Lixi Zhou

lixi.zhou@outlook.com | Homepage | Google Scholar

Profile

Ph.D. candidate in Computer Engineering with expertise in integrating database management systems with machine learning techniques and hands-on industry experience in deep learning.

- Experienced in system development, performance optimization, and integrating AI to enhance database systems.
- Self-motivated and task-oriented, with experience in contributing to open-source projects and ownership.
- Published in top-tier conferences (VLDB, SIGMOD, ICDE), tackling real-world DB and AI integration challenges.

Professional Experience

Ph.D. Researcher, Arizona State University July 2020 - Present

Co-Optimization for SQL Queries and AI/ML Model Inferences in Database Systems

- Designed a multi-layer intermediate representation for inference query and an MCTS-based query optimizer, leveraging transformed-based Query2Vec model for cost-aware query optimization.
- Integrated the proposed optimization techniques into Meta's open-source execution engine, Velox, accelerating AI-powered SQL workloads by up to 400× on real-world movie recommendation, retailing, LLM, and retrieval augmented generative (RAG) workloads with proposed co-optimization rules.
- Automated system deployment and configuration with Docker and AWS. The manuscript is under review.

Research Scientist-Hybrid Cloud Internship, IBM May 2023 - Aug. 2023

Model Inference Parallelization on ONNX-MLIR Deep Learning Compiler

- Proposed and implemented OpenMP-based parallelization within the MLIR ecosystem to optimize model inference for the open-source DL compile, ONNX-MLIR, enabling scalable and efficient execution.
- Led and collaborated with community developers to integrate the optimization seamlessly into the compiler, ensuring compatibility with existing optimization rules without additional overhead for developers or end-users.
- Achieved $5-22\times$ inference speedup in production environments for LLMs such as RoBERTa, GPT-2, and BERT. The works were rigorously tested, validated, and successfully merged into the production branch.

Hybrid Cloud Software Engineer Internship, IBM May 2022 - Aug. 2022

Privacy-Preserving Redaction of Diagnosis Data through Source Code Analysis

- Developed a static analysis approach to identify the sensitivity of the log message by tracking the data flow from log statements to the data sources, enhancing privacy protection in enterprise applications.
- Improved precision and recall by 33% and 24% on average, surpassing regex and AI-based methods.

Selected Publication

- **[SIGMOD 2025]** Hong Guan, Lei Yu, **Lixi Zhou**, Li Xiong, Kanchan Chowdhury, Lulu Xie, Xusheng Xiao, and Jia Zou. Privacy and Accuracy-Aware AI/ML Model Deduplication.
- **[ICDE 2024] Lixi Zhou**, K. Selçuk Candan, and Jia Zou. 2024. DeepMapping: Learned Data Mapping for Lossless Compression and Efficient Lookup. [PDF]
- **[EDBT 2024] Lixi Zhou**, Qi Lin, Kanchan Chowdhurya, Saif Masood, Alexandre Eichenberger, Hong Min, Alexander Sim, Jie Wang, et al. Serving Deep Learning Model in Relational Databases. [PDF]
- **[SSDBM 2023] Lixi Zhou**, Lei Yu, Jia Zou, and Hong Min. 2023. Privacy-Preserving Redaction of Diagnosis Data through Source Code Analysis. [PDF]
- **[VLDB 2022] Lixi Zhou**, Jiaqing Chen, Amitabh Das, Hong Min, Lei Yu, Ming Zhao, and Jia Zou. 2022. Serving deep learning models with deduplication from relational databases. [PDF]

Education

Arizona State University, Ph.D. in Computer Engineering 2020 Jul. - 2025 Aug.(Exp.)
Northeastern University, M.S. in Computer Systems Engineering: Software Design 2017 Sep. - 2020 May
Central South University, B.E in Measurement & Control Technology and Instrument 2013 Sep. - 2017 May

Skills

Languages/Frameworks/Tools: Python, C++, PyTorch, TensorFlow, ONNX, MLIR, Spark, Velox, Git, Docker, CI/CD