

## Lab2 Report

While working on lab2 I tested two different LLMs: llama3.1 and gemma2. I have decided that llama3.1 is better for speech based dialogue model. Gemma2 was using emojis, and I couldn't prompt it to change that. My Dialogue with the user begins with the greeting message. It was coded outside of the main loop. First message in my *messages* is "Greet the user, tell them they can ask anything. One sentence only". After testing different combinations I used *assistant* as a role for this message. Then the utterance that is generated by LLM is actually assigned to role of "system". Although it doesn't make sense according to documentation it's the option that worked the best. It is not the only problem I encountered with the usage of roles. For few hours I was trying to do make ASR\_NOINPUT state produce "I can't hear you" type of message. I was testing both "system" and "assistant" roles, also with changing the configuration in rest of the code. Finally I got a tip from my colleges that actually silent message from "user" role works. And that was true for me as well it actually dialogue system is able now to tell the user to repeat the question - role: "user", content: "Tell the user you didn't hear anything".

Another problem I encountered was that LLM was always answering first question. I realised that's because I was adding messages at the beginning of the list not the end. After changing that. LLM is able to answer the last message.