ZIQI(OLIVIA) WANG

ziqi.wang0908@rutgers.edu | +1-732-322-3515
| https://github.com/Chaplain0908 | Rutgers University, Piscataway, NJ 08854-8019

Research Interests

- **Reasoning, Agentic RAG, Deep Search, Graph Model**

Education

- **Rutgers University, New Brunswick, NJ, United State** *01/2025 - 12/2026(expected)*
  Master of Science in Computer Science , Department of Computer Science ***GPA**: 4.0/4.0*
  Core Courses: Mathematical Foundations of Data Science; Introduction to Artificial Intelligence; Database Systems for Data Science; Machine Learning; Economics and Computation; Brain Inspired Computing
- **Hebei University of Technology (Project 211), Tianjin, China** *09/2020 - 06/2024*
  Bachelor of Literature in International Education of Chinese , School of Humanities and Law ***GPA**: 3.98/4.0*
  Coursera: Differential Calculus through Data and Modeling Specialization; Object-Oriented Data Structures in C++; Advanced Modeling for Discrete Optimization; Graph Search, Shortest Paths, and Data Structures
- **2nd Prize in Graduation Scholarship,** 2023-2024 Academic Year, Hebei University of Technology *06/2024*
- **1st Class Scholarship,** 2020-2021 Academic Year, Hebei University of Technology *12/2020*

Research Experience

- **Intelligent RAG Routing for Complex Reasoning**
  Research Assistant, Supervisor:Yongfeng Zhang, Rutgers University *05/2025 - Present*
  My research aims to solve the one-size-fits-all limitation of RAG systems by designing a smart router that dynamically selects the best reasoning path, like GraphRAG, Naive RAG, or LLM-only, based on the user's query and corpus.
  - **Benchmarked GraphRAG Algorithms:** Benchmarked subgraph retrieval algorithms, revealing that fusing structure and semantics drives high hit rates and defines GraphRAG's core advantage in multi-hop reasoning.
  - **Exposed Architectural Limitations:** Experiments revealed single-architecture limitations, showing Naive RAG is as effective and more efficient than GraphRAG for single-hop queries on narrative corpora.
  - **Designed Intelligent Router:** Shifted research to designing an intelligent router that dynamically selects optimal RAG paths from query/corpus features, with preliminary experiments verifying its feasibility.

Working Experience

- **Rutgers University, New Brunswick** *09/2025 - Present*
  Teaching Assistant (CS462 Deep Learning, CS520 Introduction to AI), Computer Science Dept. *New Jersey, USA*
  - **Translated Theory to Code:** Translated core theoretical concepts from the CS 462 Deep Learning course into executable Python and PyTorch code examples for weekly recitation sessions.
  - **Taught Advanced Models:** Developed and delivered a range of coding tutorials, covering topics from foundational machine learning to advanced architectures (e.g., CNNs, GNNs, GANs) and modern learning paradigms.
  - **Derived Core Algorithms:** Conducted mathematical derivations of core models to deepen student understanding, including manually deriving the update rules for Linear Regression, Gradient Descent optimization etc.
- **Siemens Ltd., China** *05/2025 - 08/2025*
  LLM and Data Analysis Intern, Digital Industrial Software Dept. *Beijing, China*
  - **Fine-tuned Industrial LLM:** Fine-tuned an industrial-specific Llama 3 model by combining LoRA and prompt engineering, enhancing accuracy on technical terminology by 12% and reducing hallucinations by 7%.
  - **Automated Data Pipeline:** Developed an automated data pipeline using Scrapy, PyMuPDF, and MMOCR to process 500+ technical manuals, and utilized a VLM to parse diagrams, reducing manual processing time by over 80%.
  - **Built RAG Knowledgebase:** Cleaned and structured extracted multimodal data with Pandas to build an industrial knowledge base in Elasticsearch with over 2,000 entities, providing the dataset for a downstream RAG system.
- **Turingops Co., Ltd.** *04/2024 - 11/2024*
  AI Engineer Intern, Product Development Dept. *Shanghai, China*
  - **Built Intent Classifier:** Co-developed a core user intent recognition module, using prompt engineering, embedding similarity, a fine-tuned classifier, and an LLM reflection mechanism to achieve 85% stable accuracy.
  - **Optimized Query Robustness:** Contributed to a query preprocessing module by using sub-question generation, query rewriting and structural decomposition to boost average answer accuracy on complex queries by 16%.
  - **Generated User Insights:** Conducted in-depth user query analysis using BERTopic and LLM+Prompt for topic clustering, identifying high-frequency themes and knowledge gaps to provide data-driven reports for the team.

- **SOHU.com Limited** *01/2024 - 04/2024*

  Back-end Development Intern, Product Development Dept.t *Wuhan, China*

  – **Back-end Developed:** Contributed to building microservices and RESTful APIs (Spring Boot) and optimizing database persistence layers, including implementing CRUD operations with JPA.

- **Danfoss Ltd., China** *03/2023 - 12/2023*

  Strategic Intern, Security Operation Center, Danfoss Group IT *Tianjin, China*

  – **Cybersecurity Maintenance and Research:** Contributed to cybersecurity operations, including vulnerability analysis (Nessus, Tenable, Wireshark etc.) and management (MDE) for 1000+ web and device security.

Project Experience

- **A Educational Python Framework for Planning, Reasoning, and Learning Algorithms**

  Research Project, Supervisor: Wes Cowan, Rutgers University *05/2025 - Present*

  This project aims to build a self-contained educational framework for AI, featuring classic algorithms implemented from scratch in Python based on Rutgers CS462 and CS520 and inspired by *Artificial Intelligence: A Modern Approach.*

  – **Algorithm Design and Implementation:** Implemented a modular collection of AI algorithms in Python, covering planning, probabilistic reasoning, and neural learning, selectively using PyTorch for model training and fine-tuning.

  – **Problem Solving:** Modeled and solved diverse AI problems (e.g., maze navigation, N-Queens, Bayesian inference) by building clean, reusable code for environments and solvers to enable easier testing and comparison.

  – **Data Visualization and Analysis:** Visualized algorithm performance using Matplotlib (e.g., heatmaps), analyzing each method's strengths, limitations, and explainability across diverse tasks to gain deeper insights.

- **A RAG-based Automotive Knowledge Question Answering System**

  Internship Project, Supervisor: Zhongyu Liu, Turingops Co., Ltd. *04/2024 - 06/2024*

  This project aims to build a domain-specific question answering system powered by large language models, designed to handle a wide range of automotive-related queries on usage, repair, and maintenance with 90%+ accuracy and relevance.

  – **Hybrid Retrieval Engine:** Designed and implemented a hybrid retrieval engine for the Geely Auto QA dataset, combining classic (TF-IDF/BM25) and modern semantic search (Transformer Embeddings) as the QA bot's core.

  – **Multi-Stage Optimization:** Developed a multi-stage optimization pipeline, incorporating query preprocessing (intent recognition, expansion) and post-retrieval reranking to significantly improve recall and precision on complex queries.

  – **Domain-Specific Fine-tuning:** Performed local fine-tuning on the ChatGLM model for domain adaptation, combining it with prompt engineering to ultimately achieve an 82% answer accuracy in the final competition.

Skills

- **Programming Languages:** Frontend: HTML, CSS; Backend: Python, C++, Java
- **Languages:** Mandarin (Native), English (Proficient)