Chi **Xing**

Machine Learning Research · Computer System

48-50 Iona Street, Edinburgh, UNITED KINGDOM

chi.xing2025@gmail.com | openchi.life | MartinRepo | Chi Xing

Education

University of Edinburgh

Edinburgh, United Kingdom

M.Sc. in Artificial Intelligence

Sep. 2024 - Sep. 2025

- Focused on various machine learning frameworks, ranging from basic neural networks (RNN, CNN, MLP, etc.) to advanced modern frameworks (Transformers, Diffusion Models, Large Multi-Modality Models, etc.)
- Dissertation is working on accelerating and serverless-supported preference alignment techniques (such as LoRA fine-tuning, RLHF, DPO and SFT, etc.). This project is supervised by Prof. Luo Mai.

University of Liverpool

Liverpool, United Kingdom

B.Sc. in Computer Science

Sep. 2020 - Jul. 2024

- Research Point: Algorithm Design, C++/C/C#, Optimisation, Machine Learning, AI Safety, Java, Web Development
- Dissertation is focused on exploring various scheduling algorithms for modern smart grid. This project is supervised by Prof. Prudence Wong

Publication

Preference Alignment on Diffusion Model: A Comprehensive Survey for Image Generation and Editing

IJCAI 2025 (Under Review)

Author

Feb. 2025

- **Chi Xing**, et al. Preference Alignment on Diffusion Model: A Comprehensive Survey for Image Generation and Editing. Responsible for preference alignment techniques on DMs Applications section (Autonomous Driving, Robotics, Medical Imaging and Others).
- Link: Click for pdf

Presentation

N8 CIR Conference

York, United Kingdom

Presenter for <Benchmarking LLMs for reading bio-medical literatures>

Sep. 2024

- Report summer internship work on N8 CIR Conference. Presented on the challenge of 'Benchmarking LLMs for reading bio-medical literature', an impressive intersection of #DigitalHealth and #MachineLearning research themes.
- Link: Click for post

Experience

ServerlessLLM (400+ Stars)

ServerlessLLM, Github

Core Contributor

Nov. 2024 -

- Support ServerlessLLM deployment on SLURM-based HPC
- Working on accelerating/serverless-supported fine-tuning techniques on ServerlessLLM

N8 Centre of Excellence in Computationally Intensive Research

Liverpool, United Kingdom

Research Intern, Funded by EPSRC

Jun. 2024 - Sep. 2024

- **Focused on benchmarking various LLMs** for reading biomedical literature, utilizing Llama.cpp to quantize open-source models such as Llama3.1-70B, Llama3.1-405B, DBRX, and Mixtral-8x22B.
- **Developed an objective scoring system** that extracts key information from model outputs and evaluates their similarity to manually extracted data for performance benchmarking.
- **Designed a summarization method** to reduce input size, enabling the use of models with smaller context windows.
- **The work also involved comparing model performance** across different hardware platforms, including NVIDIA GH200, A100, and CPU/GPU references, and deploying LLMs on high-performance computing (HPC) architectures.

Casibase (3000+ Stars)

Casbin Community, Github

Core Contributor

Jan. 2024 - Sep. 2024

- **Expanded Casibase's capabilities** by integrating support for various LLMs, including open-source and commercial models for chat and embedding tasks.
- **Full-stack developement**, backend services in BeeGo and frontend interfaces in React.js, applying the MVC design pattern to ensure loose coupling and maintainable code.
- **Enhanced Casibase with multimodal support**, optimized output formatting, and bug fixes.
- **Optimized text splitting logic** to improve vectorized embedding for the RAG knowledge base.
- **Developed an instant messaging system** for multi-agent functionality.
- **Contributed 9,000+ lines** of code across the project.

IFLYTEK *Suzhou, China*

NLP Software Engineer *Jun. 2022 - Sep. 2022*

- **Contributed to NLP data annotation and quality assurance** for address data in the ”IFLYTEK Foresight” Police Super Brain System, focusing on improving the accuracy of location-based NLP tasks.
- **Re-labeled and refined address POI data (Point of Interest)** previously annotated by automated systems, enhancing data quality and addressing low-accuracy outputs generated by machine-based labeling.
- **Developed a program** to process and clean data points, significantly improving annotation efficiency and earning recognition from project leadership.
- **Explored entity relationship extraction techniques in NLP**, including Subject-Predicate-Object (SPO) extraction, and learned methods for building knowledge graphs from structured data, enhancing understanding of semantic representation in NLP.