

Jinhao Zhao

Ph.D. Candidate | Algorithms, Distributed Systems & Applied Machine Learning
St. Louis, MO | +1 (314) 934-7449 | jinhaoz@wustl.edu

Summary

Computer Science Ph.D. candidate with expertise in algorithms, distributed systems, and cost-aware ML pipelines, plus industry experience in e-commerce recommendation, speech recognition, knowledge graphs, and computer vision. Applies algorithmic optimization to model routing, low-latency inference, recommendation/search, and scalable AI systems.

Education

Washington University in St. Louis | *Ph.D. in Computer Science* St. Louis, MO | **Expected Nov 2026**
Tsinghua University, Yaoclass | *B.Eng. in Computer Science and Technology* Beijing, China | **Jun 2018**

Technical Skills

Languages & Tools: Python, C++, Kaldi, Git, Linux

ML / AI: Recommendation systems, uplift modeling, ASR, knowledge graphs, computer vision, time-series classification

Systems & Algorithms: Distributed systems, load balancing, scheduling, data placement, cache algorithms, latency optimization

Experience & Research

TikTok Inc. | *E-commerce Recommendation Algorithm Intern* Seattle, WA | **Jun 2026-Present**

- Collected and analyzed online data during the cold-start phase of e-commerce recommendations.
- Trained models for uplift estimation in e-commerce recommendation.

Washington University in St. Louis | *Doctoral Research (Advisor: Prof. Kunal Agrawal)* St. Louis, MO | **Aug 2021-Present**

- Designed algorithms with provable guarantees for distributed load balancing under reappearance dependencies; published at PPOPP 2023 and SPAA 2024 (**Best Paper**).
- Developed cost-aware classifier cascades for time-series streams, optimizing model execution order to reduce end-to-end runtime by **~29%** without adding new models; relevant to model routing and efficient inference.
- Analyzed semi-clairvoyant scheduling for jobs with multiple criticalities using competitive analysis; the ECRTS 2026 paper received the **Outstanding Paper Award**.

Zhuiyi.ai | *Automatic Speech Recognition Research* Shenzhen, China | **Nov 2019-Aug 2021**

- Trained and optimized Kaldi-based ASR models in Python/C++ for industrial dialogue data, reaching **98%+ accuracy** in technical conversation scenarios.
- Reduced system latency from **>500 ms to <200 ms** using high-performance computing methods.

Ruyi.ai | *ML Intern, Knowledge Graph & Recommendation* Beijing, China | **Feb-Aug 2019**

- Built knowledge-graph-based company risk detection models and financial-product recommendation systems used in banking scenarios; designed explainable outputs for experts at China CITIC Bank and Bank of China.

Selected Project

jx3bla | *Open-Source Developer* | github.com/moeheart/jx3bla **Sep 2021-Present**

- Built and maintained a user-facing MMORPG combat-log analytics platform serving **1K+ active users** and receiving **~1M logs**; delivered detailed performance data and comparison features.

Selected Publications

- "Semi-Clairvoyant Scheduling for Jobs with Multiple Criticalities." **ECRTS 2026 - Outstanding Paper Award.**
- "Distributed Load Balancing in the Face of Reappearance Dependencies." **SPAA 2024 - Best Paper Award.**
- "IDK Cascades for Time-Series Input Streams." **RTSS 2024**; "Faster Classification of Time-Series Input Streams." **ECRTS 2025.**
- "Provably Good Randomized Strategies for Data Placement in Distributed Key-Value Stores." **PPOPP 2023.**