

Zichun Yu

4500 Centre Ave, Room 1504-B, Pittsburgh, Pennsylvania, 15213

☎ 412-500-3762 | ✉ 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+), Introduction to Machine Learning (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), Fundamentals of Search Engine Technology (98, Top 3), Fundamentals of Computer Graphics (98, Top 5)

Publication

• Augmentation-Adapted Retriever Improves Generalization of Language Models as a Zero-Shot Plug-In

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)*.

Research Experience

Augmentation-Adapted Retriever Improves Generalization of Language Models as a Zero-Shot Plug-In

RESEARCH ASSISTANT; ADVISOR: PROF. ZHIYUAN LIU, NATURAL LANGUAGE PROCESSING LAB, THU

Oct. 2022 - Jul. 2023

- Proposed the augmentation-adapted retriever (AAR), which is trained with the cross-attention signals from a source language model (LM), and plugged AAR into the target LMs in an unprecedented zero-shot manner to provide outer knowledge.
- Boosted the performance of target LMs of various sizes and architectures using AAR trained only with a tiny source LM.
- Revealed that training signals yielded from different source LMs overlapped significantly, bringing AAR generalization ability.

sciDataQA: Scientific Dataset Recommendation for Question Answering

RESEARCH ASSISTANT (REMOTE); ADVISOR: PROF. SHENG WANG, WANG LAB, UNIVERSITY OF WASHINGTON, SEATTLE

Jul. 2022 - Dec. 2022

- Proposed to use LLMs for in-context question generation grounded on Gene Expression Omnibus (GEO) database.
- Improved generative performance by leveraging a retrieval augmentation approach, adopting unsupervised contrastive learning on questions, and filtering the out-of-distribution instances.
- Validated the quality of our dataset by completing manual annotations, bi-citation question similarity tests, and comparisons between multiple question answering and recommendation algorithms.

Automatic Label Sequence Generation for Prompting Sequence-to-sequence Models

RESEARCH ASSISTANT; ADVISOR: PROF. ZHIYUAN LIU, NATURAL LANGUAGE PROCESSING LAB, THU

Apr. 2021 - Apr. 2022

- Introduced label sequences into seq2seq models such as T5 to enlarge the verbalization space of prompt-based learning.
- Proposed an automatic method to generate label sequences for prompt-based few-shot learning, including beam decoding on T5 and contrastive re-ranking across target classes, with minimal manual effort in designing the templates.
- Significantly improved the results of original T5 fine-tuning (for example, increased the accuracy by 9.4% on average) and achieved comparable results with expertly designed templates and label words on most tasks.

Open-Source Projects

zCore Asynchronous Scheduler

FINAL PROJECT OF OPERATING SYSTEMS; INSTRUCTOR: PROF. YU CHEN

Feb. 2022 - Jul. 2022

- Implemented a multi-thread and multi-core asynchronous scheduler for zCore, leveraging the low cost of context switch between coroutines and low latency of asynchronous syscalls.
- Explored an expeditious way to boot U740 (a RISC-V development board) via the network, and the efficiency of zCore asynchronous scheduler greatly outperformed Linux by 10% on that board, measured by Lmbench usec.
- Merged above contributions into the main branch of zCore, receiving over 1.4k GitHub stars.

SimSLS: Simple Contrastive Learning of Similar Law Case Search

FINAL PROJECT OF SEARCH ENGINE TECHNOLOGY; INSTRUCTOR: PROF. YIQUN LIU

Feb. 2022 - Jul. 2022

- Implemented Lawformer-based contrastive learning for similar law case search.
- Proposed to normalize and take the weighted average of Lawformer-based similarity and BM25 score, achieving new SOTA (NDCG@30 = 93.8) on the LeCaRD dataset.

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

Extracurricular Activity

- Developed the primary code for the 2021 THU Agent Competition.
- Organized various sporting events for students in the Department of Computer Science and Technology.
- Helped students in need for over 100 hours.

Skills

Languages	Chinese (Native), English (Fluent), Spanish (Basic)
English Fluency	TOEFL iBT 105 (Reading 28, Listening 27, Speaking 23, Writing 27) GRE Verbal 153 (59%), Quantitative 170 (96%), Analytical Writing 3.5 (37%)
Computer	C/C++, Python, Rust, Verilog, Java, JavaScript, Assembly, Git, Bash, GDB, SQL LaTeX, Markdown, Microsoft Office