Yihao Liu
Curriculum Vitae

✆ *(+86) 18810097393*
✉ *yhliu918@gmail.com*

Education

2017–2021 Bachelor	**Peking University**, *Yuanpei College, major in Data Science*. **Total GPA** 3.70, **rank** 62/347
2021– 2026 P.H.D	**Tsinghua University**, *Institute for Interdisciplinary Information Sciences (IIIS), major in computer science*.

Experience

2019.3– 2020.1	**Peking University**, *Prof. Bin Cui* . **Auto-machine learning in medical treatment**: I developed some AutoML algorithms to extract the underlying data distribution features of various medical datasets. It searches for the model with the best potential accuracy on newly involved datasets.
2019.10– 2020.6	**Peking University**, *Prof. Zhanxing Zhu* . **Poisoning attack**: I designed an algorithm to distort the specific embedding in the feature map to achieve a backdoor attack on deep learning models.
2020.5- 2020.12	**UCLA**, *Prof. Cho-Jui Hsieh* . **IRMAdv**: In order to achieve multi-robustness under different kinds of adversarial attacks on the deep learning models, I developed a training algorithm enlightened by the invariant risk minimization algorithm.
2020.9- 2021.1	**ByteDance**. **Cold-start advertisement placement**: In the TikTok Advertising system, when we lack data points on a new advertisement, the placement strategy is hard to decide. I proposed a strategy of 'maximizing information gain' to help the system converge to an optimal strategy quickly.
2021.3- 2021.7	**DP Technology**. **Contributor of deepmd-kitv2.0**: Improving embedding process in Deepmd-kit simulation algorithm, get experience about large scope scientific computing. https://github.com/deepmodeling/deepmd-kit
2021.1- 2023.12	**TsingHua University**, *Prof. Huanchen Zhang* . **[Accepted to SIGMOD2024] LeCo**: We proposed a general compression framework LeCo to achieve compression by eliminating serial correlations among data. LeCo achieves a Pareto improvement on both space savings and random access latency.
Recent Interest	**TsingHua University**, *Prof. Huanchen Zhang* . **[Accepted to CIDR2024] Resource allocation in cloud databases**: Optimizing (computational) resource allocation according to the bi-objective optimization target, including both query latency and monetary cost.

Course Projects

2022.06	**Docking scoring function in Drug discovery** Introduce a new scoring function leveraging non-bonded interaction information which outperforms VinaRF20.
2022.06	**Distributed system (MIT 6.824)** Build a fault-tolerant distributed kv-store with linearizability.
2021.10	**Distributed database system** We build a distributed NoSQL database using Mongodb, Mongos, Redis and HDFS, with a user-friendly frontend.
2020.10	**BoolQ** A Q&A NLP task, we tried to use GPT1.0, BERT, Roberta, GPT2.0 and T5 on this task, and introduced that we can further combine bi-conditional Transformer with GPT2.0's unsupervised finetune.

Technical Skills

Programming	PYTHON, C++, C, GO, SQL, NOSQL
Systems	ROCKSDB, PARQUET, PRESTO, DUCKDB