

Voice-Chatbot Interface for Hospitality

Sayed Suhel Mohd Salim

Anjuman-I-Islam's College of Hotel & Tourism Management Studies & Research, Mumbai

Abstract: *Voice chatbots are a new and emerging technology that has the potential to revolutionize the hospitality industry. By providing a natural and intuitive way for guests to interact with hotels, voice chatbots can help to improve the guest experience, reduce costs, and increase efficiency.*

This research article explores the potential of voice chatbots for the hospitality industry. It reviews the current state of the art in voice chatbot technology, discusses the benefits and challenges of using voice chatbots in hospitality, and identifies opportunities for future research.

A chatbot is a machine that can respond to queries via a conversational user interface automatically. One of the most remarkable and exciting forms of human-computer interaction is the chatbot. Artificial intelligence (AI) devices, such as voice-based chatbots, revolutionize bidirectional human-computer communications by enabling users to interact with interactive voice response (IVR) systems using natural language. In this research, we examine voice-based chatbots as a means of mediating interactions between guests and hotels from the viewpoints of both visitors and hospitality technology providers. We created a web application for hotels that can be used to enter voice commands. A closed domain question answering (cdQA) natural language processing (NLP) solution was utilized to query the answer in the application, which was constructed using speech recognition and deep synthesis API for voice to text and text to voice conversion.

Keywords: Chatbot, Natural Language Processing, Closed Domain Question Answering, Voice Based Digital Assistants

I. INTRODUCTION

The hospitality industry is constantly evolving in order to meet the changing needs of its guests. One of the most recent trends in the industry is the adoption of new technologies, such as voice chatbots.

Voice chatbots are computer programs that can simulate conversation with humans. They are typically powered by artificial intelligence (AI) and natural language processing (NLP) technologies. Voice chatbots can be used to provide a variety of services, such as answering questions, making reservations, and providing customer support.

Voice chatbots have the potential to revolutionize the hospitality industry by providing a new and innovative way for guests to interact with hotels. Guests can use voice chatbots to book rooms, check in and out, order food and drinks, request services, and get information about the hotel and the surrounding area.

A chatbot is a programming interface that uses text or voice interactions to mimic a human's "chatter" or conversation. These days, chatbots may be found in practically every area of technology, including e-commerce, mobile assistants, smart gadgets, and customer support. This kind of software facilitates communication between users and consumers by automating talks over messaging networks. These virtual assistants, or chatbots, are helpful for basic look-up activities in business-to-business and business-to-consumer contexts. Virtual chatbot assistants are useful for saving time for support staff members and for offering a certain level of customer support in the event that the agents are unavailable. Chatbots comprehend and analyze user input, providing them with a preprogrammed response in real time. The selection of an appropriate natural language processing (NLP) engine is the most crucial step in the implementation of a chatbot. The chatbot needs a speech recognition engine, for instance, if the user communicates with it vocally. Chatbots have an application layer, a database, APIs, and a Conversational User Interface (CUI), much like ordinary apps do. Conversations can be both structured and unstructured. Although they simplify programming, chatbots designed for organized discussions are heavily scripted and limit the types of questions that users can ask. Chatbots are typically programmed and utilized in B2B applications to answer frequently asked inquiries or carry out basic, repetitive calls to action. A chatbot could provide sales representatives with a simple way to obtain phone numbers. It helps service

agents respond to frequently asked questions, which benefits service departments. A human support representative will usually take over a call or text window if the conversation becomes too complicated for a chatbot.

Early attempts to develop software that could, at the very least, momentarily trick a real person into believing they were speaking with someone else were made with chatbots like ELIZA and PARRY. An early version of the Turing test was used to measure PARRY's efficacy; testers could only correctly identify a person from a chatbot at a level consistent with making a random guess.

Since then, chatbots have advanced significantly. They require enormous amounts of data and are based on artificial intelligence (AI) technology such as deep learning, machine learning (ML) algorithms, and natural language processing. Voice recognition gets stronger at anticipating the right response the more an end user engages with the bot. Chatbots can be broadly divided into three categories:

Rule-based chatbots are the most basic kind. To provide pertinent answers, they ask the user to pick a few choices, such as by using buttons or drop-down menus. Although they are simple to use, they are slow. This method might not be the ideal choice when the knowledge base has a lot of conditions or factors.

Chatbots that are intellectually independent: These chatbots employ machine learning to pick up knowledge from the input and demands of users. These particular bots have been trained to recognize particular words and phrases, which prompts the bot to respond. With effort and experience, they teach themselves to comprehend an increasing number of queries. They identify terms or phrases and offer a predetermined response based on these terms or phrases.

(c). Chatbots driven by AI: It blends the most intelligently independent and rule-based chatbot features. These chatbots are capable of comprehending natural language and ensure that the problems they resolve follow a preset path. They can respond to arbitrary user requests at any time and change the conversational scenario as needed. These chatbots analyze and comprehend human speech, determine the appropriate response, and respond in a comprehensible manner using machine learning, artificial intelligence, and natural language processing (NLP).

In general, chatbots are regarded as one of the most promising and sophisticated forms of human-machine interaction. Applications for chatbots facilitate more seamless customer-service interactions. With less human intervention, they can improve and engage customer relationships.

Iuliia Moldavska and Dimitrios Buhalis discussed the value of voice assistants in the hotel sector in. They make it quite evident that the benefits of voice assistants in hotels exceed the drawbacks for both lodging facilities and visitors. Their research shows that voice assistants will be used extensively in the future and that voice-based human-computer interactions have a number of advantages. Although setting up technology integrations can be difficult and expensive, they have certain advantages, particularly for the hotel and tourism sectors. According to their reports, while tech-savvy consumers are less concerned about privacy while using voice assistants, visitors still enjoy the potential benefits but have concerns about usability and privacy. The results pointed to a multilingual and modulated offer path for speech technology in the hospitality sector going forward, which will eventually guarantee the technology's broader adoption across the board.

A real-world conversational AI and natural language processing system for hotel booking was examined by Li, Bai, et al.. A frame-based discourse management system that uses machine learning models for information retrieval, named entity recognition, and categorization is part of their architecture design. Tens of thousands of searches for hotels are handled daily by their chatbot, which has been implemented on a commercial basis. Additionally, they have described the many machine learning models they deployed and how to create an e-commerce chatbot for the tourism sector. According to Adam et al. [7], chatbots are important in a number of industries. They describe how user compliance with a request for service feedback in a chatbot conversation is impacted by the application of ADCs (identity, small talk, and empathy) as a typical compliance approach. They looked at the effects of vocal anthropomorphic design cues and the foot-in-the-door strategy on user request compliance in a randomized online experiment. Their findings demonstrate that users are much more likely to cooperate with a chatbot's request for service feedback when anthropomorphism and consistency are present. They have observed that the impact of anthropomorphic design signals on user compliance is mediated by social presence. In [8], Hasan et al. investigate the intents of tourists to use chatbots during service interactions in the context of upcoming overseas travel, presuming persistent social withdrawal. Their findings indicate that intents to use chatbots were positively influenced by automation, habit, social presence, and health consciousness. A few differences were noted as a result of going through a lockdown enforced by the

government. When regulating for lockdown and throughout the journey experience, chatbots' social presence and human characteristics were less important.

SYSTEM RECOMMENDED

On any closed domain system, our suggested method may identify speech through chatbots that use natural language processing (NLP). It built an internet application and leverages the most recent technological advancements. Instead of searching and navigating the website, users can use voice to communicate with the web application. The chatbot can respond to any inquiries about the hospitality industry, the hotel online application (such as how to use it or where to locate it), or the business domain.

The system has been taught to respond to inquiries pertaining to the hospitality industry. Any data set containing information about the hotel website is used to train the system. Its implementation makes it simple to retrain using any set of data.

The following modules are needed for the suggested system: a voice chatbot that can be hosted on a hotel website. Web applications must be able to record voice input and translate it into text. There are several ways to handle this. The two most popular speech to text conversions for online applications are listed below.

VOICE TO TEXT ENGINE

An open-source voice recognition and speech-to-text engine called Deep Speech uses Baidu's Deep Speech research paper to train its model, which is implemented under a Mozilla Public license. The system at its core makes use of Tensor Flow from Google. It developed two models: the language model and the acoustic model. The language model, which is incorporated as a separate model, is utilized to improve the accuracy of the transcription output in the Acoustic model, an end-to-end deep learning system. Depending on our domain, the language model can be modified.

First, the model needs to be downloaded in order to be implemented. The acoustic model, denoted as .pbmm, is trained using American English and employs tensor flow in the background. The language model, called Scorer, is helpful in raising the output's projected accuracy. This allows it to determine, for instance, which word in a given situation is grammatically correct.

The engine's initial architecture was derived from Deep Speech: Scaling up end-to-end speech recognition. It is now distinct in many ways and is built on a recurrent neural network (RNN) that has been trained to consume speech spectrograms and produce transcriptions of English text [9]. Hybrid models are used by the Deep Speech model for parallel optimization. The advantages of synchronous and asynchronous optimization are combined in hybrid parallel optimization. Although it doesn't have the erroneous gradient issue that asynchronous optimization has, it does permit the usage of several GPUs.

At the beginning of hybrid parallel optimization, the model is stored in CPU memory. Then, just like in asynchronous optimization, the current model parameters and a tiny batch of data are obtained by each of the G GPUs. The gradients for each model parameter are then computed by each GPU using the micro batch, and they are then sent back to the CPU. Unlike asynchronous optimization, this time the CPU waits for every GPU to complete its small batch before averaging all of the gradients from the G GPUs and updating the model with this average gradient.

Hybrid parallel optimization offers a number of benefits and minimal drawbacks. Hybrid parallel optimization permits the usage of many GPUs in parallel, just like asynchronous parallel optimization. Additionally, this problem does not include the wrong gradient, in contrast to asynchronous parallel optimization. Actually, a single mini-batch handled by a single GPU is G times smaller than the mini-batch that is used in hybrid parallel optimization. Hybrid parallel optimization is not flawless, though. All other GPUs will have to wait until this straggler completes its mini-batch if one GPU is taking longer than the rest to complete it. It reduces throughput. But this issue ought to be reduced if every GPU is the same brand and model.

In light of this, hybrid parallel optimization appears to have more benefits and fewer drawbacks than both synchronous and asynchronous processes. In this study, the hybrid model is employed.

SPEECH RECOGNITION API

Version 25 of the Chrome browser introduced the Web Speech API a few years ago, enabling web apps to convert audio to text. It is becoming more and more common and will dominate voice recognition in the future [10]. Through the Speech Recognition interface, the browser makes the speech recognition feature available. This interface can respond correctly when it detects voice context from an audio input (usually through the device's built-in speech recognition service). Using JavaScript, we must create a Speech Recognition object with many event handlers that will allow us to recognize speech input through the device's microphone. Using Webkit Speech Recognition, which is a feature of the browser window object, we can additionally verify the browser's compatibility. One specific collection of grammar that our program should be able to identify is represented by the Speech Grammar interface [11]. JSpeech Grammar Format is used to define grammar.

TEXT TO VOICE CONVERSION

The server will produce output in text format that must be converted to voice. Thus, we require a method to effectively translate this text into voice. Users will have an interactive experience with the voice response. Text to speech conversion comes with a plethora of possibilities. A few of these include standalone implementations that use pre-trained models like gTTS, or a multitude of cloud-hosted APIs. In this case, we examined two options: Web Speech API and gTTS.

A) USING A PRE-TRAINED MODEL, TEXT TO SPEECH ENGINE

The process of transforming text data into a vocal audio format is called text-to-speech (TTS). The software takes an input text, understands the linguistics of the language being used through natural language processing techniques, and applies logical inference to the text. The processed text moves on to the following block, where it is subjected to digital signal processing. This processed text is ultimately transformed into a speech format by a number of algorithms and modifications. Speech synthesis is included throughout the entire process. For text to speech conversion, we make use of Google's Text To Speech (gTTS) library. A very user-friendly Python module called gTTS is used to convert text to audio files, which are then sent as blob data to the client. Numerous languages, including Hindi, Hindi, French, German, and English, are supported by this API.

B) SPEECH SYNTHESIS API

An interface for the browser's Web Speech library API is called speech synthesis. The Speech Synthesis interface, a text-to-speech component that enables programs to read aloud their text content, is the means by which one can access speech synthesis. We can adjust the pitch, pace, and variety of voice kinds that appear in the synthesized voice [11].

CLOSED DOMAIN QUESTION ANSWERING (cdQA)

An NLP-based system for answering closed-domain questions is called Closed Domain Question Answering (cdQA). Anyone can ask a question in plain English on cdQA and receive a response without needing to read the pertinent internal documentation. This is the goal of the platform.

There are two types of question answering systems that come to mind: closed-domain QA (cdQA) systems and open-domain QA (ODQA) systems. Open-domain systems can only rely on general ontologies and world knowledge, and they handle queries about almost anything. One such system is DrQA, an ODQA created by Facebook Research that draws its knowledge from a sizable database of articles from Wikipedia. We can see why this approach is regarded as an ODQA since these documents cover a wide range of issues and topics. Closed-domain systems handle queries falling under a particular domain (like hospitality or medicine), and they can take advantage of domain-specific expertise by using a model that has been fitted to a database with a unique domain. The cdQA-suite was developed to make it simple for anyone to create a closed-domain quality assurance system.

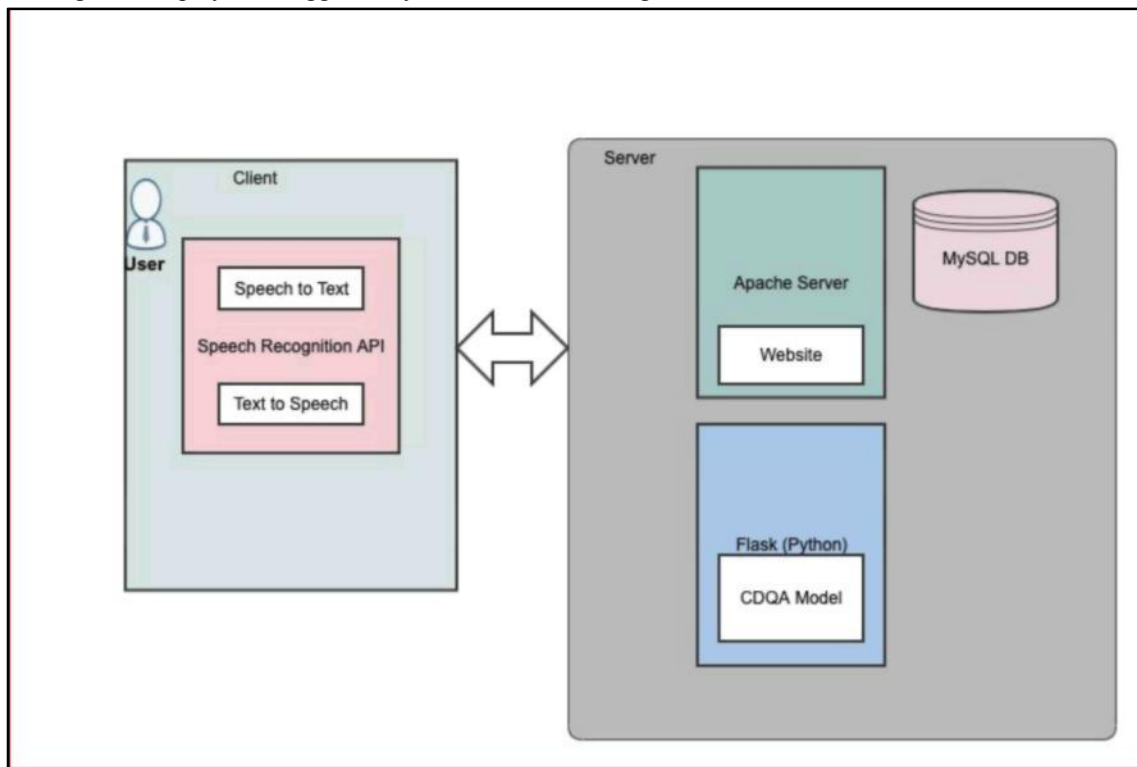
The Retriever and the Reader are the two primary building blocks of the cdQA architecture. The Retriever chooses a list of documents from the database that are most likely to have the answer when a question is submitted to the system. It is built on the same DrQA retriever that computes the cosine similarity between each document in the database and the

query sentence, and generates TF-IDF (term frequency-inverse document frequency) features based on uni- and bi-grams.

The system separates each document into paragraphs and sends them together with the question to the Reader, which is essentially a Deep Learning model that has been trained beforehand, after choosing the most likely papers. HuggingFace provided the Pytorch version of the popular NLP model BERT, which was the model utilized. Next, for each paragraph, the Reader outputs the most likely response it could find. The system's last layer, which comes after the Reader, compares the responses using an internal score function and outputs the response that is most likely based on the scores.

The Bert Stanford Question Answering Dataset (SQuAD), which has more than 100,000 question-answer pairs on more than 500 articles, is included in this pretrained model. We can turn this model into a closed cdQA system by further training it using our domain.

Taking into account factors like cost, convenience of use, accuracy, maintainability, and maintainability, the suggested system will be put into practice using the solution listed below. PHP is the server side scripting language that we utilize with the Apache server. Using Speech Recognition API, convert audio to text. To convert text to voice, we utilize the Speech Synthesis API. cdQA is used in the question-answering model. The Python Flask web server was used to host the cdQA. Figure 1 displays the suggested system architecture diagram.



II. IMPLEMENTATION AND EXPERIMENTAL RESULTS

The following are the ways in which the system has been built based on the suggested system architecture. This application contains five views. The user interface displays surrounding significant attractions, retail centers, events, and information about the many types of rooms that are offered. The website was created with the help of MySQL for the backend and HTML, CSS, JavaScript, Bootstrap, FontAwesome, and PHP and Apache for the server side. The user interface displays the availability of rooms that are configured in the database during a user search. There is a button to start a voice chat at the top of the website. The home screen is set up as follows, with a search bar on the left where users may look up rooms by entering their guest number and date of stay. When the search button is clicked, the server will get a rest API call, loading the available rooms from the database.

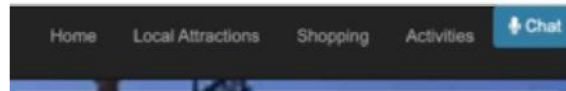


Figure 2: Home Page Menu

A) CLOSED DOMAIN QUESTION ANSWERING (cdQA)

Using PDF documents about hospitality, we retrained the model. A sample answer to the query will appear as follows. In addition to providing the answer, the system displays the document or article title and the paragraph in which the solution was located.

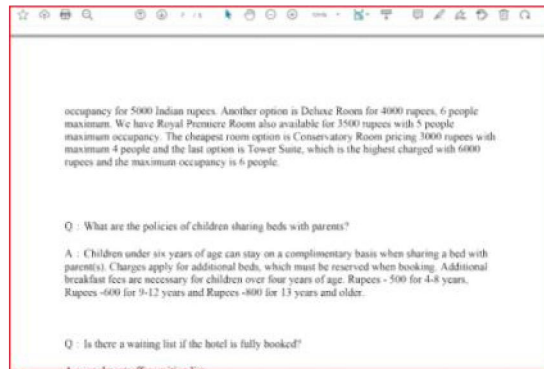


Figure 3: Sample cdQA model pdf

Between fifty and one hundred questions about hospitality were used to train the model. Every response to a question will be a paragraph. once the Flask server has been used to train, deploy, and expose the model. It provides an endpoint that may receive a string input, query the model, and return JSON data that includes the query's answer, the paragraph in which the answer was found, and the name of the document. A sample cdQA model PDF document is displayed in Figure 3.

B) VOICE CHAT BOT (EMMA)

"Emma" is the name of the voice chatbot that was created. When you select the "Chat (voice)" button at the top of the navigation panel, a chat window with the welcoming voice message "My name is Emma, your voice assistance, how can I help you today" will appear as seen below. It is possible to customize this message.

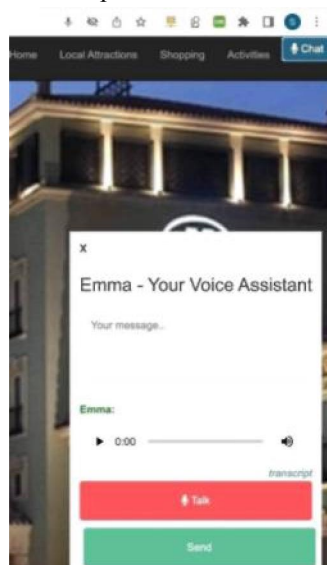


Figure 4: Voice chat bot screen

The "Talk" button causes the user to speak, and it will continue to listen for voice input until they stop. It will turn the audio input into text using the Web Recognition API and display in the text field at the top of the voice chat popup, as seen in Figure 5. A server call hosted by the cdQA model will be started when you click "send." The loading indicator will indicate the processing state, in Figure 5, to the upper left of the Talk button. Take note that the user's voice transcription was displayed on the screen in addition to text.

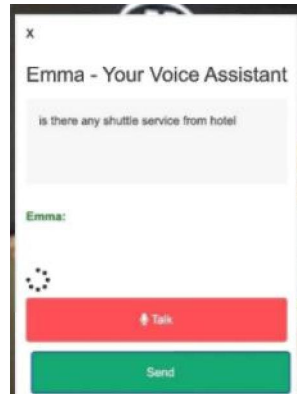


Figure 5: After user's voice input and the Send button clicked

The audio response will appear beneath the text area as seen in Figure 6 after the outcome is received. Additionally, there is a transcript button that, when clicked, expands to reveal the text.

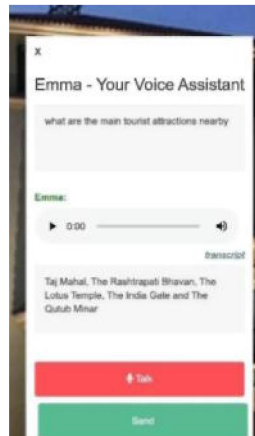


Figure 6: The voice and text response of the Chat bot

III. REVIEW OF LITERATURE

The use of chatbots in the hospitality industry has been growing in recent years. A study by Juniper Research found that the global chatbot market in the hospitality industry is expected to reach \$1.7 billion by 2023.

Several studies have explored the benefits of using chatbots in the hospitality industry. For example, one study found that chatbots can help to improve the guest experience by providing a more convenient and personalized way for guests to interact with hotels. Another study found that chatbots can help to reduce costs by automating tasks such as answering questions and making reservations.

However, there are also some challenges associated with using chatbots in the hospitality industry. One challenge is that chatbots can be difficult to develop and maintain. Another challenge is that chatbots may not be able to understand and respond to all guest queries accurately.

Scope for Study:

There is a significant scope for further research on voice chatbots for the hospitality industry. Some potential areas of research include:

Developing new and innovative ways for guests to interact with voice chatbots

Copyright to IJAR

www.ijarsct.co.in

Improving the accuracy and reliability of voice chatbots

Integrating voice chatbots with other hotel systems, such as property management systems and customer relationship management (CRM) systems

Evaluating the impact of voice chatbots on the guest experience, hotel costs, and hotel efficiency

IV. CONCLUSION

Voice chatbots have the potential to revolutionize the hospitality industry by providing a new and innovative way for guests to interact with hotels. However, there are still some challenges associated with using voice chatbots in the industry. Further research is needed to develop new and improved voice chatbot technologies, and to evaluate the impact of voice chatbots on the hospitality industry.

Chatbots are computer programs that interpret human needs and direct users toward the intended result using artificial intelligence and natural language processing. Here, the project's main goal was to use a voice-based chatbot. An algorithm on a remote server parses the spoken inquiries and searches the paper for all pertinent responses. The user receives back the most pertinent response along with an approximation of the model's confidence. Reusable generic module that receives voice input, converts it to text, and then, after receiving the outcome from the server, converts the text back to voice. JavaScript libraries and HTML are used in its implementation.

The Google Text-to-Speech (gTTS) engine and the speech-to-text engine are two examples of technologies for implementing voice-to-text and text conversion that have been researched. However, the drawback of each is that it requires its own infrastructure to be maintained. Thus, it was decided to move forward with the implementation of the speech synthesis and recognition API with event handling using a JavaScript utility layer.

According to the review, the question-answering system worked very well and consistently for texts that were less than 3000 words. Longer texts, however, caused it to become less accurate, lose track of information, and make serious errors. With the disclaimer that the summarizing model behaved strangely when given longer texts, the summaries that were supposed to assist the user in formulating inquiries also functioned as planned. The true value offered varies throughout documents. Maybe further research is needed to figure out how to solve this problem.

By using a feedback loop and more questionnaires to retrain the model, system performance and accuracy can be improved step by step. A more sophisticated system with more GPUs is needed for training with the larger dataset.

REFERENCES

- [1]. Juniper Research, "Global Chatbot Market to Reach \$7.7 Billion by 2023," 2018.
- A. G. S. Athikkal and J. Jenq, "Voice Chatbot for Hospitality," arXiv preprint arXiv:2208.10926, 2022.
- [2]. Hasan, M. R. Islam, M. H. Haque, and M. M. Hassan, "Tourist Chatbot Usage Intentions in Service Encounters Within the Context of a Future International Travel," arXiv preprint arXiv:2111.09244, 2021.
- [3]. S. A. Khan, M. S. Khan, and S. Khan, "Chatbots and Voice Assistants: Digital Transformers of the Company–Customer Interface—A Systematic Review of the Business Research Literature," *Sustainability*, 18(2), 51, 2023.
- [4]. Brsuh, Kate and Scardina, Jesse. "Chatbot", Techtargat. <https://searchcustomerexperience.techtargat.com/definition/chatbot>
- [5]. Lishchynska, Daryna. "What Are Bots? How Do Chatbots Work?",
- [6]. <https://botscrew.com/blog/what-are-bots/#:~:text=Chatbot%20or%20bot%20%E2%80%93%20is%20a,an%20instant%20pre%2Dset%20answer>
- [7]. Soeyuenmez, Anil. "Customer Service Chatbots", <https://www.messengerpeople.com/customer-service-chatbots-rule-based-vs-ai-what-you-need-to-know/>
- [8]. Selig, Jay. "The Power Of Chatbots Explained". <https://www.messengerpeople.com/customer-service-chatbots-rule-based-vs-ai-what-you-need-to-know/>
- [9]. DimitriosBuhalis and IuliiaMoldavska (2021). In-room Voice-Based AI Digital Assistants Transforming On-Site Hotel Services and Guests' Experiences. In: Wörndl, W., Koo, C., Stienmetz, J.L. (eds) *Information and Communication Technologies in Tourism 2021*. Springer, Cham. https://doi.org/10.1007/978-3-030-65785-7_3

- [10]. Bai Li, Nanyi Jiang, Joey Sham, Henry Shi, Hussein Fazal, "Real-World Conversational AI for theHotel Bookings". <https://arxiv.org/pdf/1908.10001.pdf>
- [11]. Adam, M., Wessel, M. &Benlian, A. AI-based chatbots in customer service and their effects onuser compliance. Electron Markets 31, 427–445 (2021). <https://doi.org/10.1007/s12525-020-00414-7>
- [12]. Rajibul Hasan, BernadettKoles, MustafeedZaman, Justin Paul, "The potential of chatbots intravel & tourism services in the context of social distancing", April 2021 International Journal of Technology Intelligence and Planning 13(1) DOI:10.1504/IJTIP.2021.10041470