformy testowej rozwiązującej problem in-context learning, gdzie wejściem jest plik PDF, który podlega przekonwertowaniu na formę zdatną do wyszukiwania (np. baza danych wektorowa pgvector, chromadb czy baza danych grafowa neo4j). Na podstawie tego pliku tworzona jest baza wiedzy (ang. knowledge base) z podziałem na kawałki (ang. chunks) odpowiedniej wielkości (do przetestowania wielkość - chunk size optimization). Do bazy wiedzy wiedzy można dodać różnego rodzaju metadane, aby poprawić jakość wyszukiwania. Do pliku należy pozyskać pytania z odpowiedziami (np. można wygenerować), aby zweryfikować jakość odpowiedzi udzielnych przez system. Silnikiem odpowiedzi dla tej platformy testowej jest wybrany duży model językowy (LLM). W celach badawczych można się wzorować lub użyć komponentów platform open source, jak h2oGPT czy LlamaIndex. Opcjonalnym elementem jest interfejs graficzny np. gradio.</p> <p>Literatura:</p> <ol style="list-style-type: none"> 1. Building RAG from Scratch (Open-source only!) https://docs.llamaindex.ai/en/latest/examples/low_level/oss_ingestion_retrieval.html# 2. A Cheat Sheet and Some Recipes For Building Advanced RAG https://www.llamaindex.ai/blog/a-cheat-sheet-and-some-recipes-for-building-advanced-rag-803a9d94c41b 3. Advanced RAG Techniques: an Illustrated Overview https://pub.towardsai.net/advanced-rag-techniques-an-illustrated-overview-04d193d8fec6 4. Implementing Advanced Retrieval RAG Strategies With Neo4j https://neo4j.com/developer-blog/advanced-rag-strategies-neo4j/ 5. pgvector: Embeddings and vector similarity https://supabase.com/docs/guides/database/extensions/pgvector 6. All You Need to Know about In-Context Learning https://towardsdatascience.com/all-you-need-to-know-about-in-context-learning-55bde1180610 7. How to Create a Chatbot with Gradio https://www.gradio.app/guides/creating-a-chatbot-fast
dr inż. Robert Kłopotek	Analiza podejść wielo-agentowych do odpowiadania na pytania do zbioru dokumentów	<p>Celem pracy jest stworzenie systemu bazy danych dokumentów (np. baza danych wektorowa pgvector, chromadb czy baza danych grafowa neo4j), do których można zadawać pytania (np. baza danych artykułów na określony temat). Do bazy wiedzy wiedzy można dodać różnego rodzaju metadane, aby poprawić jakość wyszukiwania. Do pliku należy pozyskać pytania z odpowiedziami (np. można wygenerować), aby zweryfikować jakość odpowiedzi udzielnych przez system. Silnikiem odpowiedzi dla tej platformy testowej jest wybrany duży model językowy (LLM).</p> <p>Chatbot/agent powinien zwracać najbardziej poprawną odpowiedź ze źródłem swojej wiedzy (który dokument lub jego część). W celach badawczych można się wzorować lub użyć komponentów platform open source, jak CrewAI, LangGraph czy LlamaIndex. Opcjonalnym elementem jest interfejs graficzny np. gradio.</p> <p>Literatura:</p> <ol style="list-style-type: none"> 1. Multi-Document Agents https://docs.llamaindex.ai/en/stable/examples/agent/multi_document_agents/ 2. Crew AI — your own minions https://medium.com/@csakash03/crew-ai-you-own-minions-9b8596ce3da3 3. Multi-Agent System — Crew.AI https://abvijaykumar.medium.com/multi-agent-system-crew-ai-3773356b8c3e 4. How to Create a Chatbot with Gradio https://www.gradio.app/guides/creating-a-chatbot-fast 5. LangGraph: Multi-Agent Workflows https://blog.langchain.dev/langgraph-multi-agent-workflows/ 6. MultiPDF Chat App https://github.com/alejandro-ao/ask-multiple-pdfs

opiekun pracy	proponowany temat pracy	uwagi, zakres pracy, literatura
dr inż. Robert Kłopotek	Rozwiązywanie problemów obliczeniowych przy użyciu agentów i dużych modeli językowych	<p>Celem pracy jest stworzenie platformy testowej rozwiązującej wybrane problemy obliczeniowe przy użyciu agentów i dużych modeli językowych. Użytkownik wprowadza materiały źródłowe (dokumenty, pliki csv) na podstawie której system dokonuje analizy (np. analiza finansowa, analiza kursów akcji). W rozwiązaniu problemu należy stworzyć kilku agentów z różnymi rolami, np. interpretowania zapytania użytkownika w języku naturalnym, generowania zapytania do bazy danych (SQL) lub bazy wiedzy, obróbka danych w Pythonie (generowanie kodu) i zwrócenie odpowiedzi, np. wykres z opisem (wygenerowanie kodu i zapis do pliku wykresu). Do zadań interpretacyjnych należy użyć dużych modeli językowych (np. GPT4 turbo, LLAMA3, LLAMA3 instruct), które będą w stanie zrealizować powierzone im zadania.</p> <p>W celach badawczych można się wzorować lub użyć komponentów platform open source, jak CrewAI, LangGraph czy Llamaindex. Opcjonalnym elementem jest interfejs graficzny np. gradio.</p> <p>Literatura:</p> <ol style="list-style-type: none"> 1. AI accelerates problem-solving in complex scenarios https://news.mit.edu/2023/ai-accelerates-problem-solving-complex-scenarios-1205 2. Multi-Agent Conversations Using AutoGen to Analyze Stock Price and Generate Chart https://djajafer.medium.com/multi-agent-conversations-using-autogen-to-analyze-stock-price-and-generate-chart-beb04866e31a 3. Multi-Agent System — Crew.AI https://abvijaykumar.medium.com/multi-agent-system-crew-ai-3773356b8c3e 4. How to Create a Chatbot with Gradio https://www.gradio.app/guides/creating-a-chatbot-fast 5. LangGraph: Multi-Agent Workflows https://blog.langchain.dev/langgraph-multi-agent-workflows/ 6. Awesome LLM-based Multi-Agents Papers https://github.com/taichengguo/LLM_MultiAgents_Survey_Papers
dr Tomasz Kulpa	Analiza stabilności w stochastycznym modelu optymalizacji rojem cząsteczek	
dr hab. Mirosław Kurkowski, prof. uczelni	Protokoły obliczeń wielostronnych bez zaufanej trzeciej strony	W literaturze znaleźć można kilka przykładów protokołów sieciowych wykonujących obliczenia wielostronne realizujące cele bezpieczeństwa bez zaufanej trzeciej strony. W pracy planuje się opisać jeden lub dwa takie protokoły oraz zasymulować ich działanie w wirtualnej rzeczywistości.
dr hab. Mirosław Kurkowski, prof. uczelni	Nowoczesne szyfry kryptografii lekkiej i ich kryptoanaliza	Proponuje się wykonanie eksperymentów obliczeniowych w zakresie bazującej na SAT kryptoanalizie szyfrów kryptografii lekkiej. Praca zawierać będzie bezpośrednie kodowanie szyfru do formuł booleowskich oraz wykonanie kryptoanalizy z tekstem jawnym i szyfrogramem. Do obliczeń zastosowane będą SAT solvery.
dr Artur Mikitiuk	Implementacja nieformalnej semantyki formalizmów logicznych ze szczególnym uwzględnieniem formalizmów używanych w answer-set programming	

opiekun pracy	proponowany temat pracy	uwagi, zakres pracy, literatura
dr hab. inż. Wojciech Stecz, prof. uczelni	Algorytm harmonogramowania projektu informatycznego uwzględniający ograniczenia dostępności zasobów	Projekt i implementacja algorytmu harmonogramowania zadań w projekcie informatycznym, który uwzględni dostępność zasobów oraz konieczność wykonania zadań w określonych oknach czasowych. Do realizacji zaprojektowanego algorytmu wykorzystany zostanie solver CPLEX CP bazujący na technice Constraint Programming firmy IBM (dostępny dla studentów i wykorzystywany na uczelni). Dla określonego rodzaju projektu informatycznego określone zostaną najważniejsze ograniczenia, jakie należy spełnić. Dyplomant opracowuje model zadania optymalizacji bazując na dostępnej literaturze przedmiotu. Dopuszczalne jest wykorzystanie modelu zaprezentowanego w bazach naukowych z e-źródeł dostępnych na UKSW (Springer lub Scopus). W takim przypadku opisany w artykule model zostanie przeniesiony na język modelowania OPL CPLEX a osiągnięte wyniki porównane z wynikami z wybranego artykułu. Celem pracy jest samodzielne opracowanie modelu dla potrzeb rozwiązania zadania harmonogramowania projektu IT oraz jego integracja z wybranym środowiskiem oprogramowania, preferowane C++ z MVS.
dr hab. inż. Wojciech Stecz, prof. uczelni	Wyznaczanie trasy lotu drona rozpoznawczego w oparciu o dostępną sieć połączeń w terenie z uwzględnieniem ograniczeń drona i czasu rozpoznania obiektów	Zadanie polega na opracowaniu modelu działania drona w terenie opisanym siecią połączeń w rozumieniu teorii grafów i sieci. Na tej sieci określone są miejsca rozpoznania obiektów wraz z wagami / premią za rozpoznanie obiektu przez drona. Dyplomant opracowuje model działania drona w postaci zadania programowania mieszanego całkowitoliczbowego a do rozwiązania wykorzystuje środowisko IBM CPLEX (dostępne dla studentów). W pracy wymagane jest porównanie opracowanej metody z algorytmami heurystycznymi. Niezbędne narzędzia i literatura są dostępne. Wymagana jest integracja rozwiązania z aplikacją napisaną w C++ lub Java (CPLEX ma dobrze opracowane interfejsy).
dr hab. inż. Wojciech Stecz, prof. uczelni	Wyznaczanie tras lotu roju dronów za pomocą algorytmu genetycznego z uwzględnieniem ograniczeń na zasięg łączności dronów	Zadanie polega na opracowaniu modelu działania kilku dronów działających w sąsiedztwie opisanym siecią połączeń w rozumieniu teorii grafów i sieci. Na sieci określone są miejsca rozpoznania obiektów wraz z wagami za rozpoznanie obiektów. Dyplomant opracowuje model działania drona w postaci algorytmu genetycznego. Dopuszczalnym jest wykorzystanie jednego z algorytmów, jaki został opisany w literaturze, po uzgodnieniu z promotorem. W takim przypadku Dyplomant będzie projektował algorytm na bazie wskazówek z artykułu, mając na uwadze wszystkie ograniczenia zadania wymienione w artykule. W pracy wymagane jest porównanie opracowanej metody z algorytmami heurystycznymi. Preferowanym językiem implementacji algorytmu jest C++ (język można zmienić).
dr hab. inż. Wojciech Stecz, prof. uczelni	Modelowanie dynamiki systemów na bazie modeli dynamicznych epidemii typu SEIR	Zadanie polega na zapoznaniu się ze sposobem modelowania dynamiki systemów, wykorzystywanym szeroko w modelowaniu procesów biznesowych od lat 60-siątych ubiegłego wieku. Opracowany model dynamiki opisać powinien jeden z przypadków epidemii (tzw. modele SEIR), które są bardzo dokładnie i szeroko pokazywane w literaturze. Do modelowania dynamiki epidemii wykorzystane zostaną równania różniczkowe i ich implementacja w języku C++ z wykorzystaniem dostępnych bibliotek numerycznych. Zadaniem Dyplomanta będzie ułożenie modelu wybranej epidemii i implementacja algorytmu rozwiązującego problem minimalizacji zakażeń w przypadku podjęcia działań sanitarnych. Wymagana literatura zawierająca model numeryczne i wskazówki do implementacji zostanie przekazana Dyplomantowi. Praca ma na celu wprowadzenie w zagadnienia wykorzystywane w konsultingu biznesowym, gdzie modelowanie dynamiczne jest szeroko stosowane.
dr hab. William Steingartner	Modelowanie semantyczne języka konkatenacyjnego/kompozycyjnego.	
dr hab. William Steingartner	Modelowanie semantyczne wybranego języka dziedzinowego.	
dr Piotr Śliwka	Modelowanie odwróconego kredytu hipotecznego	

opiekun pracy	proponowany temat pracy	uwagi, zakres pracy, literatura
dr Piotr Śliwka	Modelowanie cen nieruchomości z wykorzystaniem metod filtrowania	
dr Piotr Śliwka	Modelowanie i prognozowanie współczynników śmiertelności	
dr Piotr Śliwka	Modelowanie i prognozowanie natężenia ruchu w sieciach optycznych typu DWDM z wykorzystaniem metod probabilistyczno-statystycznych	
dr Piotr Śliwka	Charakterystyka i zastosowanie algorytmów MCMC do szacowania parametrów modeli	
dr hab. inż. Krzysztof Trojanowski, prof. uczelni	Planowanie trajektorii ruchu zespołu dronów maksymalizującej liczbę znalezionych ofiar na obszarze dotkniętym katastrofą, gdzie położenie ofiar nie jest z góry znane	
dr hab. inż. Krzysztof Trojanowski, prof. uczelni	Modelowanie heurystyczne mapy obszaru dotkniętego katastrofą na potrzeby znajdowania przez zespół dronów jak największej liczby ofiar w pierwszych minutach poszukiwań	