Zichun Yu

5000 Forbes Avenue Pittsburgh, PA 15213

✉ zichunyu@andrew.cmu.edu | ⌂ yuzc19.github.io

Education

Carnegie Mellon University (CMU)

PHD STUDENT AT LANGUAGE TECHNOLOGIES INSTITUTE   *Aug. 2023 - Present*

- GPA: 4.0/4.0
- Core Courses: Large Language Models Methods and Application (A+), Multimodal Machine Learning (A+), On-Device Machine Learning (A+), Introduction to Machine Learning (A), Probabilistic Graphical Models (A)

Tsinghua University (THU)

B.ENG. IN COMPUTER SCIENCE AND TECHNOLOGY   *Sep. 2019 - Jul. 2023*

- GPA: 3.94/4.0, Ranking: 7/202
- Core Courses: Computer Organization (100, Top 1), Operating Systems (100, Top 1), Formal Languages and Automata (100, Top 1), Probability and Statistics (100, Top 1), Foundation of Object-Oriented Programming (99, Top 3)

Publication

- **Data-Efficient Pretraining with Group-Level Data Influence Modeling** [PDF]

  Zichun Yu, Fei Peng, Jie Lei, Arnold Overwijk, Wen-tau Yih, Chenyan Xiong

  Under Review.

- **MATES: Model-Aware Data Selection for Efficient Pretraining with Data Influence Models** [PDF][Talk]

  Zichun Yu, Spandan Das, Chenyan Xiong

  In *Proceedings of The 38th Annual Conference on Neural Information Processing Systems (NeurIPS'24)*.

- **Montessori-Instruct: Generate Influential Training Data Tailored for Student Learning** [PDF]

  Xiaochuan Li, **Zichun Yu**, Chenyan Xiong

  In *Proceedings of The 13th International Conference on Learning Representations (ICLR'25)*.

- **Augmentation-Adapted Retriever Improves Generalization of Language Models as Generic Plug-In** [PDF]

  Zichun Yu, Chenyan Xiong, Shi Yu, Zhiyuan Liu

  In *Proceedings of The 61st Annual Meeting of the Association for Computational Linguistics (ACL'23)*.

- **An In-depth Look at Gemini's Language Abilities** [PDF]

  Zichun Yu*, Syeda Nahida Akter*, Aashiq Muhamed*, Tianyue Ou*, Alex Bäuerle, Ángel Alexander Cabrera, Krish Dholakia, Chenyan Xiong, Graham Neubig

  arXiv preprint arXiv:2312.11444.

- **Automatic Label Sequence Generation for Prompting Sequence-to-sequence Models** [PDF]

  Zichun Yu, Tianyu Gao, Zhengyan Zhang, Yankai Lin, Zhiyuan Liu, Maosong Sun, Jie Zhou

  In *Proceedings of The 29th International Conference on Computational Linguistics (COLING'22)*.

Honors & Awards

2022	**First Prize (Top 1 in Academics and Research)**, THU Comprehensive Merit Scholarship	
2021	**First Prize (Top 5%)**, THU Academic Excellence Scholarship	
2020	**First Prize (Top 5%)**, THU Academic Excellence Scholarship	
2020	**First Prize (Top 5%)**, THU Academic Improvement Scholarship	