

Haoxiang (Steven) Yu, Ph.D.

hxyu@utexas.edu | stevenyu.xyz

SUMMARY

Machine Learning Research Scientist specializing in Large Language Models (LLMs), multimodal AI, and distributed learning systems. Ph.D. from UT Austin with 15+ peer-reviewed publications. Currently conducting **Multimodal and Contextual AI** research at Meta Reality Labs, developing multimodal models with contextual adaptation for real-time applications. Track record of Best Paper Awards at IMWUT(Ubicomp) and other conferences.

Skills: Python, PyTorch, (m)LLM Post-training & Alignment, Distributed Learning, Resource Constrained Inference

PROFESSIONAL EXPERIENCE

Meta Platforms, Inc. | Research Scientist (Multimodal and Contextual AI) May 2025 - Present

- Conduct research and development of multimodal AI models for smart glasses and other wearable devices from concept to production and deploy on 2M+ devices
- Design the architecture of low-latency machine learning systems for resource-constrained devices, utilizing quantization and structural pruning to reduce memory footprint while maintaining SOTA inference benchmarks
- Develop multimodal large language models (mLLMs) with post-training contextual adaptation to enhance real-time interaction and situational understanding

Meta Platforms, Inc. | Software Engineer Intern (MLE - Wearables) May 2024 - Aug 2024

- Developed multimodal ML models integrating sensor data for wearable AR devices
- Designed LLM-powered pipeline leveraging foundation models to accelerate research prototyping

Toyota InfoTech. Center Co., Ltd. | Research & Development Intern/Researcher Jan 2023 - May 2025 (Seasonal)

- Designed and implemented next-generation federated and decentralized learning systems to maximize resource utilization for automated and assistive driving infrastructure
- Researched machine learning, optimization techniques, infrastructure, and data platforms to support autopilot and related systems, exploring the future of mobility. These innovations directly resulted in 4 filed US patents

The University of Texas at Austin | Graduate Teaching/Research Assistant Aug 2021 - May 2025

- Pioneered research on LLM-based reasoning for smart environments and developed novel decentralized machine learning algorithms and systems
 - Engaged in an NSF-funded project and collaborated with industry partners (such as Toyota and Tektronix)
 - **Awards: Distinguished Paper Award (IMWUT 2025), Best Paper Award (IEEE MASS 2024).**
-

EDUCATION

Ph.D., Electrical and Computer Engineering, The University of Texas at Austin Aug 2021 - Aug 2025

Research Areas: Decentralized Learning & Inference in Resource-Constrained Environments

Supervisor: Dr. Christine Julien/Dr. Haris Vikalo

M.S./B.S., Computer Science, Miami University Aug 2017 - May 2021

SELECTED PUBLICATIONS & PATENTS

1. King, Evan, **Haoxiang Yu**, Sangsu Lee, and Christine Julien. "Sasha: creative goal-oriented reasoning in smart homes with large language models" Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies
2. King, Evan, **Haoxiang Yu**, Sahil Vartak, Jenna Jacob, Sangsu Lee, and Christine Julien. "Thoughtful Things: Building Human-Centric Smart Devices with Small Language Models." arXiv preprint arXiv:2405.03821
3. **Yu, Haoxiang**, Vaskar Raychoudhury, Snehanshu Saha, Md Osman Gani, Janick Edinger, and Roger Smith, "Automated Surface Classification System using Vibration Patterns - A Case Study with Wheelchairs" IEEE Transactions on Artificial Intelligence
4. **Yu, Haoxiang**, Jingyi An, Evan King, Edison Thomaz, and Christine Julien. "Cheating off your neighbors: Improving activity recognition through corroboration" arXiv preprint arXiv:2306.06078
5. **Yu, Haoxiang**, Jie Hua, and Christine Julien, "Analysis of IFTTT Recipes to Study How Humans Use Internet-of-Things (IoT) Devices" Proceedings of the 4th Workshop on Data Acquisition To Analysis (at BuildSys). 2021
6. **Yu, Haoxiang**, Hsiao-Yuan Chen, Sangsu Lee, Sriram Vishwanath, Xi Zheng, and Christine Julien. "IDML: Incentivized decentralized machine learning", Proceedings of the 21st IEEE International Conference on Mobile Ad-Hoc and Smart Systems. 2024
7. **Yu, Haoxiang**, Evan King, Chianing Wang, and Alex Pham. "Systems and methods for distributed machine learning with less vehicle energy and infrastructure cost" Application US20250254502A1, USPTO
8. **Yu, Haoxiang**, and Chianing Wang. "Systems and methods for predicting presence of objects using decentralized data collection and map databased information compression" Application US20250053785A1, USPTO