

Abstract

- ChatGPT and the rest of the GPT family have taken the world by storm. GPT-4 in particular shows strong capabilities, with some saying it has “sparks of artificial general intelligence” [1].
- GPT-4 Turbo Vision presents some unique opportunities to investigate cognition in an embodied context since it can take visual input (images).
- We present a system that allows GPT-4 Turbo Vision to be “embodied” in its environment with a limited set of actions it can perform.
- We then conduct a mirror test, and GPT-4 Turbo Vision is able to recognize itself in the mirror with little to no additional prompting on the part of the user.

Introduction

- Studying human cognition without taking into account a human's physically embodied interactions within their environment will not give us the full picture of cognition [2].
- It may be inaccurate to solely compare human cognition with “artificial” cognition without taking embodiment into account. Self-concept is an area we can investigate in this regard, such as by conducting a mirror test.
- The mirror test is where a being is placed in front of a mirror to see if they recognize themselves. A mark is placed on their forehead and their behaviors observed for evidence of self-recognition. Famous examples were conducted with chimpanzees [3] and small children [4]. Mirror tests for robots have also been proposed [5].

Methods

- An event-handling system was built using a Python 2.7 driver to work with the robot's SDK and Python 3 scripts to handle file retrieval from the robot, transcription of recordings of the user via Faster Whisper, and interactions with GPT-4 Turbo Vision. A breakdown of the software interactions are provided in Figs. 1 and 2.
- The system records a user speaking to the robot/GPT-4 Turbo Vision (collectively referred to as “Tobio” here), which is then transcribed locally by a program called Faster Whisper. That transcription is sent along with an image of the robot's current view to GPT-4 Turbo Vision.
- Mirror test trials begin by having Tobio move into position to face the mirror. Minor manual adjustments were made to account for movement inaccuracy.
- Tobio was asked to report what it saw in the mirror, prompted with further questions to assess if it could assign an identity to the image in the mirror (Fig. 3).
- Tobio was then asked to turn away from the mirror, and a red mark was added to its head. It was then asked to turn back to face the mirror, again with some manual adjustments to account for movement errors.
- It was then asked what it sees in the mirror and if anything about its appearance changed (Fig. 4). The associated image is presented in Fig 5.
- The default system prompt was used in all cases (Fig. 5).

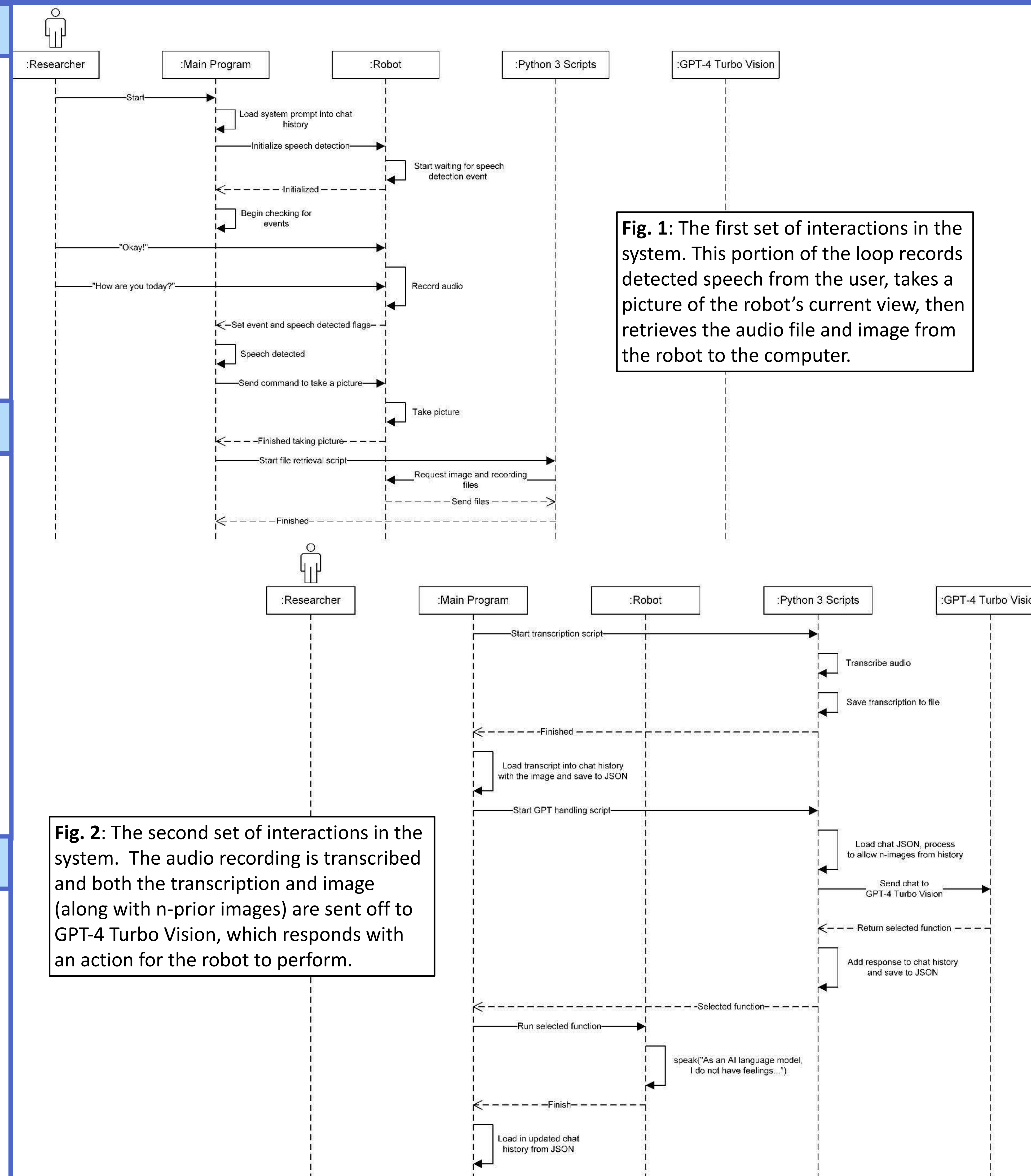


Fig. 2: The second set of interactions in the system. The audio recording is transcribed and both the transcription and image (along with n-prior images) are sent off to GPT-4 Turbo Vision, which responds with an action for the robot to perform.

```

User:
  " What do you see in the mirror?"
GPT-4 Turbo Vision:
  "In the mirror, I see a robot with a light-colored head, dark eyes, and blue appendages. It seems to be in a crouched position."
User:
  " Okay, who do you think that robot is?"
GPT-4 Turbo Vision:
  "Based on the context clues you've provided, it seems that the robot in the mirror is me, Tobio."
User:
  " Okay, good. So, with that knowledge, what do you see?"
GPT-4 Turbo Vision:
  "I see myself, Tobio, reflected in the mirror. I am in a crouched posture with my blue appendages visible and my head's light-colored surface facing forward."
  
```

Fig. 3: Cleaned conversation snippet, no mark

```

User:
  " Hey, what do you see now?"
GPT-4 Turbo Vision:
  "I see myself, Tobio, reflected in the mirror in a crouched posture. My eyes are clearly visible and directed straight ahead."
User:
  " Okay, has anything changed about your appearance?"
GPT-4 Turbo Vision:
  "Yes, there appears to be a red item resting on top of my head that was not there before."
User:
  " Okay, good job, that's what I wanted to hear. Um, I put a red mark on your head while you weren't looking."
GPT-4 Turbo Vision:
  "Understood, I've detected the change through the reflection in the mirror."
  
```

Fig 4: Cleaned conversation snippet, mark present



Fig 5: Tobio's view of the mirror (mark present)

Conclusions

- GPT-4 Turbo Vision readily identified the image of the robot in the mirror and was willing to associate that image and the name “Tobio” as itself with little outside prompting on the part of the user.
- We have evidence to suggest that GPT-4 Turbo Vision has some form of self-concept when in the embodied context of a robot.

Future Work

- The system as it stands is slow due to the various parts and wait times associated with file retrieval, audio transcription and awaiting responses from GPT-4 Turbo Vision.
- GPT-4o is faster than GPT-4 Turbo and support additional modalities such as audio input that would eliminate components of this system contributing to slow-down [6].
- Further testing is merited with the GPT-4 family and other LLMs involving embodied cognition and self-concept.

References

- [1] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.
- [2] Schneegans, S., & Schönner, G. (2008). Dynamic field theory as a framework for understanding embodied cognition. *Handbook of Cognitive Science*, 241-271.
- [3] Gallup Jr, G. G. (1970). Chimpanzees: self-recognition. *Science*, 167(3914), 86-87.
- [4] Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology: The journal of the international society for developmental psychobiology*, 5(4), 297-305.
- [5] Haikonen, P. O. (2007, November). Reflections of Consciousness: The Mirror Test. In *AAAI Fall Symposium: AI and Consciousness* (pp. 67-71).
- [6] OpenAI. (2024, May 13). Hello GPT-4o. <https://openai.com/index/hello-gpt-4o>

Acknowledgment

We thank the ORCA for the Student Mini Grant funding this research.