Ziyang Luo

Curriculum Vitae

☎ *+86 18925068648*
✉ *cszyluo@comp.hkbu.edu.hk*
⌂ *https://chiyeunglaw.github.io/*

Education

2021 – 2025	**Ph.D. in Computer Science**, Hong Kong Baptist University, Hong Kong SAR, China.
2019 – 2021	**M.A. in Language Technology**, Uppsala University, Uppsala, Sweden.
2015 – 2019	**B.Sc. in Information and Computing Science**, Sun Yat-sen University, Guangzhou, China.

Research Interests

- **Code Intelligence**
- **(Large) Language Models**
- **Multimodal Machine Learing**

Representative Works

- **Code Intelligence and LLMs**: *WizardCoder: Empowering Code Large Language Models with Evol-Instruct*. [2024/01/04] We released WizardCoder-33B-V1.1, the **SOTA OSS Code LLM** on the EvalPlus Leaderboard. [2023/08/26] We released WizardCoder-Python-34B-V1.0, which achieves the 73.2 pass@1 and **surpasses GPT4 (2023/03/15), ChatGPT-3.5, and Claude2** on the HumanEval. [2023/06/16] Our WizardCoder-15B-V1.0 achieves 57.3 pass@1 score on HumanEval, more than 20 points higher than the SOTA open-source LLMs. ICLR, 2024.
- **Multimodal ML**: *LexLIP: Lexicon-Bottlenecked Language-Image Pre-Training for Large-Scale Image-Text Retrieval*. Our LexLIP introduces a new paradigm for image-text retrieval, which outperforms CLIP with **5.8x** faster retrieval speed and **19.1x** less index storage memory. ICCV 2023.
- **Language Models**: *Positional Artefacts Propagate Through Masked Language Model Embeddings*. We are the **first** to discover the outlier neurons phenomenon in the Transformer-based LMs. This is widely used to quantize transformers in recent works. ACL 2021.

Selected Publications

- Hongzhan Lin, **Ziyang Luo**, Wei Gao, Jing Ma, Bo Wang, and Ruichao Yang. *Towards Explainable Harmful Meme Detection through Multimodal Debate between Large Language Models.*. (**WWW**). 2024.
- **Ziyang Luo**, Can Xu, Pu Zhao, Qingfeng Sun, Xiubo Geng, Wenxiang Hu, Chongyang Tao, Jing Ma, Qingwei Lin and Daxin Jiang. *WizardCoder: Empowering Code Large Language Models with Evol-Instruct*. Proceedings of the Twelfth International Conference on Learning Representations. (**ICLR**). 2024.
- **Ziyang Luo**, Pu Zhao, Can Xu, Xiubo Geng, Tao Shen, Chongyang Tao, Jing Ma, Qingwei Lin and Daxin Jiang. *LexLIP: Lexicon-Bottlenecked Language-Image Pre-Training for Large-Scale Image-Text Retrieval*. Proceedings of the IEEE/CVF International Conference on Computer Vision. (**ICCV**). 2023.

- Hongzhan Lin*, **Ziyang Luo***, Jing Ma and Long Chen. *Beneath the Surface: Unveiling Harmful Memes with Multimodal Reasoning Distilled from Large Language Models*. Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP**). 2023.
- Hongzhan Lin, Pengyao Yi, Jing Ma, Haiyun Jiang, **Ziyang Luo**, Shuming Shi, and Ruifang Liu. *Zero-Shot Rumor Detection with Propagation Structure via Prompt Learning*. Proceedings of the 2023 AAAI Conference on Artificial Intelligence. (**AAAI**). 2023.
- **Ziyang Luo**, Yadong Xi, Rongsheng Zhang, GongZheng Li, Zeng Zhao and Jing Ma. *Conditioned Masked Language and Image Modeling for Image-Text Dense Retrieval*. Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (**EMNLP**). 2022.
- Zhiwei Yang, Jing Ma, Hechang Chen, Hongzhan Lin, **Ziyang Luo**, and Yi Chang. *A Coarse-to-fine Cascaded Evidence-Distillation Neural Network for Explainable Fake News Detection*. Proceedings of the 29th International Conference on Computational Linguistics. (**COLING**). 2022.
- **Ziyang Luo**, Yadong Xi, Jing Ma, Zhiwei Yang, Xiaoxi Mao, Changjie Fan and Rongsheng Zhang. *DecBERT: Enhancing the Language Understanding of BERT with Causal Attention Masks*. Findings of the 2022 North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL**). 2022.
- **Ziyang Luo**, Yadong Xi and Xiaoxi Mao. *Smoothing with Fake Label* . Proceedings of the 30th ACM International Conference on Information and Knowledge Management (**CIKM**). 2021.
- **Ziyang Luo**, Artur Kulmizev and Xiaoxi Mao. *Positional Artefacts Propagate Through Masked Language Model Embeddings*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (**ACL**). 2021.

Working Experiences

Internship

Ongoing	**Research Intern**, Language Technology Lab, Alibaba DAMO Academy.
2023.05-2024.01	**WizardCoder Core Contributor**, WizardLM Large Language Model Teams, Microsoft Research Lab - Asia, Work on Code Intelligence and Large Language Models research.
2022.11-2024.01	**Research Intern**, Data, Knowledge, and Intelligence Group, Microsoft Research Lab - Asia, Work on Multimodal ML and Large Language Models research.
2022.01-07	**Research Intern**, NLP Research Group, Fuxi AI Lab, NetEase, Inc.
2021.01-07	**Research Intern**, NLP Research Group, Fuxi AI Lab, NetEase, Inc.
2020.05-08	**Engineer Intern**, Jovi Chatbot Group, VIVO Communication Technology Co. Ltd.

Academic Services

- **Reviewer:**
  ACCV 2022, ECCV 2024, EMNLP 2022-23, AAAI 2023-24, ACL 2023, ACL Rolling Review 2023-24, Instruction Workshop @NeurIPS 2023
- **Student Volunteer:**
  EMNLP 2022