Shiyuan (Eric) Zhou

+1 (425) 449-3966 | szhou20@uci.edu | Google Scholar | Github

SUMMARY

Ph.D. candidate in Computer Science at UC Irvine's Information Systems Group, working at the intersection of **AI and data management**. My research focuses on building **reliable and efficient LLM-based systems** that reason over data and automate complex analytical workflows.

PUBLICATIONS

[1] SALUS: Automated Auditing of NL-to-SQL Benchmarks through Weak Supervision of Multi-Agent Output.
Shiyuan Zhou, *Ashwin Gerard Colaco, Sainyam Galhotra, and Sharad Mehrotra* **SIGMOD 2027 (Forthcoming)**

[2] MADRAG: Multi-Agent Debate with Retrieval-Augmented Generation for Training-Free Analytic Essay Scoring.
Ali Keramati, ***Shiyuan Zhou****, Sharad Mehrotra, and Mark Warschauer* **NLP4DH @ ACL 2026**

[3] WorkLoadWeaver: Realistic Synthetic Query Workloads for Adaptive Database Systems.
Keming Li, Nada Lahjouji, Ashwin Gerard Colaco, ***Shiyuan Zhou****, Pratyoy Das, Guangxue Zhang, Sahil Makhijani, Vishal Chakraborty, Sarvesh Pandey, and Sharad Mehrotra* **SynthAI @ SIGMOD 2026**

[4] HEPHA: A Mixed-Initiative Image Labeling Tool for Specialized Domains.
Shiyuan Zhou*, Bingxuan Li, Xiyuan Chen, Zhi Tu, Yifeng Wang, Yiwen Xiang, and Tianyi Zhang* **IUI 2025**

[5] Aligning Documentation and Q&A Forum through Constrained Decoding with Weak Supervision.
Rohith Pudari, ***Shiyuan Zhou****, Iftekhar Ahmed, Zhuyun Dai, and Shurui Zhou* **ICSME 2023**

[6] Rapid Image Labeling via Neuro-Symbolic Learning.
Yifeng Wang, Zhi Tu, Yiwen Xiang, ***Shiyuan Zhou****, Xiyuan Chen, Bingxuan Li, and Tianyi Zhang* **KDD 2023**

SELECTED RESEARCH EXPERIENCE

University of California, Irvine — Information Systems Group Sep. 2024 – Present
Advisor: *Dr. Sharad Mehrotra* and *Dr. Sainyam Galhotra* (Cornell University)

- **SALUS:** Proposed an automatic auditor for text-to-SQL benchmarks. For a set of benchmark tasks, SALUS aggregates signals from agent outputs into high-confidence weak correctness labels, then uses them to supervise SQL-feature-based decision trees that learn which agents to trust for different query types, enabling annotation-error detection without manually labeled training data.
- **SALUS:** Achieved **0.919 error-detection F1** versus 0.839 for existing auditor on a set of manually audited text-to-SQL tasks; Determined **over 30% annotation error rates** in existing text-to-SQL benchmark.
- **AI Data Agent:** Developed an agentic system for heterogeneous data analysis that ranked among the top-performing systems on DataAgentBenchmark and KramaBench.
- **AI Data Agent:** Built an offline data profiler and hierarchical task-network (HTN) planning workflow in which agents recursively decompose analytical queries into executable tasks; intermediate task failures are stored in shared agent memory to trigger task-tree pruning.
- **WorkLoadWeaver:** Developed a three-agent pipeline to generate synthetic workload specifications from database schemas and query templates, modeling user behavior, parameter dependencies, and potential workload shifts.

University of California, Irvine — Digital Learning Lab Sep. 2025 – Present
Advisor: *Dr. Mark Warschauer*

- **MADRAG:** Proposed a training-free multi-agent evaluation workflow that combines criterion-specific advocate–skeptic debates with judge-based scoring, demonstrated on essay assessment.
- **MADRAG:** Developed a score-aware retrieval-augmented generation (RAG) pipeline that retrieves semantically similar examples across score levels to calibrate LLM judges and reduce middle-score bias.

Purdue University — Interactive Intelligent Systems Lab May 2022 – Feb. 2025
Advisor: *Dr. Tianyi Zhang*

- **RAPID:** Developed a neuro-symbolic image-labeling technique that infers interpretable labeling rules from visual features through inductive logic programming, achieving **over 80% accuracy** across four domain-specialized tasks.
- **RAPID:** Proposed multi-criteria active learning that combines labeling-rule disagreement and sample diversity to prioritize images for human correction, refining labeling rules with fewer annotations.

- **HEPHA:** Developed an end-to-end human-in-the-loop labeling system built on RAPID, integrating automated rule inference, visual rule editing, and targeted recommendations to help users refine labeling rules.
- **HEPHA:** Evaluated the system in a 16-participant user study, achieving **over 87% accuracy** across four tasks while reducing labeling time by approximately **15–38%** compared with a ResNet-based labeling tool.

University of Toronto — FORCOLAB Sep. 2022 – Jun. 2023

Advisor: *Dr. Shurui Zhou*

- **DOSA:** Developed an LLM-based retriever that maps Stack Overflow questions to documentation by fine-tuning on passages paired with their section and subsection names and constraining predictions to valid section–subsection paths.
- **DOSA:** Achieved **98%** precision and **94%** recall on Python and **98%** precision and **99%** recall on Flask for category-level documentation retrieval.

SELECTED INDUSTRY EXPERIENCE

Zhejiang University Institute of Computer Innovation — Volkswagen Team Jun. 2023 – Aug. 2023

Machine Learning Engineer Intern

- **Auto Parts Anomaly Detection:** Developed an automotive assembly inspection system for Volkswagen that used production-line cameras to monitor multiple installations and localize assembly errors within each captured image.
- **Auto Parts Anomaly Detection:** Addressed the scarcity of defective training samples by implementing PatchCore-based one-class anomaly detection with pretrained patch-level features, detecting previously unseen installation errors without part-specific model training or fine-tuning.
- **Auto Parts Anomaly Detection:** Generated spatial anomaly maps to highlight abnormal regions and associate localized anomalies with affected components.
- **CCTV Search:** Developed a video search system that enabled workers to identify when personnel, equipment, or other target objects first appeared in long surveillance recordings and returned a five-second evidence clip.
- **CCTV Search:** Leveraged SegGPT to support a large and diverse set of searchable objects without category-specific model training, combined with coarse-to-fine frame sampling to localize the earliest occurrence.

Basebit Jun. 2021 – Aug. 2021

Data Scientist Intern

- Applied Funk SVD matrix factorization to model highly sparse patient-provider interaction data for a healthcare recommendation system, learning latent representations from limited historical visit records.
- Developed a hospital analytics and visualization module, transforming administrative datasets into visual reports to support data exploration and operational analysis by hospital stakeholders.
- Built an internal patent application workflow portal using HTML, CSS, and JavaScript, providing research teams with a centralized interface to navigate and track the application process.

EDUCATION

Ph.D. Candidate, Computer Science, University of California, Irvine Sep. 2024 – Present

Advisor: *Dr. Sharad Mehrotra* | **GPA: 4.0/4.0**

Selected Courses: Distributed System by *Dr. Nalini Venkatasubramanian*; DBMS by *Dr. Chen Li*; Graphical Models by *Dr. Erik B. Sudderth*; Algorithm by *Dr. Michael Shindler*; Information Retrieval by *Dr. Iftekhar Ahmed*

Honours B.Sc., Data Science Specialist, University of Toronto Sep. 2019 – Jun. 2023

Selected Courses: Machine Learning by *Dr. Roger Grosse*; Deep Learning by *Dr. Jimmy Ba* and *Dr. Bo Wang*; DBMS by *Dr. Diane Horton*; Statistical Data Science by *Dr. Anna Goldenberg* and *Dr. Meredith Franklin*

TECHNICAL SKILLS

Programming Languages: Python, SQL, C, Java, R, JavaScript, Bash
Databases & Data Systems: PostgreSQL, MySQL, SQLite, MongoDB, Apache Kafka, Apache Spark, PySpark
Retrieval & Vector Search: FAISS, pgvector, ChromaDB
ML & LLMs: PyTorch, Hugging Face Transformers, Sentence Transformers, PEFT (LoRA), vLLM, scikit-learn
Agent Frameworks: LangChain, LangGraph, AutoGen
Cloud & Infrastructure: AWS, Docker, Linux, Git, GitHub Actions
Web & Engineering: React, Flask, HTML/CSS, pytest