

# Shihao LIANG

shihao.liang0828@163.com | +86-17627812352

No.70 Zhengchongxia Street, Sandun Xingzheng Village, Minzhong Town, Zhongshan, Guangdong, 528400, P.R.China

## EDUCATION

### Tianjin University

#### Bachelor's Degree of Engineering in Computer Science and Technology

Tianjin, China

Sep. 2018-Jul. 2022

- Average Score for the 3<sup>rd</sup> and the 4<sup>th</sup> academic year (out of 100): **89.33** and **88.10**
- Average Score for four academic years (out of 100): **86.15**

**IELTS: Overall 7.0** (Listening 7.5, Reading 7.5, Speaking 6.0, Writing 6.0)

Oct. 2021

**Technical Skills:** Python, Pytorch, Paddle, Linux, NLP, QA, HTML, JS, C++, Java, SQL, MySQL.

## RESEARCH EXPERIENCES

### Natural Language Processing Lab at Tsinghua University - Research Assistant

Beijing, China

#### - *QASnowball: An Iterative Bootstrapping Framework for High-Quality Question-Answering Data Generation*

submitted to *ACL 2023* (under review). First Author. Supervisor: Han Xu

June. 2022- Jan. 2023

- Introduced a iterative bootstrapping framework that can continually generate high-quality and large-scale QA data.
- Facilitated QA models: training on our generated data achieves comparable results to using supervised data, and pre-training on the generated data and then fine-tuning on supervised data can achieve better performance.
- Equipped the [CPM-Ant+](#) with question-answering ability by providing our data as pre-training data, supporting the [CPM-Bee](#) the same way with data of better quality.

#### - *Interactive Web Search for Chinese Long-form Question Answering*

submitted to *ACL 2023* (under review). Co-author. Supervisor: Qin Yujia

Aug. 2022- Jan. 2023

- Introduced IWSQA, the first Long-form question answering (LFQA) dataset, with retrieved information based on interactive web search. Collected 5500 question-answer pairs, together with the supporting facts and human behaviors, with web search behaviors recorded.
- Fine-tuned PLMs to imitate human behaviors for information retrieval and to restructure the collected facts into coherent answers. Generated answers that are no worse than human-written ones in approximately 30% and 50% of the cases on our dataset and an out-of-distribution Chinese QA dataset, respectively.
- Online [demo](#) available. Be pre-launched on [Zhihu](#) with the ability to generate a logical coherent, semantically fluent and knowledge-rich answer to the hot issues on Zhihu. Datasets and models will be released in the future.

#### - *Unified Instructions*

under preparation. Supervisor: Qin Yujia

Dec. 2022- Present

- To mitigate the format gap from different instruction-tuning benchmarks. Improve multi-task zero-shot performance on cross-benchmark by a large margin.
- Introduce chain-of-thoughts (COT) informations to current Instruction-tuning Paradigm.
- Unifying all the instruction-tuning benchmarks to our proposed instruction paradigm.

## PROFESSIONAL EXPERIENCES

### Beijing Baidu Network Technology Co., Ltd.

Beijing, China

#### Algorithm Intern, Natural Language Processing Department

Dec. 1, 2021- Apr. 8, 2022

- Deep Question-Answering: Optimization of the first result of search engine. Leveraged levenshtein distance, term weight calculation based on LSTM to mine hard negatives from user query log. Applied the hard negatives to model training to improve semantic relevance ranking models' accuracy based on ERNIE;
- Model distillation: Assisted in the application of question-answering technology in search engine in industrial projects. Distilled a model with 12 transformer encoder layers to a 2-layer model with satisfying efficiency and performance.

### Beijing Zhongke Wenge Tech. Ltd.

Beijing, China

#### Algorithm Intern, R & D Department of Tianhu Product Center

Aug. 14- Sep. 30, 2021

- License plate privacy protection project: Built and trained Yolo-v5 model to recognize the plate in photos and videos;
- Data augmentation: Enhanced data diversity by rotating and flipping the plate photos to improve the models' robustness.

## HONORS & AWARDS

**First Prize** in Paper Reproduction Challenge of China Society of Image and Graphics (**Top 1/40**) - Nov. 2021

**Advanced Individual** in Academic Progress of TJU (**Top 3%**) -Dec. 2020

**Scholarship for Academic Progress (Top 2%)** -Dec. 2020

**Merit Student** in Social Practice of TJU -Apr. 2019

**Advanced Individual** of Sun Sports (**Top 3%**) -Dec. 2020

**Outstanding Member** of TJU Art Troupe -Dec. 2018

## LEADERSHIP & VOLUNTARY EXPERIENCES

### Director, Dance Club of Peiyang Art Troupe

2020-2021

- Held TJU's annual special event with nearly tens of thousands of viewers.

### Winter Social Practice

Jan. -Feb. 2019

- Delivered speeches to introduce Tianjin University at my high school, wrote promotion articles and designed posters, and established WeChat group to provide application and admission consulting for over 100 students.

### Volunteer, Zhongshan Taiping Church

Mar. -Nov. 2020

- Made audio of sermons based on the recordings provided by pastors and preachers to help local believers participate in online church activities during the Covid-19 control and prevention periods.