

YUANKAI FAN

Postdoctoral Research Fellow

 github.com/Kaimary

 +86 13681689643

 kaimary1221@gmail.com, fanyuankai@fudan.edu.cn

 Fudan University, Shanghai, China



I am currently a postdoc at the School of Computer Science, Fudan University, collaborating with Prof. X.Sean Wang. My Ph.D thesis "Research on Generative Translation Method of Natural Language Query to Database" was assessed as excellent. My research interests lie in natural language processing (NLP), artificial intelligence for database systems (AI4DB), and data governance, specifically, (1) building advanced natural language interfaces for databases and knowledge graphs (a.k.a. NL2SQL and KBQA), interactive and conversational AI systems; (2) natural language understanding includes semantic parsing, machine translation, and text generation; (3) large language model (pre-)training, parameter-efficient fine-tuning, prompting, agent systems, and network quantization; (4) data cleaning, schema matching, and data integration.

EDUCATION

PH.D. IN SOFTWARE ENGINEERING, FUDAN UNIVERSITY
with Education Program Sponsor from Microsoft

2017.9 – 2023.1

B.S. IN COMPUTER SCIENCE&TECHNOLOGY, YANSHAN UNIVERSITY

2009.9 – 2013.6

RESEARCH PUBLICATIONS

CONFERENCE

- ① Y.Fan, et al, MetaSQL: A Generate-then-Rank Framework for Natural Language to SQL Translation. In *40th International Conference on Data Engineering* , 1765-1778. (**ICDE-2024, A+ Ranking**)
- ② T. Ren, Y. Fan, et al, PURPLE: Making a Large Language Model a Better SQL Writer. In *40th International Conference on Data Engineering* , 15-28. (**ICDE-2024, A+ Ranking**)
- ③ Y.Fan, et al, GAR: A Generate-and-Rank Approach for Natural Language to SQL Translation, In *39th International Conference on Data Engineering* , 110-122. (**ICDE-2023, A+ Ranking**)
- ④ Y.Fan, et al, GenSQL: A Generative Natural Language Interface to Database Systems. In *39th International Conference on Data Engineering* , 3603-3606. (**ICDE-2023 Demo, A+ Ranking**)
- ⑤ Y.Fan, et al, An Integrated Interactive Framework for Natural Language to SQL Translation. In *24th International Conference on Web Information Systems Engineering* , 643-658. (**WISE-2023, A Ranking**)
- ⑥ Y.Fan, et al, GAR++: Natural Language to SQL Translation with Efficient Generate-and-Rank. In *8th APWeb-WAIM International Joint Conference on Web and Big Data* , 411-427. (**APWeb-2024, A Ranking**)
- ⑦ Y.Fan, et al, Grounding Natural Language to SQL Translation with Data-Based Self-Explanations. In *51th International Conference on Data Engineering*. (**ICDE-2025, A+ Ranking**) [**Under Revision**]
- ⑧ T. Ren, K. Chen, Y. Fan, et al, REDSQL: A Data-Aware SQL Refiner for Natural Language to SQL. In *International Conference on Very Large Data Bases* . (**VLDB-2025, A+ Ranking**) [**Under Review**]

JOURNAL

- ① Y.Fan, et al, A Confidence-based Knowledge Integration Framework for Cross-Domain Table Question Answering. In *Knowledge-based Systems* . (**KBS-2024, IF 7.2**)
- ② H. Chen, F. Ye, Y. Fan, et al, Staged Query Graph Generation based on Answer Type for Question Answering over Knowledge Base. In *Knowledge-based Systems* , 253:109576. (**KBS-2022, IF 8.8**)
- ③ T. Yu, Z. He, Z. Yang, F. Ye, Y. Fan, et al, Zebra: A Novel Method for Optimizing Text Classification Query in Overload Scenario. In *World Wide Web* , 26(3):905-931. (**WWWJ-2022, IF 3.7**)

MISC.

- ① Y. Fan, Z. He, X. Wang, the New Trend in Natural Language Interfaces for Databases - Opportunities and Challenges in the Era of Large Language Models. *Communications of the CCF*. (**CCCF-2023**)

PROJECTS

- **LLM-POWERED AUTONOMOUS DATA INTEGRATION SYSTEM** 2023.5 – 2024.3
Major Duty: system Design, algorithm research and development, writing report
Python LMQL
- **GENERATIVE NATURAL LANGUAGE INTERFACE FOR DATABASE SYSTEM** 2022.1 – 2023.1
 Cooperator Unit: China UnionPay Merchant Services Co.,LTD.
Major Duty: algorithm research, writing report
Python Pytorch AllenNLP
- **HIGHLY EFFICIENT BIG DATA INTELLIGENT INTERACTIVE GUIDE SYSTEM** 2017.9 – 2019.1
 Cooperator Unit: Sun Yat-sen University, Shanghai Baosight Software Co., Ltd.
Major Duty: system design, functional development, writing report
Spring Java
- **EXPLORATORY BIG DATA ANALYSIS SYSTEM** 2017.9 – 2018.5
Major Duty: toolset development
Spark Scala

PATENTS

- ❶ Z. He, X. Wang, **Y. Fan**, T. Ren, 生成数据库的查询语句集合的方法、训练排序模型的方法, 2023.5.15, China, 202310545980.2
- ❷ Z. He, X. Liu, **Y. Fan**, T. Ren, 基于自然语言生成结构化查询语言的方法和装置, 2023.6.16, China, 202310749648.8
- ❸ Z. He, X. Liu, **Y. Fan**, T. Ren, J. Dai, R. Yi, H. Fan, Y. Jing, X. Wang, 数据集的处理和SQL-自然语言翻译模型的测试方法和装置, 2023.10.13, China, 202311322798.7

INDUSTRIAL EXPERIENCES

- MICROSOFT (SHANGHAI, CHINA)** 2016.5 – 2023.2
Support Escalation Engineer
• Act as an exclusive full-stack engineer in the team. My expertise encompass *Windows drivers*, *Windows Hardware Lab Kit*, *Azure IoT/Cognitive Services/Machine Learning Service*.

- DELL EMC (SHANGHAI, CHINA)** 2015.6 – 2016.5
Software Engineer (*Vendor*)
• Project: Develop the cloud-native distributed file system for Atmos storage product.

- SHUHUI SYSTEM TECHNOLOGY CO., LTD. (SHANGHAI, CHINA)** 2013.7 – 2015.5
Software Engineer
• Project: Develop add-ins (C++ program) for AutoCAD software.

MISCELLANEOUS EXPERIENCES

- HONORS AND AWARDS**
- Outstanding Paper Award of the 8th China Conference on Intelligent Technology and Big Data 2023
 - Second Prize in the SPDB-Baidu Geek Algorithm Challenge, NL2SQL Task 2020
 - Student Scholarship Third Prize 2017

ACADEMIC ACTIVITIES

- DASFAA (A Ranking) Conference Reviewer 2024

SKILLS

- Familiar with Python and C++ development
- Familiar with commonly used machine-learning algorithms and deep-learning algorithms
- Familiar with popular deep-learning libraries (AllenNLP, PyTorch, Huggingface) and optimization tools (DeepSpeed)
- Have some knowledge with time-series methods