

Hongyi Liu

Email: hliu794@wisc.edu

ABOUT

As a PhD student of University of Wisconsin-Madison, I have great interest in machine learning, especially those involved with natural language processing (NLP). As the rapid development of large language models (LLMs), I am determined to explore the potential of language models and apply certain techniques to conquer various related real-world problems.

EDUCATION

University of Wisconsin - Madison (UWM)

2024 - 2029 (*Expected*)

Ph.D. in Computer Science

Shanghai Jiao Tong University (SJTU)

2020 - 2024

Bachelor of Engineering in Computer Science

Member of **ACM Class**, a selective CS program for the top 5% students.

- Major GPA: 3.88/4.30 (11/35)
- English Proficiency: TOEFL 104/120, CET-6 648/710

PUBLICATION

Named Entity Recognition Under Domain Shift via Metric Learning for Life Sciences *NAACL2024*

Hongyi Liu, Qingyun Wang, Payam Karisani, Heng Ji

Open-Domain Text Evaluation via Contrastive Distribution Methods *Preprint*

Sidi Lu, Hongyi Liu, Asli Celikyilmaz, Tianlu Wang, Nanyun Peng

Multilingual machine translation with large language models: Empirical results and analysis

Preprint

Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Shujian Huang, Lingpeng Kong, Jiajun Chen, Lei Li

RESEARCH EXPERIENCE

PLUS Lab @ University of California, Los Angeles

Supervised by Prof. Nanyun Peng

May 2023 - Feb 2024

- Mainly focus on evaluation metric of open ended machine generation.

Blender Lab @ University of Illinois Urbana-Champaign

Supervised by Prof. Heng Ji

May 2023 - Oct 2023

- Mainly focus on information extraction and transfer learning techniques in life science domains.

UCSB NLP Group @ University of California, Santa Barbaran

Supervised by Prof. Lei Li

Jul 2022 - April 2023

- Mainly focus on In-context Learning (ICL), machine translation, multi-task or multi-lingual potential of language models.

RESEARCH PROJECTS

Named Entity Recognition Under Domain Shift via Metric Learning for Life Sciences

- Classical information extraction tasks are challenging LLMs, especially for life science domains.
- We explore to boost the knowledge transfer ability from the biomedical domain to the chemical domain.

Open-Domain Text Evaluation via Contrastive Distribution Methods

- Build up a (meta) distribution over generative language models as evaluation metrics assessing language models' generation ability in regards to open-domain generation tasks.

Multilingual machine translation with large language models: Empirical results and analysis

- In-context learning (ICL) has been proved promising in NLU tasks but ICL performance on NLG tasks has not been sufficiently investigated.
- We conduct massive experiments in an attempt to compare ICL abilities in machine translation of neural models with varying structures and scales.

Java-and-C-like Language Compiler (Java, Course Project)

Github Repo can be found [here](#).

RISC-V CPU of Tomasulo Framework (Verilog, Course Project)

Github Repo can be found [here](#).

TEACHING

Teaching Assistant of COMP SCI 400 Fall

instructed by [Prof. Gary Dahl](#) & [Prof. Florian Heimerl](#)

Sep 2024 - Dec 2024

TA at University of Wisconsin - Madison

Major Teaching Assistant of Data Structure 2022 Spring

instructed by [Prof. Yong Yu](#)

Feb 2022 - Jun 2022

TA at Shanghai Jiao Tong University

Teaching Assistant of Program Design 2021 Fall

instructed by [Prof. Huiyu Weng](#)

Sep 2021 - Jan 2022

TA at Shanghai Jiao Tong University

HONORS AND AWARDS

Zhiyuan Honorary Scholarship, top 5% students

2021, 2022, 2023

SKILLS

- **Programming Languages:** C/C++ , Python , Java , Golang
- **Hardware:** Verilog
- **Math:** Calculus, Linear Algebra, Algebraic Structures, Mathematical Logic, Computational Complexity, Graph and Combinatorics