Yichuan Song

songyc@connect.hku.hk | Phone: (+86) 18501183737 | LinkedIn: /lovestaiga
The University of Hong Kong, HK SAR

EDUCATION

Bachelor - The University of Hong Kong Expected 05/2026

Department of Computer Science

• Overall GPA:	**3.57**	
• Research interests:	**Natural Language Processing, Energy Based Models, Model Explainability**	
• Awards and Honors:	Dean's Honours List	2021– 2022
	2nd Prize in RoboMaster University Championship 2022	08/2022
	3rd Prize in RoboMaster University League 2022	08/2022

RESEARCH AND PROJET

Research on Energy-based Masked Diffusion Reasoning (code, demo) 02/2025 – Current

Research Assistant; Supervised by Yilun Du, Harvard University

- Parameterized on masked diffusion model, train a sequence of annealed energy-based models with pseudolikelihood, and perform iterative Gibbs sampling to unmask during inference.
- Initially tested the training efficiency and inference performance on simple discrete reasoning tests like Sudoku

Research on Introspective Reasoning (code) 08/2024 – Current

Research Assistant; Supervised by Dr. Lingpeng Kong, University of Hong Kong, HKU NLP lab

- Probed on the hidden states from LLM to predict the accuracy of specified future reasoning steps, performed comprehensive analysis on several math, logic, and commonsense reasoning tasks.
- Tested latent-space intervention with the probing guidance in a plug-and-play manner.
- Plan to perform probing-guided rethinking and pausing in long-CoT to optimize the test-time scaling curve.
- To be submitted to EMNLP2025

Research on Model Merging 06/2024 – 08/2024

Research Assistant; Mentored by Yufeng Du, University of Illinois at Urbana-Champaign, Hao Peng's Lab

- Currently aimed to merge long-context and math reasoning capabilities into a model.
- Applied DARE as delta parameter dropping scheme for removing the redundant changes due to fine-tuning.
- Merged the reasoning capability from Eurus-7B-SFT to the base model Mistral-7B-v0.1.

Research on Probing Transformers (code, demo) 05/2024– Suspended

Research Assistant; Supervised by Dr. Xujie Si, University of Toronto

- Aimed to propose a new methods for interpreting the internal explainability of Transformer using probes.
- Verified the robustness of probes for the OthelloGPT, a pretrained minGPT for playing Othello games, using Marabou as the verification tool.
- Deduced some interesting properties on the robustness of the OthelloGPT's decoder and the probes by probing and logical expressions on Othello rules.
- Plan to generalize to more popular language models with Transformer encoders and their tasks.

Research on Improving DPO with Token Masks (code) 05/2024 – 06/2024

Research Assistant; Supervised by Jipeng Zhang, Hong Kong University of Science and Technology, Tong Zhang's Lab

- Aimed to improve the efficiency of DPO by selectively calculating the most relevant token's logits.
- Introduced masked tokens as inputs to the DPO training for Pythia-1B-deduped on UltraFeedback dataset.
- Explored some mechanisms on finding the most relevant token's logits used for masking
- Planned to test the efficiency of maskedDPO compared to the original methods.

ACADEMIC ACTIVITIES

Casual Helper for HKU Archaeology Team in the 2023 Summer's Excavation in Armenia 02/2023 – 05/2023

- Built GUIs and applied APIs for Canon cameras to automatically take photos of the shards and then store them in a file system.

Student Assistant for Course Engg1340 in HKU 01/2022 – 04/2023

EXTRACURRICULAR ACTIVITIES

- Member of HKU RoboMaster Team, Mechanical Group 09/2021 – 07/2023
- Guest in Revive Tech Asia 2022 and VXCON 2022 08/2022 – 09/2022
- Volunteer in Scaleup Impact Summit 2022 09/2022