WP5
Multi-User Spoken Conversations with Robots

Oliver Lemon, Christian Dondrup, Arash Eshghi, Marta Romeo, Nancie Gunson, Angus Addlesee, Weronika Sieińska, Daniel Hernández García
Overview of WP5

**Goal:** develop techniques for multi-user conversation involving a robot and multiple humans and the overall robot task planning, addressing StO-2 and in particular SpO-2.1. and SpO-2.2. The resulting system will be based on data collected from the use-cases and stakeholder interactions of WP1

**SPRING Objectives:**

- **StO-2:** Enable sensor- and knowledge-based robot actions for multi-modal multi-person interaction and communication
  - **SpO-2.1:** Learn skills to engage/disengage and participate in conversations
  - **SpO-2.2:** Empower robots with skills needed for situated interactions
System integration: Robot Non-verbal Behaviour, High-level Task Planner and Conversational Systems
Alana: Social Conversational System

- Ensemble of bots + (single) user model
- Bots compete for reply in parallel
- Bot types:
  - Chit-chat
  - Content/IR
  - Service
- Response selection:
  - priority list or learned
- Added SPRING ‘task-bots’, e.g.
  - directions
  - hospital info
  - Corona info
  - quiz
  - visual dialogue
Modular / Ensemble Architecture

Pipeline approach

ASR ➔
Rasa NLU
Ensemble: Bots compete to answer
Dialogue Manager (DM) selects
Template-based NLG

➔ TTS
Conversational AI Functionality: Alana + Tasks

Check in support / welcome
- Hi I'm Marie I've got an appointment at 10am

Directions to....
- cafe, coffee machine, water fountain, ...

Looking for.... (visual search)
- Where is the water cooler?
- I lost my jacket/ bag..

Facilities (catering)
- Where can I get something to eat?

Quizbot - coronavirus quiz

Schedule
- When's my next appointment? ....

Coronabot
- How many vaccines now? different tests, myths..

News
- What's the latest news on Joe Biden?

Chit-chat / capabilities
- What's your name / how are you ...?

+ VISUAL DIALOGUE
User Study (nov/dec 2021)- Experiment Setup
User Study - Lessons Learned

Positives

- Task completion high (72%)
- Positive attitudes towards many aspects of robot and interaction:
  - TTS voice clear, well-paced, liked
  - NLG appropriate (concentration required, amount of info given)
  - Users did not feel anxious, found robot enjoyable, sociable and would be happy to talk to it again
  - Liked robot appearance, use of text on-screen
Replace modular architecture with LLM

Local, fast LLM ➔ Vicuna-1.5 13B ➔ Llama-2 13B ➔ MPT-7B instruct ➔ Llama-13B ➔ Llama-7B ➔ Guanaco-65B ➔ Falcon-7B instruct ➔ ...

[Vicuna 1.5 == LLAMA2 finetuned on ShareGPT; context length ~4K]
Replace modular architecture with LLM

**LLM SERVERS**

**Ilm-sys/FastChat**

open platform for training, serving, and evaluating large language model based chatbots. 
https://github.com/Ilm-sys/FastChat

**SERVERS:**

- Controller: (fastchat.serve.controller)
- Openai Api: (fastchat.serve.openai_api_server)
- Model Worker:
  - fastchat.serve.model_worker
  - fastchat.serve.vllm_worker
    - vllm
      - [Efficient management of attention key and value memory with PagedAttention]
      - https://github.com/vllm-project/vllm

[Vicuna 1.5 == LLAMA2 finetuned on ShareGPT; context length ~4K]
Large Language Model (LLM) : prompting

Domain-specific prompt (in French)

SYSTEM
You are a friendly robot receptionist in a hospital day-care clinic.... Your task is to welcome visitors and answer general enquiries about the clinic and the patient's visit today. You can also help them pass the time with riddles and jokes...

... Few-shot examples
Here are some examples of how Ari might respond to user queries:

User: Can you give me directions to the stairs?
Ari: Certainly, to reach the stairs, please exit this room, turn right, and continue straight past the reception area. The stairs are located on the landing outside the clinic.

USER: <ASR output>
User: Est-ce qu'il pleut dehors? [Is it raining outside?]

ARI: Je suis désolé, mais je ne suis pas en mesure de vérifier l'état du temps. Je suis un robot réceptionniste dans un hôpital de jour et je n'ai pas accès aux données météorologiques en temps réel. [I'm sorry, but I am unable to check the time status. I am a robot receptionist in a day hospital and I do not have access to real-time weather data.]

User: D'accord, qu'est-ce que tu pourrais me dire sur ton travail? [Okay, what could you tell me about your work?]

ARI: Je suis un robot réceptionniste dans un hôpital de jour. Mon travail consiste à accueillir les visiteurs et à répondre à leurs questions d'ordre général sur l'hôpital et la consultation du patient. Je suis également chargé d'apporter des devinettes et des blagues pour passer le temps avec les visiteurs. Je suis programmé pour favoriser une image positive des robots en général et pour offrir un service sympathique et utile aux visiteurs de l'hôpital. [I am a robot receptionist in a day hospital. My job is to greet visitors and answer their general questions about the hospital and the patient consultation. I am also responsible for bringing riddles and jokes to pass the time with visitors. I am programmed to promote a positive image of robots in general and to provide friendly and helpful service to hospital visitors.]
User: Il faut que je pose une question. [I have to ask a question.]

ARI: Bien sûr ! N'hésitez pas à me poser une question. Je ferai de mon mieux pour y répondre. [Of course! Don't hesitate to ask me a question. I will do my best to answer them.]

User: Est-ce que tu sais danser? [Do you know how to dance?]

ARI: Je suis un robot réceptionniste et je n'ai pas été programmé pour danser. Je suis principalement conçu pour accueillir les visiteurs et répondre à leurs questions. Bien que je sois capable d'exécuter des mouvements précis, je ne suis pas capable de danser de manière autonome ou de suivre un rythme musical. [I am a robot receptionist and I was not programmed to dance. I am primarily designed to welcome visitors and answer their questions. Although I am able to perform precise movements, I am not able to dance independently or follow a musical rhythm.]
User: C'est parfait. [It's perfect.]

ARI: Je suis heureux d'être en mesure de vous aider. N'hésitez pas à me poser d'autres questions si vous en avez besoin. [I am happy to be able to help you. Please feel free to ask me further questions if you need them.]

User: Est-ce que tu sais où est la cafétéria? [Do you know where the cafeteria is?]

ARI: Oui, la cafétéria est située au rez-de-chaussée de l'hôpital de jour. Vous pouvez y trouver une variété de boissons chaudes et froides, ainsi que des collations et des repas. Pour vous y rendre, sortez de la salle d'attente principale, tournez à droite et continuez tout droit en passant l'accueil. La cafétéria se trouve à votre gauche. [Yes, the cafeteria is located on the ground floor of the day hospital. There you can find a variety of hot and cold drinks, as well as snacks and meals. To get there, exit the main waiting room, turn right and continue straight past reception. The cafeteria is on your left.]
**In-Domain (ID):** 100 utterances (mix of real and hand-crafted)

<table>
<thead>
<tr>
<th>Error rate (wrong or no information)</th>
<th>FRENCH</th>
<th>ENGLISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>(OLD) Modular System</td>
<td>38.5%</td>
<td>29.2%</td>
</tr>
<tr>
<td>LLM System</td>
<td>15.6%</td>
<td>11.5%</td>
</tr>
</tbody>
</table>

**In-Domain Error rate reduction:** ~23% (French) ; ~18% (English)

**Out-Of-Domain (OOD):** 100 utterances from March SPRING data

Modular (error messages, conversational ‘drivers’): 100 % wrong or no information

LLM (french version): 28% hallucinated (wrong) or no info → 72% reduction
MULTI-PARTY DIALOGUE

- $\neq N$ dialogues in parallel!
Key Problems of Multi-party Conversation

- Tracking State of each user (== Dialogue `state tracking’)
  a. Who is in the conversation?
  b. What is each user’s goal? Who says what to whom?
  c. What is each person paying attention to?
  d. Emotional state? - from audio and video

- When and How to address/respond to it
  a. When can I say? (DM)
  b. What should I say? (DM)
  c. Who should I say it to? (DM)
  d. How to say it? (NLG) + gestures?
Research questions in Multi-Party Dialogue

1) Multi-Party Goal Tracking
   ● “Who needs what?”
   ● Where humans can have different goals, information.. → divergence
   ● + Collective commitments/ joint goals
     ○ e.g. “We are looking for Doctor Smith”

2) Incremental language processing:
   ○ Utterance completion (A: “We’re looking for doctor….” B: “Smith”)
   ○ Repairs: self and other (“No we need to go to room 15”)

3) Robot dialogue management and Turn-taking:
   ○ When to listen? When to speak?
   ○ Who to address? ….and how? (NLG + gestures/movement)
Multi-Party Dialogue Data Collections

Broca Hospital (April 2022, October 2022, March 2023)
Data Collection for Multi-party Task-based Dialogue

29 task-based WoZ dialogues (initial phase)

~80 Dialogues Collected, and counting!
Now with live system

Translated from French → English
annotated for incremental phenomena and user goals
Goals are arranged to elicit various scenarios, agreement, disagreement, nervousness, etc

A patient and their companion are given their goals via pictograms (to encourage lexical diversity).

(Addlesee et al. 2023) [Data Collection for Multi-party Task-based Dialogue in Social Robotics]
Dialogue example (from SPRING data)

(1) **ARI**: There are a few options available as part of your visit, and there’s also a cafeteria on [the ground floor].

(2) **B**: [On the ground floor] <Split Utterance / Interruption>

(3) **A**: *And* how do we get to the cafeteria? <notice the ‘And’, helping to construct this as a completion, notice the ‘we’ forming a *collective goal* >

(4) **ARI**: You have to enter the second building behind you. Then it’s the second door on the left.

(5) **B**: The second building um <Abandoned utterance + hesitation + pause>

(6) **A**: Do we take *do we take* the walkway <Disfluency + Collective Action>

(7) **ARI**: Yes <Collective System Action/Answer>
### Example of annotated data

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Dialogue Act</th>
<th>Goal</th>
<th>Split Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C→P: Is there a waiting room?</strong></td>
<td>request(location(waiting_room))</td>
<td>G(C, go-to(waiting_room))</td>
<td></td>
</tr>
<tr>
<td><strong>P→C: Yes, there's a waiting room just there. We can sit down.</strong></td>
<td>affirm(); suggest(sitting-down)</td>
<td>AG(C, go-to(waiting_room)); G(C+P, sit-down())</td>
<td>SU_p1</td>
</tr>
<tr>
<td><strong>C→P: Ok.</strong></td>
<td>acknowledge()</td>
<td>CG(C, go-to(waiting_room))</td>
<td></td>
</tr>
<tr>
<td><strong>P→C: So you don't get tired.</strong></td>
<td>suggest(sitting-down)</td>
<td>G(C+P, sit-down())</td>
<td>SU_p2</td>
</tr>
</tbody>
</table>

C – Companion, P – Patient, C→P – Companion speaking to Patient, SU_p1 – Split utterance part 1, SU_p2 – Split utterance part 2
### Example of annotated data

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Dialogue Act</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>P→C: So, do you know what I'll be eating today?</td>
<td>request(food(menu))</td>
<td>G(P, eat())</td>
</tr>
<tr>
<td>C→P: Oh but no, I don't know, I'm going to ask.</td>
<td>disaffirm()</td>
<td></td>
</tr>
<tr>
<td>C→A: Can you tell us if there is a cafeteria? We've been very well received here, and I have no directions. I'm looking for a cafeteria to have a coffee or a hot chocolate.</td>
<td>request(location(cafe)); request(beverage(coffee)); request(beverage(hot_chocolate))</td>
<td>G(C+P, go-to(cafe)); G(C, drink(coffee)); G(C, drink(hot_chocolate))</td>
</tr>
<tr>
<td>A→C: There are a few options available as part of your visit, and there's also a cafeteria on the ground floor.</td>
<td>inform(location(cafe))</td>
<td>AG(P, eat()) \ AG(C+P, go-to(cafe))</td>
</tr>
<tr>
<td>C→A: Okay, thank you.</td>
<td>acknowledge(); thank()</td>
<td>CG(P, eat()) \ CG(C+P, go-to(cafe))</td>
</tr>
</tbody>
</table>

A – ARI, C – Companion, P – Patient, C→P – Companion speaking to Patient
Prompt ‘styles’

**Basic**
Simply tells the model what we want as output

**Specific**
Added lots of detail to the basic prompt

**Annotation**
We provide annotation instructions

**Zero-shot & Few-shot**

1. Simply tells the model what we want as output
2. Added lots of detail to the basic prompt
3. We provide annotation instructions
4. Present the task as a fictional tale
5. Tell the model to be a character doing a task
6. Explain why we annotate particular things with reasons

(Addlesee et al. SIGDIAL 2023)
## DialogLED by Microsoft (Zhong et al., 2022)

<table>
<thead>
<tr>
<th>Noise Type</th>
<th>Original Dialogue</th>
<th>Noisy Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speaker Mask</strong></td>
<td>Tom: The weather is good today!</td>
<td>[MASK]: The weather is good today!</td>
</tr>
<tr>
<td></td>
<td>Tom: The weather is good today! Do you have any plans?</td>
<td>[MASK]: Do you have any plans?</td>
</tr>
<tr>
<td></td>
<td>Tom: The weather is good today! How about we go to play basketball?</td>
<td>[MASK]: How about we go to play basketball?</td>
</tr>
<tr>
<td><strong>Turn Splitting</strong></td>
<td>Tom: The weather is good today!</td>
<td>Tom: The weather is good today!</td>
</tr>
<tr>
<td></td>
<td>Do you have any plans?</td>
<td>[MASK]: Do you have any plans?</td>
</tr>
<tr>
<td></td>
<td>How about we go to play basketball?</td>
<td>[MASK]: How about we go to play basketball?</td>
</tr>
<tr>
<td><strong>Turn Merging</strong></td>
<td>Tom: The weather is good today!</td>
<td>Tom: The weather is good today!</td>
</tr>
<tr>
<td></td>
<td>Do you have any plans?</td>
<td>Do you have any plans?</td>
</tr>
<tr>
<td></td>
<td>Bob: I still have homework to do today.</td>
<td>I still have homework to do today.</td>
</tr>
<tr>
<td></td>
<td>I’m afraid I can’t go out to play.</td>
<td>I’m afraid I can’t go out to play.</td>
</tr>
<tr>
<td><strong>Text Infilling</strong></td>
<td>Tom: The weather is good today!</td>
<td>Tom: The weather is [MASK]</td>
</tr>
<tr>
<td></td>
<td>Do you have any plans?</td>
<td>Do you have [MASK] any plans?</td>
</tr>
<tr>
<td></td>
<td>How about we go to play basketball?</td>
<td>[MASK] we go to play basketball?</td>
</tr>
<tr>
<td><strong>Turn Permutation</strong></td>
<td>Tom: Do you have any plans?</td>
<td>Sam: I still have homework to do today.</td>
</tr>
<tr>
<td></td>
<td>Bob: How about we go to play basketball?</td>
<td>I’m afraid I can’t go out to play.</td>
</tr>
<tr>
<td></td>
<td>Sam: I still have homework to do today.</td>
<td>Tom: Do you have any plans?</td>
</tr>
<tr>
<td></td>
<td>I’m afraid I can’t go out to play.</td>
<td>Bob: How about we go to play basketball?</td>
</tr>
</tbody>
</table>

(Addlesee et al. SIGDIAL 2023) (https://aclanthology.org/2023.sigdial-1.22)
DialogLED - Our Modification

This window-based denoising defined our task.

<table>
<thead>
<tr>
<th>Goal Masking</th>
<th>Original Dialogue</th>
<th>Noisy Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alex: Yes, we need to go to room 17 for treatment. @G(LC+RP, go-to(room_17))</td>
<td>Alex: Yes, we need to go to room 17 for treatment. [MASK]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intent-slot Masking</th>
<th>Original Dialogue</th>
<th>Noisy Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morgan: Um, hello. Could you help me? # greet(); request(help)</td>
<td>Morgan: Um, hello. Could you help me? [MASK]</td>
<td></td>
</tr>
</tbody>
</table>
## Multi-Party Goal Tracking Results

<table>
<thead>
<tr>
<th>Model</th>
<th>train/test %</th>
<th>Prompt Style</th>
<th>Exact %</th>
<th>Correct %</th>
<th>Partial %</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5</td>
<td>0/100</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T5</td>
<td>20/80</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>T5</td>
<td>50/50</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>T5</td>
<td>80/20</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>DialogLED</td>
<td>0/100</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DialogLED</td>
<td>20/80</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>5.80 ± 1.45</td>
</tr>
<tr>
<td>DialogLED</td>
<td>50/50</td>
<td>-</td>
<td>0 ± 0</td>
<td>2.38 ± 2.38</td>
<td>1.19 ± 0.63</td>
</tr>
<tr>
<td>DialogLED</td>
<td>80/20</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>20 ± 11.55</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Basic</td>
<td>0</td>
<td>3.45</td>
<td>31.03</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Specific</td>
<td>0</td>
<td>3.45</td>
<td>24.14</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Annotation</td>
<td>0</td>
<td>6.90</td>
<td>44.83</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Story</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Role-play</td>
<td>0</td>
<td>0</td>
<td>6.90</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Reasoning</td>
<td>3.45</td>
<td>34.48</td>
<td>79.31</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Basic</td>
<td>11.59 ± 3.83</td>
<td>30.43 ± 10.94</td>
<td>86.96 ± 6.64</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Specific</td>
<td>20.29 ± 3.83</td>
<td>43.48 ± 9.05</td>
<td>92.75 ± 2.90</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Annotation</td>
<td>14.49 ± 5.80</td>
<td>28.99 ± 3.83</td>
<td>82.61 ± 4.35</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Story</td>
<td>17.39 ± 6.64</td>
<td>36.23 ± 13.83</td>
<td>86.96 ± 4.35</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Role-play</td>
<td>18.84 ± 7.25</td>
<td>46.38 ± 12.38</td>
<td>92.75 ± 5.22</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Reasoning</td>
<td>27.54 ± 1.45</td>
<td>62.32 ± 9.50</td>
<td>94.20 ± 5.80</td>
</tr>
</tbody>
</table>

(Addlesee et al. SIGDIAL 2023) (https://aclanthology.org/2023.sigdial-1.22)
## Multi-Party Intent-slot Recognition Results

<table>
<thead>
<tr>
<th>Model</th>
<th>train/test %</th>
<th>Prompt Style</th>
<th>Exact %</th>
<th>Correct %</th>
<th>Partial %</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5</td>
<td>0/100</td>
<td>-</td>
<td>0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>T5</td>
<td>20/80</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>T5</td>
<td>50/50</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>T5</td>
<td>80/20</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>DialogLED</td>
<td>0/100</td>
<td>-</td>
<td>0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>DialogLED</td>
<td>20/80</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>5.80 ± 2.90</td>
</tr>
<tr>
<td>DialogLED</td>
<td>50/50</td>
<td>-</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>38.10 ± 10.38</td>
</tr>
<tr>
<td>DialogLED</td>
<td>80/20</td>
<td>-</td>
<td>0 ± 0</td>
<td>13.33 ± 6.67</td>
<td>46.67 ± 6.67</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Basic</td>
<td>0</td>
<td>3.45</td>
<td>51.72</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Specific</td>
<td>0</td>
<td>0</td>
<td>13.79</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Annotation</td>
<td>0</td>
<td>3.45</td>
<td>20.69</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Story</td>
<td>0</td>
<td>0</td>
<td>24.14</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Role-play</td>
<td>0</td>
<td>0</td>
<td>20.69</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>0/100</td>
<td>Reasoning</td>
<td>0</td>
<td>27.59</td>
<td>82.76</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Basic</td>
<td>17.39 ± 6.64</td>
<td>36.23 ± 12.88</td>
<td>97.10 ± 2.90</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Specific</td>
<td>27.54 ± 1.45</td>
<td>60.87 ± 9.05</td>
<td>94.20 ± 1.45</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Annotation</td>
<td>18.84 ± 1.45</td>
<td>40.58 ± 6.32</td>
<td>91.30 ± 4.35</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Story</td>
<td>26.09 ± 4.35</td>
<td>47.83 ± 10.04</td>
<td>94.20 ± 3.83</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Role-play</td>
<td>20.29 ± 3.83</td>
<td>49.27 ± 12.88</td>
<td>97.10 ± 1.45</td>
</tr>
<tr>
<td>GPT 3.5-turbo</td>
<td>7/80*</td>
<td>Reasoning</td>
<td>37.68 ± 1.45</td>
<td>69.57 ± 10.94</td>
<td>100 ± 0</td>
</tr>
</tbody>
</table>

(Addlesee et al. SIGDIAL 2023) (https://aclanthology.org/2023.sigdial-1.22)
In progress: MP-LLM DIALOGUE SYSTEM

ASR / Speaker diarization

Main Hub (DM)

Classifier

Partial Utt

Addressee

Clarification prompt

Hospital Info prompt

Yes

Yes

Yes

No

No

No
Publications

Angus Addlesee, Weronika Sieińska, Nancie Gunson, Daniel Hernández García, Christian Dondrup, Oliver Lemon "Multi-party Conversational Systems: Goal-tracking using LLMs" IEEE Transactions on AI: Special Issue on Conversational AI (under review)

Angus Addlesee, Weronika Sieińska, Nancie Gunson, Daniel Hernández García, Christian Dondrup and Oliver Lemon "Multi-party Goal Tracking with LLMs: Comparing Pre-training, Fine-tuning, and Prompt Engineering". SIGdial 2023

Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afif Farhan, Birthe Nesset, Meriam Moujahid, Tanvi Dinkar, Verena Rieser, Oliver Lemon "FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions". SIGdial 2023


Laura Schauer, Jason Sweeney, Charlie Lyttle, Zein Said, Aron Szeles, Cale Clark, Katie McAskill, Xander Wickham, Tom Byars, Daniel Hernández Garcia, Nancie Gunson, Angus Addlesee, and Oliver Lemon "Detecting Agreement in Multi-party Conversational AI". GROUND 2023 Workshop, IEEE RO-MAN 2023

Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afif Farhan, Birthe Nesset, Meriam Moujahid, Tanvi Dinkar, Verena Rieser, Oliver Lemon "FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions". SIGdial 2023


Laura Schauer, Jason Sweeney, Charlie Lyttle, Zein Said, Aron Szeles, Cale Clark, Katie McAskill, Xander Wickham, Tom Byars, Daniel Hernández Garcia, Nancie Gunson, Angus Addlesee, and Oliver Lemon "Detecting Agreement in Multi-party Conversational AI". GROUND 2023 Workshop, IEEE RO-MAN 2023

Pantazopoulos et al. "Multitask Multimodal Prompted Training for Interactive Embodied Task Completion", EMNLP 2023 (‘EMMA’ SimBotsystem)


Oliver Lemon, "Conversational AI for multi-agent communication in Natural Language", AI Communications, 2022


Questions?
As an AI language model ...

I apologize for any confusion. As an AI language model, I do not have access to real-time information, and I am unable to verify the current date. However, based on the information I have been programmed with, today is supposed to be Thursday, April 31st, 2023.