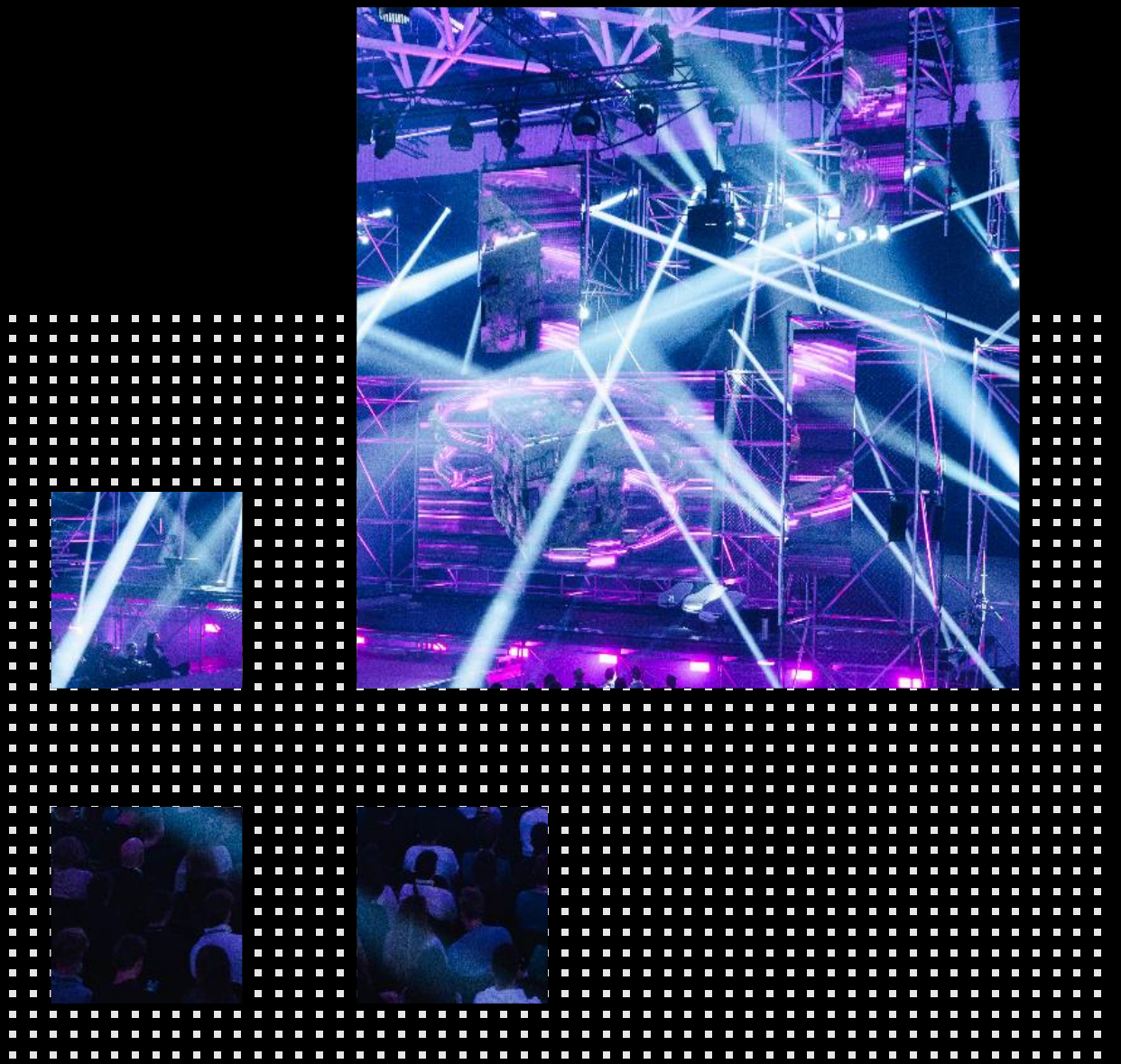
helbling

Voice Assistant Enhancement
Helbing x START

Deep Dive

19.03.2025, St. Gallen
Pascal Berger



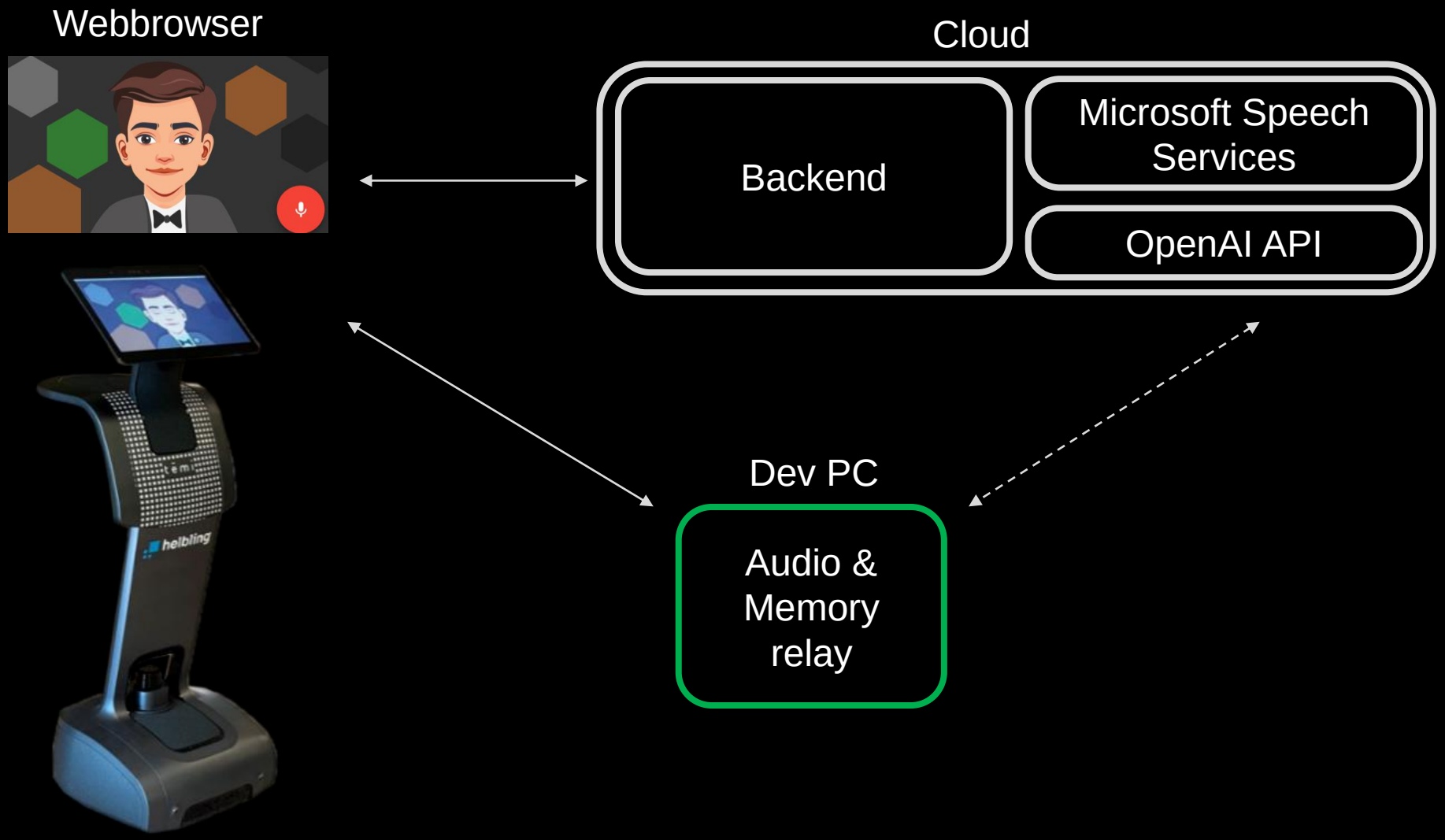
Agenda

- State of Product
- Infrastructure Overview
- Data Flow: Speech to text
- Data Flow: Memories
- Memories Integration
- Challenge Topics
- Resources
- Questions

State of Product

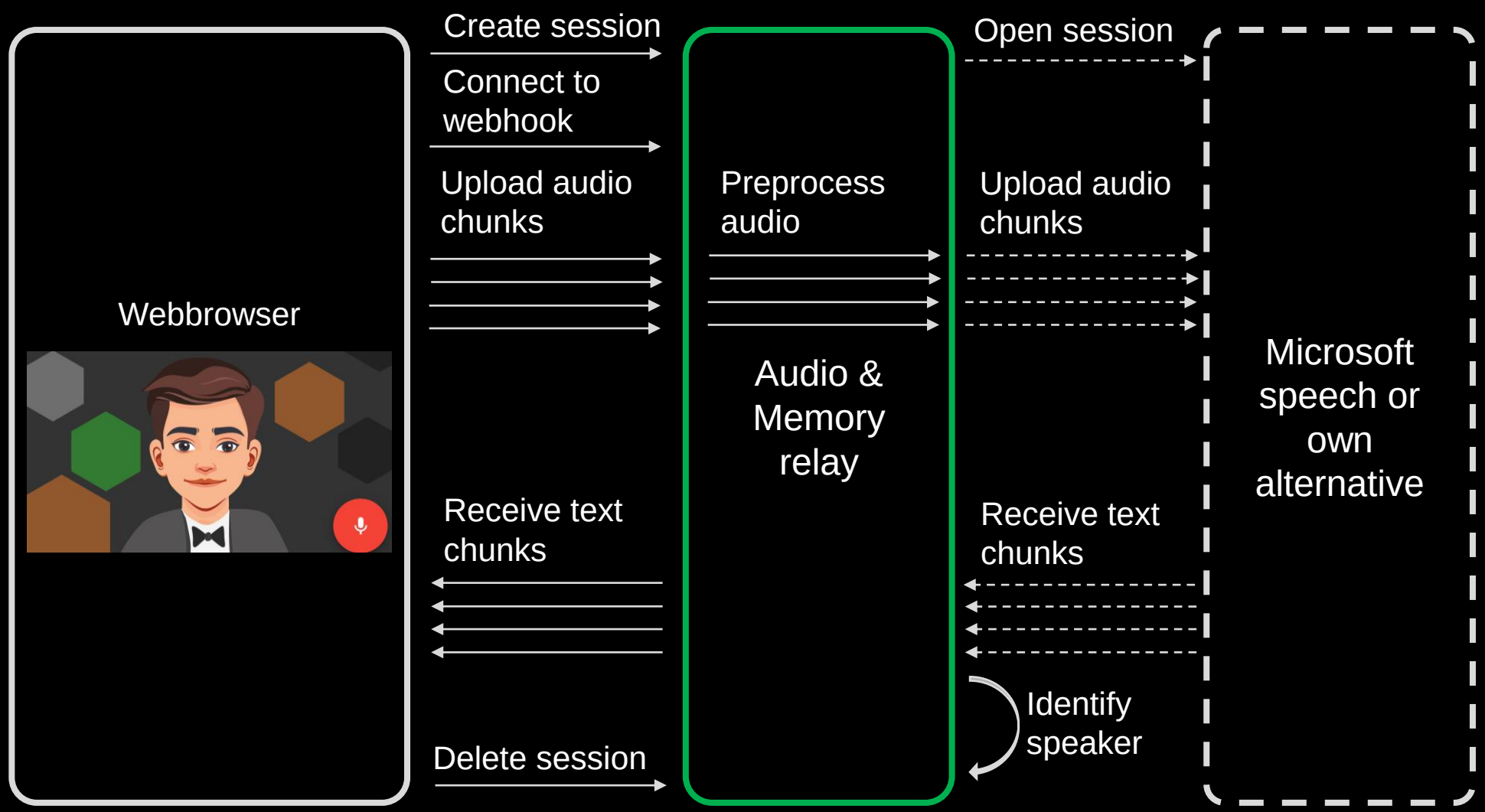
- Chatbot acting as a waiter in a restaurant.
- Working well:
  - Complex conversations (gpt-4o)
  - Restaurant context understanding and meta informations
  - Robot movement interactions
  - User Interface
- Not working well:
  - Voice understanding in noisy environments
  - Voice understanding with voices in the background
  - Memory of past conversations (not implemented)

Infrastructure Overview



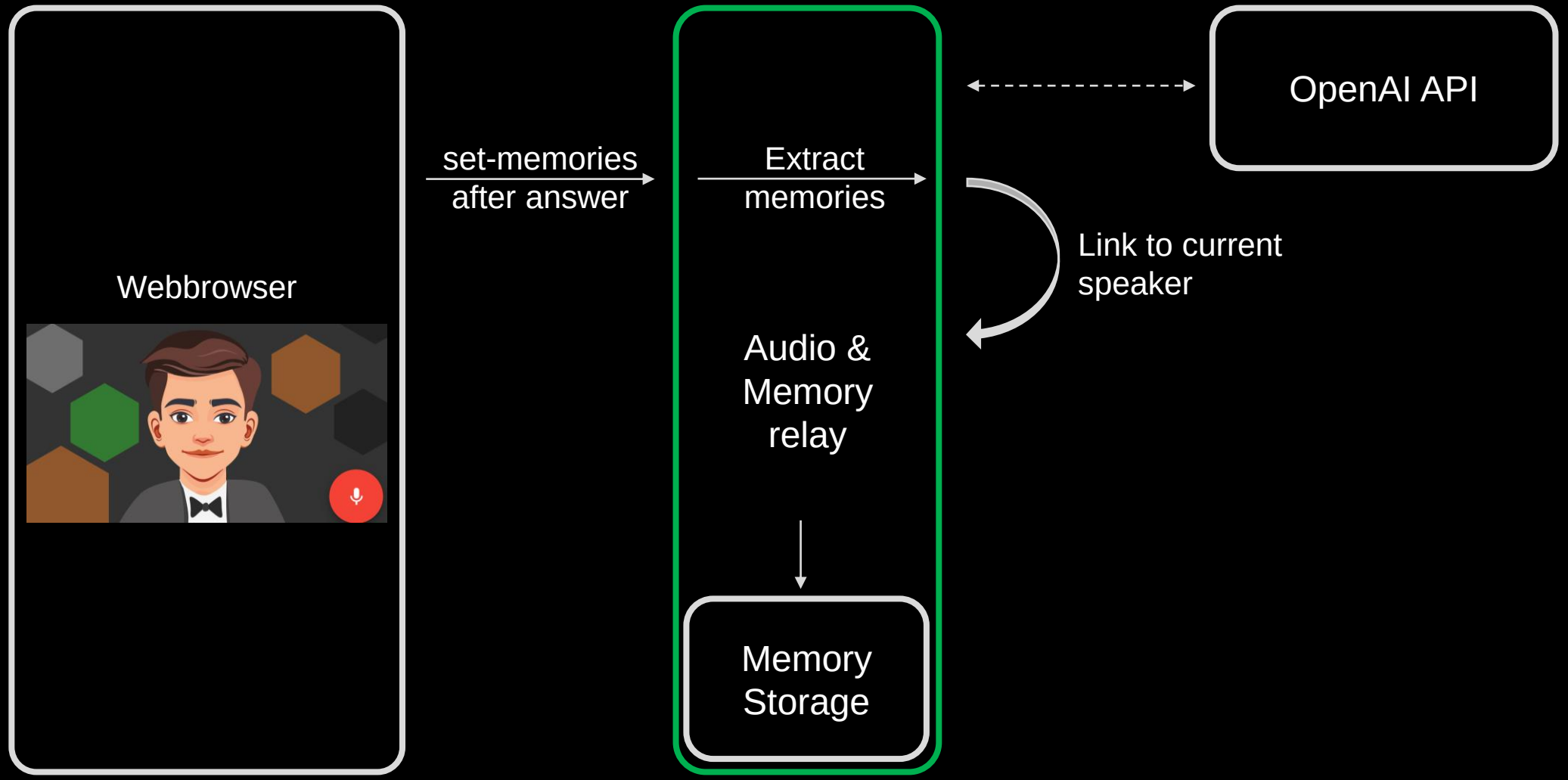
Dataflow

Speech to text



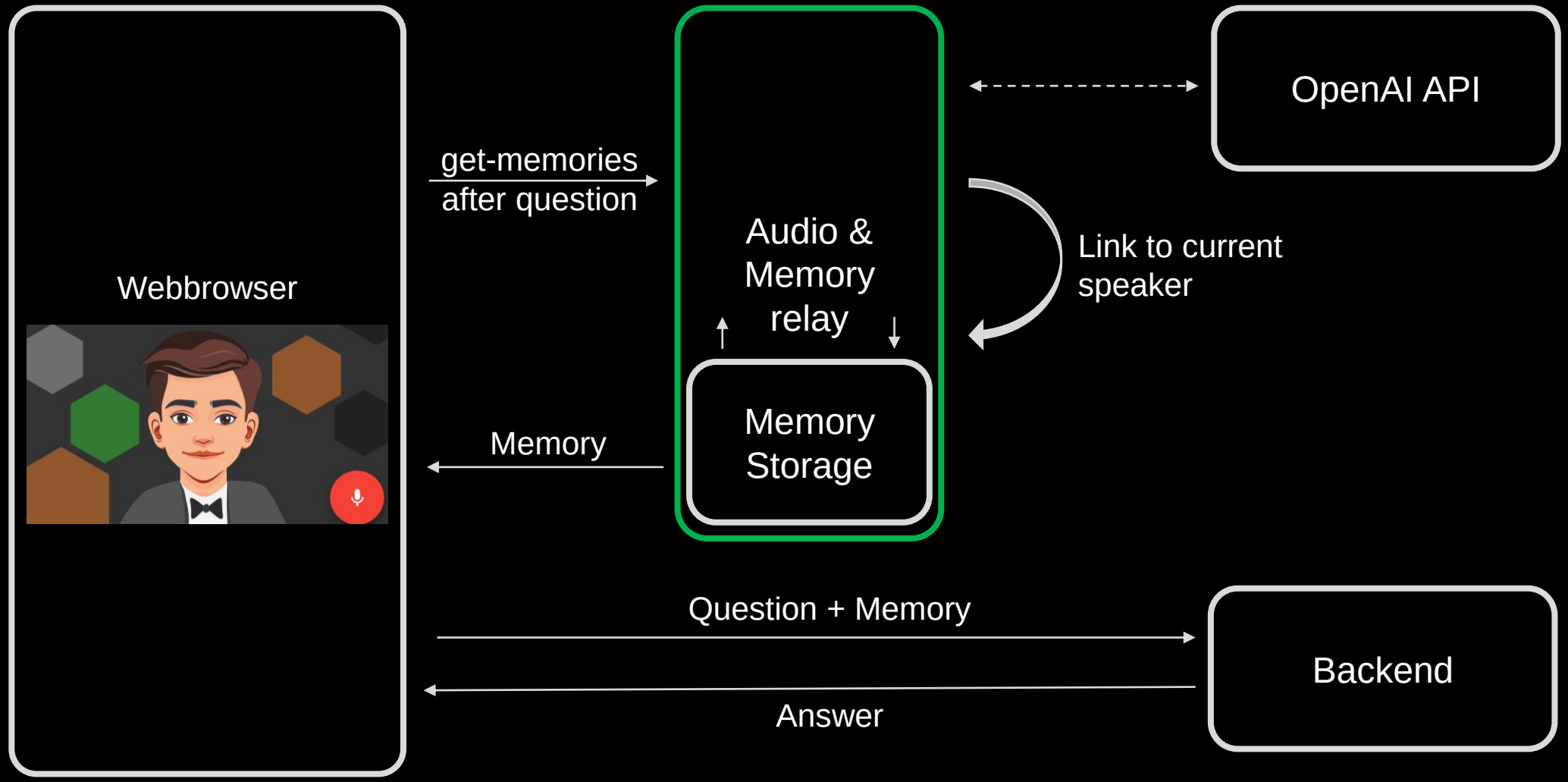
Dataflow

Set-memories



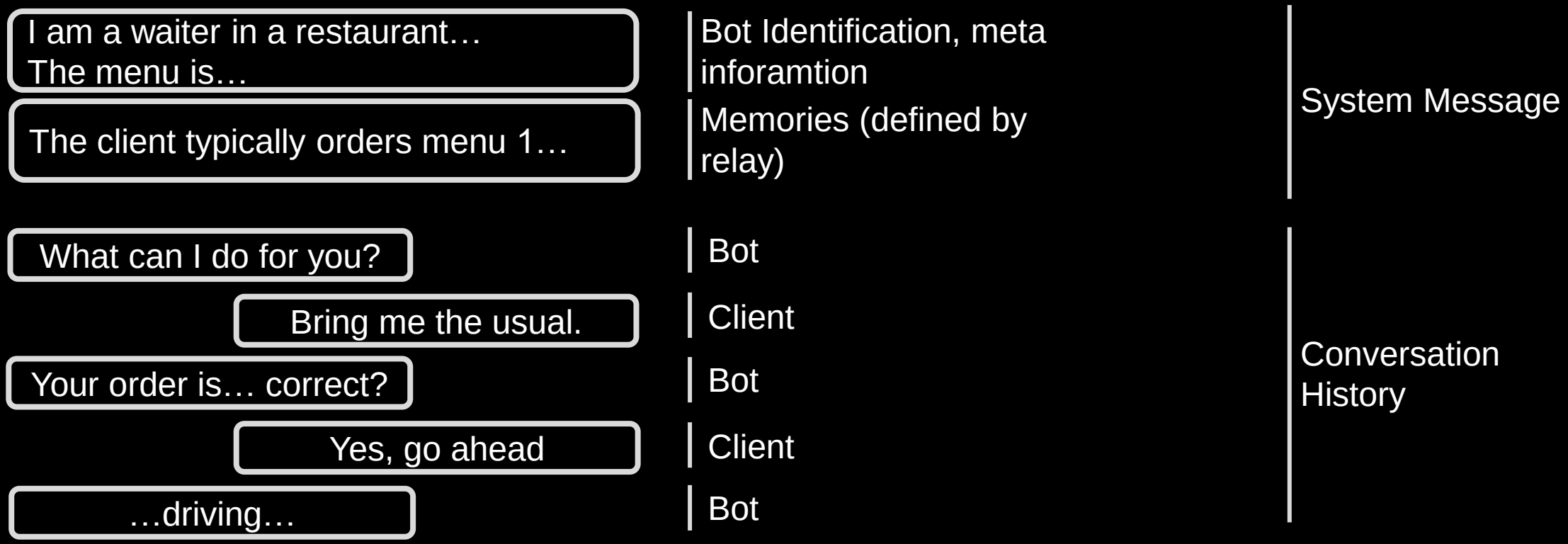
Dataflow

Get-memories



Memories integration

Chat Format



Challenge Topics

- Quality (50%)
  - The quality of the solution is key. The main voice does not have to be completely isolated, but it should work well in the speech to text process. We identify different tasks:
    - Isolation of the main voice over background noise and voices. (15%)
    - Labelling of different voices. (10%)
    - Extraction and storing of conversational information “memories” per voice. (10%)
    - Injection of memories into conversation for emotional and informational enhancement of conversation. (10%)
    - Safety & Privacy of stored information. (5%)
- Presentation (20%)
- Performance (20%)
  - The input preprocessing should work in real-time on sliced audio samples in the cloud or on a tablet. It should not take more than 0.5s per sample on consumer grade hardware.
- Business (10%)
  - What are the privacy and safety requirements for such a solution?
  - What is the environmental impact of the software?

Resources

- Webapp access & sample code «relay.py» with Microsoft Speech, no enhancement or memories
- Swagger documentation of «relay.py» if you want to start from scratch
- API Keys for Azure Speech and OpenAI API
- README containing all information needed to setup and run «relay.py»
- Helbling professionals at the Hack Booth and on Discord

Demo

voiceoasis.azurewebsites.net

Questions?