Chaohao Yang

Email: cyang140@jh.edu Website: https://louiscroix.github.io

EDUCATION

Johns Hopkins University Aug. 2025 – Expected May 2027

- **Master of Science in Engineering in Computer Science**

The Chinese University of Hong Kong, Shenzhen Sept. 2021 – Jun. 2025

- **Bachelor of Engineering in Computer Science and Engineering**
- **Cumulative GPA:** 3.74/4.00 **Major GPA:** 3.86/4.00
- **Core Coursework:** Python Programming, C++ Programming, Calculus, Linear Algebra, Discrete Mathematics, Statistics, Data Structures, Computer Architecture, Operating System, Database, Parallel Programming, Software Engineering

University of California, Berkeley (USA) GLOBE Visiting Student Program Jan. 2024 – May 2024

- **GPA** 4.0/4.0 **Coursework**: Algorithm, Machine Learning, Artificial Intelligence

PUBLICATIONS

- Han, Y., Yang, C., Chen, C., Wang, X., & Sun, R. (2025). *Q-Adam-mini: Memory-efficient 8-bit quantized optimizer for large language model training*. ICML 2025 Workshop ES-FoMo-III.
  [co-first author]
- Yang, C., & Ding, C. (2024). Learning word embedding with better distance weighting and window size scheduling. https://doi.org/10.48550/arXiv.2404.14631
  [first author]
- Han, R., Peng, T., Yang, C., Wang, B., Liu, L., & Wan, X. (2023). Is information extraction solved by ChatGPT? An analysis of performance, evaluation criteria, robustness and errors. https://arxiv.org/abs/2305.14450v1
  [third author, cited **more than 140 times**]

RESEARCH EXPERIENCE

Research Assistant (Advisor: Associate Professor *Ruoyu Sun*)

The Chinese University of Hong Kong, Shenzhen Sept. 2024 – May 2025

Topic: Developing a quantized version for the emerging deep learning optimizer **Adam-mini**

- Pre-trained and fine-tuned multiple **Llama** models (**60M-8B** parameters) on the **C4**, **MMLU**, and **GSM-8K datasets** based on **AdamW** and **Adam-mini** optimizers and the **Low-Rank Adaptation (LoRA)** technique to provide experimental baselines
- Developed the quantized Adam-mini by inserting quantization and dequantization nodes, performing varied quantization schemes on different parameters, and adopting **stochastic rounding** to ensure effective optimizer updates
- Compared the performance of the quantized Adam-mini with baselines under the same settings, showed that the quantized Adam-mini achieved **comparable performance** with only **15%** of the optimizer state memory of AdamW

Research Assistant (Advisor: Presidential Chair Professor *Chris Ding*)

The Chinese University of Hong Kong, Shenzhen Jun. 2023 – Jan. 2024

Topic: Introducing distance information into the **Word2Vec** word embedding model

- Put forward the epoch-based dynamic window size strategy for the **Skip-gram** model to sample more from context words that are closer to the center word in a more stable manner, thus taking distance into consideration
- Conducted the formulated learnable distance-related weights for the average pooling of the **CBOW** model, combining the prior knowledge about distance information with the posterior adjustments of distance weights, taking both the modeling effect and adaptability into account

- Achieved **15.3%** accuracy improvement for CBOW and **2.5%** for Skip-gram on the **Google analogy test set**, demonstrating the effectiveness of the two proposed methods

Research Assistant (Advisor: Assistant Professor *Benyou Wang*)

The Chinese University of Hong Kong, Shenzhen Dec. 2022 – Jun. 2023

Topic: An empirical study on the information extraction ability of mainstream **large language models (LLMs)**

- Collected and cleaned **over 20** high-quality information extraction task datasets to evaluate the information extraction ability of LLMs including **GPT-4 Turbo** and **GPT-3.5 Turbo**, of which **16** datasets were selected for experiments
- Designed multi-threaded Python programs to efficiently send information extraction prompts to LLMs and collect responses, then monitored the program's execution
- Calculated **precision**, **recall**, **F1**, and other statistics for the information extraction ability of LLMs as evaluation metrics based on the collected responses

INTERNSHIP EXPERIENCE

Performance R&D Group Intern

Tencent TiMi Studio Group May 2024 – Aug. 2024

- Refined the term tables for Chinese-to-English, English-to-French, and English-to-German translations to support the overseas release of Tencent's **Honor of Kings (HOK)** and **Delta Force (DF)**
- Successfully trained translation models for the two games using **Llama-3-70B**, enhancing automation and efficiency in language localization for their international launches
- Introduced the **retrieval-augmented generation (RAG)** technique to the models to address term translation challenges and improve model performance, achieving **translation edit rates (TER) below 15** for HOK and **below 10** for DF, in line with company targets

CAMPUS ACTIVITIES

Committee Member of Mathematical Modeling Club

Mathematical Modeling Club Nov. 2022 – Jun. 2025

- Advised for the detailed process of club activities, such as site selection and participant guidance
- Offered information technology support for activity logistics, including automatic arrangements of personnel and activity-related software maintenance

Undergraduate Student Teaching Fellow

Graduate course: Artificial Intelligence Sept. 2023 – Dec. 2023

- Designed **5 homework projects** for the course, covering multiple aspects including word embedding and pre-trained language models, and scored the project completion of all **92 students** in the course
- Provided students with **10 tutorials** about the implementation details of AI models covered in lectures

AWARDS

- **Dean's List Certificates** in three academic years from 2021 to 2024 for excellent academic performance
- **Undergraduate Research Award** in 2023 for excellent research achievements
- **Student Union Certificate of Honor** in the 2022-2023 academic year for outstanding contribution to the Mathematical Modeling Club

SKILLS

- **Programming Languages:** Python, C, C++
- **Deep Learning Libraries:** PyTorch, OpenAI API, Hugging Face Libraries
- **Other Research Libraries:** NumPy, Pandas, NLTK