LARGE LANGUAGE MODELS IN HUMAN-MACHINE INTERACTION

Ömer Emre Mutlu

emre.mutlu@rwth-aachen.de

Abstract

Dieses Dokument dient als Anleitung und als Vorlage für die schriftlichen Ausarbeitungen für Seminare am *Institut für Mensch-Maschine-Interaktion (MMI)*. Einsteiger in LaTeX sollten parallel mit einem Ausdruck dieser Vorlage und dem Quellcode arbeiten, um so leicht entspechende Befehle für ein gewünschtes Layout zu finden. Helfen dieses Dokument sowie angegebene Literatur nicht weiter, stehen natürlich die Betreuer für Fragen bereit.

Keywords: LaTeX, mmiSeminar-Style, Vorlage

INHALTSVERZEICHNIS

1 Introduction & Motivation

In this section I will be talking about the general topic. Give introduction to what Human Machine Interaction is, what LLMs are and how their use can aid in HMI. I think I will be taking the *Conversational_User_Interfaces_for_Astronauts* paper as the base reference.

2 Survey Method & Corpus

Very short section, I will be generally discussing how i have made the research paper selection. How i have classified the papers, my methodology etc. maybe 1-2 paragraphs

3 LLM-Driven Interaction Technologies

A more interesting section for me as I work at Sagel AI regarding realtime telephony agents which can take agentic actions. I will give a general overview of the general Multimodal LLM Interaction Technologies.

3.1 Text-Based Interfaces, Planning, and Tool Use

This is plain text interfaces similar to Chat GPT or LLama or Deepseek.

3.2 Audio Pipelines: Speech-to-Text $\rightarrow$ LLM $\rightarrow$ Text-to-Speech

This is how speech to speech interfaces were being done a few years ago by translating input speech into text, feed it into LLM and then turn the resulting output into speech again. Its a solid approach, however quite slow. Easier to integrate with RAG though.

3.3 Speech-to-Speech LLMs

This is the state of the art regarding Speech to Speech man machine interactions ever since the launch of Open Ai Realtime models. works basically by taking input speech in divided chunks and feeding them into the model as they come. The model starts outputting directly after the first few chunks of audio comes in, giving it a near realtime feel during conversation. Very fast, however cognitivly a lot dumber than the top of the line models such as GPT4o or GPT5. Quite hard to integrate with RAG but not impossible. There are a few papers I saw regarding this but never done it myself, currently working on this at work.

3.4 Vision-Language(-Action) Models

I have learned of this sort of model family not so long ago, and while looking for papers for this seminar I came across these again. Basically this model family can draw context from visual streams such as video feeds and take action. I need to read more on this to give better explanation but certainly very interesting in context of this seminar topic.

4 Domain Case Studies

In this section I will be dicussing select papers that are relevant to this seminar topic. Exactly which papers are to be determined. Potential research direction is given below.

4.1 Aerospace and Remote Science

Aerospace, Deep Sea wayages etc.

Psychological and Cognitive Impacts

I am sure in long isolated voyages the use of a somewhat human feeling interface is a lot better from a Psychological aspect rather than a GUI. I have found a paper regarding this. Its certainly something i would like to involve in this paper.

4.2 Emergency and Environmental Response

4.3 Industrial Operations

5 Comparative Analysis and Metrics

I will be comparing the methods I have found out in Doman Case Studies section.

6 Safety, Trust, and Deployment Constraints

Safety is a huge topic regarding the use of AI in any system that may endanger lives. And hallucinations are a huge issue. I will be looking into more papers in this direction. At work we had huge problems with our customer because the realtime Speech models hallucinate on the very extreme end. Several times they were making things up regarding our customers data. Ours is a CRM system in worst case our customers would have to deal with angry clients, such a thing can result in deaths in Aerospace etc.

7 Conclusion

This section will be my conclusion.

Literaturverzeichnis