Exploring and Expanding Multimodal Large Language Models (LLMs) With Human-Centric Approaches

Sriram Ganapathy, EE (http://www.leap.ee.iisc.ac.in/sriram/)
Prathosh A P, ECE (https://sites.google.com/view/prathosh/home)

In the recent years, large language models with multi-modal capabilities have made significant advancements in various tasks and skills. In some applications, the models already outperform human abilities. However, these models have also shown fundamental deficiencies in dealing with multi-modal data, like spatial reasoning or audio question-answering, where they significantly underperform with respect to human abilities. This has fueled renewed interest in debate on the state of human versus artificial intelligence. Further, the LLMs require significantly more compute power and resources restricting their wide-spread deployment.

In this project, we will explore the following research questions

1. What are the blindspots of the current multi-modal LLMs for audio-visual tasks, where they significantly lag humans abilities.
2. What lessons can we derive from human learning principles in structuring the learning/training of multi-modal LLMs to attain parity or achieve super-human skills
3. How can models continually learn with humans through in-context learning
4. How do we build more resource efficient models that can reduce the huge computational demands that plague the current LLMs.

The project will explore the use of Vaani data and other publicly available data resources to address these questions, along with open source large language models like LLaMa and Gemma.

Pre-requisite - Interest in mathematical models, coding and machine learning with an attitude of curiosity and hard-work.

More reading

[1] Zhang, Duzhen, et al. "Mm-llms: Recent advances in multimodal large language models." *arXiv preprint arXiv:2401.13601* (2024).

[2] Lu, Sheng, et al. "Are Emergent Abilities in Large Language Models just In-Context Learning?." *arXiv preprint arXiv:2309.01809* (2023).