

JIALE LAO

 Website

 Google Scholar

 Github

 jiale@cs.cornell.edu

≡ EDUCATION

Cornell University, Department of Computer Science

August 2024 - May 2029 (expected)

PhD in Computer Science

Ithaca, NY

- Advisor: Prof. Immanuel Trummer

Sichuan University, Department of Computer Science

August 2020 - June 2024

B.Eng in Software Engineering

Sichuan, China

- Advisor: Prof. Mingjie Tang and Prof. Jianguo Wang (Purdue University)

≡ PUBLICATIONS

GPTuner: A Manual-Reading Database Tuning System via GPT-Guided Bayesian Optimization

- Jiale Lao, Yibo Wang, Yufei Li, Jianping Wang, Yunjia Zhang, Zhiyuan Chen, Wanghu Chen, Mingjie Tang, Jianguo Wang

- VLDB 2024,  **SIGMOD Research Highlight Award 2024** 

A Demonstration of GPTuner: A GPT-based Manual-Reading Database Tuning System

- Jiale Lao, Yibo Wang, Yufei Li, Jianping Wang, Yunjia Zhang, Zhiyuan Chen, Wanghu Chen, Yuanchun Zhou, Mingjie Tang, Jianguo Wang
- SIGMOD 2024

GPTuner: An LLM-Based Database Tuning System

- Jiale Lao, Yibo Wang, Yufei Li, Jianping Wang, Yunjia Zhang, Zhiyuan Chen, Wanghu Chen, Mingjie Tang, Jianguo Wang
- SIGMOD Record 2025

Demonstrating SQLBarber: Customized and Realistic SQL Workloads via Large Language Models

- Jiale Lao, Immanuel Trummer
- SIGMOD 2025

ToxicSQL: Migrating SQL Injection Threats into Text-to-SQL Models via Backdoor Attack

- Meiyu Lin, Haichuan Zhang, Jiale Lao, Renyuan Li, Yuanchun Zhou, Carl Yang, Yang Cao, Mingjie Tang
- SIGMOD 2026

PathBee: A Generic Optimization Framework for Efficient Distance Labeling

- Jiale Lao, Yinghao Tang, Tingfeng Lan, Mingjie Tang, Yuanchun Zhou, Jianguo Wang
- In submission, ICDE 2026

Water: A Workload-Adaptive Knob Tuning System

- Yibo Wang, Jiale Lao, Chen Zhang, Cehua Yang, Yuanchun Zhou, Jianguo Wang, Mingjie Tang
- In submission, SIGMOD 2026

QUITE: A Query Rewrite System Beyond Rules with LLM Agents

- Yuyang Song, Hanxu Yan, Jiale Lao, Yibo Wang, Yufei Li, Yuanchun Zhou, Jianguo Wang, Mingjie Tang
- Under revision, SIGMOD 2026

SQLBarber: Leveraging Large Language Models to Generate Customized and Realistic SQL Workloads

- Jiale Lao, Immanuel Trummer
- Under revision, SIGMOD 2026

SemBench: A Benchmark for Semantic Query Processing Engines

- Jiale Lao, Andreas Zimmerer, Olga Ovcharenko, Tianji Cong, Matthew Russo, Gerardo Vitagliano, Michael Cochez, Fatma Ozcan, Gautam Gupta, Thibaud Hottelier, H. V. Jagadish, Kris Kissel, Sebastian Schelter, Andreas Kipf, Immanuel Trummer
- In submission, VLDB 2026

☰ AWARDS

- **SIGMOD Research Highlight 2024** (only ten papers are selected in the database community)

☰ RESEARCH EXPERIENCE

Automatic Optimization of Database with Large Language Models ☞

May 2023 – Present

Advisors: Prof. Jianguo Wang (Purdue); Prof. Mingjie Tang (SCU)

Project Leader

- Designed and implemented GPTuner, a novel manual-reading database tuning system that leverages domain knowledge automatically and extensively to enhance the knob tuning process.
- Developed an LLM-based data pipeline, a prompt ensemble algorithm, a workload-aware and training-free knob selection strategy, and a Coarse-to-Fine Bayesian Optimization framework.
- Experimentally evaluated GPTuner under different benchmarks, metrics and DBMS. Compared to the state-of-the-arts, GPTuner identifies better configurations in **16x** less time on average and achieves **30%** performance improvement over the **best-performing** alternative.
- Project outcome: a research paper accepted by **VLDB 2024** and selected as **SIGMOD Research Highlight 2024**, a demonstration accepted by **SIGMOD 2024**.

Benchmarking Semantic Query Processing Engines ☞

May 2025 – Present

Advisors: Prof. Immanuel Trummer (Cornell University)

Project Leader

- A collaboration project, which originated from a Dagstuhl Seminar, was completed with the participation of several institutions, including Cornell University, Google, MIT, the University of Michigan, BIFOLD and TU Berlin, the University of Technology Nuremberg, and Vrije Universiteit Amsterdam.
- Built SemBench, a systematic benchmark designed to evaluate and compare recent LLM-powered multi-modal data systems in realistic settings. It includes six well-defined use cases that integrate structured data (tables) with unstructured modalities (text, images, audio), enabling the assessment of systems' ability to process complex semantic queries with ground truth validation.
- Evaluated 3 academic systems and 1 industrial systems (Google BigQuery) on SemBench, and analyzed their design and performance limitations to reveal insights for future system development.
- Project outcome: An open-source benchmark SemBench, a benchmark paper in submission to VLDB 2026.

Customized and Realistic SQL Workload Generation ☞

September 2025 – Present

Advisors: Prof. Immanuel Trummer (Cornell University)

Project Leader

- Built SQLBarber, a system based on LLMs to generate customized and realistic SQL workloads.
- Implemented (1) a declarative interface for users to effortlessly generate customized SQL templates, (2) an LLM-powered pipeline augmented with a self-correction module that profiles, refines, and prunes SQL templates based on query costs, and (3) a Bayesian Optimizer to efficiently explore predicate values and identify a set of queries that satisfy the target cost distribution.
- Project outcome: a demonstration accepted by SIGMOD 2025, a paper under revision of SIGMOD 2026.

Distance Indexing Optimization via Graph Neural Networks ☞

October 2022 – Present

Advisors: Prof. Jianguo Wang (Purdue); Prof. Mingjie Tang (SCU)

Project Leader

- Developed PATHBEE, a generic optimization framework to achieve efficient distance labeling.
- Provided a solid theoretical analysis to reveal a performance degradation factor shared by existing methods, proved it is NP-hard to find the optimal vertex traversal order, and identify the best-performing ranking method via a formal modeling of indexing process.
- Developed an effective GNN-based approach to rank the vertices, proposed a novel sampling strategy to further enhance this approach.
- Extensive experiments on 26 real-world datasets shows that PATHBEE achieves substantial reductions in indexing time (up to **21.49 times**), index size (up to **5.78 times**), and query time (up to **2.18 times**).
- Project outcome: a paper in submission to **ICDE 2026**.

Programming Languages: Python, SQL