

陈帝羲 Dixi Chen

电话: +86-18580344667 | Email: cdxcqu2014@163.com

<https://chen-dixi.github.io/about> | <https://github.com/Chen-Dixi>



Education

Beijing University of Posts and Telecommunications (211)

09/2018-06/2022

- Master of Engineering (**Recommended for admission**), Computer Science and Technology
- **National Scholarship, November 2020**

Chongqing University (985)

09/2014-06/2018

- Bachelor of Engineering, Computer Science and Technology

Experience

Kuaishou Kwaishop – AI Infrastructure Center – R&D Engineer

07/2022 ~ 03/2026

Retrieval-Augmented Generation (RAG) Knowledge Base

Spearheaded the development of an internal RAG knowledge base platform from scratch. Built a self-developed RAG framework integrating multiple data sources including online documents and spreadsheets with intelligent document parsing. Supported Graph RAG knowledge graph retrieval and hybrid retrieval (Elasticsearch vector + BM25). Designed core pipelines including query rewriting, ReRank, and context enhancement. Served 10+ business lines including e-commerce, commercialization, and customer service, processing millions of retrieval requests daily. This system laid the foundation for subsequent Video Conference Assistant (meeting notes, to-do lists, knowledge Q&A) and MemorySys (similar to Mem0) services.

AI Workflow Orchestration Platform

Designed and implemented an AI workflow orchestration system with LLM capabilities, integrating GPT, Qwen, Claude and other models. Enabled visual AI capability orchestration with debuggable and diagnosable runtime preview: node execution results, execution time tracking, and exception stack recording. Served multiple business lines with 300+ orchestrated applications created, automating hundreds of workflows

Kwaishop API Gateway

Developed multi-dimensional API grayscale release mechanism. Designed API grayscale routing instrumentation and built monitoring dashboards to help detect change risks, preventing multiple change incidents.

Microsoft STCA & MSRA – Research Intern

12/2020 ~ 03/2022

Research: Fitted density functional theory (DFT) using neural networks. Explored TorchScript and CUDA to accelerate neural network inference speed.

Development: Participated in Microsoft Carina enterprise dialogue system development. Responsible for asynchronous dialogue data storage, NRules rule engine dialogue flow customization, and ES + Kibana user behavior visualization analysis.

Other Projects

Nano-Bruce: Minimal Agent Runtime Harness

[Personal Project]

Multi-turn agent-loop implementation in TypeScript (tool calling, streaming output, skill loading), including streaming events and tool execution chains, supporting multiple LLM providers. Engineered to align with Pi's runtime approach for systematic understanding of agent harness implementation. Github: <https://github.com/Chen-Dixi/nano-bruce>

2024 Kuaishou Hackathon – 1st Place

Developed e-commerce live streaming video translation technology. Responsible for the NLP module, implementing real-time speech recognition, language translation, and TTS for live audio streams. Coordinated with CV face-swap and lip-sync modules to complete an end-to-end real-time video processing pipeline.

NEEDLE: Necessary Elements of Deep Learning

[Personal Project]

A mini deep learning framework developed by [Carnegie Mellon University \(CMU\)](https://www.cmu.edu) for its "Deep Learning Systems" course.

Implemented a PyTorch-like deep learning training framework supporting automatic differentiation, common loss functions, Adam optimizer, and classic models including RNN, LSTM, and Transformer. Built matrix operations module running on both CPU and GPU devices, using matrix tiling, multi-threaded cooperative reading, and cache reuse in CUDA programs to accelerate matrix multiplication. Github 地址: <https://github.com/Chen-Dixi/dlsys-hw4>

Publications

Learning Likelihood Estimates for Open Set Domain Adaptation, 2020 IEEE International Conference on Multimedia and Expo (ICME), CCF-B, (Oral Presentation)

Haiyang Zhang; **Dixi Chen**; Liang Liu, Date of Conference: 6-10 July 2020