Yufei Gu

The Hong Kong University of Science and Technology (Guangzhou) | Email: ygu167@connect.hkust-gz.edu.cn
GitHub: Yufei-Gu-451 | Google Scholar: Yufei Gu | (+44) 07529968601 | (+86) 13122366091 | Homepage

EDUCATION

The Hong Kong University of Science and Technology (Guangzhou) - HKUST(GZ)

PhD in Data Science and Analytics

September 2025 - June 2029

University College London (UCL)

M.Eng. Computer Science [First Class Honour (71/100)]

September 2020 - June 2024

Acabridge College, Shanghai

AS & A-Levels [$4A^*$]

September 2018 - June 2020

INTERNSHIP

The Hong Kong University of Science and Technology (Guangzhou)

Research Assistant (Full-Time)

September 2024 - August 2025

Understanding and improving efficiency for LLM pretraining.

University College London Hospital (UCLH)

Research Assistant (Full-Time)

June 2024 - August 2024

Data collection and survival analysis for SBRT treatments.

Fudan University, NLP Group

Research Assistant (Remote)

June 2023 - June 2024

Biologically-plausible learning algorithms for ANNs/SNNs.

PUBLICATIONS

[ICLR' 2026] Late-to-Early Training: LET LLMs Learn Earlier, So Faster and Better

Second Author, HKUST-GZ [2025.1 - 2025.10]

- We propose that early LLM layers can be guided by later or lately trained layers, which can significantly accelerate training convergence.

[ICML' 2025] Investigating the Overlooked Hessian Structure: From CNNs to LLMs

Co-first Author, HKUST-GZ [2024.10-2025.4]

- We report the existence of a power-law Hessian structure in CNNs and LLMs with a maximum-entropy interpretation, and discovered that the structure can effectively predict generalization performance.

[ICLR' 2024] Unraveling the enigma of double descent: An in-depth analysis through the lens of learned feature space

First Author, UCL x Fudan University [2023.6 - 2023.9]

- We demonstrated a statistical relationship between a neural network's interpolation strategy on noisy data and its generalization performance.

[Neural Networks] SpikeCLIP: A contrastive language-image pre-trained spiking neural network

Fourth Author, Fudan University [2023.6 - 2024.6]

- We propose a two-step training algorithm for the SpikeCLIP model to effectively align multimodal features using spiked representations.

PROJECTS

[GitHub] Mano: Restriking Manifold Optimization for LLM Training

Personal Project, HKUST-GZ [2025.12 - Present]

- We proposed a SOTA optimizer, Mano, which is inspired by traditional manifold optimization techniques and improves the Pareto efficiency of LLM pretraining beyond AdamW and Muon.

[GitHub] Principal Spectral Regularization for LLM Pretraining

Joint Project, HKUST-GZ x CCF-Baidu Fund [2024.10 - 2025.10]

- We propose a novel principal spectral regularization method that selectively penalizes the dominant gradient components to investigate partial gradient orthogonalization in pretraining LLMs.

Biomedical Question-Answering System with LLM & RAG

Group Project, UCL [2024.1 - 2024.4]

- Established a Biomedical QA System by Retrieval Augmented Generation through a query-retrieval model built on PubMed.