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, Explanation-Guided Learning, Large Language Models Distillation, Large Language Models Evaluation, Multimodal Large Language Models, Medical Imaging

Publications

[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***, 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.

Work Experience

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, and prepared findings for submission to a top-tier NLP conference.

Guotai Junan Securities Beijing, China
Quantitative Analyst Intern *May 2019 – Aug 2019*

- Analyzed 30 years of China's quarterly GDP using time-series decomposition techniques like STL and SEATS, while correlating GDP trends with major economic events.
- Designed an LSTM network with Keras to predict the Shanghai Composite Index, fine-tuning parameters for optimal performance, and achieved a 43% reduction in MAE compared to the ARIMA model.

Research Experience

Explanation-Guided Large Language Models Efficient Distillation
Supervisor: Prof. Liang Zhao, Department of Computer Science, Emory University *Sep 2023 – Feb 2024*

- Developed the ELAD framework, achieving efficient knowledge distillation from large language models (LLM) to smaller models through active learning, balancing annotation costs and model performance.
- Introduced a sample selection method based on generative explanations, which accurately identifies and prioritizes samples with high uncertainty in active learning, significantly enhancing the efficiency of knowledge distillation.
- Proposed a customized explanation correction technique, enabling the teacher LLM to specifically detect and correct reasoning errors in student models, improving the quality and reliability of distillation.

Enhance Image Recognition Performance via Multi-Annotated Explanation Supervision.
Supervisor: Prof. Liang Zhao, Department of Computer Science, Emory University *Oct 2022 – Mar 2023*

- Developed an innovative framework for explanation supervision trained in a multi-task manner, leveraging class labels and integrating multiple explanation annotations, dynamically weighted for each annotator for optimal results.
- Introduced a new generative model designed to fill in missing annotations, utilizing variational inference that adapts to the individual characteristics of each annotator during annotation generation.
- Proposed a unique alignment mechanism integrated into the generative model to learn the alignment between annotations and annotators during training, transforming the inference challenge into a linear sum assignment problem.

Improve model predictability through Explanation-guided Supervision and Data Augmentation.
Supervisor: Prof. Liang Zhao, Department of Computer Science, Emory University *Aug 2022 – Feb 2023*

- Introduced a novel framework that integrates explanation supervision with adversarial-trained data augmentation, enhancing image augmentation through iterative interplay.
- Developed an "annotation-to-image" generator with dual decoders, capturing distinct foreground and background patterns for realistic, multi-mapping image generation.