

Langlin Huang

Email: h.langlin@wustl.edu | LinkedIn: langlin-huang | Tel: 206-643-8598 | GitHub: shrango.github.io

EDUCATION

•Washington University in St. Louis (WashU)

St. Louis, U.S.

Ph.D. Student in Computer Science

Aug. 2024 - Jun. 2029 (expected)

•Institute of Computing Technology, Chinese Academy of Sciences (ICT/CAS)

Beijing, China

M.E. in Computer Technology -GPA: 3.75/4.0, Outstanding Dissertation

Sep. 2021 - Jun. 2024

•University of International Business and Economics (UIBE)

Beijing, China

B.E. in Data Science and Big Data Technology -GPA: 3.74/4.0 (Rank:2/147)

Sep. 2017 - Jun. 2021

RESEARCH INTERESTS

- Language Model Reasoning & Efficient Model Training and Generation
- Multilingual Language Model

PUBLICATIONS

•PosS: Position Specialist Generates Better Draft for Speculative Decoding

Langlin Huang, Chengsong Huang, Jixuan Leng, Di Huang, Jiaxin Huang

Submitted to ICLR 2026 [Paper] [Code]

•MoCE: Adaptive Mixture of Contextualization Experts for Byte-based Neural Machine Translation

Langlin Huang, Mengyu Bu, Yang Feng

NAACL 2025 oral [Paper] [Code]

•Integrating Multi-scale Contextualized Information for Byte-based Neural Machine Translation

Langlin Huang, Yang Feng

ACL findings, 2024 [Paper][Code]

•Enhancing Neural Machine Translation with Semantic Units

Langlin Huang, Shuhao Gu, Zhuocheng Zhang, Yang Feng

EMNLP findings, 2023 [Paper][Code]

•CrossWordBench: Evaluating the Reasoning Capabilities of LLMs and LVLMs with Controllable Puzzle Generation

Jixuan Leng, Chengsong Huang, Langlin Huang, Bill Yuchen Lin, William W Cohen, Haohan Wang, Jiaxin Huang

COLM 2025 [Paper][Code] [Dataset]

•Divide, Reweight, and Conquer: A Logit Arithmetic Approach for In-Context Learning

Chengsong Huang, Langlin Huang, Jiaxin Huang

ICLR 2025 workshop [Paper][Code]

•Efficient test-time scaling via self-calibration

Chengsong Huang, Langlin Huang, Jixuan Leng, Jiacheng Liu, Jiaxin Huang

Preprint [Paper][Code]

•BayLing: Bridging Cross-lingual Alignment and Instruction Following through Interactive Translation for Large Language Models

Shaolei Zhang, Qingkai Fang, Zhuocheng Zhang, Zhengrui Ma, Yan Zhou, Langlin Huang, Mengyu Bu,

Shangdong Gui, Yunji Chen, Xilin Chen, Yang Feng

Preprint [Paper] [Code]

•Automatic Construction of a Depression-Domain Lexicon Based on Microblogs: Text Mining Study

Genghao Li, Bing Li, Langlin Huang, Sibing Hou

JMIR medical informatics, 2020, Vol 8. Jun. 2020 [Paper]

INDUSTRY EXPERIENCE

•Search Relevance Team, ByteDance

Jan. 2024 - Apr. 2024

Algorithm Engineer

Proposed an adapter module to bridge the relevance model used in production with an LLM. The adapter transformed representations from the traditional model into LLM-compatible hidden states, enabling:

- Joint utilization of pretrained traditional models and powerful LLM reasoning capabilities.
- Robust handling of out-of-vocabulary and multilingual inputs.
- A representation model plus LLM framework extensible to user-isolated data or multimodal input scenarios.

LANGUAGES PROFICIENCY

Master:Python, Shell, LaTeX, C++

Proficient:JAVA, R, C

COMMUNITY SERVICES

Top-reviewer: NeurIPS 2025

Reviewer: NeurIPS, ICLR, COLM, ACL, ACL Rolling Review, TALLIP