Shao Wang

(+86)135-0587-3867 | shaowang@sjtu.edu.cn | github.com/Electronic-Waste

Education

Shanghai Jiao Tong University | *Computer Science M.Eng.*   2024.09 - 2027.03
Shanghai Jiao Tong University | *Software Engineering B.Eng.*   2020.09 - 2024.06

Open Source

Kubeflow | *WG AutoML/Training Maintainer*

- **LLMOps:** Led the design and implementation of Kubeflow LLM Trainer V2, which streamlines LLM fine-tuning on Kubernetes by: 1) Hiding complex Kubernetes configurations from data scientists via designing CRDs for different personas 2) Unleash data scientists from the burden of manually setting up environment by providing editable workflow templates and a simpel yet efficient Python API
- **Mentorship:** Acted as a Mentor for several Kubeflow GSoC'25 projects, leading a three-member team and coordinating cross-time-zone collaboration to support students in accomplishing their project goals.
- **Maintenance:** Participate in the daily maintenance routine for *kubeflow/trainer* and *kubeflow/katib*

Selected Publications & Talks

RAG and Fine Tuning with Kubeflow

- *Francisco Javier Arceo,* **Shao Wang***, KubeCon + CloudNativeCon NA 2025*
- I presented this work in replace of Andrey Velichkevich, who was unable to talk due to some reasons

Streamline LLM Fine-tuning on Kubernetes with Kubeflow LLM Trainer

- **Shao Wang***, Andrey Velichkevich, Kubeflow Summit NA 2025*

Industrial Experience

Databend Labs | *Cloud Platform Intern*   2024.10 - 2025.04

- **Operator:** Developed an open-source operator for databend on Kubernetes from scratch, providing open-source users with a flexible yet efficient approach to orchestrate databend clusters on Kubernetes.
- **Observability:** Integrated Prometheus for enhanced system visibility on the on-premise cloud platform.

Goolge Summer of Code | *Contributor, CNCF - Kubeflow*   2024.05 - 2024.09

- **Push-based Metrics Collection:** Built a push-based training metrics system to replace Katib's sidecar approach, which reduced performance/resource overhead and improved metrics collection latency.
- **SDK:** Developed a simple Python API for metric reporting, minimizing format errors and improving robustness.
- **E2E Testing:** Added e2e tests for hyperparameter tuning in GitHub Actions to enhance CI coverage.

Bondi Tech Ltd. | *Infrastrcture Intern*   2024.01 - 2024.04

- **Investigation:** Evaluated time series databases for OLAP analysis of market data and delivered benchmark reports.
- **Quote Storage:** Build a new storage system from scratch on Kubernetes in replace of the current CSV-based storage system, which enhances the performance of data analysis business by 10x times.

Skills

- **Programming Language**: **Golang**, **Python**, C/C++, Java
- **Language**: CET6 Score: 605/710