

Article

An Exploratory Study on a Reinforcement Learning Prototype for Multimodal Image Retrieval Using a Conversational Search Interface

Abhishek Kaushik ^{1,*}, Billy Jacob ^{2,†} and Pankaj Velavan ²¹ CeADAR, School of Computer Science, University College Dublin, D04 V2N9 Dublin, Ireland² School of Business, Dublin Business School, D02 WC04 Dublin, Ireland; billy.jacob33@gmail.com (B.J.); gvelavan1@gmail.com (P.V.)

* Correspondence: abhishek.kaushik@ucd.ie

† These authors contributed equally to this work.

Abstract: In the realm of information, conversational search is a relatively new trend. In this study, we have developed, implemented, and evaluated a multiview conversational image search system to investigate user search behaviour. We have also explored the potential for reinforcement learning to learn from user search behaviour and support the user in the complex information seeking process. A conversational image search system may mimic a natural language discussion with a user via text or speech, and then assist the user in locating the required picture via a dialogue-based search. We modified and improved a dual-view search interface that displays discussions on one side and photos on the other. Based on the states, incentives, and dialogues in the initial run, we developed a reinforcement learning model and a customized search algorithm in the back end that predicts which reply and images would be provided to the user among a restricted set of fixed responses. Usability of the system was validated using methodologies such as Chatbot Usability Questionnaire, System Usability Scale, and User Experience Questionnaire, and the values were tabulated. The result of this usability experiment proved that most of the users found the system to be very usable and helpful for their image search.

Keywords: conversational search; image search; user experience; usability; reinforcement learning



Citation: Kaushik, A.; Jacob, B.; Velavan, P. An Exploratory Study on a Reinforcement Learning Prototype for Multimodal Image Retrieval Using a Conversational Search Interface. *Knowledge* **2022**, *2*, 116–138. <https://doi.org/10.3390/knowledge2010007>

Academic Editor: Gwanggil Jeon

Received: 1 October 2021

Accepted: 8 February 2022

Published: 28 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Web search has become an inevitable activity in the day-to-day lives of people. Various pioneers in the field of search engines, comprising of Google, Bing, DuckDuckGo, and more, have made revolutionary improvements in the search process. Image search is one key aspect of the web search results which helps users gain a better picture of what they are looking for. With the plethora of information available across the internet, it is a strenuous task to provide relevant information to the end user. Another challenge is the user's shortage of knowledge of the topic about which they are querying. For efficient image retrieval to take place, the user should be able to fully describe what they want to know [1]. To overcome the above-mentioned challenges, an alternative model for search interaction is gaining momentum. In this model, the user communicates with an agent that seeks to help their search activities. Conversational search is a very engaging method of image retrieval as it simulates the way in which people converse with each other and find the required data [2–5].

A conversational image search system should facilitate the user in improving their query incrementally [6] until the image requested by the user is found. In this process, the conversational agent enables the user to learn about their image of interest by incrementally aiding them in developing their image search query within a dialogue, enabling them to move towards meeting their image need. This way of engagement with an image search

system can potentially reduce cognitive load by assisting the user in building a query that describes their image need in detail over multiple conversational steps.

Over the past few years, search engines have made great improvements in their capability to accurately understand natural language queries and even respond to follow-up queries which depend on the previous searches for context. While this directly describes human–human conversation and can be sufficient for simple questions about images, one of the key challenges in the field of human–computer interaction is the potential unpredictable variation in the user input, and what constitutes a valid response for the end user [7]. For more complex or exploratory search queries, it is not feasible for the agent to comprehend what the user is trying to achieve. There is no way of incrementally building the query along the conversation to retrieve the best images at the end [4,8].

In this paper, we extend a prototype multi-view image search interface [9] to a search engine API. The interface incorporates a conversational image search assistant named “Ovian”, which means “painting artist” in Tamil (a language spoken in India). The user provides the input search query and the retrieved images are displayed through an extended conventional graphical image search interface. The conversational agent helps the user to find the intended image by proactively putting out the fixed number of relevant questions to extract the intention of the user using the reinforcement algorithm. Reinforcement learning methodology has been adopted to train the conversational agent on responses it should deliver to the end-user. The user can converse directly with the image search system while receiving directions from the search assistant, both to advise them to build their query and to guide their interaction with retrieved content.

We have introduced a custom image search algorithm which filters out the relevant images from all the images retrieved through the Wikipedia API and presents it to the user. The response generation from the chatbot is handled by the reinforcement algorithm which has been trained with numerous episodes to understand the user request and reply accordingly.

1.1. Motivation

The motivation behind the study is to explore the potential of conversational interfaces in multimodal information retrieval. Information searching is a difficult process both from the user’s perspective and the logical end of the information retrieval process. It is difficult to obtain relevant results without refining the search several times, and the results also have to be verified as relevant by the user. Traditional or single-shot searches may result in increased cognitive load and frustration for the user throughout the search procedure, which might lead the searcher to abandon the search process without receiving required information [1,10,11]. We want to investigate the user experience of the interactive conversational search process in this study, which was implemented utilizing the proposed interface presented in the later portion of this article.

Moreover, studies related to conversational search are majorly focused on two dimensions:

1. Data-driven approaches (Questions-Answering systems) [8,12,13], which necessitate the development of expensive and time-consuming language models, extensive computation, and a lack of the exploratory nature of the search process.
2. Rule-based systems, which are inflexible and prone to error, and users must learn how to use them [5].

To provide the best possible solution, we proposed a reinforcement learning rule-based approach. This approach does not require a language model which is huge and robust to errors, as the model will be penalized for making mistakes and rewarded for following the right path. Additionally, most of the search bots or conversational agents are text-based systems [1], which lack the component of multi-modality. Multi-modal data can ease the user learning experience, whereas reading too much text to seek information can expand the cognitive burden on the user or the searcher.

Given the foregoing considerations, we advocate highlighting the multiview interface, which enables modality other than text, such as pictures, in the complicated information

seeking process. This research highlights the most exciting discoveries, such as the application of reinforcement learning in conversational search to record user behaviour and the promise of picture retrieval in addressing information demands, among others. Furthermore, the majority of conversational bots are assessed using either a single usability metric or empirical approaches such as precision and recall [8]. Unlike previous studies, this one is evaluated not only using the empirical evaluation while training the reinforcement learning model, but also three different usability metrics. All of these interactive usability metrics were compared among themselves to investigate the uniformity of the users' experience while using the chatbot. Our problem statement and research questions are discussed in the next section.

Research Question

In this section, the research questions framed based on the problem statement have been discussed in detail. Our problem statement is, "With the increase in the use of chatbots across multiple domains, there is an increased need to find an effective hybrid approach that can be deployed along with an efficient image information retrieval system to help the user realise his objectives by providing an interactive environment". We have used reinforcement learning to train a chatbot, which solves the issues with data-driven training and rule-based approaches. We do not have proper datasets for training image retrieval chatbots, and the condition-based approach is very tightly coupled, which allows it to answer only a certain set of questions. In this study, we also investigate the user's interactive experience on multiple interactive usability metrics. Based on this, we have framed the research questions.

- Exploratory Research Question: "How can reinforcement learning be used for improving the search experience of the user?"
- Comparative Research Question: "Are multiple interactive usability metrics associated, and do they follow a consistent pattern based on user reactions when using the multimodal interface?"

The next sections overview the recent works, the prototype multi-view interface for conversational image search, review of appropriate images, search algorithm, reinforcement learning, user engagement, and conclusion.

2. Literature Review

The recent work section comprises of the following components, viz., search interface, conversational search, image search, and conversational search interface.

2.1. Search Interface

As there is abundant information available on the internet, it is absolutely essential to have a proper interface to access this information in an efficient manner. The user interface designed by Sunayama et al. [14] addresses the need for an efficient search process by restructuring the part of user's query to identify their hidden interests. The need for an ideal and efficient interface laid the seeds for the development of the modern day search engine. Sandhu et al. [15] conducted a survey on their redesigned user interface based on the Wikipedia search interface. The novel idea of their research work lies in the provision of interactive information retrieval interface design. Suraj Negi et al. [16] suggested that, even with vast advancements in speed and availability, most of the search engine interface still relies on textual layouts to convey results. They aimed at developing a user interface with hand gesture recognition with the idea that an interface controlled by gestures will make user interaction more engaging. This led to the design of a search interface with an average gesture recognition time of 0.001187 seconds and an accuracy of 94 percent. Compared to textual results, charts can be more interactive and easily understandable for the end users. Hence, Marti Hearst et al. [17] conducted a study in which the participants report their preference for viewing visualizations in a chat-style interface when answering questions about comparisons and trends. The key insights of this study revealed that nearly 60

percent of the participants opted for additional visualizations and charts in addition to the regular textual replies in the chat interface. This led to the need for a better designing of the chat interfaces with interactive features. Another significant limitation of the conventional search interface is the need for certain logical combination of keywords as input from the user. The study by Tian Bai et al. [18] aimed to overcome this using an improved Gated Graph Neural Network (GGNN) model in which the database entities and relations are encoded. They used the database values in the prediction model to identify and match table and column names to automatically synthesize an SQL statement from a query expressed in a NL (Natural Language) sentence. The study conducted by D. Schneider et al. [19] presented an innovative UX (User Experience) for a search interface which can be used to query semantically enriched flash documents. The major advantage of this approach is that the search interface is more powerful for traversing across the multimedia content in the web, while the conventional search engines provide these results only in a textual format. Another interesting study in the multimedia information retrieval domain conducted by Silvia Uribe et al. [20] revealed the usage of the special M3 ontology in their semantic search engine called BUSCAMEDIA, which presents the multimedia content to the users based on their network, device, and context in a dynamic manner.

The study conducted by Yin Lan et al. [21] presented an intelligent software interface which was capable of retrieving emergency disaster events from the internet. The core idea of this study focuses on retrieving the metadata based on the H-T-E (Hazard, Trigger, and Event) query expansion approach to achieve high relevance of ever updating event series. The research conducted by Ambedkar Kanapala et al. [22] had addressed the issue of lack of availability of legal information online. They developed an interactive user interface which indexes both the court proceedings and the information on legal forums online. The study conducted by Susmitha Dey et al. [23] presented a new search interface which assists the users by providing them results which are domain specific. The users are allowed to choose more than one domain to refine their search. The prototype search interface developed was able to provide better results due to its ability to fragment the information based on the domain of interest.

2.2. Conversational Search

Conversational search interfaces that allow for intuitive and comprehensive access to digitally stored information over the web remain an ambitious goal for many organizations and institutions all around the world. The study conducted by Heck et al. [24] proposed two novel components for conversational searches: dynamic contextual adaptation of speech recognition and fusion of users speech and gesture inputs. Using a conversational chatbot in B2C-based business may increase customer satisfaction and retention as it is available 24/7. At the same time, however, it is also essential that the chatbot perform efficiently compared to human personnel. It is also necessary for conversational search agent to understand their users' emotions and produce replies that match their emotional states. They used maximum-a-posteriori (MAP) unsupervised adaptation to leverage the visual context. Visual context helps the users search for a specific hyperlink's relative position or an image on the page, such as "click on the top one" or "click on the first one". Kaushik and Jones [4] reported a study examining the behaviour of users when using a standard web search engine which was designed to enable them to identify opportunities to support their search activities using a conversational agent. The study conducted by Fergencs and Meier [25] compared the conversational search user interface of a medical resource centre database with its graphical-website-based search user interface in terms of user engagement and usability. Conversational chatbots can also be built as an android application so that users can search for any information on the go. The survey conducted by Fernandes et al. [26] showed the different technologies used in various chatbots all around the world. Currently, the understanding is too limited to exploit the Conversational Search. The report generated by Anand et al. [27] summarised the overview of invited talks and findings from the seven working groups which cover the definition, evaluation,

modelling, explanation, scenarios, applications, and prototype of Conversational Search and proposed to develop and operate a prototype conversational search system called “Scholarly Conversational Assistant”.

2.3. Image Search

The work by Grycuk [28] presented a novel software framework for the retrieval of images. The framework was multilayered and designed using ASP.NET Core Web API, C# language, and Microsoft SQL Server. The framework is based on content-based information retrieval. The system was designed to retrieve similar images to the query image from a large set of indexed images. The first step relies on automatically detecting objects, finding salient features from the images, and indexing them by database mechanisms.

The study conducted by Pawaskar and Chaudhari [29] proposed a unique web image re-ranking framework that learns the semantic meaning of images, both online and offline, with numerous query keywords. The research conducted by Meenaakshi and Shaveta [30] proposed an efficient content-based search using two methodologies, namely text-based and feature vector-based ability. One of the significant parts in image search is the database which contains the corpus of images. The research conducted by Nakayama et al. [31] proved that Wikipedia is an authentic source for knowledge extractions as it covers a wide range of fields such as Arts, Geography, History, Science, Sports, and Games. As part of our research, we have used Wikipedia as the image corpus. One type of image search is tag search, which returns all images tagged with a specific keyword. However, most of the time the keywords are ambiguous.

The study conducted by Lerman et al. [32] proposed a probabilistic model that takes advantage of keyword information to discover images contained in the search results. The research conducted by Xu et al. [33] suggests that precision in image search is not a satisfying metric because the labels used to annotate images are not always relevant to the images. They proposed a re-ranking solution classifying the visual features into three categories based on relevance feedback with genetic algorithm. Smelyakov et al. [34] suggested a model of service for searching using the image. They used colour content comparison, texture component comparison, object shape comparison, SIFT signature, perceptible hash, Haar features, and artificial neural networks to achieve a sufficient reliability of image search.

2.4. Conversational Search Interface

Searching for any information requires interaction with the system. Conversational search interfaces fall under such techniques. The study conducted by Kaushik et al. [5] introduced a prototype multi-view search interface to a search engine API. The interface combines a conversational search assistant with an extended standard graphical search interface. The research conducted by Kia et al. [35] proposed an open domain conversational image search system using the TREC CaST 2019 track. They first processed the text, followed by text classification, and then finally found the effect of the previous questions on the current question. Usually, a conversational search interface has a limited scope so that it solves a particular issue and increases efficiency and usability. The GUApp proposed by Bellini et al. [36] is an agent with such limited scope. It is a platform for job-postings search and recommendation for the Italian public administration. Another example is the conversational chat interface for stock analysis proposed by Lauren and Watta [37]. They used Slack as the platform for interaction and RASA framework for Natural Language Understanding and dialogue management. In order to design a better conversational chatbot which would help the users by giving a variety of information in a natural and efficient manner, Filip Radlinski et al. [3] had taken into consideration the necessary properties in their study. These properties include the user revelation, system revelation, mixed initiative, memory, and the set retrieval. They have proposed a theoretical model by taking into consideration the above properties carefully.

The study conducted by McTear [38] explores how the conversational search interface is relevant today and identified some key takeaways such as usage of conversational search interface in messaging apps, personalised chat experience, and search interfaces which learn from previous experiences. Another interesting fact is that, currently, most of the conversational agents are passive [39] and provide a list of replies, which in turn puts more cognitive pressure on the end users. The study conducted by Dubiel et al. [39] investigated the effect of active search support and summarisation of search results in a conversational interface. When the uses of conversational search interface increases, the scope of creating tools, datasets and academic resources also increases. This idea was discussed by Balog et al. [40] and they listed some research steps, obstacles, and risks in storing the research data. Some of the obstacles and risks are, viz., privacy and retention rules, stability and reproducibility, usage volume, etc.

2.5. Conversational Search Evaluation

With a large amount of chatbots being introduced into the market, there arises the need to evaluate them and measure their standard in understanding as well as meeting the customer's needs. The study conducted by Forkan et al. [41] evaluated and compared the various cloud-based chatbots. Average response time, fallback rate, comprehensive rate, precision, and recall are some of the evaluation metrics used in this study. The research by Ayah Atiyah et al. [42] aimed at measuring the naturalness of interaction of a chatbot. They evaluated a chatbot which acts as a personal chat assistant for a financial assistant and revealed that the machine-human chatbot interaction is highly competitive to human performance in terms of naturalness of interaction.

The study conducted by Laila Hidayatin et al. [43] measured the results of combining the cosine similarity metric with query expansion techniques. The need to evaluate the usability of any system was stressed by John Brooke [44] in his research work. He proposed a system usability scale (SUS) that measures the need for training, support, and complexity of any system using 10 questions. Holmes et al. [45], in their research, evaluated a conversational healthcare chatbot based on the conventional evaluation methods such as System Usability Scale (SUS), User Experience Questionnaire (UEQ), and Chatbot Usability Questionnaire (CUQ). To evaluate a conversational agent, a study should first identify the relevant metrics that should be calculated to complete the evaluation exercise. The study by Abd-Alrazaq et al. [46] identified the technical metrics used by all the previous studies to evaluate healthcare chatbots. Common metrics are response speed, word error rate, concept error rate, dialogue efficiency, attention estimation, and task completion. The study found that currently there is no standard method used to evaluate health chatbots, and the medium for the evaluation is usually questionnaires and user interviews. The research led by Enrico Coiera et al. [47] compared six questionnaires to measure the user experience of the interactive systems. The questionnaires were useful in evaluating the hedonic, aesthetic, frustration, and pragmatic dimensions of UX. Simultaneously, they also identified that most of the questionnaires failed to measure the enchantment, playfulness, and motivation dimensions of the system. Hence, they proposed the use of multiple scales to have a complete understanding of any interactive system.

The study conducted by Sensuse et al. [48] evaluated ELISA, the automated answering machine for users' queries based on the DeLone and McLean model of information system as reference. In this research, they have evaluated the chatbot based on features such as information quality, system quality, user satisfaction, etc. The team lead by Kerstin Denecke [49] has developed a mobile application 'SERMO' integrated with a chatbot which uses techniques from cognitive behaviour therapy (CBT) to help mentally ill people deal with thoughts and feelings. The study carried out by Supriyanto et al. [50] evaluated the user interface of a chatbot on a holiday reservation system based on the keystroke-level model. The next section describes the details of the prototype.

To summarize the outcome learned from recent studies discussed in the above sections: After reviewing current studies on the conversational search interface, we discovered that

there has not been much research on search interfaces that can conduct both conversation and search at the same time. Simple chatbot apps lack information space, which might slow down learning and leave the searcher with a negative experience. We would want to examine the multiview interface for conversational search support picture retrieval, which was inspired by the work of Kaushik et al. [5]. Furthermore, studies on conversational search focus on either question and answer systems or on utilizing people as intermediary search agents who complete the search process for the searcher. Both types of research lack the searcher's continual dialogue engagement and incremental exploratory learning while conducting the discussion during the information seeking process. Furthermore, most conversational search systems are assessed using usability measures, leaving room to investigate and compare them using other scientifically established criteria. Furthermore, the image retrieval system, which is integrated in a conversational context with a self-learning reward-based system, has never been explored and addressed in the multiview search interface. Taking into account all of the preceding considerations, we are motivated to invest in the following field.

3. Prototype Multi-View Interface for Conversational Image Search

Our prototype multi-faceted interface for conversational image search is shown in Figure 1. This interface includes the following components:

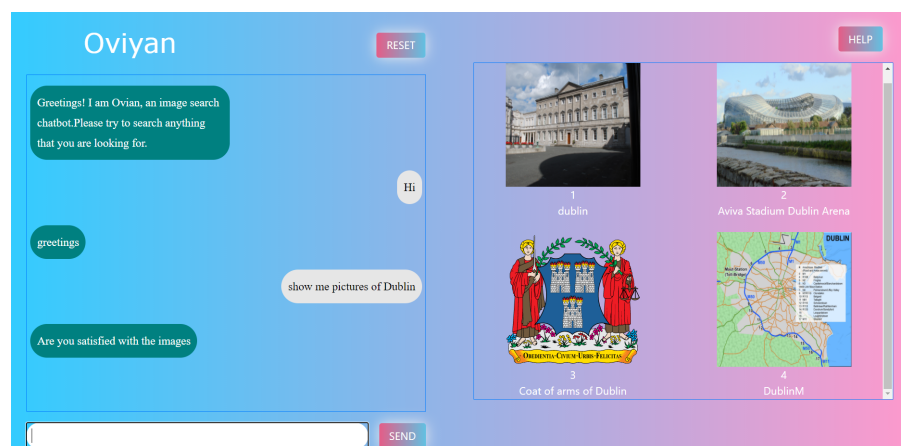


Figure 1. Chatbot Interface.

- **Chat Display**—Shows the conversational dialogue between the agent and the user. This is the place which shows the text that went to and fro between the user and the agent.
- **Chat Box**—The user enters the input text here. The text provided here is used by the agent to provide the results.
- **Image/Video Box**—This is the place where the retrieved images and videos are displayed. The agent shows four images and videos each time and the numbering of the images is also provided inside this box.
- **Send Button**—This button can be used either by pressing the enter button or by a mouse click. This is how the user provides the input text to the agent.

The chatbot interacts with the user using the conversation box and returns the response in the image box. The chatbot comprises of various actions including greeting the user, responding with answers directly, and searching for related images if the user is unsure about the subject they are searching. The functionality of the chatbot has also been further improved using various image retrieval (IR) techniques such as query expansion and gaining feedback from the user on the quality of the results.

- **Query Expansion:** Query expansion was carried out by asking the user to select more than one image from the resultant image. The labels of the selected images are

retrieved and provided as input to the entity extraction module, and the Wikipedia search API and image results were provided to the user.

- **User Feedback:** User feedback is carried out by the chatbot where the agent asks the user whether he is happy with the results provided. If the user enters no, he is provided with more similar results. If the user enters yes, he is provided with an option to do query expansion.

The next sub-sections overview the prototype system implementation, dialogue strategy, workflow, and user engagement.

3.1. System Implementation

The developed model has two sections, viz., the web interface and the logical system.

- **Web interface:** The web interface has been developed using Flask, an extensible web micro framework for building web applications with Python, HTML, CSS, and Bootstrap.
- **Logical System:** The logical system is responsible for the conversation and the image search management. The keyword from the input message is extracted and given as input to the Wikipedia API for searching images.
- **Operations:** The conversation accepts the following commands from the user.
 - Any text with the entity or keyword about which the user needs pictures.
 - Yes/No—This response is accepted by the bot from the user as a feedback on the image if it is relevant.
 - 1,2,3,4—the number of the image that is relevant to the one the user is looking for. This helps the search agent to come up with better images based on the user preference.
 - Get video—This message is received by the bot and four videos are displayed to the end user using the YouTube API.
 - End—This message is acknowledged by the chat bot that the user is satisfied with this input and looking to proceed for next topic.

3.2. Dialogue Strategy and Taxonomy

In searching, while the scope of the user tasks can generally be well defined, the topics are very diverse and unstructured. This poses unique challenges not faced by standard conversational systems which perform clearly defined tasks of limited scope [1]. After careful examination, we have classified the dialogue strategy into three phases. These three phases comprise the strategy to obtain the keyword from the user, return image response or a video response, and try to continue the conversation until the user aborts intentionally. The agent can seek feedback from the user regarding the satisfaction with results. If the user is not satisfied the agent will try to come up with better image results. If the topic searched by the user is unavailable, then the agent tries to come up with more relevant search results. Using the above-mentioned strategies, the agent tries to keep the chat interactive and provide more relevant image responses to the user. The dialogue strategy is structured as follows:

- **Phase 1:** In this initial phase, the chat-bot will try to greet the user and ascertain the query request using causal conversation.
- **Phase 2:** This phase begins when the user starts asking questions to the chatbot. The chatbot attempts to collect feedback from the user about the provided image results. If the user is not satisfied with the image, then the chatbot will try to show a more appropriate image.
- **Phase 3:** This phase begins at the end of phase 2, wherein the chatbot tries to end the conversation by asking the user to reply 'End' to stop the loop and proceed to the next query.

3.3. System Workflow

The system workflow mainly has the two sections, viz., conversation management and image search management, as shown in Figure 2:

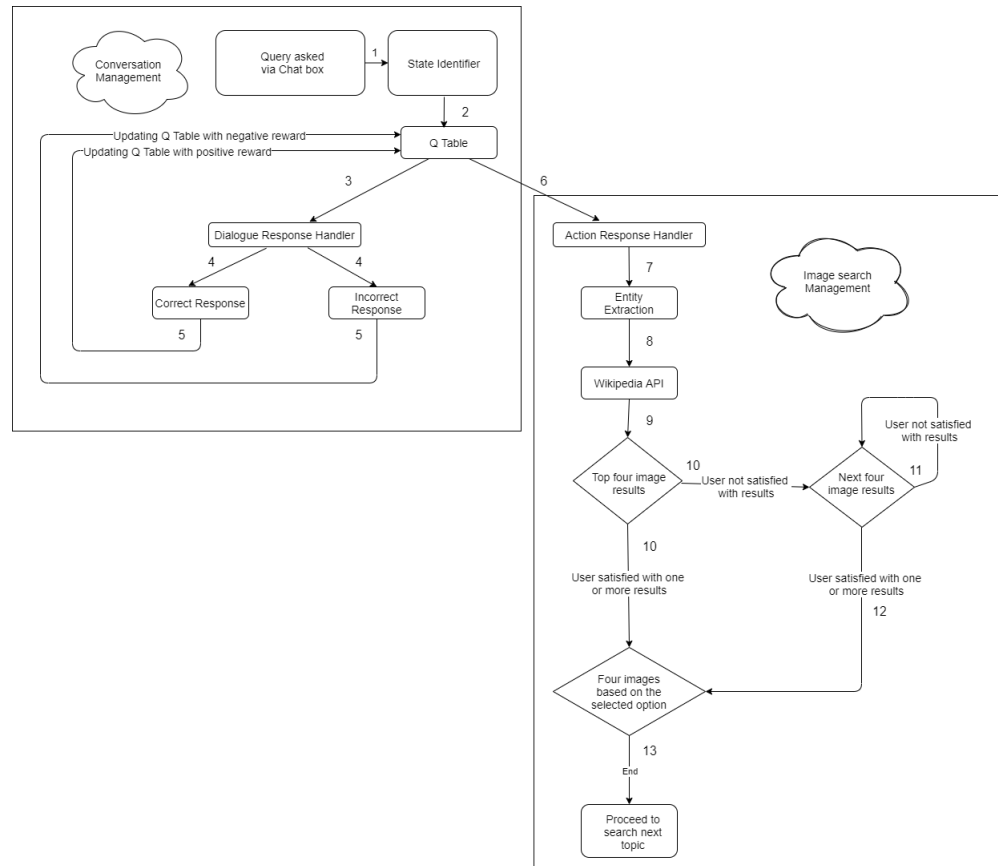


Figure 2. Workflow of the conversation model.

3.3.1. Conversation Management

The conversation management comprises the State Identifier. Once the user gives any input, the state identifier compares the text with a list of text such as Yes/No, (1,2,3,4), get video, and more, as discussed in Section 3.1. Based on this comparison, the state identifier finds the relevant state.

For example, the decision of whether to consider the input string as either a greeting intent or the search intent is taken by the state identifier. If it is a greeting intent, it replies to the user with a greeting and the user can proceed to query his desired question. If it is a search intent, then it is given as input to the entity extraction. The categorisation is conducted by verifying the input string against a list of greeting words. If the input matches the greeting words then it is categorised as a greeting intent, and if the input string does not match any of the greeting words then it is categorised under the search intent.

3.3.2. Image Search Management

The image search management comprises the entity extraction and Wikipedia API. The entity extraction module uses two open source libraries to extract the keywords from the input string. The output from the entity extraction module is concatenated and then given as input to the Wikipedia search API.

- The Gensim library is used to remove the stop-words from the input string.
- The Spacy library uses the en_core_web_sm model to extract the nouns as keywords from the input string.

The keywords from the conversation are extracted using the Entity extraction component and then the extracted words are given as input to the Wikipedia API for image search. For example, Question: Show me something about Dublin. Output from entity extraction: Dublin. If the chatbot is not able to identify any keyword from the user input through entity extraction, it prompts the user to provide a valid input. The image search management is responsible for displaying relevant images. Here, the image search box displays the top four images from the search results of Wikipedia API. The images are listed using numbers. If the users are looking for images related to any image that has been displayed, then giving the number as input to the chat bot will enable the chat bot to look for more specific results. The image search extracts images from Wikipedia pages of the given keyword, and only the images that contain the keyword are being extracted and shown to the user. An average of twenty relevant images will be shown for every search. The suggestions about the relevance of the image will be enquired through the conversation management, which works together with the image search in sync.

1. **Image search algorithm:**
Our prototype conversational agent adopts a logic-based system to filter out images from the collection of images provided through the Wikipedia API. From the user input, the search keyword is extracted and used as input to the Wikipedia search API, which retrieves all the Wikipedia page titles based on the search keyword. These titles are then used as input to retrieve relevant Wikipedia pages. Then the images from these pages are extracted along with their labels and compared with the search keyword. If it matches, then the first twenty images are used for display purpose to the user. The entire flow has been shown graphically in Figure 3. We have taken care of all the exceptions that might occur during the search process. The algorithmic flow of the image search logic is as shown in Algorithm 1.
2. **User navigation:**
To make the experience for the user to be more lively, both the conversation tab and the image search tab work together seamlessly. Whenever the user interacts with the chatbot through the conversation tab, he/she will be able to view the image results immediately on the image search tab, which makes it convenient for the user to adapt to the working of the chatbot. The numbering of the images greatly helps the user to specify the images which they are more interested in. To incorporate this methodology, the query expansion concept has been applied on the images fetched from the Wikipedia API.
3. **User Relevance Feedback:**
In the study, it was evident that sometimes the user is not able to find the most appropriate image on the first search. Hence, to help them in finding the image of interest, we have implemented the feedback mechanism, wherein the chatbot will record the number of the image that is of interest to the user, and once again search the images based on the new keyword from the image selected by the user.

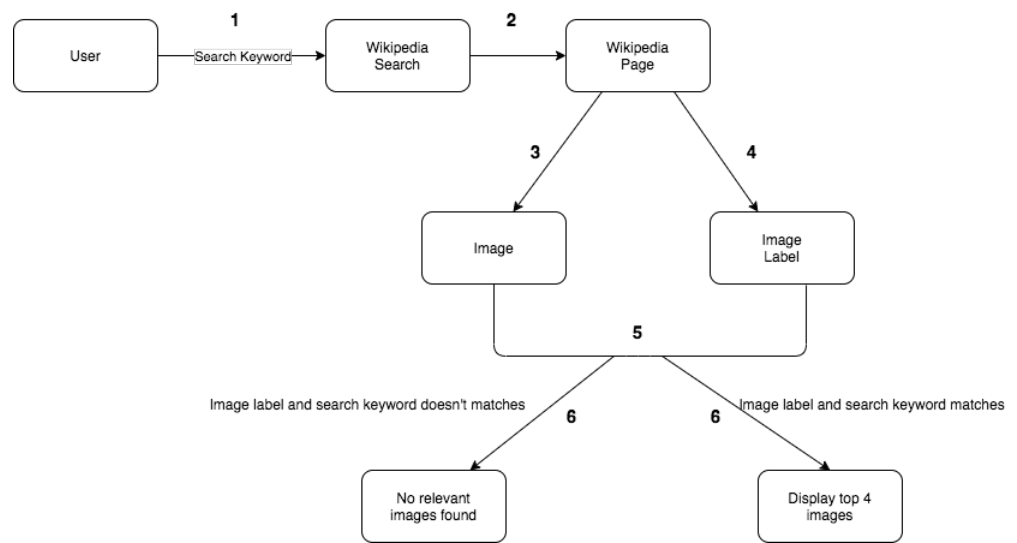


Figure 3. Search algorithm.

Algorithm 1 Image search

- 1: Search the input text using the Wikipedia search API and retrieve all the relevant topics.
 - 2: **for** $topics = 1, 2, \dots, N$ **do**
 - 3: Retrieve all the pages by providing the topics as input to the Wikipedia Page API.
 - 4: **for** $Page = 1, 2, \dots, N$ **do**
 - 5: Retrieve all the images using the Wikipedia Image API.
 - 6: **for** $Image = 1, 2, \dots, N$ **do**
 - 7: get the image label
 - 8: **if** image label matches input text **then**
 - 9: add the image to the image list
 - 10: **end if**
 - 11: **end for**
 - 12: **end for**
 - 13: **end for**
-

4. Reinforcement Learning Modelling

Reinforcement learning is a domain of machine learning concerned with how software agents ought to take actions in an environment to maximize the reward provided by the environment. It is applied in various software, games, machines, and robots to find the best possible behaviour or path it should take in a specific situation. The reinforcement learning module holds three elements which are states, actions, and rewards. The objective of the model is to predict which action to choose based on the current state to receive the maximum reward. There are various types of machine learning models, namely the Q-Learning, SARSA (State-Action-Reward-State-Action), etc.

Q-learning is a model-free reinforcement learning algorithm, that learns a policy telling an agent what action to take under what circumstances. It is called off-policy because the updated policy is distinct from the initial behaviour policy. In different words, it estimates the reward for future actions and appends a value to the new state without actually following any greedy policy. Therefore the conversation agent in our system initially provides wrong actions, and then after multiple routes, it starts to give the correct

actions based on the current states. There are four important components in the Q-Learning model, namely State, Action, Reward, and Q-table. We developed the conversational agent with five states and five possible actions for each state. Based on the actions in each state, the rewards will be provided, which is updated in the Q-table.

- State: In the conversational environment created, the state can be described as the possible replies that the user can provide during a conversation. The “Ovian” environment contains five states. They are:
 - Greeting;
 - Question;
 - Affirmative response;
 - Video query;
 - Negation,1,2,3,4;
 - End.
- Actions: In the conversational environment created, the action can be described as the possible replies that the agent can provide during a conversation. The “Ovian” environment contains five Actions. They are:
 - Greeting.
 - Are you satisfied with the images?
 - If you want more specific search results, enter the image number/numbers? Reply with “end” to complete this search.
 - Sorry, I am unable to find anything relevant to this topic. Please try any other topic.
 - Thank you for your insight. Please continue with your search.
- Rewards: Reward functions define how the conversational agent ought to behave. In other words, they have the regulating content, stipulating what you want the agent to accomplish. In general, a positive reward is provided to encourage certain agent actions, and a negative reward to discourage other actions. A well-designed reward function leads the agent to maximize the expectation of the long-term reward. In any environment, both continuous or discrete reward values can be provided. In this prototype, if the agent behaves as per the expectation, then a positive reward of 10 is provided, and if the agent behaves not as expected, then a negative reward of 10 is provided. The rewards were provided based on the expectation of the way the conversation agent should work. For example, if the user greets the agent, the expectation of the agent is to greet the user back. If the expectation is satisfied, then the reward +10 is granted and if the expectation is not satisfied then the reward –10 is granted.
- Q-table: Q-Table is just a fancy name for a simple lookup table where we determine the maximum anticipated future rewards for action at each state. Each Q-table score will be the maximum expected future reward that the agent will earn if it takes that action at that state. The rewards collected in each step are accumulated and it is used to enrich the Q-table during the training or the learning phase. When the learning phase is completed, the reinforcement model uses the enriched table to choose the action for that state.

4.1. Training Phase

In the training stage (as shown in Figure 4), the agent returns random actions for every state and tries different combinations of replies for each state and input. The reward is provided by the reward handler in the environment for each action taken by the agent. Using the rewards the q -value is calculated using the formula given below.

$$q\text{-value} = (1 - \alpha) * \text{old-}q\text{-value} + \alpha * (\text{reward} + \gamma * \text{maxvalue}) \quad (1)$$

where

$$\alpha = 0.1 \text{ and } \gamma = 0.6.$$

After the q -value is calculated, the new q -value is updated in the Q-table. Likewise, when more conversations happen, the agent investigates various combinations of replies and actions, and thereby enriches the Q-table with the q -values. Training of the prototype conversational agent was performed using two methodologies, namely Open-AI Gym and expert users.

4.1.1. Open-AI Gym

Open-AI Gym is a toolkit for developing and training reinforcement learning algorithms. This is an open-source interface for reinforcement learning tasks. The Open-AI Gym offers various environments or playgrounds of games all packaged in a Python library, to make Reinforcement learning environments available and easy to access from your local computer. Available environments range from Pacman and cart-pole, to more complex environments such as mountain car, etc. The Open-AI Gym does not provide an environment for a conversational agent. Thus, as a part of this project we developed a custom gym environment “GymChatbot-v0” for our conversational agent which simulates the role of the user who explores the images. The input conversations are configurable and can be customized as per our requirement. The gym training process was carried out. Gym training is the process of training the chatbot using the Open-AI gym. Here, the gym environment will simulate the activity of the user by providing the agent with pre-defined conversations. When the agent returns the relevant conversations and images, a positive reward of +10 is provided by the gym environment, and when irrelevant results are provided by the agent, a negative reward of -5 is awarded by the gym environment.

4.1.2. Implementation Phase

Once the training phase is completed and all the actions are investigated for the various states, the Q-learning model starts to provide the action based on the max value in the enriched Q-table for a particular state. Now, the actions provided by the agent are suitable for the states, compared to the initial learning phase. Thus, without implementing any conditional operators, the agent has learned to provide actions to any particular state in the implementation phase, based on the training provided in the learning phase. The flow of the trained phase is as shown in Figure 4. Here, the step 4 is where the trained model differs from the training phase. In the trained phase, the agent selects the action with the highest Q-value instead of collecting rewards.

The evaluation phase comes after the implementation phase. The assessment procedure is separated into two parts: (a) empirical evaluation and (b) user evaluation. The model will be evaluated empirically based on the training and learning rates, which are discussed in depth in the next section. As a result, we were able to train using the best, most reinforced learning model for deployment. Once the model has been deployed with the multi-view interface, it is assessed using three separate user assessment metrics in order to understand the user’s true and practical expectations with the interface. Following the part on empirical assessment, the specifics of the user evaluation and experiment are discussed.

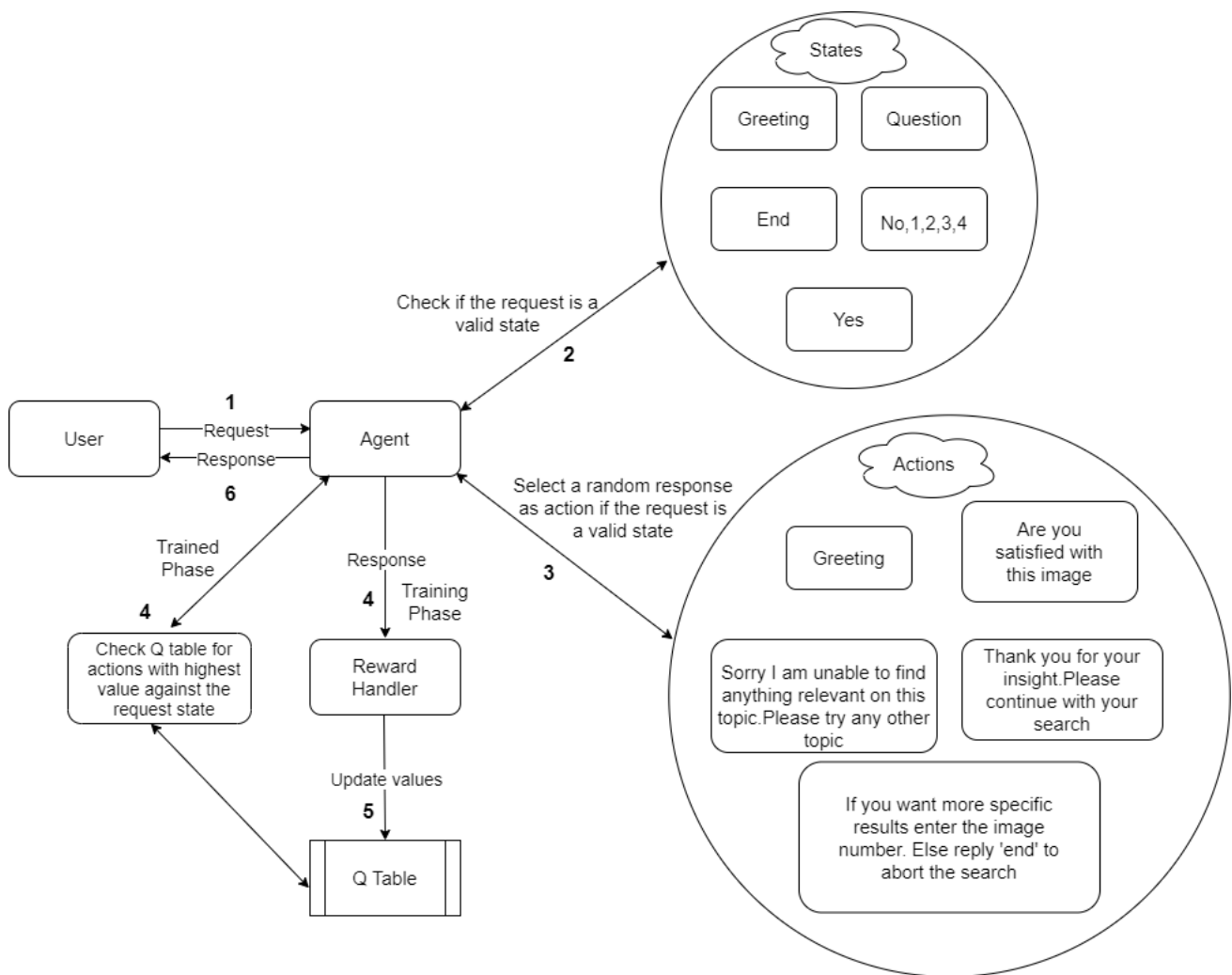


Figure 4. User training image.

4.1.3. Empirical Model Evaluation

To evaluate the performance of the chatbot using the feedback from the users, initially we need to train the reinforcement learning model, and then deploy the trained model in the chatbot environment so that it can converse effectively with the user. We have opted to train the model using the Open AI gym environment. One of the key factors in training the model is to keep it cost effective, which means using the optimal number of episodes to train the model. One episode refers all the states that comes in between an initial-state and a terminal-state. Therefore, as part of this research, the reinforcement learning agent was trained through three different episode count (100, 250, and 500), and the corresponding results were plotted as shown in Figure 5. From the three images, we concluded that the agent had collected negative rewards in the initial stages and gradually moved towards higher rewards. Upon comparison, we concluded that 250 episodes would be cost effective. The next step is to find the optimal epsilon value. Epsilon value is used when a certain action is selected based on the Q-values we already have for all the available actions for a specific state. Epsilon value can be between 0 and 1. For example, when the value of epsilon is 0, which is pure greedy, we are always selecting the highest q -value among all the q -values for a specific state. It is the value which determines whether the reinforcement learning agent should follow a greedy policy, an explore policy, or an optimal policy which has the effect of both the flavours. As a part of this research, we compared three different epsilon values (0.2, 0.5, and 0.7), and used them to find the rewards accumulated by the agent as the number of episode increases. A graph is created with the number of episodes

on the x-axis and number of rewards accumulated in the y-axis, and the corresponding results are plotted as shown in Figure 6. From the figure, we can clearly understand that epsilon values 0.5 and 0.7 are the optimal values, because when we used epsilon values of 0.5 and 0.7, the rewards accumulated gradually and increased as the number of episodes increased. In the case of epsilon value 0.1, sometimes there were negative rewards even after 50 episodes. As our environment has very limited number of states, both 0.7 and 0.5 gave similar results. Therefore, we opted for epsilon value of 0.5. The trained Q-table was then deployed and 10 different users were given 2 tasks each from a collection of 10 unique tasks. Latin square method was used to reduce systematic error.

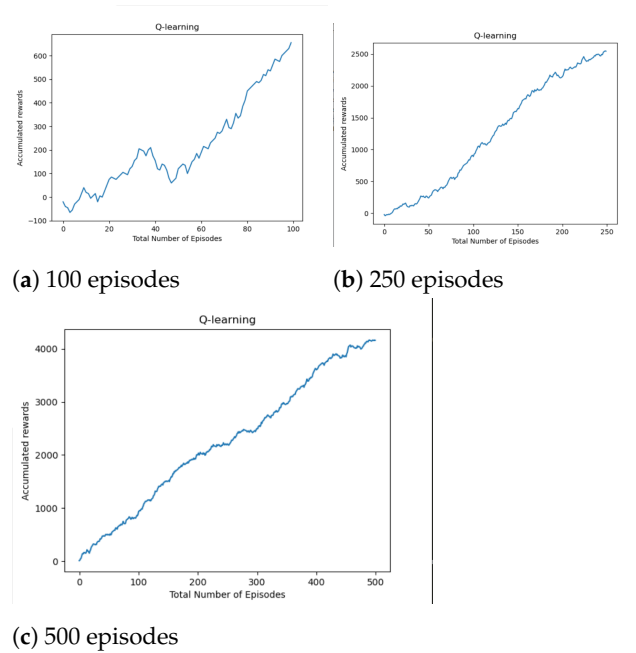


Figure 5. Rewards for various number of episodes.

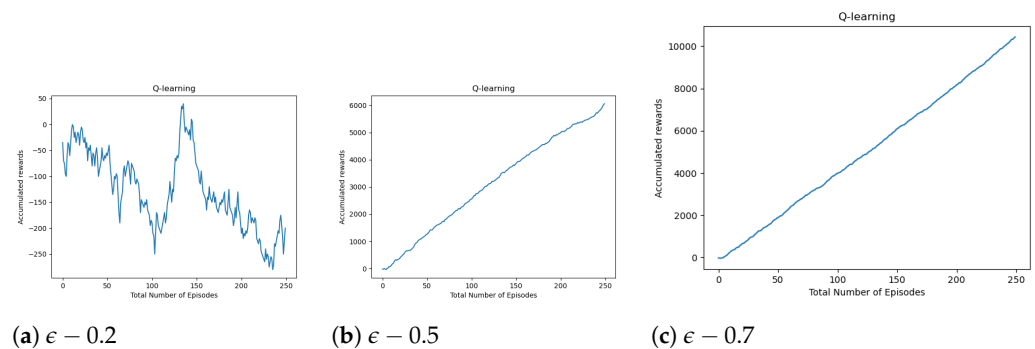


Figure 6. Accumulated rewards for various epsilon values.

5. Experimental Procedure And User Evaluation

We conduct empirical evaluation to identify the optimum training model, based on comparison of outcomes with different values of epsilon. We then propose to deploy the model and collect user input on three common user evaluation metrics, CUQ, UEQ, and SUS. We also look at the correlation between all of the indicators to see how consistent the user reaction is. For example, if a user scores a high UEQ score but a low CUQ score, this indicates a difference in the user’s response and leads to the conclusion that the response was not taken seriously. We identified a significant positive correlation between all of the indicators, implying consistency in user replies. The specifics of the experiments and their outcomes are explained in the next section. A pilot study was conducted as a

first methodology to collect user feedback and improve the conversational agent before deploying it for extensive testing. Three subjects have participated in pilot study. The feedback observed from them was used to enhance search interface from the main user study.

After the pilot study, controlled experiments were carried out to learn about the conversational agent's user experience. The entire web application, which was developed using the Python flask framework, was deployed on our personal computer using the tomcat server.

Ten search tasks were selected from the UQV100 test [51] search task selections, which consists of 100 search tasks from the TREC 2013 and 2014 Web tracks. The ten search tasks have been classified according to their level of cognitive complexity based on the Taxonomy of Learning [52].

The sample of search task is shown in Figure 7. Ten subjects were selected from various domains and were given two individual tasks, each from the set of ten tasks. Pairs of search tasks for each session were selected using a Latin square method to avoid the sequence effect [53].

You have heard quite a lot about cheap computing as being the way of the future, including one recent model called a Raspberry Pi. You start thinking about buying one, and wonder how much they cost

Figure 7. Example Search task from UQV100 test collection.

5.1. Questionnaire

While doing their search activity, the subjects had to fill an online questionnaire in a Google form. For each search task, the subject filled out a questionnaire divided into two sections:

- Basic Demography Survey: Subjects entered their assigned user ID, age, occupation, task ID to be undertaken.
- Post-Search Usability Survey: Post-search feedback from the user including three metrics: SUS, CUQ, and UEQ.

The subjects were requested to join a zoom call, and they were given access to the computer on which the web application was deployed. Then they were asked to interact with the agent to complete the assigned tasks. As they started to converse with the agent, the agent provided relevant images, videos, and replies. The users' feedback was received and archived for further improvements and calculating the various scores. The flow of conversations is shown in Figure 4. We have followed all the guidelines outlined by the research ethics committee guidelines prior to beginning user feedback step. A sample video of the entire workflow of the chatbot can be accessed using the link (<https://www.youtube.com/watch?v=AyTOdWZeSTg> (accessed on 11 September 2021)).

5.1.1. Chatbot Usability Questionnaire

The CUQ is based on the chatbot UX principles provided by the ALMA Chatbot Test tool, which assess personality, on-boarding, navigation, understanding, responses, error handling and intelligence of a chatbot [45]. The CUQ is designed to be comparable to SUS except it is bespoke for chatbots and includes 16 items. The CUQ scores were calculated out of 160 (10 for each question) using the formula in Equation (2), and then normalized by multiplying by 1.6 to give a score out of 100, to permit comparison with SUS. We took the mean and the median value for all the tasks and we found the values 77.54 and 81.25, respectively. The scores show us that the chatbot is highly usable due to its simple interface and conversation-driven functionality.

$$CUQ = \left(\left(\sum_{n=1}^m 2n - 1 \right) - 5 \right) + \left(25 - \left(\sum_{n=1}^m 2n \right) \right) * 1.6 \quad (2)$$

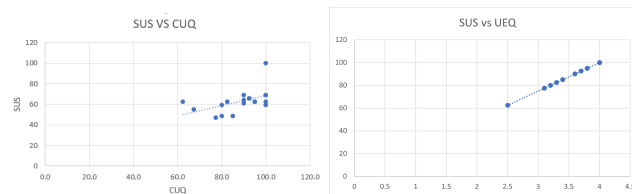
where $m = 16$ and $n =$ individual question score.

5.1.2. System Usability Scale

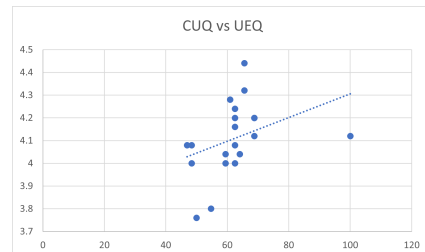
SUS is comprised of ten validated statements, covering five positive aspects and five negative aspects of the system. Participants score each question out of five [44]. Final scores are out of 100 and may be compared with the SUS benchmark value which is 68. The SUS score calculation spreadsheet was used to calculate SUS scores out of 100. The SUS questionnaire comprises of 10 questions which are valued from 0 to 4 each. Then, the summation of the individual scores is multiplied by 2.5. As adding them would only give a score out of 40, and as we want a score in the possible range of 0 to 100, the score is multiplied by 2.5. This can be interpreted from Equation (3). The lowest value was 62.5 and the highest value was 100. The median value was 90 and the mean value was 89.625. Correlation plots were drawn comparing the scores retrieved from the SUS, CUQ, and UEQ, as shown in Figure 8.

$$SUS = 1/n \sum_{i=1}^n norm \sum_{j=1}^m \begin{cases} q_{i,j}^{j-1}, q_{i,j}^{mod2>0} \\ 5 - q_{i,j}, otherwise \end{cases} \tag{3}$$

where $n =$ number of subjects (questionnaires), $m = 10$ (number of questions), $q_{(i,j)} =$ individual score per question per participant, and $norm = 2.5$.



(a) Scatter plot of CUQ and (b) Scatter plot of SUS and UEQ scores



(c) Scatter plot of CUQ and UEQ scores

Figure 8. Scatter Plots.

5.2. User Experience Questionnaire

The UEQ is a fast and reliable questionnaire to measure the user experience of interactive products. By default, the UEQ does not generate a single score for each participant, but instead provides six scores, one for each attribute [45]. The UEQ score calculation excel sheet was used to calculate the UEQ score. It scores the UI on six qualities, attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. The scores given by the users were from a scale of 1 to 7, and the cumulative collection of all the scores were collected and plotted in the graph shown in Figure 9. The chatbot scored highly in all UEQ scales. Scores were well above the benchmark and are presented graphically in Figure 9. The scores for each scale were all above +2 except the attractiveness of the chatbot, suggesting that, in general, participants were satisfied with the Ovian chatbot user experience.

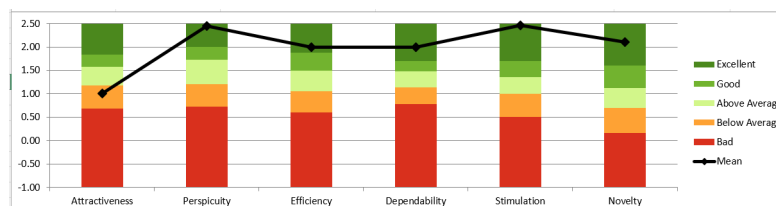


Figure 9. UEQ results on a scale of -1 to 2.5 .

6. Results and Analysis

In this section, we elaborate on the results and learning from the study conducted. We evaluated this image search interface on three standard different usability metrics invigorated by the evaluation framework developed by Kaushik et al. [2]. During the study, we have explored major exploratory research questions (RQ). The findings of this research question are as follows:

6.1. Findings

1. How can reinforcement learning be used for improving the search experience of the user?

In this study, we used the basic model of reinforcement learning using Q-learning algorithm. We evaluated the model empirically during training and quantitatively through user evaluation during testing. The scores obtained on all three metrics (SUS, CUQ, and UEQ) by testing the trained interface have outperformed the baseline scores discussed in the section. The score indicated the potential of reinforcement techniques on the conversational search concepts. Based on the score obtained by the user study and observing the positive high correlation (as shown in the scatter plot) among the metrics, we can infer that Q-learning technique in a conversational setting could be a potential approach in complex information-seeking tasks.

2. Are multiple interactive usability metrics associated, and do they follow a consistent pattern based on user reactions when using the multimodal interface?

Based on the input supplied by users after using the search interface, this graph (Figure 8) depicts the similarity of the responses mentioned by users. This confirms that consumers have comparable experiences after interacting with the UI. As seen in Figure 8, all of the responses are significantly connected and highly related. After analysing the scores of different metrics, it was possible to infer that users had a very pleasant experience when looking for information, and the hybrid approach proved beneficial with the picture retrieval system. This suggests an intriguing conclusion for further investigation. This interface might be further tested based on the user's cognitive load and knowledge expansion while using this interface for searching.

The learning points from the study are discussed in the following section.

6.2. Learnings

This section highlights some of the key learning points that were achieved through this research work.

- Possibility of using a Reinforcement learning for the incremental search process
This knowledge is derived from our examination of RQ1. As observed from the metrics evaluation score, the users who used this system have reported better interactive and usability experiences while seeking information on the cognitive complex task. Searching for images to satisfy the information need has enhanced the difficulty of the task by limiting the information mode and space. Users can only satisfy their information requirements by solely relying on the images provided by the search interface. Including all these challenges, the users' observations were positive and this directed to the potential of using a Q-learning-based reward system in the process of search, which can capture the user's search behaviour.

- **Combination of Images and Videos**
This understanding stems from our consideration of RQ1 and RQ2. As mentioned earlier, this interface is restricted to image search and video search to satisfy user's information needs. Based on the user evaluation, they find it rather interesting to fulfil their needs from the video and images without reading through long documents. Satisfying information needs through long documents can increase the cognitive load while accessing too much information. Image search could be used to reduce the cognitive load during the search, which needs to be further investigated. Based on the feedback in this investigation, the initial results have pointed in the same directions.
- **Conversational Search Problems and its Potential Solutions**
This comprehension arises from our examination of RQ1 and RQ2. The common problem faced by researchers who are working in the area of conversational search is the lack of availability of a data set that can completely capture the user's search behaviour. Creating a similar dataset is very expensive in terms of effort and time. Another challenge is dealing with high language models, which can capture contextual meaning, but not the patterns of user behaviour. The study conducted by Kaushik et al. [1,2] clearly indicated the factors considered during conversational search, which in general are missing in the heavy language models. In contrast, the approach mentioned in this paper is not completely dependent on heavy language models or huge data sets, but rather provides unique and novel solutions to capture user behaviour using reinforcement learning techniques. This could also encourage researchers to think about the concept of Explainability when dealing with conversational search bots.

7. Discussion

The study conducted in this paper shows us the possible extensions and directions to the concepts of conversational search. The key finding of the paper is to explore the implications of using reinforcement learning in the light of dialogue strategy. The study conducted by Kaushik [5] focused on the rule-based and machine-learning-based approaches to conversational search, and this study extends a step further by casting the dialogue strategy into a reward-based Q-learning system that can capture the user's seeking behaviour. The system proposed by Kaushik et al. was majorly focused on text-based search [5]. In contrast, the system discussed in the study is based on an image retrieval system in conversational settings. The studies conducted in the past related to conversational search were broadly divided into four dimensions: existing conversational agents (search via Alexa, Siri, etc.) [54], human experts [55,56] (such as a librarian who will search for you), Wizard-of-Oz approach methods [57–59] (similar to human experts, but the searcher is not aware that they are dealing with a human being; the searcher would use a computer machine to search, which is supported by a human agent), and rule-based or machine learning-based conversational interfaces [5,8,60] (systems completely based on data modelling or dialogue strategy). Our approach is very different in comparison to the approach discussed above, where a reward-based system was trained to assist users in their search behaviour.

The other aspect that needs to be discussed is the evaluation of the conversational system. Studies related to conversational search focused majorly on qualitative measures [57,58,61,62] (task completions, SUS, or cognitive load) and empirical measures [8,60,63] (precision, recall, and F1 score). Kaushik et al. [2] introduced the new framework for evaluation of conversational search, which is based on six different factors. Following this investigation, we evaluated the system based on different usability metrics and found the correlation between them, presenting them in a scatter plot to show the similarity between the subjects' grading the systems on user experience.

The approaches used here took a considerable amount of time for the users to evaluate the chatbot, and the usage of evaluation techniques apart from this would be time consuming and discourage the users from doing it effectively. After careful consideration of the various literature works and discussion with the school of psychology, we have concluded that increasing the cognitive load for the user is against the purpose of the

chatbot, as the main goal is to reduce the cognitive load for the user. Overall, this study extends the research not only in conversational search, but also in the approach of using a multi-view conversational search interface used for image and video retrieval. However, the conversational agent can understand any user's search query, but it cannot understand all the user's replies at specific states. For example, when the agent asks the user whether he is satisfied with the results, he can answer either yes or no. Predefined replies are one limitation of the agent and an area for improvement. The recommendation here would be to use entity extraction in all states of the environment to accommodate all the user's queries and replies.

8. Conclusions and Future Work

As a part of this work, we have introduced a multi-faceted user interface for conversational image search. A reinforcement learning model was used to determine the action (replies) provided by the agent in each state. Our conversational search application has a significant difference compared to the current state-of-the-art in the field of conversational search. The unique aspect of this study is that there is no conditional operator or rule-based approach used to influence the decisions made by the agent. The actions taken by the agent were solely based on the state of the environment.

This project was developed to implement and evaluate the elements of reinforcement learning within the interface of a conversational agent used for image search, and investigate the pros and cons. The current prototype can be used to retrieve only images and videos. This can be modified in future endeavours where the knowledge retrieved can be a mixture of images, videos, and text. To train this reinforcement model, we have created an Open-AI Gym which can be used in future research for training a model.

The SUS, UEQ, and the CUQ scores were calculated and found to be above average. This infers that the users found the application easy to use and had a very good user experience. From this research work, we conclude that reinforcement-learning-enabled conversational agents can be implemented to answer search queries in various use cases. This study is a user pilot study, and the results will be used to evaluate further with more subjects, and the states' rewards can be refined with the feedback gained from observations of the user study. The future prospective of this study could be to perform comparative studies between two different versions of the system, one implemented with the help of reinforcement learning, and the other without. The advantages and limitations of a reinforcement-learning-based system could be investigated in comparison with deep-learning-based and rule-based systems.

Author Contributions: The following work has been categorized under different authors name. Conceptualization: A.K. and B.J.; Methodology: P.V.; Validation: A.K. and B.J.; Formal analysis: A.K.; Investigation: B.J. and P.V.; Data curation: B.J.; writing—original draft preparation: B.J. and A.K.; Writing—review and editing: P.V.; Visualization: P.V.; Supervision: A.K.; Project administration: A.K.; Funding acquisition: NA. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was exempt from ethical review and clearance since it did not include any personal information about the subjects or a focus on any specific group or gender. The study is classified as an open survey study by the appropriate institute "Dublin Business School.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Sargam Yadav for proofreading and providing suggestions for the improvement of the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaushik, A. Dialogue-Based Information Retrieval. *European Conference on Information Retrieval*; Springer: Berlin, Germany, 2019; pp. 364–368.
2. Kaushik, A.; Jones, G.J. A Conceptual Framework for Implicit Evaluation of Conversational Search Interfaces. Mixed-Initiative Conversational Systems workshop at ECIR 2021. *arXiv* **2021**, arXiv:2104.03940.
3. Radlinski, F.; Craswell, N. A theoretical framework for conversational search. In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, Oslo, Norway, 7–11 March 2017; ACM: New York, NY, USA, 2017; pp. 117–126.
4. Kaushik, A.; Jones, G.J.F. Exploring Current User Web Search Behaviours in Analysis Tasks to be Supported in Conversational Search. In Proceedings of the Second International Workshop on Conversational Approaches to Information Retrieval (CAIR'18), Ann Arbor, MI, USA, 8–12 July 2018; Volume 16, pp. 453–456.
5. Kaushik, A.; Bhat Ramachandra, V.; Jones, G.J.F. An Interface for Agent Supported Conversational Search. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, 14–18 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 452–456. [\[CrossRef\]](#)
6. Liu, Z.; Chu, W.W. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *Inf. Retr.* **2007**, *10*, 173–202. [\[CrossRef\]](#)
7. Brandtzaeg, P.B.; Følstad, A. Chatbots: User changing needs and motivations. *Interactions* **2018**, *25*, 38–43. [\[CrossRef\]](#)
8. Arora, P.; Kaushik, A.; Jones, G.J.F. DCU at the TREC 2019 Conversational Assistance Track. In Proceedings of the Twenty-Eighth Text REtrieval Conference, Gaithersburg, MD, USA, 13–15 November 2019; Voorhees, E.M., Ellis, A., Eds.; National Institute of Standards and Technology (NIST): Gaithersburg, MD, USA, 2019; Volume 1250.
9. Kaushik, A.; Loir, N.; Jones, G.J. *Multi-View Conversational Search Interface Using a Dialogue-Based Agent*; Springer International Publishing: Cham, Switzerland, 2021.
10. Kaushik, A. Examining the Potential for Enhancing User Experience in Exploratory Search using Conversational Agent Support. Ph.D. Thesis, Dublin City University, Dublin, Ireland, 2021.
11. Sharma, M.; Kaushik, A.; Kumar, R.; Rai, S.K.; Desai, H.H.; Yadav, S. Communication is the universal solvent: Atreya bot—An interactive bot for chemical scientists. *arXiv* **2021**, arXiv:2106.07257.
12. Dalton, J.; Xiong, C.; Callan, J. CAsT 2019: The Conversational Assistance Track overview. In Proceedings of the Twenty-Eighth Text REtrieval Conference, TREC, Gaithersburg, MD, USA, 13–15 November 2019; pp. 13–15.
13. Dalton, J.; Xiong, C.; Callan, J. TREC CAsT 2019: The Conversational Assistance Track Overview. *arXiv* **2020**, arXiv:2003.13624.
14. Sunayama, W.; Osawa, Y.; Yachida, M. Search interface for query restructuring with discovering user interest. In Proceedings of the International Conference on Knowledge-Based Intelligent Electronic Systems, Proceedings, KES, Adelaide, SA, Australia, 31 August–1 September 1999; pp. 538–541.
15. Sandhu, A.K.; Liu, T. Wikipedia search engine: Interactive information retrieval interface design. In Proceedings of the 2014 3rd International Conference on User Science and Engineering: Experience, Engineer, Engage, i-USER 2014, Shah Alam, Malaysia, 2–5 September 2015; pp. 18–23.
16. Negi, S.; Joseph, S.; Alemao, C.; Joseph, V. Intuitive User Interface for Enhanced Search Experience. In Proceedings of the 2020 3rd International Conference on Communication Systems, Computing and IT Applications, Mumbai, India, 3–4 April 2020; pp. 115–119.
17. Hearst, M.; Tory, M. Would You Like A Chart with That? Incorporating Visualizations into Conversational Interfaces. In Proceedings of the 2019 IEEE Visualization Conference, Vancouver, BC, Canada, 20–25 October 2019; pp. 36–40.
18. Bai, T.; Ge, Y.; Guo, S.; Zhang, Z.; Gong, L. Enhanced Natural Language Interface for Web-Based Information Retrieval. *IEEE Access* **2020**, *9*, 4233–4241. [\[CrossRef\]](#)
19. Schneider, D.; Stohr, D.; Tingvold, J.; Amundsen, A.B.; Weil, L.; Kopf, S.; Effelsberg, W.; Scherp, A. Fulgeo-towards an intuitive user interface for a semantics-enabled multimedia search engine. In Proceedings of the 2014 IEEE International Conference on Semantic Computing, Newport Beach, CA, USA, 16–18 June 2014; pp. 254–255.
20. Uribe, S.; Álvarez, F.; Menéndez, J.M. Personalized adaptive media interfaces for multimedia search. In Proceedings of the 2011 International Conference on Computational Aspects of Social Networks, Salamanca, Spain, 19–21 October 2011; pp. 195–200.
21. Shuoming, L.; Lan, Y. A study of meta search interface for retrieving disaster related emergencies. In Proceedings of the 9th International Conference on Computational Intelligence and Security, Emeishan, China, 14–15 December 2013; pp. 640–643.
22. Kanapala, A.; Pal, S.; Pamula, R. Design of a meta search system for legal domain. In Proceedings of the 2017 4th International Conference on Advanced Computing and Communication Systems, Coimbatore, India, 6–7 January 2017; pp. 1–5.
23. Dey, S.; Abraham, S. Personalised and domain specific user interface for a search engine. In Proceedings of the 2010 International Conference on Computer Information Systems and Industrial Management Applications, Krakow, Poland, 8–10 October 2010; pp. 524–529.
24. Heck, L.; Hakkani-Tür, D.; Chinthakunta, M.; Tur, G.; Iyer, R.; Parthasarathy, P.; Stifelman, L.; Shriberg, E.; Fidler, A. Multi-Modal Conversational Search and Browse. In Proceedings of the First Workshop on Speech, Language and Audio in Multimedia, Marseille, France, 22–23 August 2013; pp. 6–20.
25. Fergencs, T.; Meier, F. Engagement and Usability of Conversational Search—A Study of a Medical Resource Center Chatbot. In *International Conference on Information*; Springer: Cham, Switzerland, 2021; pp. 1–8.

26. Fernes, S.; Gawas, R.; Alvares, P.; Femandes, M.; Kale, D.; Aswale, S. Survey on Various Conversational Systems. In Proceedings of the International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 24–25 February 2020; pp. 1–8.
27. Anand, A.; Cavedon, L.; Joho, H.; Sanderson, M.; Stein, B. Conversational Search (Dagstuhl Seminar 19461). *Dagstuhl Rep.* **2020**, *9*, 34–83. [CrossRef]
28. Grycuk, R.; Scherer, R. Software Framework for Fast Image Retrieval. In Proceedings of the International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, Poland, 26–29 August 2019; pp. 588–593.
29. Pawaskar, S.K.; Chaudhari, S.B. Web image search engine using semantic of Images’s meaning for achieving accuracy. In Proceedings of the International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 9–10 September 2016; pp. 99–103.
30. Munjal, M.N.; Bhatia, S. A Novel Technique for Effective Image Gallery Search using Content Based Image Retrieval System. In Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 25–29.
31. Nakayama, K.; Nakayama, K.; Pei, M.; Erdmann, M.; Ito, M.; Shirakawa, M.; Hara, T.; Nishio, S. Wikipedia as a Corpus for Knowledge Extraction. *Wikipedia Min.* **2018**, 1–8.
32. Lerman, K.P.A.; Wong, C. Personalizing Image Search Results on Flickr. *AAAI07 Workshop Intell. Inf. Pers.* **2007**, 12.
33. Xu, W.; Zhang, Y.; Lu, J.; Li, R.; Xie, Z. A Framework of Web Image Search Engine. In Proceedings of the 2009 International Joint Conference on Artificial Intelligence, Hainan, China, 25–26 April 2009; pp. 522–525.
34. Smelyakov, K.; Sandrkin, D.; Ruban, I.; Vitalii, M.; Romanenkov, Y. Search by Image. New Search Engine Service Model. In Proceedings of the International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S T), Kharkiv, Ukraine, 9–12 October 2018; pp. 181–186.
35. Kia, O.M.; Neshati, M.; Alamdari, M.S. Open-Domain question classification and completion in conversational information search. In Proceedings of the International Conference on Information and Knowledge Technology (IKT), Tehran, Iran, 22–23 December 2020; pp. 98–101.
36. Bellini, V.; Biancofiore, G.M.; Di Noia, T.; Di Sciascio, E.; Narducci, F.; Pomo, C. GUapp: A Conversational Agent for Job Recommendation for the Italian Public Administration. In Proceedings of the IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Bari, Italy, 27–29 May 2020; pp. 1–7.
37. Lauren, P.; Watta, P. A Conversational User Interface for Stock Analysis. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 5298–5305.
38. McTear, M.F. The Rise of the Conversational Interface: A New Kid on the Block? Future and Emerging Trends in Language Technology. In Proceedings of the International Workshop on Future and Emerging Trends in Language Technology, Seville, Spain, 30 November–2 December 2016; pp. 38–49.
39. Atiyah, A.; Jusoh, S.; Alghanim, F. Evaluation of the Naturalness of Chatbot Applications. In Proceedings of the Third International Workshop on Conversational Approaches to Information Retrieval (2020), Amman, Jordan, 9–11 April 2019; p. 7.
40. Balog, K.; Flekova, L.; Hagen, M.; Jones, R.; Potthast, M.; Radlinski, F.; Sanderson, M.; Vakulenko, S.; Zamani, H. Common Conversational Community Prototype: Scholarly Conversational Assistant. Informaion retrieval. *arXiv* **2020**, arXiv:2001.06910.
41. Forkan, A.R.M.; Jayaraman, P.P.; Kang, Y.B.; Morshed, A. ECHO: A Tool for Empirical Evaluation Cloud Chatbots. In Proceedings of the 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing, Melbourne, VIC, Australia, 11–14 May 2020; pp. 669–672.
42. Atiyah, A.; Jusoh, S.; Alghanim, F. Evaluation of the Naturalness of Chatbot Applications. In Proceedings of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, Amman, Jordan, 9–11 April 2019; pp. 359–365.
43. Hidayatin, L.; Rahutomo, F. Query Expansion Evaluation for Chatbot Application. In Proceedings of the ICAITI 2018—1st International Conference on Applied Information Technology and Innovation: Toward A New Paradigm for the Design of Assistive Technology in Smart Home Care, Padang, Indonesia, 3–5 September 2018; pp. 92–95.
44. SUS: A Quick and Dirty Usability Scale. Available online: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale (accessed on 5 August 2021).
45. Holmes, S.; Moorhead, A.; Bond, R.; Zheng, H.; Coates, V.; McTear, M. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces. In Proceedings of the 31st European Conference on Cognitive Ergonomics: “Design for Cognition”, Belfast, UK, 10–13 September 2019; pp. 207–214.
46. Abd-Alrazaq, A.; Safi, Z.; Alajlani, M.; Warren, J.; Househ, M.; Denecke, K. Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, 1–10. [CrossRef]
47. Kocabalil, A.B.; Laranjo, L.; Coiera, E. Measuring user experience in conversational interfaces: A comparison of six questionnaires. In Proceedings of the 32nd International BCS Human Computer Interaction Conference, Belfast, UK, 2–6 July 2018.
48. Sensuse, D.I.; Dhevanty, V.; Rahmanasari, E.; Permatasari, D.; Putra, B.E.; Lusa, J.S.; Misbah, M.; Prima, P. Chatbot Evaluation as Knowledge Application: A Case Study of PT ABC. In Proceedings of the 2019 11th International Conference on Information Technology and Electrical Engineering, Pattaya, Thailand, 10–11 October 2019.
49. Denecke, K.; Vaaheesan, S.; Arulnathan, A. A Mental Health Chatbot for Regulating Emotions (SERMO)—Concept and Usability Test. *IEEE Trans. Emerg. Top. Comput.* **2020**, *9*, 1170–1182. [CrossRef]

50. Supriyanto, A.P.; Saputro, T.S. Keystroke-level model to evaluate chatbot interface for reservation system. In Proceedings of the International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Bandung, Indonesia, 18–20 September 2019; pp. 241–246.
51. Bailey, P.; Moffat, A.; Scholer, F.; Thomas, P. UQV100: A Test Collection with Query Variability. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, 17–21 July 2016; ACM: New York, NY, USA, 2016; pp. 725–728. [[CrossRef](#)]
52. Krathwohl, D.R. A revision of Bloom’s taxonomy: An overview. *Theory Into Pract.* **2002**, *41*, 212–218. [[CrossRef](#)]
53. Grant, D.A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.* **1948**, *45*, 427. [[CrossRef](#)] [[PubMed](#)]
54. Hoy, M.B. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Med. Ref. Serv. Q.* **2018**, *37*, 81–88. [[CrossRef](#)] [[PubMed](#)]
55. Trippas, J.R.; Spina, D.; Cavedon, L.; Joho, H.; Sanderson, M. Informing the Design of Spoken Conversational Search: Perspective Paper. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA, 11–15 March 2018; ACM: New York, NY, USA, 2018; pp. 32–41. [[CrossRef](#)]
56. Trippas, J.R.; Spina, D.; Cavedon, L.; Sanderson, M. How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis. In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, Oslo, Norway, 7–11 March 2017; ACM: New York, NY, USA, 2017; pp. 325–328. [[CrossRef](#)]
57. Ghosh, D.; Foong, P.S.; Zhang, S.; Zhao, S. Assessing the Utility of the System Usability Scale for Evaluating Voice-based User Interfaces. In Proceedings of the Sixth International Symposium of Chinese CHI, Montreal, QC, Canada, 21–22 April 2018; ACM: New York, NY, USA, 2018; pp. 11–15. [[CrossRef](#)]
58. Avula, S.; Chadwick, G.; Arguello, J.; Capra, R. SearchBots: User Engagement with ChatBots during Collaborative Search. In Proceedings of the 2018 Conference on Human Information Interaction&Retrieval, New Brunswick, NJ, USA, 11–15 March 2018; ACM: New York, NY, USA, 2018; pp. 52–61.
59. Avula, S. Wizard of Oz: Protocols and Challenges in Studying Searchbots to Support Collaborative Search. In Proceedings of the SIGIR 2nd International Workshop on Conversational Approaches to Information Retrieval, Ann Arbor, MI, USA, 12 July 2018; pp. 111–116.
60. Kaushik, A.; Bhat Ramachandra, V.; Jones, G.J.F. DCU at the FIRE 2020 Retrieval from Conversational Dialogues (RCD) task. In Proceedings of the FIRE 2020 Proceeding, Hyderabad, India, 16–20 December 2019.
61. Avula, S.; Arguello, J.; Capra, R.; Dodson, J.; Huang, Y.; Radlinski, F. Embedding Search into a Conversational Platform to Support Collaborative Search. In Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, Glasgow, Scotland, UK, 10–14 March 2019; pp. 15–23.
62. Trippas, J.R.; Spina, D.; Cavedon, L.; Sanderson, M. Crowdsourcing User Preferences and query Judgments for Speech-Only Search. In Proceedings of the 1st SIGIR Workshop on Conversational Approaches to Information Retrieval, Tokyo, Japan, 7–11 August 2017.
63. Kiesel, J.; Bahrami, A.; Stein, B.; Anand, A.; Hagen, M. Toward Voice Query Clarification. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1257–1260. [[CrossRef](#)]