SoRAIM Introduction & Demo

GOAL: Context-aware medical assistant conversational AI

Kirubakaran Ramamoorthy
H2020 SPRING Project, INRIA RobotLearn
February 2024
CHARACTERISTICS OF AN APPROPRIATE CONVERSATIONAL AI

• Context: Domain specific agent and verbal interaction
  – Relevant to its role
  – Short response time
  – Multilingual capabilities
  – Privacy and security
  – Integration in a large robotic application

• Interaction with ASR and TTS
  – Error with ASR listening
  – Short output for TTS
ADDED REQUIREMENTS

• To know when to talk and when not to
  – While person is talking
  – While navigating

• Integration with other modules
  – Navigation to known places
  – Head movements to look at the speaker

• Retaining necessary context while maintaining short response time during long conversations

• Safety – stop navigation to avoid collisions
Models Used

Nvidia Riva 2.7 ASR:
- Conformer-CTC architecture
- Combination of self-attention modules
- & convolution modules

TTS:
- Multilingual
- Different voices
- Built-in on the robot
Models Used

ALANA NLU + RASA

HERIOT WATT UNIVERSITY

Kirubakaran R (INRIA RobotLearn)

ARI's Conversational AI

February, 2024
Models Used for today

LLM: Vicuna 13B
- LLAMA model
- Finetuned & packaged by LMSYS

LangChain:
- Framework for developing applications with LLMs
- Provides extended functionalities
ARCHITECTURE

Four Channel Microphone

Controllers

Speaker

Speaker_id, Voices

Sound direction

ASR

Navigation

Behavior Generator

Interaction Management/ Trigger word detection

Prompt Generation

Response, Language

Response

Response

TTS

LLM

Prompt

Kirubakaran R (INRIA RobotLearn)

ARI's Conversational AI

February, 2024
DEMO VIDEOS
CHALLENGES

- ASR problems
  - Sound reverberations
  - Incomplete dialogues
  - Motor noises

- High GPU consumption by LLM (Use quantized models)
- Identification b/w different persons/voices is not reliable enough
- Can’t differentiate b/w human-human interactions and human-robot interactions
- Dealing with hallucinations
- Use more LLMs to guess user intentions (navigation / general query)
QUESTIONS

THANK YOU !!!