Yiyang Bian

Ph.D. Student in Computer Science
Department of Engineering
University of California, Riverside

(+1) 216-354-5035
ybian011@ucr.edu
linkedin.com/in/yiyangbian

Education

2024 – Present **Ph.D. in Computer Science, University of California, Riverside**.
Research Focus: Spatial & Big Data Management, Machine Learning

2021 – 2024 **M.S. in Computer Science, Case Western Reserve University**.
Thesis: *Extending Collaborative Filtering For ML Model Recommendation.*
Relevant Coursework: Data Mining, Machine Learning, Advanced Algorithm

2017 – 2021 **B.S. in Computer Science, Central China Normal University**.
Relevant Coursework: Database, Algorithm Design, Software Development

Employment History

2022.3 – 2024.8 **Software Engineer Intern**, *BioInVision Inc.* Cleveland, OH, USA

- Developed and deployed a comprehensive **web-based service** to transform the company's **2D/3D biomedical imaging data** into an interactive Web Service, significantly enhancing collaboration efficiency with end-users.
- Utilized **Python Flask** for back-end development, ensuring seamless data processing and API integration, while leveraging **JavaScript** and **React** for the front-end to deliver a responsive and user-friendly interface.
- Integrated **Azure Cosmos Database** as the cloud database solution, enabling scalable and secure data storage and retrieval for high-volume biomedical image datasets.
- Implemented advanced 3D image manipulation features (rotation, scaling, translation) using **Web Gl**, allowing users to intuitively interact with model scanned images for real-time feedback and analysis.

2021.1 – 2021.7 **Software Engineer Intern**, *Kingsoft Cloud*. Beijing, China

- Contributed to the **Smart City Beijing** project by processing and organizing over **100,000 municipal Excel and Word files**, enabling efficient data storage and analysis for government departments.
- Developed back-end services using **Java** and **Spring Boot**, automating data extraction, transformation, and integration with Kingsoft Cloud **MySQL** and **SQL Server**.
- Designed **RESTful API** to support seamless front-end integration, enabling real-time data visualization and analytics for smart city applications.
- Optimized back-end performance with **MyBatis Plus** for database operations and **Maven** for streamlined dependency management, improving development efficiency and scalability.

Project

2024.9 – 2024.12 **Training a Question-Answering Dialogue Model**

- Designed and trained a 1-billion parameter question-answering dialogue model, implementing a comprehensive pipeline for dataset preparation, tokenizer training, instruction fine-tuning, and evaluation.
- Utilized datasets such as **C4**, **Gutenberg**, **Books3**, and **Wikipedia** for tokenizer training, ensuring efficient vocabulary representation.
- Conducted **instruction fine-tuning** on curated datasets to enhance the model's ability to process complex queries and deliver accurate, context-aware responses.
- Evaluated the fine-tuned model using **MMLU benchmarks**, ensuring high performance on diverse linguistic tasks.
- Developed scalable training pipelines using **PyTorch** and **HuggingFace Transformers**, and deployed a user-friendly **Gradio** interface for interactive model testing.

Research Publications

- **Y. Bian**, "EXTENDING COLLABORATIVE FILTERING FOR MACHINE LEARNING MODEL RECOMMENDATION" in *OhioLink*, 2024. **[View Online]**
- M. Wang, S. Guan, H. Ma, **Y. Bian**, et al., "Selecting top-k data science models by example dataset" in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023. **[View Online]**
- M. Wang, S. Guan, H. Ma, S Guan **Y. Bian**, et al., "ModesNet: Performance-Aware Top-k Model Search Using Exemplar Datasets," in *Very Large Data Base Endowment Inc (VLDB)*, 2024. **[View Online]**
- M. Wang, S. Guan, H. Ma, **Y. Bian**, et al., "Selecting top-k data science models by example dataset," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023. **[View Online]**
- M. Wang, H. Ma, A. Daundkar, S Guan, **Y. Bian** et al., "Crux: Crowdsourced materials science resource and workflow exploration," in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM)*, 2022. **[View Online]**

Skills

Programming Languages	Python, Java, C++, JavaScript, C, MATLAB
Databases	MySQL, PostgreSQL, SQLSever, MongoDB.
Web Dev	Spring Boot, MyBatis Plus, Maven, Keras, Azure (Virtual Machine, MongoDB)
Tools & Frameworks	HTML, CSS, JavaScript, Apache Web Server, Tomcat Web Server.
Data Analysis & Machine Learning	Hadoop, Spark, TensorFlow, PyTorch, Scikit-learn