Lisen Dai
LinkedIn: www.linkedin.com/in/lisen-dai-tech
Phone: (646)629-8802
Email: lisen.dai.work@gmail.com

Experience

OM Partners USA, Inc 04/2023-Current
Platform Engineer/Software Engineer Atlanta, GA

- C++ Build Tool: Designed and developed a high-performance, debuggable build tool centered on CMake.
  - Provided rich presets and templates allowing flexible configuration for different build targets.
  - Supported full build lifecycle: code generation, Incredibuild-powered compilation, linking, and installation.
  - Implemented a scalable and high-performance build cache service using CCache, NGINX, Redis, and NFS, enabling concurrent access by hundreds of developers and CI pipelines.
  - Integrated into the CI pipeline, replacing the legacy build tool.
  - Developed scripts and logic to manage third-party C++ libraries via vcpkg, including customized ports and integration with a private ProGet server for package hosting.
- Clipboard Helper: Developed A clipboard utility app to securely share clipboard data between local laptops and development virtual machines.
  - Developed for both Windows and Linux platforms using C# using SignalR Hubs.
  - Developed features such like clipboard history, image copy-paste, pinned content, system tray extension.
- Observability Instrumentation Implementation
  - Extended the internal network protocol to support trace propagation, without introducing overhead.
  - Added HTTP instrumentation support to track service interactions.
  - Wrote automation scripts for configuring and deploying the OpenTelemetry Collector.
  - Deployed ELK clusters with Kubernetes and configured dashboards for visualizing observability data.
- Data Infrastructure
  - Developed a tracked daily user behavior data pipeline using on Azure Blob and Databricks.

TikTok 05/2022-08/2022
Software Engineer Intern Mountain View, CA

- Implement a visual-similarity feed moderation model, reducing false positive rate by 3.4%.
  - Trained a Siamese Network in PyTorch for visual similarity detection to recall cases to moderate.
  - Indexed millions of embeddings of known policy-violating images using FAISS for quick search.
- Deployed feature extraction and moderation logic as a scalable Thrift-based RPC API, supporting low-latency inference and integration with upstream services.
  - Exported the trained model to ONNX format for cross-platform deployment.
  - Built an offline data pipeline using SQL and Hadoop to regularly ingest and analyze flagged content for retraining and continuous improvement.

ByteDance 01/2021-08/2021
Software Engineer Intern Beijing, China

- Domain-Specific Model: Selected and mined upstream features; trained and deployed an XGBoost model to moderate short-video posts in the specific pattern in the moderation pipeline. The model achieved high recalling rate and save budget by reducing human moderators' workload.
- Multi-modal Models: Trained and deploy a multi-modal model to select the most representative frame from a video. The model helped the moderators make fewer mistakes.
- Computation Efficiency: Improve customized operators with C++ and CUDA.

Postsecondary Education

Columbia University in the City of New York
09/2021-12/2022 **Master of Science**
New York, NY Electrical Engineering (GPA: 3.8/4)

Dalian University of Technology
09/2016-06/2020 **Bachelor of Engineering**
Dalian, China Electrical Information Engineering (GPA: 3.81/4)

Publications

- T. Yang, **L. Dai**, Z. Liu, X. Wang, M. Cheng, Y. Tian, X. Zhang. CLIPErase: Efficient Unlearning of Visual-Textual Associations in CLIP. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2025).
- T. Yang, **L. Dai**, Z. Liang, Z. Liu, Y. Tian, X. Zhang. SaSR-Net: Source-Aware Semantic Representation for Audio-Visual Question Answering. In Proceedings of the Association for Computational Linguistics (EMNLP 2024).