

LARGE LANGUAGE MODELS IN HUMAN-MACHINE INTERACTION

Ömer Emre Mutlu

emre.mutlu@rwth-aachen.de

Abstract

This document serves as a template for the seminar write-up on Large Language Models in human-machine interaction. It will be replaced with the final abstract once the paper is complete.

Keywords: Large Language Models, Human-Machine Interaction, Seminar

CONTENTS

1	Introduction & Motivation	3
2	Survey Method & Corpus	3
3	LLM-Driven Interaction Technologies	3
3.1	Text-Based Interfaces, Planning, and Tool Use	3
3.2	Audio Pipelines: Speech-to-Text → LLM → Text-to-Speech	3
3.3	Speech-to-Speech LLMs	3
3.4	Vision-Language(-Action) Models	3
4	Domain Case Studies	4
4.1	Aerospace and Remote Science	4
4.2	Emergency and Environmental Response	4
4.3	Industrial Operations	4
5	Comparative Analysis and Metrics	4
6	Safety, Trust, and Deployment Constraints	4
7	Conclusion	4

1 Introduction & Motivation

In this section I will be talking about the general topic. I will give an 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 that 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 ChatGPT or LLaMA or DeepSeek.

3.2 Audio Pipelines: Speech-to-Text → LLM → Text-to-Speech

This is how speech-to-speech interfaces were being done a few years ago by translating input speech into text, feeding it into an LLM, and then turning the resulting output into speech again. It is 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 OpenAI Realtime models. It 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 come in, giving it a near realtime feel during conversation. Very fast, however cognitively a lot dumber than the top-of-the-line models such as GPT-4o or GPT-5. 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 a better explanation but certainly very interesting in context of this seminar topic.

4 Domain Case Studies ---

In this section I will be discussing 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. It is 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 in the Domain 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 the worst case our customers would have to deal with angry clients, and such a thing can result in deaths in aerospace etc.

7 Conclusion ---

This section will be my conclusion.

Bibliography ---

- [Ame96] American Mathematical Society. *AMS- \LaTeX Version 1.2 User's Guide*, November 1996.
- [GMS94] Michael Goossens, Frank Mittelbach, and Alexander Samarin. *The \LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, USA, 1994. ISBN 0-201-54199-8.
- [Knu84] Donald E. Knuth. *The \TeX book*, Volume A of *Computers and Typesetting*. Addison-Wesley, 1984.

- [Kop00] Helmut Kopka. *LaTeX Einführung*. Addison-Wesley, Reading, Massachusetts, USA, 3. edition, 2000. ISBN 3-827-31557-3.
- [Lam94] Leslie Lamport. *LaTeX: A Document Preparation System*. Addison-Wesley, Reading, Massachusetts, USA, 2. edition, 1994. ISBN 0-201-52984-1.
- [OPHS00] Tobias Oetiker, Hubert Partl, Irene Hyna, and Elisabeth Schlegl. *The Not So Short Introduction to LaTeX2e, Version 3.16*, September 2000.