

# Jianqiao Lu 陆剑桥

---

## About me

Currently pursuing a PhD under the supervision of [Prof. Zhiyi Huang](#), with a primary research focus on Large Language Models and Online Matching Theory.

I highlight my proficiency in fine-tuning and inference techniques for large models, which includes:

- Primarily responsible for evaluating open-source LLMs in two benchmark tasks (math & code), familiar with the usage of publicly available large models (e.g., qianwen, vicuna, chatglm, etc.), including their structures, prompt design, and specific inference details.
- Utilizing 'accelerate' and its plugin technologies like 'deepspeed zero' and 'fsdp' to partition training of billion-parameter models on GPU(V100, A100 & A800) chips.
- Employing 'vllm' to accelerate model inference. E.g., reducing the inference time for 750K test tasks from 16 hours to 2 hours.

## Email

jqlu@cs.hku.hk

## News

- **Sep 2023:** (First Author) ["Improving End-to-End Speech Processing by Efficient Text Data Utilization with Latent Synthesis"](#), was accepted to **EMNLP 2023**. (A approach that enhances E2E speech processing models by converting textual data into pseudo acoustic representations via Fixed-projection/Diffusion training, significantly improving performance on ASR and SLU tasks.)

## Under Review

### Large Language Models

- **Oct 2023:** (First Author) ["SELF: Self-Evolution with Language Feedback"](#), was released on arXiv. (A self-evolution framework for autonomous self-improvement of large language models through self-feedback and iterative self-refinement.)
- **Jan 2024:** (First Author) ["YODA: Teacher-Student Progressive Learning for Language Models"](#), was released on arXiv. (A learning framework that mimics human progressive learning to enhance language model fine-tuning.)
- **Jan 2024:** ["Planning, Creation, Usage: Benchmarking LLMs for Comprehensive Tool Utilization in Real-World Complex Scenarios"](#), was released on arXiv. (Responsible for all model testing in the paper including code and prompt design. A benchmark that evaluates LLMs' ability to plan, create, and apply tools in complex real-world scenarios, offering a comprehensive assessment beyond simple query benchmarks.)
- **Jan 2024:** (Equal Contribution) ["MHPP: Exploring the Capabilities and Limitations of Language Models Beyond Basic Code Generation"](#), will be released soon on arXiv. (Responsible for all model testing in the paper including code and prompt design. A benchmark to rigorously assess LLMs' code

generation abilities, revealing significant gaps in existing benchmarks and providing new insights into LLMs' strengths and weaknesses.)

## Online Matching

- **Jan 2024:** (Equal Contribution) "*Online Matching Meets Sampling Without Replacement*", will be released soon on arXiv. (This paper presents groundbreaking analyses demonstrating that sampling without replacement significantly enhances the competitiveness of online matching algorithms, offering both empirical evidence and novel theoretical insights.)

---

## Office

LG101, Chow Yei Ching Building, Department of Computer Science, The University of Hong Kong.

[[trivia](#)]