Yaohan Guan

⚲ Baltimore, USA   ✉ yguan19@jhu.edu   ☎ +1-(410)-624-6290

Education

Johns Hopkins University	*Aug. 2024 – Now*
Ph.D. in Electrical Engineering	
Johns Hopkins University	*Aug. 2022 – May. 2024*
M.S. in Electrical Engineering	
Tsinghua University	*Sept. 2018 – Jul. 2022*
B.E. in Automation	

Research Interests

Large Language Models, Multimodal Emotion Recognition, Voice Privacy

Publications

[1] Thomas Thebaud*, Anna Favaro*, **Yaohan Guan***, *et al.*, and Najim Dehak. "Multimodal Emotion Recognition Harnessing the Complementarity of Speech, Language, and Vision," *In Proceedings of the International Conference on Multimodal Interaction (ICMI)*, 2024.

Research Experiences

LLM in Figure Generation   *Oct. 2024 – Now*
Advisor: Prof. Daniel Khashabi, Johns Hopkins University
- Aimed to explore LLMs' capabilities in generating the most critical visual element (Figure 1) for academic papers by tool calling, and push their limits to outperform other state-of-the-art text-to-image models.
- Constructing the benchmark from ACL papers, and creating a novel LLM Figure 1s generation pipeline by generating figure layouts, and iteratively rendering the layouts using diffusion models, emojis, and flowcharts.

LLM for Table Understanding   *Oct. 2024 – Now*
Advisor: Prof. Philipp Koehn, Johns Hopkins University
- Designed a framework for table understanding with dual LLMs, where one LLM generates prompts to guide the other LLM for executing tasks.
- Enhancing the task-executing LLM performance through Chain-of-Table reasoning, and enabling the prompt-generating LLM to self improve via reinforcement learning.

Voice Anonymization   *Sep. 2024 – Now*
Advisor: Prof. Najim Dehak, Johns Hopkins University
- Developing a real-time Speaker De-Identification (SDID) system to prevent Speaker Identification (SID) and classification of target human trait (e.g., age, gender, and linguistic affinity) and state attributes (e.g., emotion expressions) from streaming audio by a speaker.
- Utilizing Probabilistic Linear Discriminant Analysis (PLDA) to generate distinct pseudo-speaker embeddings that differ significantly from both real speakers and previously generated pseudo-speakers; Planning to extend to advanced generative models like conditional GAN and DDPM.

Multimodal Emotion Recognition   *Jul. 2024 – Aug. 2024*
Advisor: Prof. Najim Dehak, Johns Hopkins University
- Ideated a novel approach to recognize emotions from core affective label intensities and appraisal dimensions, by fusing state-of-the-art text, speech, and vision models.
- Uncovered layer-specific emotion information in embeddings extracted by Transformer architectures, recognized the emotions by multimodal models, and benchmarked cross-modal fusion techniques. Improved the performance by $\sim$ 12% in Unweighted Average Recall and 0.4 in Concordance Correlation Coefficient.

Awards

Ph.D. Fellowship, Johns Hopkins University	*2024*
Comprehensive Excellence Scholarship, Tsinghua University	*2019*