# TAN HAOCHEN

(+852)91460941 ⋄ haochetan2-c@my.cityu.edu.hk

## EDUCATION

**City University of Hong Kong** *September 2020 - Now*
PhD in Computer Science

**City University of Hong Kong** *September 2018 - October 2019*
MSc in Electronic Information

**University of Electronic Science and Technology of China** *September 2013 - June 2017*
Bachelor of Electronic Science and Technology

## RESEARCH INTERESTS

Large Language Model(LLM), Retrieval/Web Augmented Language Model, Long Text Generation.

## INTERNSHIP

**Noah's Ark Lab. Huawei, Shenzhen** July 2023 - Now
*Intern*

· R&D on Pangu-Web re-search mode.
· R&D on retrieval augmented language model.

**SenseTime, Beijing / Hong Kong** November 2021 - October 2022
*Intern*

· Meeting Summarization: Developed a Chinese meeting summrization dataset and an English meeting summarization framwork.
· LLM: Developed the prototype of Sensetime LLM.

**Noah's Ark Lab. Huawei, Hong Kong** March 2019 - Oct 2019
*Intern*

· Developed a real-time machine translation demo for internal use, which translate the result extracted from ASR system.

**City University of Hong Kong** November 2019 - November 2020
*Research Associate*

· A lightweight image segmentation framework.
· A Multi-turn dialogue rewriting system.

**Analog Devices, Beijing** March 2018 - Sep. 2018
*Intern*

· Development of DSP ADN8080.

## PUBLICATION

**Haochen Tan**, Zhijiang Guo, Zhan Shi, Lu Xu, Zhili Liu, Yunlong Feng, Xiaoguang Li, Yasheng Wang, Lifeng Shang, Qun Liu, and Linqi Song. Proxyqa: An alternative framework for evaluating long-form text generation with large language models. In *Main Conference of the ACL*, 2024

**Haochen Tan**, Zhijiang Guo, Lu Xu, Zhan Shi, Xiaoguang Li, Xin Jiang, Qun Liu, and Linqi Song. Serial vs parallel: Towards the efficiency of planning and action in multi-role llm system. 2024. Under Submission

**Haochen Tan**, Han Wu, Wei Shao, Mingjie Zhan, Zhaohui Hou, Ding Liang, and Linqi Song. Reconstruct before summarize: An efficient two-step framework for condensing and summarizing meeting transcripts. In *Main Conference of the EMNLP*, 2023

Han Wu, **Haochen Tan**, Mingjie Zhan, Gangming Zhao, Shaoqing Lu, Ding Liang, and Linqi Song. Learning locality and isotropy in dialogue modeling. In *ICLR*, 2023

Han Wu, **Haochen Tan**, Kun Xu, Shuqi Liu, Lianwei Wu, and Linqi Song. Zero-shot cross-lingual conversational semantic role labeling. In *Findings of NAACL*, 2022

**Haochen Tan**, Wei Shao, Han Wu, Ke Yang, and Linqi Song. A sentence is worth 128 pseudo tokens: A semantic-aware contrastive learning framework for sentence embeddings. In *Findings of the ACL*, 2022

Kun Xu, **Haochen Tan**, Linfeng Song, Han Wu, Haisong Zhang, Linqi Song, and Dong Yu. Semantic Role Labeling Guided Multi-turn Dialogue ReWriter. In *Main Conference of the EMNLP*, 2020

**Haochen Tan**, Huimin Shi, Mingquan Lin, J. David Spence, Kwok-Leung Chan, and Bernard Chiu. Vessel wall segmentation of common carotid artery via multi-branch light network. In *SPIE Medical Imaging: Image Processing*, 2020. (Oral)