Jiatao Li

✉ leejames@stu.pku.edu.cn
🖰 leejamesss.github.io
⬡ leejamesss

Education

2022-Present **Information Management Department**, **Peking University**.
B.S. in Data Science (GPA Rank 1st in Department)
Research advisor: Prof. Xiaojun Wan

Research Interest

Natural Language Processing, Machine Learning

Publications

[1] **SMART-RAG: Selection using Determinantal Matrices for Augmented Retrieval**, *ArXiv Preprint*.
Jiatao Li, Xinyu Hu, Xiaojun Wan

[2] **Evaluating Self-Generated Documents for Enhancing Retrieval-Augmented Generation with Large Language Models**, *ArXiv Preprint*.
Jiatao Li, Xinyu Hu, Xunjian Yin, Xiaojun Wan

Selected Awards and Honors

Apr 2023 **Top1%**, ***Kaggle** HousePrice Prediction Competition*.

May 2023 **Rank2**, *2023 Peking University Hackerthon*.

Jun 2023 **Top10**, *CSDN Blog Star City Track, Beijing (**150w+** Total Visits)*.

Aug 2023 **Track & Finals Rank1**, *2023 **Tianchi** BMW 3rd Hackathon - Virtual Collaborative Co-Creation*.

Sep 2023 **May 4th Scholarship**, *(**Highest-level Scholarship** in Peking University, **125**/65k+)*.

Sep 2023 **Merit Student Pacesetter**, *Peking University Merit Student Pacesetter,2022-2023(**Top1%**)*.

Oct 2023 **Finalist**, *2023 Greater Bay Area (Huangpu) International Algorithm Case Competition - Advanced Enhancement of Large Language Model Comprehensive Abilities (Pazhou Laboratory & Fudan University)* .

Nov 2023 **Annual Person of the Year Nomination**, *2023 Peking University Annual Person of the Year Nomination (**Youngest Nominee**, **45**/65k+) Blog spotlight in Chinese*.

Jan 2024 **Top87**, *CSDN Blog Star of the Year*.

Projects

Apr-Sep 2023   **The Basic Textbook of Information Resource Management**, *PKU Data Mining Group*.
- The compilation of the enterprise information resource management section
- Write enterprise information resource management content in simple and easy-to-understand language.

Apr-Jun 2023   **Top1%**, *Kaggle HousePrice Prediction*.
- Course Project for Python Data Analysis
- EDA:Monitor variable,Group the variables
- Feature Engineering:Encode,PCA
- ML:Multi-Model-Selection,Grid-Search
- Gradio Application

Aug 2023   **Track & Finals Rank1**, *2023 Tianchi BMW Hackerthon*.
- Develop Algorithem for Gluing Gesture Error Detecting
- ChatTest for Gluing Theory Tests using ChatGPT
- Full Stack System for Learning feedbacks
- Digital Human for Learning Feedback

Sep-Dec 2023   **Final-List**, *2023 Greater Bay Area (Huangpu) International Algorithm Case Competition - Advanced Enhancement of Large Language Model Comprehensive Abilities (Pazhou Laboratory & Fudan University)*.
- LLMs are Reasoning Teachers: Using GPT4 as Teacher Model,Moss-7b as Student Model
- Moe-Lora:Set Gating for Lora Expert
- RAG:RAG with Google Search API
- DataSet de-duplication:Simhash
- Prompt Tuning:Both for Teacher Model(GPT) and Student Model(Moss7b)
- Construction of positive samples (using GPT4) and negative samples(using Moss7b) for two-choice hallucination questions

Jan-Mar 2024   **Comprehensive Domain Knowledge Graph of AI**, *PKU Data Mining Group*.
- Building the AI Enterprise Knowledge Graph Section
- Use web crawlers to crawl the official websites of AI companies.
- NER with Bert-CRF,LSTM-CRF etc.
- Extract Knowledge Graph using LLMs.
- Store data in Neo4j.

Positions of responsibility

Monitor   Information Management Class 1, Undergraduate Program, 2022

Minister   Department of Practice and Career Development

Manager   Comprehensive Coordination Department of Peking University Student Union

Manager   Office of International Cooperation, Peking University

Extra-curricular Activities
.
- I was one of the ambassadors of the 2020 World School Games.
- I was the conductor of the Peking University December-Ninth Choir.
- I am also a member of the school's Linux club.
- As the captain of the Information Management Department basketball team, I led the team to participate in the Peking University Freshman Cup Basketball Tournament.
- I once won a silver medal in the Information Management Department's 3V3 basketball tournament.

Skills, Interest & Others

Programming: Python, C/C++, Java, LATEX, SQL, HTML, CSS, JavaScript, TypeScript, PHP

Frameworks & Tools: Langchain, Pytorch, Tensorflow, Scikit-learn, Git, React, Vue, Bootstrap, Mysql, ELK, jsp, Springboot, Vmware, Docker, Maven, Xshell, CAD, Solidworks, Sketchup, Arduino

Languages: Mandarin (native), English (proficient, Silver Medal in the 21st Century Cup English Speech Contest, Fujian Province & Second Prize in the National Finals of the 21st Century Cup English Speech Contest), Hokkien (native)

Interests: Basketball, Swimming, Table Tennis, Badminton, Workout