

Hongjin Su

☎ (+852) 6571-1874 ✉ hjsu@cs.hku.hk 🏠 <https://hongjin-su.github.io/>

EDUCATION

The University of Hong Kong, Hong Kong *Sep 2022 - Present*
Ph.D. student in Computer Science
Advisor: Tao Yu

The Chinese University of Hong Kong, Hong Kong *Sep 2018 - Aug 2022*
B.Sc. in Computer Science

RESEARCH INTERESTS

- **Language Model Agents**: [4,5, 9, 10]
- **Embedding model and information retrieval**: [3, 6, 8]
- **Domain adaptation** [1, 2, 3, 7]

PUBLICATIONS (Google Scholar, * denotes equal contribution.)

- [10] Spider 2.0: Evaluating Language Models on Real-World Enterprise Text-to-SQL Workflows
Fangyu Lei, Jixuan Chen, Yuxiao Ye, Ruisheng Cao, Dongchan Shin, **Hongjin Su**, Zhaoqing Suo, Hongcheng Gao, Wenjing Hu, Pengcheng Yin, Victor Zhong, Caiming Xiong, Ruoxi Sun, Qian Liu, Sida Wang, Tao Yu
ICLR, 2025
- [9] Learn-by-interact: A Data-Centric Framework for Self-Adaptive Agents in Realistic Environments
Hongjin Su, Ruoxi Sun, Jinsung Yoon, Pengcheng Yin, Tao Yu, Serkan Ö. Arik
ICLR, 2025
- [8] BRIGHT: A Realistic and Challenging Benchmark for Reasoning-Intensive Retrieval
Hongjin Su*, Howard Yen*, Mengzhou Xia*, Weijia Shi, Niklas Muennighoff, Han-yu Wang, Haisu Liu, Quan Shi, Zachary S. Siegel, Michael Tang, Ruoxi Sun, Jinsung Yoon, Serkan O. Arik, Danqi Chen, Tao Yu
ICLR, 2025
- [7] EvoR: Evolving Retrieval for Code Generation
Hongjin Su, Shuyang Jiang, Yuhang Lai, Haoyuan Wu, Boao Shi, Che Liu, Qian Liu, Tao Yu
EMNLP, 2024
- [6] Generative representational instruction tuning
Niklas Muennighoff, **Hongjin Su**, Liang Wang, Nan Yang, Furu Wei, Tao Yu, Amanpreet Singh, Douwe Kiela
ICLR, 2025
- [5] Lemur: Harmonizing Natural Language and Code for Language Agents
Yiheng Xu*, **Hongjin Su***, Chen Xing*, Boyu Mi, Qian Liu, Weijia Shi, Binyuan Hui, Fan Zhou, Yitao Liu, Tianbao Xie, Zhoujun Cheng, Siheng Zhao, Lingpeng Kong, Bailin Wang, Caiming Xiong, Tao Yu
ICLR Spotlight, 2024
- [4] OpenAgents: An Open Platform for Language Agents in the Wild
Tianbao Xie*, Fan Zhou*, Zhoujun Cheng*, Peng Shi*, Luoxuan Weng*, Yitao Liu*, Toh Jing Hua, Junning Zhao, Qian Liu, Che Liu, Leo Z. Liu, Yiheng Xu, **Hongjin Su**, Dongchan Shin, Caiming Xiong, Tao

Yu
COLM, 2024

- [3] One Embedder, Any Task: Instruction-Finetuned Text Embeddings
Hongjin Su*, Weijia Shi*, Jungo Kasai, Yizhong Wang, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A. Smith, Luke Zettlemoyer, Tao Yu
ACL Findings, 2023
- [2] Selective Annotation Makes Language Models Better Few-Shot Learners
Hongjin Su, Jungo Kasai, Chen Henry Wu, Weijia Shi, Tianlu Wang, Jiayi Xin, Rui Zhang, Mari Ostendorf, Luke Zettlemoyer, Noah A. Smith, Tao Yu
ICLR, 2023
- [1] Taming Pre-trained Language Models with N-gram Representations for Low-Resource Domain Adaptation
Shizhe Diao, Ruijia Xu, **Hongjin Su**, Yilei Jiang, Yan Song, Tong Zhang
ACL, 2021

HONORS AND AWARDS

Faculty Dean's List, 2019, 2020, 2021, 2022
Wu Chung Scholarship, 2021
Pang Ching Cheung Scholarship, 2020
Elite Stream Scholarship, 2019, 2020

PROFESSIONAL SERVICE

Reviewer:

- ICLR 2024, 2025
- Neurips 2024
- EMNLP 2024
- NAACL 2025
- AAAI 2025

TEACHING EXPERIENCE

Essential skills for undergraduates: Foundations of Data Science, The University of Hong Kong Fall 2022
Teaching Assistant, Instructor: Tao Yu

Introduction to data structures and algorithms, The University of Hong Kong Fall 2023
Teaching Assistant, Instructor: Kao Chi Ming

Programming technologies and Computer programming, The University of Hong Kong . . . Spring 2025
Teaching Assistant, Instructor: Chenxiong Qian, T.W. Chim