Shaohuang Wang

Email: sh.wang4067@gmail.com | Tel: +86 181-1615-2720
Homepage: nicedata.eu.org | Github: wdzhwsh4067
Address: Boda Campus, Xinjiang University, Urumqi City, China

Education

Now – 2022.09 | **Xinjiang University**, China | **M.Eng. in Computer Science**
GPA: 3.5/4.0 **Core Course**: Machine Learning, Computer Networks, Data Structures
Thesis: Research on scientific and technological information recommendation via LLM
Research Interest: LLM, RAG, SFT, Fine-tuning, Recommend System

2020.06 – 2016.09 | **Shanghai University of Engineering Science**, China | **B.Eng. in Vehicle Engineering**
GPA: 3.1/4.0
Awards: National Scholarship(Top1%) First-Class Scholarship(Top3%)

Publications

[1] **Wang, S**. et al. "**Bypassing LLM Safeguards: The In-Context Tense Attack Approach.**" International Conference on Computer Engineering and Networks, 2406.12243 (2024). (Accepted)
[2]**Wang, S**. et al."**CherryRec: Enhancing News Recommendation Quality via LLM driven Framework.**" ICASSP(2025).(Under Review)
[3]Liang, Y & **Wang, S**. et al. "**LLaMA-MoT: A Cost-Effective Framework for Visual-Linguistic Instruction Tuning Based on Multi-Head Adapters and Chain-of-Thought.**" ESWA (2024).(Under Review)
[4]**Wang, S**. et al."**An agile construction method of instruction fine-tuning dataset based on semi-structured data.**" Patent (2024).(Submited)
[5]**Wang, S**. et al."**Finite element analysis of modular automotive body based on Ansys.**" Guangxi Journal of Light Industry (2020).(Accapted)
[6]**Wang, S**. et al."**Buffer connecting device for vehicle.**" Patent (2020).(Accapted)
Still in possession of 8 patents, along with various other publications.

Research Experience

Now – 2022.09 | **Domain Information Tracking and Processing Project** | **Developer@NLPIR Lab**
• Responsible for the development of the algorithm tool layer, including data collection, review and correction, dynamic selection, keyword extraction, and briefing generation algorithms.
• Utilized Elasticsearch and MySQL databases for data storage and processing, optimizing data query and analysis processes.
• Achieved rapid system deployment and front-end and back-end separation design through Docker, simplifying operations and maintenance and enhancing system maintainability.
***Technology Stack**: Python, Elasticsearch, MySQL, Docker, Vue.JS, FastAPI*

Now – 2023.04 | **Doc2QA Framework for Large Language Model SFT Datasets** | **Developer@NLPIR Lab**
• Designed and released a comprehensive dataset for QA instruction fine-tuning using semi-structured data, providing a valuable resource for future research.
• Developed a novel framework, "Doc2QA" based on Large Language Models (LLMs) to generate question-answer pairs from semi-structured data such as HTML, DOC, and PDF.
***Technology Stack**: Python,Llama-factory, Vllm, FastAPI, Docker, JavaScript*

Skills

Proficient in Coding: Pyhton, FastAPI, Elasticsearch, Docker, Vue.Js, Nginx
Model Training in AI/ML:PyTorch, TensorFlow, Llama-index, Vllm,Llama-factory
Simulation and Design: AutoCAD, CATIA, SolidWorks, ANSYS, 3DMax,Adobe Photoshop/Illustrator
Languages: Chinese(native), English(IELTS: 6.5,with L: 6.5 R: 7.5 W: 6.0 S: 6.0), Japanes(JLPT-N2)