Yifei Zhang

Email: yifeizhang556@gmail.com | Tel: (919) 638-0069 | Personal Page: yifeizhangcs.github.io/

Education

Ph.D. in Computer Science, Emory University — Aug 2022 – May 2026 (expected)
M.S. in Data Science, Columbia University — Sep 2020 – Feb 2022
Master of Engineering Management, Duke University — Aug 2019 – May 2020
B.E. in Engineering Mechanics, Dalian University of Technology — Sep 2015 – Jun 2019

Research Interests

Explainable AI, LLMs Distillation, LLMs Evaluation, Language Agents, Multimodal LLMs

Selected Publications

[KDD 2025] **Yifei Zhang**, James Song, Siyi Gu, Tianxu Jiang, Bo Pan, Guangji Bai, and Liang Zhao. *Saliency-Bench: A Comprehensive Benchmark for Evaluating Visual Explanations.* The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Datasets and Benchmarks Track.

[ACL 2024] **Yifei Zhang**, Bo Pan, Chen Ling, Yuntong Hu, and Liang Zhao. *ELAD: Explanation-Guided Large Language Models Active Distillation.* Findings of The 62nd Annual Meeting of the Association for Computational Linguistics.

[IJCAI 2024] **Yifei Zhang**, Bo Pan, Siyi Gu, Guangji Bai, Meikang Qiu, Xiaofeng Yang, and Liang Zhao. *Visual Attention Prompted Prediction and Learning.* International Joint Conference on Artificial Intelligence.

[ICCV 2023] **Yifei Zhang**, Siyi Gu, Yuyang Gao, Bo Pan, Xiaofeng Yang, and Liang Zhao. *MAGI: Multi-Annotated Explanation-Guided Learning.* The 36th International Conference on Computer Vision.

[KDD 2023] Siyi Gu*, **Yifei Zhang*** (equal contribution), Yuyang Gao, Xiaofeng Yang, and Liang Zhao. *ESSA: Explanation Iterative Supervision via Saliency-guided Data Augmentation.* The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Research Track.

Work Experience

Amazon Alexa AI — Boston, MA, US
Applied Scientist Intern — *May 2025 – Aug 2025*

- Proposed and developed **ATOD-Eval**, a holistic evaluation framework for advanced task-oriented dialogue, featuring agentic memory with dual-store design and turn-level goal status tracking.
- Constructed **A-TOD**, a large-scale synthetic dataset capturing multi-goal concurrency, dependencies, long-horizon memory, asynchrony, and proactivity, enabling reliable benchmarking.
- Conducted extensive experiments showing significant improvements over strong LLM- and memory-based baselines on goal detection, status tracking, and dependency handling.

Amazon AGI — Boston, MA, US
Research Scientist Intern — *Jun 2024 – Sep 2024*

- Developed a novel multitask framework integrating contrastive reward-style outputs with Likert scale ratings to enhance the evaluation of LLM-driven smart speaker interactions.
- Designed an innovative method for generating synthetic preference data using LLMs, addressing the scarcity of training data and improving evaluation accuracy in speaker-based environments.
- Successfully deployed the multitask evaluation framework in production models for smart speaker systems.

Research Experience

Agentic Explainable Reasoning for Joint Lung Cancer and Cardiovascular Risk Assessment
Supervisors: Prof. Liang Zhao, Department of Computer Science, Emory University — *Nov 2024 – Present*

- Designed a framework leveraging low-dose chest CT (LDCT) to jointly assess lung cancer and CVD risk with agentic explainable reasoning.
- Built an indicator reasoning module to extract pulmonary findings (e.g., emphysema) and map them to cardiovascular pathways.
- Integrated a heart-centered CVD module conditioned on lung risk and indicator reasoning for cross-disease knowledge transfer.
- Validated on screening cohorts, achieving statistically significant gains in CVD risk discrimination over image-only baselines.

Multimodal Explanation-Guided Learning with Large Language Models
Supervisors: Prof. Liang Zhao, Department of Computer Science, Emory University — *Mar 2024 – Oct 2024*

- Developed the MEGL framework to integrate visual and textual explanations, enhancing classification accuracy and model interpretability.
- Proposed the Saliency-Driven Textual Grounding (SDTG) method to align visual and textual explanations through multimodal learning.
- Introduced the Visual Explanation Distribution Consistency loss to address incomplete annotations, leveraging multimodal large language models (LLMs) for robust explanation generation.

Skills

Programming: Python, C/C++, MATLAB, Java, Shell
Frameworks: PyTorch, TensorFlow, Transformers, DeepSpeed
Systems: Linux, Git, Slurm, AWS