Differences between conversational and query-response approaches in multimodal interplay with a robot

YANCHO BOZHILOV, University of Twente, The Netherlands

Over the past half-decade, there has been tremendous innovation surrounding the advancement of AI-powered conversational agents that have the ability to simulate human-like interactions. These advancements have enabled users to engage with intelligent systems using natural language, facilitating the accomplishment of various tasks. This research undertakes a comparative examination of conversational versus query-response techniques within the scope of a multimodal human-robot interaction. For the purpose of this comparison, two analogous AI agents were developed: both designed to aid users in finding video data using voice-based commands but differing in their interaction style. The experiment results suggest that the conversational approach might provide a higher level of efficiency within the interaction, more reliable results, and a more engaging user experience in comparison to the query-response approach. These findings enhance our understanding of the potential implications and applications of conversational agents in the domain of human-robot interactions and potentially broaden their scope for beneficial use in other diverse contexts.

Additional Key Words and Phrases: Furhat, Conversational AI, Query-response, RASA, Robots, Knowledge graphs, AI agents

1 INTRODUCTION

In recent years, forms of Artificial Intelligence (AI) have been seamlessly integrated into our everyday lives, providing valuable assistance in a wide array of sectors such as education, customer service, healthcare, and security [10]. Artificial Intelligence can be defined as a computational system showing intelligent behavior, thus having the ability to perform cognitive functions comparable to a human mind [7].The utilization of AI across most domains often involves enhancing the ability of users to search and filter necessary information. To achieve this, AI agents are created as autonomous systems that can execute tasks on a user's behalf and are capable of learning or adapting to their environment [1].

The most common way to facilitate an interaction between the user and an AI-enhanced system is to implement the traditional query-response framework. Systems that solely rely on this framework include for example Google Search, where the user interacts with the system by describing their desired interest using natural language, after which they receive a plethora of online resources [10]. Despite the fact that this technique provides an effective way to engage with the system, it requires the user to manually filter through large amounts of data to obtain the relevant information they desire. Moreover, this approach heavily relies on the user's ability to effectively frame their interest for accurate results [6].

With advances in AI, an alternative approach has emerged in the form of conversational agents. Unlike their traditional counterparts, these agents facilitate interactive human-computer exchanges, leveraging human-like text or speech to simulate dialogues. This allows the user to interact using natural language and receive responses in the same form [6]. The user and AI agent have the capability to engage in multiple rounds of semantically cohesive conversation through a natural language dialogue. This facilitates effective interaction, allowing the system to understand the user's needs by directly providing further information that helps the user redefine their request or asking appropriate questions helping to clarify their interests [14].

There are a variety of conversational system embodiments, from chatbots and personal assistants to social robots. These systems are designed to mimic conversational interactions with users, but it is important to note that they can also operate in a query-response format based on the user's specific requests and needs. Systems that make use of both communication strategies include ChatGPT, Siri, and Google Assistant. Moreover, both Google Assistant and Siri utilize knowledge graphs to enhance their capabilities. They leverage the properties of the graph to establish connections between objects stored in their databases, thereby improving their functionalities [12].

In this paper, the researcher conducts an analysis to compare and contrast the two above-mentioned approaches - conversational and query-response - specifically in the context of users seeking audiovisual materials. Audiovisual data refers to information that can be perceived simultaneously both audibly and visually. Examples of audiovisual data include television broadcasts, movies, and other online video content.

To give weight to this study, the researcher has developed two custom AI agents. These agents share a lot in common, with the only significant difference being their interaction styles which aid users in their search for desired videos.

Both agents operate through the same robotic interface and employ spoken natural language to communicate with users as well as using the same audiovisual materials from a database that organizes and represents its contents using a knowledge graph structure. To effectively identify the potential strengths and weaknesses of each system, a study has been conducted with two independent groups, each consisting of eleven participants.

The remainder of this paper is organized as follows. The second section will discuss the problem statement and research question. Next, a summary of related works will follow, after which the methodology and approach of the study will be outlined. The fifth section presents the results. Then, a discussion will follow, and the conclusion will end the paper.

TScIT 39, July 7, 2023, Enschede, The Netherlands

2 PROBLEM STATEMENT AND RESEARCH QUESTION

While interactions with AI-enhanced systems have traditionally been based on the query-response strategy, recent advancements have introduced agents that make use of the conversational approach. Despite their potential, there is limited research comparing these methods of interaction, especially within the context of a virtual robot assisting users in finding videos from an audiovisual database.

This study aims to address this research gap by comparing these interaction techniques and looking at their strengths, weaknesses, and impact. The research will focus on how the conversational approach, which provides additional context about discovered videos, influences user experience and productivity compared to the traditional query-response method.

The problem statement will lead to the following research question:

- What are the differences and advantages between conversational and query-response approaches in the context of a multimodal interplay with a robot, and how do the differences influence user experience, engagement, and overall effectiveness of the interaction?

3 RELATED WORK

The following section provides background information on the current state of research.

My research combines three topics, namely conversational AI, knowledge graphs, and helper robots. Previous research focuses rather on only one of these topics but not on linking all of these together. Wilcock and Jokinen [12] are one of the few researchers exploring that. They examined how advanced knowledge graphs can be developed using a graph database and performed an experiment with the robot Furhat and Rasa conversational AI. The focus of my research is to address the gaps highlighted by Wilcock and Jokinen, specifically the insufficient amount of research conducted on the integration of conversational AI, knowledge graphs, and robots.

Other researchers focused on implementing a social robot in domains, like education or health and psychology. The study by James Kennedy et al. [5] examines the effectiveness of a socially interacting robot as a language tutor for children. This study discovered that conversational interaction approaches work better for robots than query-response approaches when dealing with children. They divided the children randomly into two groups for the test, where each group engaged with either a robot that used a conversational approach or the query-response method. Although the study is specific to language tutoring, the researchers found a significant increase in motivation, engagement, and gains in learning outcomes for the group that interacted with the conversational robot. Another similar study by Saerbeck et al. [8] investigated the impact of varying the degree of social supportiveness of a robot that played the role of a tutor. They discovered that a socially supportive robot with a conversational approach was more successful in fostering engagement and learning than a less socially supportive robot with a query-response method.

Even if we can prove the effectiveness of the conversational approach for robots, we cannot guarantee that most implementations of this approach will be successful. The difficulties of educating robots to behave proactively are discussed in the study by Garrell et al. [3]. Two major obstacles to creating robots that exhibit this behavior are teaching them to anticipate human preferences and respond to human needs without explicit instructions. The authors suggest a method where the robot first studies the behavior of the human and what interactions he or she makes with the environment before implementing the proactive approach. They evaluated their proposed approach by conducting a user study where participants interacted with a robot in a simulated home environment. The outcomes from this approach demonstrate that the robot was successful in anticipating the user's preferences and aims. In the paper published by Sera Buyukgoz et al. [2], a similar approach is tested. Furthermore, they introduce an alternative method where the robot focuses on future threats and opportunities and acts to mitigate them. One example they give is the robot offering you an umbrella when it is supposed to rain.

There are various studies exploring the concept of conversational AI. Some of them compare them to traditional query-response agents. While both can be used to achieve the goal of information retrieval in an easily accessible and fast way, conversational agents show further benefits. They actively ask the user appropriate questions, thus understanding their needs better and a conversation is being built up by multiple rounds of asking where the system gets to know the user and their needs get clarified. It avoids an inefficient and inconvenient search process. They also improve learning outcomes and enjoyment of studying [13, 14]. Furthermore, the study done by Sakirin et al. [9] showed that users preferred the more intuitive and user-friendly interface (normal language conversation) on conversational systems leading to higher rates of satisfaction. Also, the results were more accurate in comparison to traditional systems. An important point they make is that a conversational agent should improve the user experience. This perspective is shared by Schmitt et al. [10], who found that using a user-centered conversational agent increased the trust in information, enjoyment, and correctness of information-retrieval tasks compared to a non-user-centered and query-response agent.

The majority of studies on this subject have focused on an isolated scenario where the robot only needs to interact with one person at a time. Gabriel Skantze [11] proposed a systematic way to tackle the problem of a robot interacting with multiple people at a time through adaptive turn-taking.

My research will pick up on the findings of the previously mentioned researchers and extend the scope by incorporating a virtual robot that employs a query-response and conversational strategy. These approaches will be enhanced through the utilization of knowledge graph data obtained from metadata retrieval.

4 METHODOLOGY AND APPROACH

To research the conversational and query-response approaches and subsequently answer the research question of this thesis, two comparable systems were built and tested using two independent groups of participants, each consisting of eleven adult individuals.

The systems have been created to achieve the same task of aiding users in finding videos in relation to their interests. Interactions between the user and the system are made using spoken input and output. To avoid bias, they have been implemented with the same core principles and techniques, which will be explained in detail in the following subsections. Additionally, section 4.6 provides an overview of how the participant study was conducted and what participant data was collected.

4.1 Robot interface

To facilitate the multimodal interaction between the participants and the system, the Furhat Visual SDK was used. This software development kit facilitates the development of human-like robot interfaces that have the capability to display a variety of facial expressions, speech synthesis customizations, and other functionalities that make the robot interface suitable for this research [4].

This SDK provides a wide range of pre-defined faces, voices, and other customizable features. In the participant study, both groups interacted with the robot face 'Jamie' displayed in Figure 1, and heard the voice named 'Matthew'.

To access the capabilities of Furhat for this experiment, their Remote API was used in combination with the Python programming language.

Fig. 1. Robot interface of 'Matthew'

4.2 Example flows and specifications of agent actions

For both systems, there are variations in the wording of responses given by the robot. A simple example flow is given for the two systems in the following subsections.

4.2.1 Query-response agent example flow.

Participant: Hello.
Agent: Hello there. My name is Matthew, and I can help you find videos regarding a variety of topics. Please describe what video you are looking for.
Participant:: I want to find a video about athletics.
Agent: Let me see what I can find.
Agent action: Opens a new tab on the primary browser with the created HTML file in Figure 2.
Agent: These are the top 5 videos I found. Just let me know which one you would like to see by saying the number it corresponds to.
Participant: I would like to see video number five.
Agent: Give me a second to open it for you.
Agent action: Downloads and opens the chosen video in a new browser tab as shown in Figure 3.
Agent: If you would like to find a different video, just start describing it. Otherwise, you can tell me to end our conversation.

Fig. 2. Top 5 video results

Fig. 3. Chosen video

4.2.2 Conversational agent example flow.

Participant: Hello.
Agent: Hello there. My name is Matthew, and I can help you find videos regarding a variety of topics. Please describe what video you are looking for.
Participant: I want to find a video about athletics.
Agent action: Opens a new browser tab containing both word clouds displayed in Figure 4.
Agent action: Opens a new browser tab containing a geo-location map with the most relevant locations marked as displayed in Figure 5.
Agent: Hmm, interesting topic. To help you find exactly what you are looking for I have created two word- clouds with the most relevant keywords that relate to your search. As well as a map with the most relevant locations marked. You can include any extra details in your search by utilizing the word clouds and map that are available.

Participant: I would like to add stadiums and Amsterdam
Agent: Alright, I think I understand exactly what you are looking for. Give me a second to open it for you.
Agent action: Downloads and opens video in primary browser as shown in Figure 3.
Agent: If you would like to find a different video, just start describing it. Otherwise, you can tell me to end our conversation.

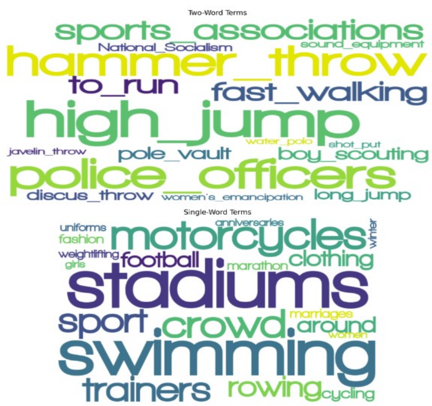

Fig. 4. Generated word clouds

Fig. 5. Generated map with the most popular locations marked

4.3 Audiovisual materials used in the experiment

When creating the video library for this experiment, two key sources were used, namely the Netherlands Institue for Sound & Vision's dataset 'Catalogue Open Data' and Open Beelden, a Dutch online platform that provides access to historical audiovisual materials. The former provides metadata in a graph-based format using RDF (Resource Description Framework) for each object in their dataset. This format enhances the ability to find related videos corresponding to a topic of interest. To access and manipulate this metadata, the SPARQL NISV Media Catalog API is used.

The metadata information extracted from Sound & Vision is only available in Dutch; thus, a translation is needed after identifying the user's interest. To speed up this process, I make use of the above-mentioned API to harvest all relevant keywords related to downloadable audiovisual content. After which, I translate them using the Python library 'deep_translator' and store them locally in a text file with a dictionary structure. This significantly lowers the number of unnecessary queries to the NISV Media Catalog API and eliminates the need of using the deep_translator modules for translating words on the fly which improves the speed of the system.

4.4 Natural Language Processing and user inputs

First, the open-source conversational AI framework called RASA is utilized to identify the user's intent behind their request to the agent using the Natural Language Understanding (NLU) capabilities of the framework. Subsequently, the dialogue management functionalities are used to decide on the appropriate response or action that the agent should take in each step of the conversation. It is important to note that depending on the results given received by the audiovisual database the action or response chosen by RASA might be changed.

Next, the Python library called NLTK is used to preprocess the user input by tokenizing the sentence provided and removing predefined stopwords, which are explicit words that will be deleted from the user input and will not be considered as keywords. Due to the fact that the user might search for bike or bicycles, but in the metadata, only the term bikes appears, a custom user keyword expansion module is developed to handle this issue. The module integrates the inflection functionality of the inflect library and the synonym extraction feature of NLTK's WordNet. This results in a dictionary that maps each keyword to its plural or singular form and generates synonyms to the original keyword. Thus resolving the problem mentioned above.

In case there are no precise keyword matches, the fuzzy-wuzzy library comes into play. The preprocessed and expanded user keywords are compared against the keyword dictionary using a fuzzy string matching algorithm which is based on the Levenshtein distance. This technique calculates the distances and determines the similarity between the user-extracted keywords and known terms contained in the metadata. If the specified threshold ratio is satisfied, the keyword will be matched; otherwise, if no words meet the threshold, the agent will say that he does not have videos on that topic.

By combining all of these approaches, the accuracy and efficiency of intent recognition and keyword matching are improved to a sufficient degree for the context of this thesis.

4.5 Differences between the query-response and conversational configurations

Both agents have been deliberately designed with limited differences to isolate and understand the precise effects of providing a word cloud and a geo-location map rather than automatically giving the best-matched results. This is one example of how an agent could be more conversational in the context of achieving the described task.

A disadvantage of using the query-response approach in this context is that if the audiovisual repository the agent uses is relatively small, delivering satisfactory results to the user becomes much more difficult. For example, if the user searches for videos related to 'children painting flowers' there might be no relevant results even though the database might contain children engaging in other forms of painting. In this case, the system shows the user a video relating to a subset of the desired search terms, which might result in the most important term for the participant being removed.

In contrast, the conversational agent might also need to remove a term but will gather the essence of the search and provide additional information to the user about what related topics are available, which could result in finding at least an acceptable video result.

4.6 Participant study

After developing the two agents, a participant experiment was conducted. This experiment was done to determine which system enhanced user experience, engagement, and the overall effectiveness of the interaction, thus answering our research question.

In this experiment, 22 individuals were split up in half to either use the query-response or the conversational system. The aim of the study was to find videos concerning a topic that is included in the video database, e.g., athletics or equestrians.

Before starting the experiment, the participants were given limited explanations on how to use the system, with the exception of defining what a word cloud map represents for the conversational group. They were told to speak with a normal voice speed and volume in a microphone and follow the further instructions the agent gave them.

The participants were informed about which video database is linked to the system and that searching for more modern words or topics could lead to no results. Apart from that, no restrictions on the choice of topic were given, and the participant could decide freely what topic they wanted to explore. Furthermore, every participant could decide on their own how many searches and different topics they would ask for.

After the interaction with the agent, a follow-up survey was sent. It included questions about the overall experience with general AI agents, e.g., Siri and ChatGPT, overall experience with the AI agent they used in the experiment, and finally, their demographic. The survey can be seen in the Appendix, and the results will be discussed in the following section.

5 RESULTS

The participant experiment and the follow-up survey led to interesting insights. Twenty-two individuals with different nationalities (German, Romanian, Greek, Spanish, Latvian, Dutch, Indonesian, Italian, North Macedonian, and Georgian), 11 for each approach, participated in the study. Nearly all of them are bachelor's or master's students aged 19-30, and one person is full-time employed. As mentioned before the full survey can be seen in the appendix of this paper.

5.1 Participant's experience and knowledge of general AI agents

The survey's first section dealt with the participant's overall experience and knowledge about general AI agents. When asked to rely on AI for help, like using AI for finding information and satisfaction with the results, the mean overall of 22 participants using a 5-point Likert Scale was 3.63, 3.95, and 3.82, respectively.

Furthermore, most participants knew many of the popular AI agents that are currently available, the data for this is presented in Figure 7. This shows that the participants are familiar with a variety of AI agents and have seen most of them being used at least once. There were various purposes mentioned for using AI agents. The majority of participants use it for education (86.4%), general information (81.1%), and research (72.7%). This section was added to identify the overall knowledge of the participants about AI agents because this could influence their experience with the custom AI agent they interacted with.

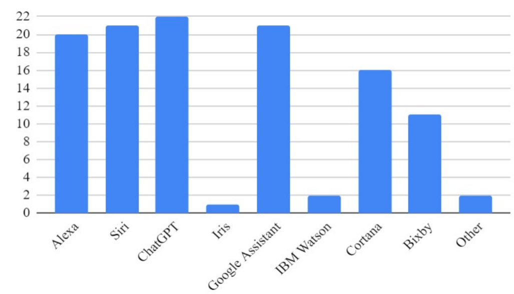

Fig. 6. Known AI agents

5.2 Participant's experience with the custom AI agents

The second section of the survey contained questions about the experience the participant had when using either the query-response or conversational AI agent.

Generally, while observing participants of both groups, it became apparent that the query-response group did not follow the video provisions as well as the conversational one. Even though both groups received the same explanation before the interaction with the system, the query-response participants kept requesting either more modern videos that were outside of the scope of the available audiovisual material or requested videos with multiple keywords for which no specific video contained all of them. This was due to the fact that the system found the most relevant videos based on the provided information and did not ask the participants to clarify or modify their search which is just the inherent nature of the query-response approach.

When analyzing the results regarding how personal the interaction on a 5-point Likert Scale (1 = not personal at all, 5 = very personal) was with the two agents, a significant difference in favor of the conversational approach is visible and can be seen in Figure 7. A reason for this is that the participants are used to query-response systems, especially in the context of finding videos online. Even though a virtual human-like robot was talking to them, that was not enough to create a personal interaction on its own. Important to point out is that neither one of the systems received 5 points. Even including facial gestures, like an eyebrow raise or smile, does not change the fact that a robot, being non-human, lacks personality.

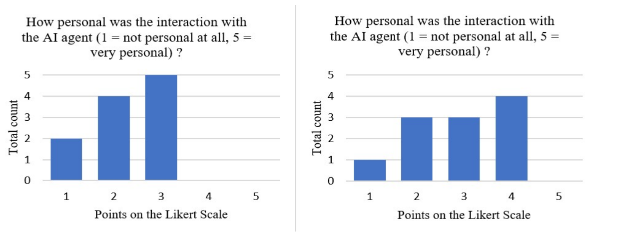

Fig. 7. Ratings of personal interaction in the query-response (left) and conversational group (right)

Another aspect being looked into is whether the video shown is related to the search topic. When it did for all respondents in the conversational approach, it was only the case for 63 % of the query-response group. As explained in Section 4, the conversational agent generated a word cloud which provided the participants with words that were connected to their initial input. With this additional feature, the video shown related to at least part of the search. The query-response agent, in contrast, did not give further suggestions. This resulted in only a subset of the top videos relating to the topic the user asked for. On top of that, while the titles of the videos may not have included their desired search terms, the content of the videos could still contain elements that are relevant to the participant's input. This is especially apparent if the participant requested a niche topic.

An AI system should solve the task of helping the user [14]. This was tested using a 5-point Likert Scale (1 = not helpful at all, 5 = very helpful). As the experiment shows, it did so for the conversational agent, giving a mean of 4.09, but only an intermediate score of 2.64 for the query-response agent. This can be clarified when examining the question of whether it is easy to find the desired results. In the former group, the majority of the participants agreed with this statement, whereas only 27.3% did so in the latter group.

Not only should an AI agent help users, but it should also provide reliable outputs [10]. To investigate this functionality, the survey asked about the user's rating of the provided results. Figures 8 and 9 show the participant's answers for the query-response and conversational approach, respectively.

Again, the advantage of the conversational approach comes into play. The additional interaction through providing the word clouds and geolocation map resulted in the agent understanding the user's preferred results much more reliably.

When posed with the question of whether the user would use an AI agent capable of searching for videos on various platforms like YouTube in their daily lives, the results showed a stark contrast. In the conversational group, 81.8% of the participants expressed

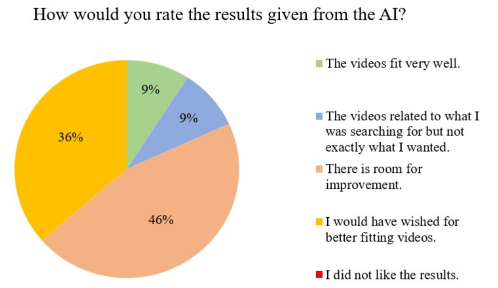

Fig. 8. Ratings of the results in the query-response group

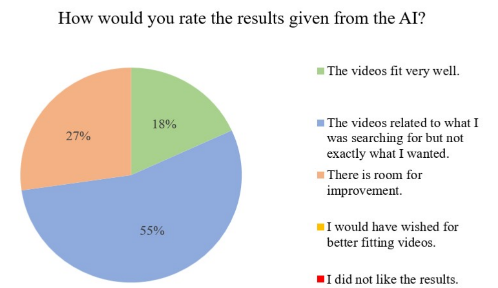

Fig. 9. Ratings of the results in the conversational group

their intention to utilize such an agent, whereas only 36.4% in the query-response group would. Two respondents in the latter group explained their reasoning by stating, 'I can already search with my voice on YouTube, so technically I can just use that.' and 'Google can do this too.'. This shows that the query-response agent provided no additional functionality than already existing systems. It showed the top 5 videos, which would also show up when searching on YouTube or Google. This explains that a conversational agent can provide additional functionalities to already existing systems.

The next valuable insights were provided by asking how satisfied the participant was with the overall experience and the rating of enjoyment while working with the agent. A 5-point Likert Scale (1 = not satisfied/enjoyable at all, 5 = completely satisfied/very enjoyable) was used for both questions, and Table 1 summarizes the means for the two groups.

-	Level of satisfaction	Level of enjoyment
Query-response	2.73	2.64
Conversational	3.09	3.09

Table 1. Means of satisfaction and enjoyment in both groups

The conversational approach scored slightly better in both aspects. As discussed in the previous sections, the conversational agent gave more relevant videos. The participant did not have to rephrase their keywords or restart the search process that much. In comparison to the previous results presented, the conversational approach did not get a high rating. A possible explanation for this is the database with its specific videos about the Netherlands, which did not fit all of the interests of some participants.

To measure the engagement level of participants, the total number of requests made to the agent was recorded. In the query-response group, the average number of requests was found to be 6.54. On the other hand, the conversational group had a total number of 8.36. Even though the participants in the query-response group needed to rephrase their questions more often than in the other group to find a fitting video for their desired topic, the number of total queries was less. An explanation for this is that the participants in the query-response group learned what the system is capable of quite quickly after the first rounds. After that, the engagement with the system declined. In contrast, the conversational agent offered more functionalities resulting in the interaction being more entertaining, thus engaging the user to run more queries to find out what other possible videos are available in the database. Another criterion the researcher wanted to look at is the total amount of time spent with the agent, but due to some participants wanting to think about their topic longer than others, this resulted in the data not being used in the study.

The participants were able to comment on what they disliked the most about the two systems. The most prevalent complaint with the conversational agent was the amount of speech. This can be seen by the following answer of a participant: 'I feel like it speaks for a little long on the follow-up for the answer.'. This is due to explanations needed for the word cloud and the geo-location map. It gives valuable insight into how the conversational aspects of the systems should be programmed to not impact the flow of the conversation. The problem is not that easy to solve because, for a multimodal interplay with a robot to be implemented for the general public, it has to speak at a rate that is digestible for a large majority of users, including children and the elderly.

Additionally, it needs to contain enough detailed information for someone new to AI agents to effectively use it. This is a problem for most speech-based AI agents. A text-based system like ChatGPT does not have this issue because the user can read the output at their own pace and can see the chat history, which aids in comprehending the context.

Another complaint was that some words in the word cloud did not seem related to the participant's initial input. A user could, for example, search for 'children', and the word cloud gives the connected term of 'smoking'. The video provided by the agent would thus include both keywords, which might lead to confusion. This occurs because the available videos consist of multiple segments, where each segment is not necessarily connected or dependent on the preceding one. In the above-mentioned example, the video first showed children, and in a later section of the same video, adults smoking cigars were displayed.

In the query-response group, participants criticized the results of their search since it did not relate to their initial keyword. As suggested by them, this could be improved by 'more interaction/questions to give better fitting videos', an approach done by the conversational agent. Another interesting aspect is that participants would have preferred a text-based query-response agent. Even though there was a robot interface, the system had the same capabilities as a regular search engine, like Google, making the robot less useful.

In both groups, the specific database was mentioned as a negative aspect. Oftentimes the participants had to rephrase their request because it was outside the scope of the video archive, which focused on videos relating to historical events in the Netherlands. This can be seen in the following participant's responses: 'If there was not a limited database, it would have been a much more interesting interaction' and 'Too many times I had to rephrase'.

Apart from stating what the users disliked, they could also comment on what they liked when using the system. In both approaches, the fast answers and results were mentioned by far the most. In the query response, two of the respondents liked that the agent provided them with multiple videos to select from. For the conversational agent, users were acknowledging 'the fact that it is interactive', thus allowing the participant to start vague and become more specific.

In conclusion, these results highlight the differences between a query-response and a conversational agent and the resulting effects these approaches have when interacting with a user.

6 DISCUSSION

In this section, I will address the limitations of this research and possible future research on the subject of this paper.

Firstly, the rather small sample of 22 participants might hinder the generalizability of the findings.

Secondly, nearly all participants were students. The advantages of easy access and low costs for data collection lead to a limited representation of the population. On top of that, the data collected concerning their knowledge of general AI agents indicated that they were already familiar with such systems and thus possibly affecting their experience with the custom AI agents.

Thirdly, the video database posed limitations on possible topics. Covering only a limited time range and videos from the Netherlands meant that the proposed search topics by the participants had to be restricted. As previously mentioned, some participants disregarded the general explanation of what videos were available, leading them to have a worse experience with the agent.

Lastly, a vast amount of time must be invested into developing and refining the system to create a truly conversational agent, even for a specific topic like the one chosen for this research. The time limitations and deadlines imposed on this research prevented the system from being fine-tuned, leading to a constrained range of possible conversational capabilities. Multiple demo systems were developed with varying functionalities before choosing this version for the participant study. Furthermore, after the first few interviews, possible major improvements were identified in both systems, but they could not be implemented due to the integrity of the already performed interviews.

Even though this paper provides valuable insights into how individuals perceive query-response and conversational AI agents, further research is needed to establish a more generalizable outcome. Another important research area related to this paper is comparing text and speech conversational agents to see if there is a preference for the user and in which tasks text or speech would be better suited.

7 CONCLUSION

This research paper examined the differences between conversational and query-response approaches in a multimodal interplay with a robot. Through a participant study using two comparable systems, this paper analyzed the outcomes and evaluated the strengths and limitations of each approach. The findings revealed several key insights. Firstly, the conversational approach implemented on the Furhat robot showed enhanced user experience, engagement, and overall effectiveness in the described context-specific human-robot interaction. Secondly, although widely used, the query-response approach exhibited limitations in providing accurate and relevant results, particularly when the database was limited in size or lacked precise matches to user queries. Furthermore, the conversational approach facilitated a better understanding of user needs and enabled the presentation of additional information and related topics, thereby improving the search process.

However, these findings are heavily influenced by the specific implementation choices described in the paper and possibly the limited demographics of the participant study. Further research is required to explore the differences between these approaches when it comes to different task contexts. Moreover, the use of more comprehensive knowledge graphs as well as further development of both systems might lead to different interesting findings.

REFERENCES

[1] Eduardo Alonso. 2002. AI and Agents: State of the Art. *AI Magazine* 23, 3 (Sep. 2002), 25. https://doi.org/10.1609/aimag.v23i3.1654

[2] Sera Buyukgoz, Jasmin Grosinger, Mohamed Chetouani, and Alessandro Saffiotti. 2022. Two ways to make your robot proactive: reasoning about human intentions, or reasoning about possible futures. (05 2022).

[3] A. Garrell, Michael Villamizar, F. Moreno-Noguer, and A. Sanfeliu. 2017. Teaching Robot's Proactive Behavior Using Human Assistance. *International Journal of Social Robotics* 9 (04 2017). https://doi.org/10.1007/s12369-016-0389-0,

[4] Kristiina Jokinen. 2022. Conversational Agents and Robot Interaction. In *HCI International 2022 - Late Breaking Papers. Multimodality in Advanced Interaction Environments*, Masaaki Kurosu, Sakae Yamamoto, Hirohiko Mori, Dylan D. Schmorrow, Cali M. Fidopiastis, Norbert A. Streitz, and Shin'ichi Konomi (Eds.). Springer Nature Switzerland, Cham, 280–292.

[5] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. Social Robot Tutoring for Child Second Language Learning. https://doi.org/10.1109/HRI.2016.7451757

[6] Bernhard Kratzwald and Stefan Feuerriegel. 2019. Putting Question-Answering Systems into Practice: Transfer Learning for Efficient Domain Customization. 9, 4 (2019). https://doi.org/10.1145/3309706

[7] Vincent C. Müller. 2021. Ethics of Artificial Intelligence and Robotics. In *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[8] Martin Saerbeck, Tom Schut, Christoph Bartneck, and Maddy Janse. 2010. Expressive Robots in Education Varying the Degree of Social Supportive Behavior of a Robotic Tutor. *Proceedings of the 28th International Conference on Human Factors in Computing Systems* 3, 1613–1622. https://doi.org/10.1145/1753326.1753567,

[9] Tam Sakirin and Rachid Ben Said. 2022. User preferences for ChatGPT-powered conversational interfaces versus traditional methods. *Mesopotamian Journal of Computer Science* (2022).

[10] Anuschka Schmitt, Thiemo Wambsganss, and Jan Marco Leimeister. 2022. Conversational Agents for Information Retrieval in the Education Domain: A User-Centered Design Investigation. 6, CSCW2 (2022). https://doi.org/10.1145/3555587

[11] Gabriel Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech Language* 67 (2021), 101178. https://doi.org/10.1016/j.csl.2020.101178

[12] Graham Wilcock and Kristiina Jokinen. 2022. Conversational AI and Knowledge Graphs for Social Robot Interaction. *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 1090–1094.

[13] Rainer Winkler and Matthias Söllner. 2018. Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. *Academy of Management Proceedings* 2018 (04 2018), 15903. https://doi.org/10.5465/AMBPP.2018.15903abstract

[14] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3269206.3271776

A APPENDIX

Survey questions

Overall experience
These questions are about your **overall experience** with general AI agents.

I rely on AI for helping me find information:
Likert: 1-5 (completely disagree – completely agree)

I like using AI for helping me find information:
Likert: 1-5 (completely disagree – completely agree)

I am satisfied with the results I get from the AI agents:
Likert: 1-5 (completely disagree – completely agree)

How often do you on average use AI agents in a week?
- ○ 1-2 days
- ○ 3-4 days
- ○ 5-6 days
- ○ 7 days

Which AI agents do you know?
- ○ Alexa
- ○ Siri
- ○ ChatGPT
- ○ Iris
- ○ Google Assistant
- ○ IBM Watson
- ○ Cortana
- ○ Bixby
- ○ Other: ____

For what purposes do you use AI agents?
- ○ Education
- ○ Research
- ○ Fun
- ○ Knowledge acquisition
- ○ General information
- ○ Customer service (e.g., agents on a booking website)
- ○ Smart home functionalities

Experience with the AI agent
The following questions relate to your experience with **the AI agent**.

How personal was the interaction with the AI agent?
Likert: 1-5 (not personal at all – very personal)

How accurate were the responses the AI agent gave you?
- ○ The responses were completely accurate.
- ○ I had to change the wording on a few questions to get the desired response.
- ○ I had to change the wording on a lot of questions to get the desired response.
- ○ I had to change my questions completely.
- ○ I had to think of new questions.

Was it easy to find the desired results?
- ○ Yes
- ○ No

Did the video(s) relate to the topic you searched for?
- ○ Yes
- ○ No

How helpful was the agent in searching for the videos?
Likert: 1-5 (not helpful at all – very helpful)

How would you rate the results given from the AI?
- ○ The videos fit very well.
- ○ The video related to what I was searching for but not exactly what I wanted.
- ○ There is room for improvement.
- ○ I would have wished for better fitting videos.
- ○ I did not like the results.

Do you think you could have gotten the same results in the same time without the AI agent? (e.g., searching on Google)
- ○ Yes
- ○ No

How satisfied were you with the overall experience with the AI agent?
Likert: 1-5 (not satisfied at all - completely satisfied)

How would you rate your level of enjoyment in working with the AI agent?
Likert 1-5 (not enjoyable at all - very enjoyable)

Would you use the agent again in this context?
- ○ Yes
- ○ No

Would you use an AI agent like the one you used for finding videos in your daily life if it has the capability to search for videos on other platforms, e.g., YouTube?
- ○ Yes
- ○ No

If not, why would you not use the agent? _____

What did you like the most about the interaction with the AI agent? - _______

What did you dislike the most about the interaction with the AI agent? - _______

Are there any improvements you can suggest? ____

Personal information
What is your gender?
- ○ Male
- ○ Female
- ○ Other

How old are you?
- ○ Under 18
- ○ 19 - 24
- ○ 25 - 30
- ○ over 30

What is your nationality? _____

What is your current occupation?
- ○ Bachelor student
- ○ Master student
- ○ Full time employed
- ○ Other: ____

What is your highest level of education completed?
- ○ Secondary education/High School
- ○ Bachelor's degree
- ○ Master's degree
- ○ Doctoral degree or higher
- ○ Other: ____