

# Ziwei Tang

☎ (+86) 199 5552 8801 | ✉ [tzw@bupt.edu.cn](mailto:tzw@bupt.edu.cn)

## Research Interests

Semantic Parsing, Database Reasoning, Large Language Model, Natural Language Processing, Representation Learning

## Education

### Beijing University of Posts and Telecommunications (BUPT)

Beijing, China

Bachelor of Engineering in Computer Science and Technology, Expected Graduation: June 2024

Sept 2020 - Present

- GPA: 3.88/4.0, 93.13/100
- Rank: 3/441
- **Courses:** Data Structures (97), Python Programming (94), Compiler Principle and Technology (99), Design and Analysis of Algorithms (96), Computer Networks (95), Principles of Database Systems (95)

## Honors

- 2021 **China National Scholarship**, Highest Honor for Academic Performance, **Top 0.2% Nationwide**
- 2022 **Merit Student**, Merit Student in BUPT, **Top 1%**
- 2021 **Silver Medal**, China Collegiate Programming Contest, Weihai Site
- 2021 **Third Prize**, China Collegiate Computing Contest-Group Programming Ladder Tournament National Final
- 2021 **Silver Medal**, China Collegiate Programming Contest for Girls 🏆

## Publications

### Tool Learning with Foundation Models <https://arxiv.org/abs/2304.08354>

Beijing, China

Tsinghua University, Renmin University of China, University of Illinois Urbana-Champaign, etc.

Jun 2023

Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, Yi Ren Fung, Yusheng Su, Huadong Wang, Cheng Qian, Runchu Tian, Kunlun Zhu, Shi Liang, Xingyu Shen, Bokai Xu, Zhen Zhang, Yining Ye, Bo Li, **Ziwei Tang**, Jing Yi, Yu Zhu, Zhenning Dai, Lan Yan, Xin Cong, Ya-Ting Lu, Weilin Zhao, Yuxiang Huang, Jun-Han Yan, Xu Han, Xian Sun, Dahai Li, Jason Phang, Cheng Yang, Tongshuang Wu, Heng Ji, Zhiyuan Liu, Maosong Sun

### DevBench: A Comprehensive Benchmark for Software Development

Beijing, China

<https://arxiv.org/abs/2403.08604>

Mar 2024

Shanghai AI Laboratory, Nanjing University, Dartmouth College, Princeton University, etc.

Bowen Li, Wenhan Wu, **Ziwei Tang**, Lin Shi, John Yang, Jinyang Li, Shunyu Yao, Chen Qian, Binyuan Hui, Qicheng Zhang, Zhiyin Yu, He Du, Ping Yang, Dahua Lin, Chao Peng, Kai Chen

## Research Experience

### DevBench: A Comprehensive Benchmark for Software Development

Beijing, China

Advisor: Dr. Bowen Li, InternLM Team, OpenMM Lab, Shanghai AI Lab (remote)

Sept 2023 - Mar 2024

- Developing a benchmarking tool to rigorously test LLMs proficiency in software development tasks.
- Actively decomposing comprehensive software engineering projects into distinct automated evaluation tasks to systematically simulate and assess software development through LLMs.
- **GitHub:** <https://github.com/open-compass/DevBench>,
- **Technical Skills:** Python, Pytorch, OpenCV, LLMs prompt engineering, LLMs evaluation.

### Natural Language Interface for Intelligent Database Interaction

Hong Kong, China

Advisor: Prof. Reynold Cheng, STAR Lab, Jockey Club, The University of Hong Kong

Jul 2023 - Present

- Explored the design and development of an innovative database interface that utilizes Large Language Models to interpret natural language queries; translate user instructions into SQL commands and craft Python scripts to produce visual, diagrammatic answers.
- Steering the evolution of a two-stage approach for query processing and data visualization, while also diligently building a custom question set to critically assess the interface's efficiency and accuracy in real-time.
- **Technical Skills:** Python, Pytorch, LLMs prompt engineering, Database, SQL.

## Intermediate Representation in Large Language Models for Table Linking

Hong Kong, China

Advisor: Prof. Reynold Cheng, Department of Computer Science, The University of Hong Kong (Summer Intern)

Jul 2023 - Sept 2023

- Introduced an E-R diagram code as an intermediate representation to enhance LLMs' schema comprehension.
- Developed a two-phase demonstration for LLMs' text-to-SQL tasks, accompanied by a dedicated table linking dataset.
- Secured the **TOP 1** performance on the leading comprehensive benchmark BIRD in August, with ChatGPT improving from 38.92% to 42.50% (+3.58%) and GPT-4 advancing from 49.58% to 50.45% on the BIRD development set.
- Technical Skills:** Python, LLMs prompt engineering, Database, SQL.

## AgentVerse: A Framework for Multi-LLM Environment Simulation

Beijing, China

Advisor: Dr. Zhiyuan Liu, Dr. Yankai Lin, Tsinghua University, Renmin University of China

Apr 2023 - Sept 2023

- Contributed to AgentVerse, streamlining the creation of custom multi-agent environments for LLMs.
- Specialized in code generation components, playing a pivotal role in developing the framework's coding capabilities.
- Achieved **top-tier performance** with the GPT + unit test integration, evidencing the framework's robustness and my commitment to excellence in code development.
- GitHub:** <https://github.com/OpenBMB/AgentVerse>, **2.5k stars**. ★
- Technical Skills:** Python, LangChain, LLMs prompt engineering.

## Tool Learning with Foundation Models

Beijing, China

Advisor: Dr. Zhiyuan Liu, Tsinghua University, Renmin University of China, University of Illinois

Aug 2022 - Aug 2023

Urbana-Champaign, etc.

- Taught ChatGPT to use tabular tools for table oriented tasks such as table manipulation, calculation, QA and chart making.
- Composed a suite of table oriented tools, such as table filtering, aggregation, charting, statistical analysis, etc.
- Built interactive environment for coordination of ChatGPT and tools. Notably, ChatGPT's performance surged from 54.8% (without tools) to 60.7% (zero-shot, +5.9%) and 85.2% (few-shot, +30.4%), illustrating the model's heightened capability when equipped with our tools.
- Research on tool learning for classic table benchmarks. **The first international work on comprehensive bank of global tasks and tools for LLMs tool learning.**
- GitHub:** <https://github.com/OpenBMB/BMTools>, **2.7k stars**. ★
- Paper:** **Tool Learning with Foundation Models**. ArXiv abs/2304.08354 (2023).
- Technical Skills:** Python, LangChain, LLMs prompt engineering, table processing.

## Sentiment Analysis of Short Text Based on LSTM

Beijing, China

Advisor: Prof. Zhitao Dai, College of Ye Peida, Beijing University of Posts and Telecommunications

Jul 2021 - Apr 2022

- Led the team to explore LSTM network on Chinese and English short text sentimental analysis.
- Constructed a comprehensive Chinese-English bilingual dataset comprising over 50,000 student conversation sentences, potentially the **largest dataset** tailored for student-oriented sentiment analysis.
- Constructed a bilingual sentiment analysis system and achieved accuracy of 82% on pure English text and 73% on bilingual occasion.
- Technical Skills:** Python, Pytorch, Markdown.

## Recognition and Reconstruction of Hybrid Sorting Objects

Beijing, China

Advisor: Dr. Haiyuan Li, School of Automation, Beijing University of Posts and Telecommunications

Oct 2020 - Apr 2021

- Proposed an SSD-based technique for recognizing hybrid objects, paving the way for subsequent reconstruction.
- Compiled a comprehensive hybrid commodity recognition dataset, featuring over 8,000 images from hybrid environments and more than 100 commodities with high-quality annotations.
- Achieved an impressive recognition accuracy of 84% in hybrid environments and 87% against solid backgrounds, demonstrating **human-equivalent** capability.
- Technical Skills:** Python, Pytorch, OpenCV, Markdown.

## Skills

**Programming** Python (Pandas, PyTorch, NumPy, Scikit-learn, Matplotlib), C/C++, Java, SQL, QT, Pascal.

**Miscellaneous** Git, Linux, Android,  $\text{\LaTeX}$ (Overleaf/Markdown), familiar with OpenAI APIs.

## Referees

### Dr. Yuntian Deng

the University of Waterloo

Assistant Professor at Waterloo CS.

Computer Science

Citations: 4602

### Dr. Zhiyuan Liu

Tsinghua University

Associate Professor at NLP Lab, Department of Computer Science, THU.

Department of Computer Science

Citations: 36,500