Haofei Yu

haofeiy@cs.cmu.edu | haofeiyu.me | github | linkedin | twitter

Education

Carnegie Mellon University   2022/08 – 2024/05 (Expected)
Master of Science in Intelligent Information Systems, GPA **4.14**/4.33   Pittsburgh, PA, USA
- Teaching Assistant for 11-777 Multimodal Machine Learning (2023 Fall)
- A+ Coursework: Natural Language Processing, Advanced Topics in Multimodal Machine Learning, etc

Zhejiang University   2018/09 – 2022/06
B.Eng in Computer Science and Technology (with Honors), GPA **3.96**/4.00   Hangzhou, China
- Rank: 7/134, Provincial Scholarship (top 5%)
- 95+/100 Coursework: Discrete Methematics and Application, Object-Oriented Programming, Computer Networks, etc

Publications

[1] **Haofei Yu**, Uri Alon, Graham Neubig. *Racoon: Ranked Code Generation.* Under review.

[2] Xuhui Zhou*, Hao Zhu*, Leena Mathur, Ruohong Zhang, **Haofei Yu**, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, Maarten Sap. *SOTOPIA: Interactive Evaluation for Social Intelligence in Language Agents.* Under review.

[3] **Haofei Yu**, Paul Pu Liang, Ruslan Salakhutdinov, Louis-Philippe Morency. *Mixture of Multimodal Interaction Experts.* NeurIPS UniReps workshop.

[4] **Haofei Yu***, Cunxiang Wang*, Yue Zhang, Wei Bi. *TRAMS: Training-free Memory Selection for Long-range Language Modeling.* Findings of EMNLP 2023.

[5] Leonie Weissweiler*, Valentin Hofmann*, Anjali Kantharuban, Anna Cai, Ritam Dutt, Amey Hengle, Anubha Kabra, Atharva Kulkarni, Abhishek Vijayakumar, **Haofei Yu**, Hinrich Schuetze, Kemal Oflazer, David R Mortensen. *Counting the Bugs in ChatGPT's Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model.* EMNLP 2023.

[6] Cunxiang Wang*, **Haofei Yu***, Yue Zhang. *RFiD: Towards Rational Fusion-in-Decoder for Open-Domain Question Answering.* Findings of ACL 2023.

[7] Chiyu Song*, Hongliang He*, **Haofei Yu**, Leyang Cui, Pengfei Fang, Zhenzhong Lan. *Uni-Encoder: A Fast and Accurate Response Selection Paradigm for Generation-Based Dialogue Systems.* Findings of ACL 2023.

Work Experience

Apple - Siri & Information Intelligence   2023/05 – 2023/08
Machine Learning Intern   Seattle, WA, USA
- Enhanced user satisfaction for Siri's long-range question answering and focused on complex, real-world user queries.
- Engineered a hierarchical prompting system to disambiguate and accurately respond to challenging questions.
- Chosen to present to Senior Director Robby Walker, showcasing intern project's potential to augment Siri.

Tencent - Tencent AI lab   2022/02 – 2022/08
Research Intern   Shenzhen, China
- Spearheaded a project to optimize inference efficiency in long-range language models, reducing perplexity on wikitext-103 by 0.19 through an innovative memory selection method without training[4].

Research Experience

Carnegie Mellon University - Language Technologies Institute   2022/09 – 2023/11
Research Assistant   Pittsburgh, PA, USA
- Worked with Prof. Graham Neubig for ranking-augmented code generation[1], Prof. Ruslan Salakhutdinov and Prof. Louis-Philippe Morency for social intelligence benchmark[2] and multimodal interaction[3].

Westlake University - School of Enginnering   2021/02 – 2022/02
Research Assistant   Hangzhou, China
- Worked with Prof. Zhenzhong Lan and Prof. Yue Zhang on open-domain natural language processing tasks including open-domain question answering[6] and open-domain dialogue generation[7].