Nan Zhang

njz5124@psu.edu | zn1010.github.io

EDUCATION

Pennsylvania State University, University Park, PA 08/2020-Present

- Ph.D. in Informatics (*Co-supervised by Prasenjit Mitra and Rui Zhang*) GPA: 3.91/4.0
- **Research Interests:** LLMs Compression, RAG, Summarization, NLP, Machine Learning

Georgia Institute of Technology, Atlanta, GA 08/2018-05/2020

- M.S. in Computational Science and Engineering GPA: 3.69/4.0

Worcester Polytechnic Institute (WPI), Worcester, MA 08/2013-05/2017

- BSc. in Computer Science & Industrial Engineering (double major) GPA: 3.5/4.0 (Honors with Distinction)

PUBLICATION

* denotes equal contribution

[11] **Nan Zhang**, Yusen Zhang, Prasenjit Mitra, Rui Zhang. When Reasoning Meets Compression: Benchmarking Compressed Large Reasoning Models on Complex Reasoning Tasks. Preprint.

[10] **Nan Zhang**, Prafulla Kumar Choubey, Alexander Fabbri, Gabriel Bernadett-Shapiro, Rui Zhang, Prasenjit Mitra, Caiming Xiong, Chien-Sheng Wu. SiReRAG: Indexing Similar and Related Information for Multihop Reasoning. International Conference on Learning Representations (ICLR), 2025.

[9] Jiangshu Du, Yibo Wang, Wenting Zhao, Zhongfen Deng, Shuaiqi LIU, Renze Lou, Henry Peng Zou, Pranav Narayanan Venkit, **Nan Zhang**, Mukund Srinath, Haoran Ranran Zhang, Vipul Gupta, Yinghui Li, Tao Li, Fei Wang, Qin Liu, Tianlin Liu, Pengzhi Gao, Congying Xia, Chen Xing, Cheng Jiayang, Zhaowei Wang, Ying Su, Raj Sanjay Shah, Ruohao Guo, Jing Gu, Haoran Li, Kangda Wei, Zihao Wang, Lu Cheng, Surangika Ranathunga, Meng Fang, Jie Fu, Fei Liu, Ruihong Huang, Eduardo Blanco, Yixin Cao, Rui Zhang, Philip S. Yu, Wenpeng Yin. LLMs Assist NLP Researchers: Critique Paper (Meta-)Reviewing. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[8] Ryo Kamoi, Yusen Zhang, **Nan Zhang**, Jiawei Han, Rui Zhang. When Can LLMs Actually Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs. Transactions of the Association for Computational Linguistics (TACL), 2024.

[7] Ryo Kamoi, Sarkar Snigdha Sarathi Das, Renze Lou, Jihyun Janice Ahn, Yilun Zhao, Xiaoxin Lu, **Nan Zhang**, Yusen Zhang, Haoran Ranran Zhang, Sujeeth Reddy Vummanthala, Salika Dave, Shaobo Qin, Arman Cohan, Wenpeng Yin, Rui Zhang. Evaluating LLMs at Detecting Errors in LLM Responses. Conference on Language Modeling (COLM), 2024.

[6] **Nan Zhang**, Yanchi Liu, Xujiang Zhao, Wei Cheng, Runxue Bao, Rui Zhang, Prasenjit Mitra, Haifeng Chen. Pruning as a Domain-specific LLM Exractor. Findings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL Findings), 2024.

[5] Yusen Zhang, **Nan Zhang**, Yixin Liu, Alexander Fabbri, Junru Liu, Ryo Kamoi, Xiaoxin Lu, Caiming Xiong, Jieyu Zhao, Dragomir Radev, Kathleen McKeown, Rui Zhang. Fair Abstractive Summarization of Diverse Perspectives. Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2024.

[4] **Nan Zhang***, Connor Heaton*, Sean Timothy Okonsky, Prasenjit Mitra, Hilal Ezgi Toraman. PEaCE: A Chemistry-Oriented Dataset for Optical Character Recognition on Scientific Documents. The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING), 2024.

[3] Guangji Bai, Zheng Chai, Chen Ling, Shiyu Wang, Jiaying Lu, **Nan Zhang**, Tingwei Shi, Ziyang Yu, Mengdan Zhu, Yifei Zhang, Carl Yang, Yue Cheng, Liang Zhao. Beyond Efficiency: A Systematic Survey of Resource-Efficient Large Language Models. Preprint.

[2] **Nan Zhang**, Yusen Zhang, Wu Guo, Prasenjit Mitra, Rui Zhang. FaMeSumm: Investigating and Improving Faithfulness of Medical Summarization. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[1] **Nan Zhang**, Shomir Wilson, Prasenjit Mitra. STAPI: An Automatic Scraper for Extracting Iterative Title-Text Structure from Web Documents. Language Resources and Evaluation Conference (LREC), 2022.

PROFESSIONAL EXPERIENCE

Research Intern, Salesforce AI Research, Palo Alto, CA 06/2024-09/2024

Supervised by: Prafulla Kumar Choubey, Alex Fabbri, Gabriel Bernadett-Shapiro, Chien-Sheng (Jason) Wu

- Proposed a novel indexing method for improving the question anwering performance of RAG pipelines
- Work has been accepted to ICLR 2025: "SiReRAG: Indexing Similar and Related Information for Multihop Reasoning"

Research Intern, NEC Labs America, Princeton, NJ 05/2023-08/2023

Supervised by: Yanchi Liu

- Compressed Large Language Models (LLMs) for efficient deployment in the enterprise environment
- Work has been accepted to Findings of NAACL 2024: "Pruning as a Domain-specific LLM Exractor"

Student Intern, BNP Paribas, New York, NY 10/2016-12/2016

- Devised the front-end via AngularJS and constructed the SQL database in Microsoft SQL Server

Software Development Intern, Fidelity Investments, Smithfield, RI 05/2016-08/2016

- Used Spring Framework to redesign a web application and created JUnit Test Cases for its Java utility classes

Media Mentions

- Aqsa Younas Rana. "Revolutionizing Medical Summaries: How Penn State's New AI Framework Promises Accuracy and Reliability" **(BNN, Feb. 22, 2024)**
- Mary Fetzer. "Improving efficiency, reliability of AI medical summarization tools" **(Penn State News, Feb. 21, 2024)**

TALKS AND PRESENTATIONS

- "Medical Summarization on Factuality and Readability with Large Language Models". Poster, The 10th Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2023). April 22, 2023
- "Transformer-based Optical Character Recognition Model for Chemistry Scholarly Papers". Poster, Penn State University ICDS Fall 2022 Symposium. October 12, 2022.

SERVICES

- Program Committee of SUKI (Structured and Unstructured Knowledge Integration) @ NAACL 2022
- Reviewer, ACL Rolling Review Dec. 2022
- Review Committee of CoNLL 2023 and 2024
- Reviewer, AI4Research Workshop @ IJCAI 2024
- Reviewer, EMNLP 2023, NeurIPS 2024, ICLR 2025, COLM 2025