RUIYANG QIN

Address: 222 Cushing Hall, Notre Dame, IN 46556 Email: [rqin@nd.edu]
Homepage: https://ruiyangqin2016.github.io/ Phone: (+1)404-740-6640

RESEARCH INTERESTS

Edge AI, Emerging Technologies, Large Language Model (LLM)

Objectives: *My research focuses on edge AI and emerging technologies. I conduct my research from two directions. First, given the difficulties of deploying advanced AI models like LLMs on edge platforms, I utilize emerging technologies like (in/near-memory computing) to optimize and accelerate these AI models' on-device learning and inference, to create the next-generation advanced edge AI (i.e. LLM on edge) regarding personalization and data privacy. Additionally, to better enable emerging technologies to assist edge AI development, I co-design the advanced-edge-AI-friendly emerging technologies such as accelerating the large-scale CiM circuit simulation via ML methods.*

EMPLOYMENT

SUNY at Buffalo Buffalo, New York
Visiting Ph.D. student, Computer Science and Engineering Aug. 2024 - Present
Supervisor: Jinjun Xiong

University of Notre Dame Notre Dame, Indiana
Ph.D. Student, Computer Science and Engineering Aug. 2022 - Present
Advisor: Yiyu Shi

EDUCATION

Georgia Institute of Technology Atlanta, Georgia
M.S. in Computer Science 2021
B.S. in Computer Science 2020

REVIEW SERVICES

- IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)
- IEEE Intelligent Systems
- Conference on Neural Information Processing Systems (NeurIPS)
- Association for the Advancement of Artificial Intelligence (AAAI)

AWARDS

- Edison Innovation Fellow, 2024, USA
- DAC Young Fellow, 2024, USA
- DAC Young Fellow, 2023, USA

TEACHING

- CSE 30151 - Theory of Computing, Notre Dame, Spring 2023
- CSE 30246 - Database Concepts, Notre Dame, Fall 2022
- CS 6476 - Computer Vision, Georgia Tech, Spring 2021

RESEARCH GRANT

- Participate as PhD student, "On-device Large Language Model Personalization with Algorithm-Hardware Co-design for Healthcare Applications", National Science Foundation, $569,269

RESEARCH GIFT

- Gemma Academic Program (GCP) by Google, 07/21/2024 - 07/21/2025, $15,000

PUBLICATIONS

Conferences

- **Ruiyang Qin**, Zheyu Yan, Dewen Zeng, Zhenge Jia, Dancheng Liu, Jianbo Liu, Zhi Zheng, Ningyuan Cao, Kai Ni, Jinjun Xiong, Yiyu Shi, *Robust Implementation of Retrieval-Augmented Generation on Edge-based Computing-in-Memory Architectures*, The 43st IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2024 (acceptance rate 24.2%)
- **Ruiyang Qin**, Jun Xia, Zhenge Jia, Meng Jiang, Ahmed Abbasi, Peipei Zhou, Jingtong Hu, Yiyu Shi, *Enabling On-Device Large Language Model Personalization with Self-Supervised Data Selection and Synthesis*, The 61st IEEE/ACM Design Automation Conference (DAC), 2024 (acceptance rate 21.8%)
- **Ruiyang Qin**, Yuting Hu, Zheyu Yan, Jinjun Xiong, Ahmed Abbasi, Yiyu Shi, *FL-NAS: Towards Fairness of NAS for Resource Constrained Devices via Large Language Models*, The 29th Asia and South Pacific Design Automation Conference (ASP-DAC), 2024(acceptance rate 31%)
- Haozheng Luo, **Ruiyang Qin**, Chenwei Xu, Guo Ye, Zening Luo, *Open-Ended Multi-Modal Relational Reasoning for Video Question Answering*, The 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2023
- Zhiding Liang, Zhixin Song, Jinglei Cheng, Zichang He, Ji Liu, Hanrui Wang, **Ruiyang Qin**, Yiru Wang, Song Han, Xuehai Qian, Yiyu Shi, *Hybrid gate-pulse model for variational quantum algorithms*, The 60th IEEE/ACM Design Automation Conference (DAC), 2023 (acceptance rate 22.7%)

Journal

- **Ruiyang Qin**[*], Ryan Cook[*], Kai Yang, Ahmed Abbasi, David Dobolyi, Salman Seyedi, Emily Griner, Hyeokhyen Kwon, Robert O. Cotes, Zifan Jiang, and Gari D. Clifford, *Language Models for Online Depression Detection: A Review and Benchmark Analysis on Remote Interviews*, ACM Transactions on Management Information Systems (TMIS)
- Marialena Bevilacqua, Kezia Oketch, **Ruiyang Qin**, Will Stamey, Xinyuan Zhang, Yi Gan, Kai Yang, Ahmed Abbasi, *When Automated Assessment Meets Automated Content Generation: Examining Text Quality in the Era of GPTs*, ACM Transactions on Information Systems (TOIS)

Under Review

- **Ruiyang Qin**[*], Dancheng Liu[*], Zheyu Yan, Zhaoxuan Tan, Zixuan Pan, Zhenge Jia, Meng Jiang, Ahmed Abbasi, Jinjun Xiong, Yiyu Shi, *Empirical Guidelines for Deploying LLMs onto Resource-constrained Edge Devices*, The 38th Conference on Neural Information Processing Systems (NeurIPS), 2024