

Jitong Zhou

Website: jitongzhou.com

Linkedin : linkedin.com/in/jitong-zhou

EDUCATION

- **Stanford University** Stanford, CA
M.S. in Computer Science; GPA: 4.01/4.3 *Expected Jun 2025*
- **Carnegie Mellon University** Pittsburgh, PA
B.S. in Computer Science, B.S. in Statistics; GPA: 3.92/4.0 *May 2023*
- **TA experiences:** Distributed Systems (Stanford 244b, CMU 15440), Scalable Machine Learning (CMU 10605)

EXPERIENCE

- **Artificial Intelligence Engineer Intern** Sunnyvale, USA
LinkedIn *Jun 2024 - Sep 2024*
 - Developed an LLM-powered evaluation pipeline that achieved 98% accuracy in detecting nonfactual outputs, utilizing a combination of proprietary APIs and fine-tuned open-source models.
 - Engineered a synthetic dataset generation system that enabled fine-grained benchmarking, enhancing the evaluation pipeline and contributing to the fine-tuning of open-source LLMs.
- **Software Engineer Intern** Los Angeles, USA
rabbit inc. *Jun 2023 - Sep 2023*
 - Led end-to-end development of RESTful APIs using Docker, optimizing resource-efficient deployment and enhancing API responsiveness through asynchronous processing and stateless architecture.
 - Developed OpenAPI documentation and collaborated with machine learning engineers and UI/UX specialists to design a user-friendly web interface, ensuring seamless integration.
- **Machine Learning Engineer Intern** Beijing, China
Keep *May 2022 - Aug 2022*
 - Designed, implemented, and optimized a GNN architecture to learn user and workout embeddings, achieving a 0.70 Pearson correlation on the pretext task for pre-training, enhancing downstream intelligent workout plan recommendations.
 - Engineered and preprocessed a dataset using SQL and Hadoop MapReduce, including data cleaning, feature extraction, and transformation, to optimize the GNN pipeline.
- **Software Engineer Intern** Beijing, China
Meituan *May 2021 - Nov 2021*
 - Conducted statistical analysis on the stale gradient problem in large-scale asynchronous distributed training of sparse models, developing TensorFlow kernels and ops in C++ and Python to improve training stability and model accuracy.
 - Achieved a 0.1% average AUC increase in click-through rate (CTR) models for ad recommendations.

PROJECTS

- **Digital Human:** A web-based application that seamlessly transforms static images of human subjects into dynamic video presentations, powered by a pipeline of large machine learning models.
- **Distributed Bitcoin Miner:** With Go, Implemented a failure-recoverable distributed Bitcoin Miner simulator under the server-client communication protocol.
- **Fair ML Job Scheduler:** Built a scheduler with Python and Gurobi solver to optimize long-term efficiency and fairness for ML training jobs with elastic resource requirements on a shared cluster.

PROGRAMMING SKILLS

- **Programming Languages:** Python, C++, Go, C, Java, SQL, R, Javascript/HTML/CSS, Bash
- **Tools:** Docker, TensorFlow, PyTorch, Apache Spark, Hadoop MapReduce, AWS, Git, Linux, React, Flask, Node.js, MySQL, Numpy, Pandas