Yuzhu Mao
https://yz-mao.github.io
Email: myz20@tsinghua.org.cn
GitHub: https://github.com/yz-mao

Education

- **Tsinghua University** — Beijing, China
  M.S. in Data Science and Information Technology; GPA: 3.85 / 4.00 — *Sep. 2020 – Jun. 2023*
- **Wuhan University** — Wuhan, China
  B.E. in Information Security; GPA: 3.87 / 4.00 (Graduate with Distinction) — *Sep. 2016 – Jun. 2020*

Research Experience

- **Tsinghua-UC Berkeley Shenzhen Institution (TBSI)** — Shenzhen, China
  Research Assistant, Advisor: Prof. Wenbo Ding — *Sep. 2023 - Jun. 2024*
  - **Efficient Large Language Models (LLMs)**: Explored quantization and low-rank adaptation techniques to reduce the memory and computational costs of LLMs for deployment on resource-constrained platforms.
  - **Example Code Contributions**: https://github.com/yz-mao/Quantized-LoRA-Finetuning-of-GPT-2
- **Tsinghua University** — Beijing, China
  Thesis-based Master Student, Advisor: Prof. Wenbo Ding & Prof. Yang Liu — *Aug. 2020 - Jun. 2023*
  - **Efficient and Reliable Federated Learning**: Developed adaptive quantization and sparsification methods to reduce communication overhead in federated learning systems.
- **Wuhan University** — Wuhan, China
  Honored Undergraduate with Distinct Undergraduate Thesis of Wuhan University — *Jan. 2020 - Jun. 2020*
  - **Transformer-based Text-to-Speech Synthesis and Detection**: Developed an LSTM-based algorithm utilizing inter-frame audio features to achieve over 90% accuracy in detecting synthesized speech.

  Second prize, The National Information Security Contest for College Students — *Jan. 2019 - Aug. 2019*
  - **GAN-based Deepfake Video Generation and Detection**: Utilized temporal and spatial features from video streams for deepfake video detection, achieving over 96% accuracy.

Work Experience

(Three of the Top 10 Internet Technology Companies in China)

- **Alipay.com Co., Ltd.** — Hangzhou, China
  Researcher & Senior Engineer, Alipay Technology Group — *Jul. 2024 - Present*
  - **Artificial Intelligence Generated Content (AIGC)**: Participated in building the BaiLing foundation model for AIGC production and evaluation.
  - **Mobile Large Language Models (Mobile LLMs)**: Quantized and deployed fine-tuned large language models on mobile devices for personalized and real-time user responses.
- **Meituan Technology** — Shenzhen, China
  Research Intern, UAV Group, Manager: Dr. Tianjian Chen — *Jun. 2022 - Sep. 2022*
  - **Texture Generation**: Empowered a 2D to 3D platform for large-scale UAV simulations by creating realistic textures from 2D images.
- **Tencent Technology** — Shenzhen, China
  Algorithm Engineer Intern, Robotics-X Lab, Manager: Dr. Cheng Zhou — *Jun. 2021 - Sep. 2021*
  - **Optimization Acceleration with GPU**: Equipped an open-source JAX-based rigid body dynamics algorithm library with GPU support, enabling faster and more efficient computations.
  - **Code Contributions**: https://github.com/Tencent-RoboticsX/jbdl

Key Courses

- **Graduate Studies** — Sep. 2020 - Jun. 2023
  - Learning from Data, Optimization Theory and Machine Learning, Advanced Signal Processing: *4.0/4.0*
- **Undergraduate Studies** — Sep. 2016 - Jun. 2020
  - Data Structures, Probability Theory and Statistics, Operating Systems, Database Principles and Security, Pattern Recognition: *4.0/4.0*

Publications

(* denotes equal contribution)

REFEREED JOURNAL ARTICLES

[1] SAFARI: Sparsity-enabled Federated Learning with Limited and Unreliable Communications
Yuzhu Mao*, Zihao Zhao*, Meilin Yang, Le Liang, Yang Liu, Wenbo Ding, Tian Lan, Xiao-Ping Zhang
*IEEE Transactions on Mobile Computing (**TMC**)*, 2023.

[2] AQUILA: Communication-efficient Federated Learning with Adaptive Quantization in Device Selection Strategy
Zihao Zhao*, **Yuzhu Mao***, Zhenpeng Shi, Yang Liu, Tian Lan, Wenbo Ding, Xiao-Ping Zhang
*IEEE Transactions on Mobile Computing (**TMC**)*, 2023.

[3] Towards Efficient Communications in Federated Learning: A Contemporary Survey
Zihao Zhao, **Yuzhu Mao**, Yang Liu, Linqi Song, Ye Ouyang, Xinlei Chen, Wenbo Ding
Journal of the Franklin Institute, 2023.

[4] Communication-efficient Federated Learning with Adaptive Quantization
Yuzhu Mao, Zihao Zhao, Guangfeng Yan, Yang Liu, Tian Lan, Linqi Song, Wenbo Ding
*ACM Transactions on Intelligent Systems and Technology (**TIST**)*, 2022.

CONFERENCE PROCEEDINGS

[5] FL-TAC: Enhanced Fine-tuning in Federated Learning via Low-rank, Task-specific Adapter Clustering
Siqi Ping*, **Yuzhu Mao***, Yang Liu, Xiao-Ping Zhang, Wenbo Ding
*International Conference on Learning Representations (**ICLR**) Workshop on Large Language Model (LLM) Agents*, 2024.

[6] FormerReckoning: Physics Inspired Transformer for Accurate Inertial Navigation
Jiaqi Li, Chenyu Zhao, **Yuzhu Mao**, Xinlei Chen, Wenbo Ding, Xiaoyang Qu, Jianzong Wang
*International Workshop on Physics Embedded AI Solutions in Mobile Computing (**MobiCom** Picasso Workshop)*, 2024.

UNDER REVIEW

[7] Enhancing Parameter Efficiency and Generalization in Large-scale Models: A Regularized and Masked Low-rank Adaptation Approach
Yuzhu Mao, Siqi Ping, Zihao Zhao, Yang Liu, Wenbo Ding
arXiv preprint arXiv:2407.12074, 2024.

Awards and Honors

- Tsinghua Graduate Scholarship for Excellent Academic Performance (2020-2021 and 2021-2022, First-class, **Top 3%**)
- Wuhan University Scholarship for Outstanding Undergraduates (2020, **Top 3%**)
- **National Cyber Security Scholarship** (2019, **Top 1%**)
- **National Scholarship** (2018, **Top 1%**)
- Wuhan University Scholarship for Overseas Exchange (2018, **Top 3%**)
- Wuhan University Scholarship for Outstanding Students (2016-2017, 2017-2018, and 2018-2019, First-class, **Top 3%**)

Programming Skills

- **Tools**: PyTorch, TensorFlow, Git, Linux, SQL **Languages**: Python, C, C++, MATLAB, LaTex

Language Skills

- **English**: IELTS Overall Band 8.0; TOEFL 109 (R28, L29, S24, W28)