Project Title: Conversational Failure Analysis for Arabic Speech-First Multimodal LLMs

Project Description: This project investigates failure modes in Arabic speech-first multimodal LLM (e.g., Fanar) conversations (e.g., dialect variation, code-switching, disfluencies, turn-taking, repairs, and ASR error cascades). The intern will analyze where and why the system breaks, build a small evaluation set, and propose targeted improvements to increase conversational robustness and quality.

Project Type: Research and development

Duties/Activities: The students will be asked to work on one or more of the following:

- Curate and label a small set of Arabic spoken dialogue samples (dialect/style/phenomena tags)
- Run systematic evaluations and error analyses across conditions (clean vs. ASR transcripts, dialects, code-switch)
- Design metrics and a testing pipeline; report key failure patterns with examples

Required Skills:

- Knowledge of AI/Machine learning/NLP
- Fluent in Python

Preferred Intern Academic Level: We accept all levels: PhD, MSc, Senior undergrad students enrolled in CS, CSE.

Learning Opportunities: You will gain hands-on experience in speech-first LLM evaluation, conversational analysis for Arabic, and building robust benchmarks and diagnostics. Strong contributions may lead to a research publication or an internal benchmark/tool adopted by the team.

Expected Team Size: 2-3 people

Mentors

Shammur Absar Chowdhury (https://shammur.one/)
Firoj Alam (https://firojalam.one/)
Zein Shikh Ali
Hunzalah Hassan Bhatti