A Systematic Review of Chatbot Model and chatbot technologies

Faluyi Joshua Oluwadamilola
Department of computer science
Federal University of Technology, Akure

Akinwonmi Akintoba Emmanuel
Department of computer science
Federal University of Technology, Akure

Corresponding email: Oluwadamilolajoshua43@gmail.com

ABSTRACT

This paper presents a comprehensive review of the chatbot models and Chatbot techmologies. Chatbot technology has transformed the way businesses and institutions interact with their customers due to its enhanced user engagement, 24/7 availability and data driven insight in decision making. The paper examines the past works on chatbot application, from their earliest days as simple text-based interfaces to their current advanced capabilities in natural language processing and artificial intelligence. The study also highlights the critical technological advancements that have driven the growth of the chatbot technology, such as AI and machine learning.. The review of chatbot technology reveals the enormous potential of chatbots in revolutionizing customer and individual-institution interactions and providing new opportunities for businesses and institutions to enhance customer engagement and experience.

Keywords: Chatbot model; Chatbot technologies; machine learning; Natural Language processing (NLP)

1. Introduction

Chatbot technology has been used in different sectors and industries like education, business, marketing, health care, information retrieval, e-commerce and entertainment. It has helped enhance customer-oriented services by improving efficiency and optimizing user experience (Misischia *et al.* 2022). In consideration of these factors, Artificial Intelligence (AI) has emerged as a prominent factor in delivering customer services to web users, primarily through the utilization of chatbots (Adamopoulou and Moussiades, 2020).

A chatbot is a computer program or an AI-based application designed to simulate human conversation through text or voice interactions (Bani and Singh, 2017). Chatbots are programmed to understand user queries and respond in a conversational manner, providing information, assistance, or performing specific tasks. They can be integrated into various platforms, such as websites, messaging applications, or mobile apps, and are used to automate customer support, provide personalized recommendations, and facilitate interactive experiences for users (Abdul-Kader and Woods, 2015).

A bot is a piece of software created to perform minor but repetitive tasks either automatically or upon command, mimicking the behavior of a human user (Zeifman, 2017). The word bot is an extract from the word “robot” which is considered a programmable machine that can execute actions automatically (Fischer *et al*, 2016).

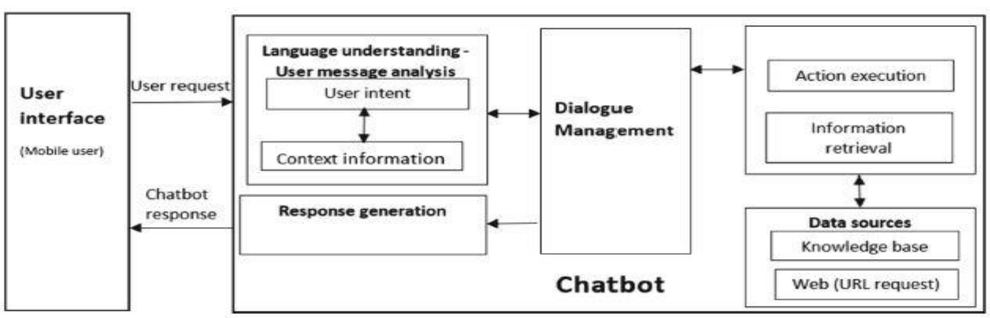

Figure 1.1: Architecture of a Chatbot (Adamopoulou and Moussiades, 2020).

The figure 1 shows the basic architecture of a chatbot system, depicting information request from user and chatbot response.

Chatbot is of two major categories; these are the rule-based and AI-based. A rule based chatbot make use of a tree-like flow in responding to queries by providing branching questions for users to choose from, why they get an answer for it (Weizenbaum, 1966). While an AI-based chatbot replicate human conversation utilizing a machine learning technique with ability to link one question to another (Serban *et al*, 2017).

1.2 Rule-based Chatbot

Adamopoulou and Moussiades (2020) describes a chatbot as a software application that is used to conduct a chat conversation on-line which can be via text or text-to-speech, alternatively providing direct contact with a live human agent.

The rule based chatbot are patterned to answer queries based on designated rules. When dialoguing, the bot follows specific rules to chat with a user. Responses are limited to compute information assigned to the rule (Mnasri, 2019). The rule-based chatbot performs simple tasks. However, to have it perform complex task requires writing many rules which can be time consuming.

Some advantages of rule-based chatbot are;

i. Faster to train and less expensive
ii. Easily integrated
iii. Can include interactive element and media
iv. Accountable an secure

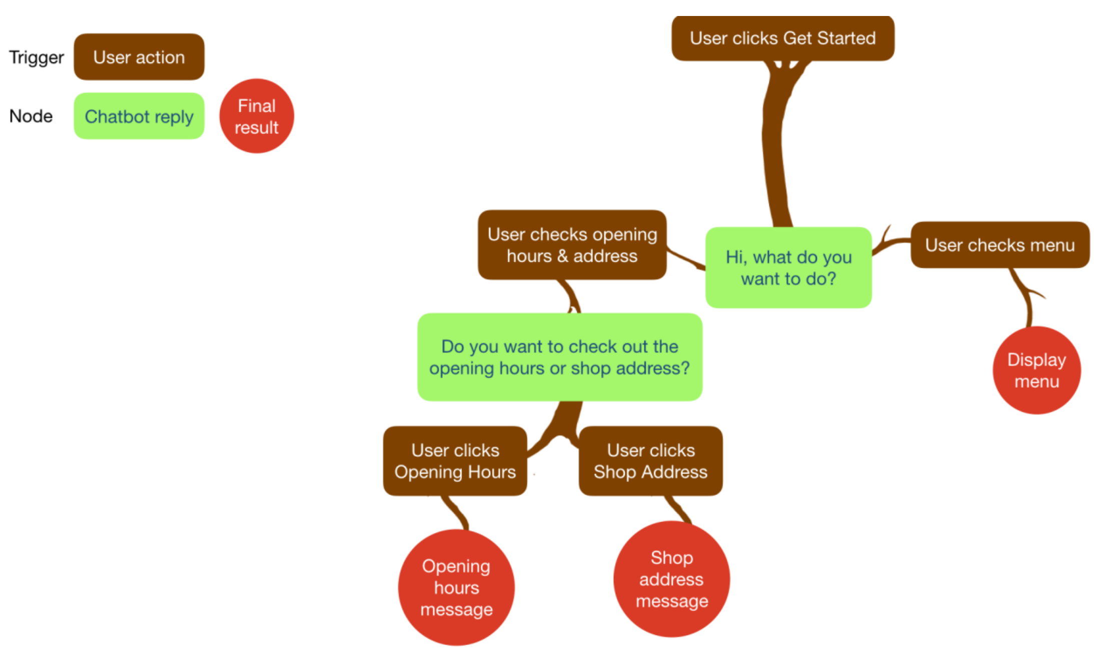

Figure 1.2: Stella Platform, 2021

1.3 AI-based Chatbot

An Artificial Intelligent (AI) based chatbot is a human-bot conversation that uses machine learning technique to understand the intent and context of a question before formulating a response; it has an ability to generate responses to complicated questions using natural-language responses (Senthilkumar and Chowdhary, 2019).

The AI-based chatbots are trained through machine learning algorithm that can train the model using training data set. Through the use of these machine learning algorithm, it is not needed to manually code new pattern matching rules, which allows for more flexibility of chatbot and also not dependent on domain specific knowledge. (Caldarni and McGarry, 2022).

Some advantages of AI-based chatbot are;

i. Learn from gathered information
ii. Improves as more data gets in
iii. Learn and understand behavioural patterns
iv. Broader range of decision making skills

ALICE chatbot (Artificial Linguistic Internet Computer Entity) by Wallace (2009) is an AI based bot which was inspired by ELIZA. It was an award winning open source natural language artificial intelligence chatting robot which is based on natural language understanding and pattern matching and uses AIML (Artificial Intelligence Mark-Up Language) to form responses to queries (Marietto *et al*., 2013).

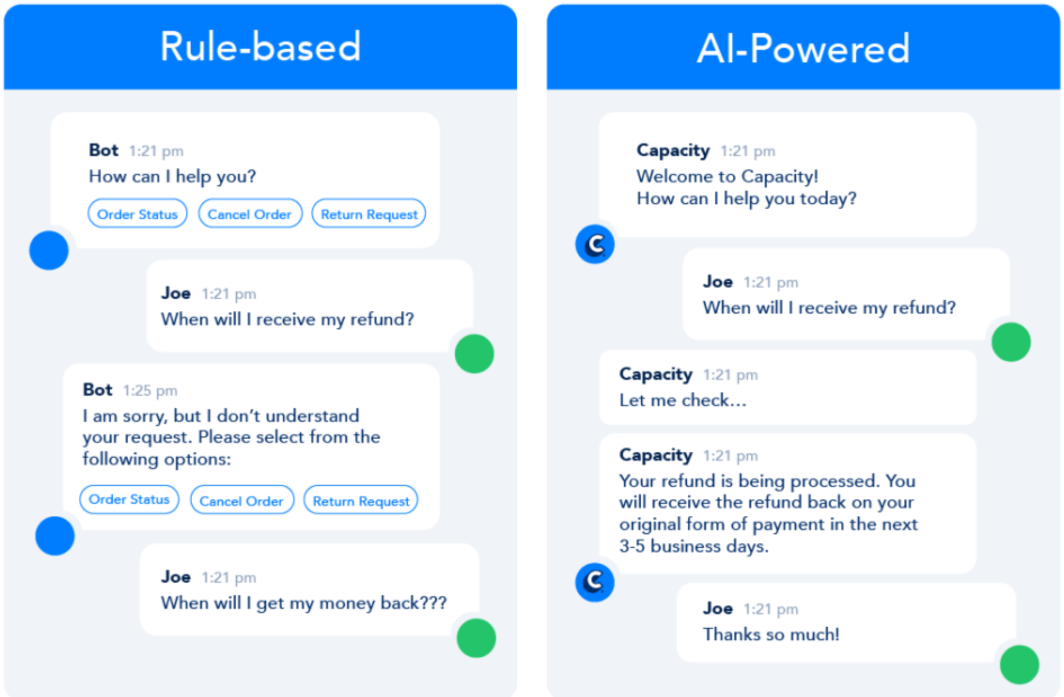

Figure 1.3: Team Capacity (2020)

Figure 1.3 describes the communication pattern of a chatbot system that is rule based and AI-based. In the rule based pattern of communication, a predefined approach of communication is seen which is based on a user making a selected request from given prompts. On the other hand, which is the AI based, communication between user and the chatbot is interactive, which is as a result of underline machine learning technique.

Chatbot as described earlier have been utilized in different sectors to aid service request delivery, and have evolved over the years. A review of this evolution will be considered in this paper work, coupled with the technology that drives them and their limitations.

2.0 Literature survey and related works on chatbot model

The literature review aims to provide an overview on the research on chatbot technology, By examining a wide range of studies and scholarly articles, this review seeks to shed light on the advancements, challenges, and future directions in chatbot implementation.

Chatbots serves as software agents used to interact between a computer and a human in natural language, and it has been adapted for usage by various sectors (Bani and Singh, 2017). In the same way people use language for human communication, chatbots use natural language to communicate with human users (Skjuve and Brandzaeg, 2019). The major aim of chatbot creation was to resemble a human being in the way they perform interaction, trying to make user think that they are conversing with a human aiding in resolving queries (Sharma, 2017).

2.1 Related works

Chatbots technologies have been applied in different domains, and it has been a great advantage to these domains in terms of attending to repetitive request from customers or clients almost immediately (Kumar and Ali, 2020). They also help to search for information internally to respond to customers (Kumar and Ali, 2020).

According to a research by Adam *et al* (2021), chatbot technology was utilized for customer support. Based on the researched work, AI-based chatbots in customer service are discussed and their effects on user compliance were considered. Utilizing live chat interfaces to engage with customers has gained immense popularity as a real-time customer service solution mostly in e-commerce environments. Clients employ these chat services to acquire information, such as product details, or seek assistance for resolving technical issues (Adam *et al*, 2021).

Customer service providers face a crucial challenge of striking a balance between service efficiency and service quality (Curran and Meuter, 2005). However, the adoption of real-time chat services has revolutionized customer service, turning it into a dynamic two-way communication process and this transformation has had a profound impact on factors like trust, satisfaction, repurchase behavior, and word-of-mouth intentions among customers (Mero, 2018). Customer support chatbot has been stated to save 80% routine questions for businesses (Reddy, 2017). In 2022, the implementation of chatbots was projected to contribute significantly to businesses, yielding cost savings of over $8 billion annually in customer-supporting expenses (Reddy, 2017). This astonishing figure represents a substantial leap from the modest $20 million in estimated savings back in 2017 (Adam *et al*, 2021).

Thus, while the opportunity and advantage is considered, Scherer et al in 2015 noted aforetime that firms should not shift completely towards customer self-service channels completely, especially not at the beginning of a relationship with a customer (Scherer et al. 2015), because the absence of a personal social actor in a transaction online can translate into loss of sales (Scherer et al, 2015).

Mikael Yang in 2015 founded manychat in 2015,a chatbot software company that helps businesses create chatbots for Facebook Messenger, this was considered a major chatbot technology that aid customer service in their request or for purchases (Gibbons, 2018).

ManyChat, a chat marketing platform, employs a range of technologies for its website and services. According to Crunchbase (2019), the platform utilizes HTML5, jQuery, and Google Analytics for web development and analytics purposes. Moreover, ManyChat ensures mobile compatibility and a seamless user experience through the implementation of Viewport Meta, IPhone / Mobile Compatible, and Apple Mobile Web Clips Icon technologies (Crunchbase, 2019). ManyChat lacks a built-in website widget, necessitating users to rely on external tools to integrate their bot into their website (Joren, 2023).

Al-Zubaide and Issa (2011) designed OntBot; an ontology based approach to model and operates chatbots. The need for this system is to guarantee functional correctness of a system under the absence of expected data. It used appropriate mapping technique to transform ontologies and knowledge into relational database and then use that knowledge to drive its chats.

This approach overcomes a number of traditional shortcomings namely the need to learn and use chatbot specific language such as AIML, and the use of non-matured technology. OntBot has the additional power of easy user's connections using their natural language, and the seamless support of different application domains. This gives a need for the system scalability and interoperability properties to be evaluated, as it could be time consuming when mapping out data to arrive at a particular response.

Yin *et al* (2017) presented DeepProbe, which is a generic information directed interaction framework chatbot, which is built around an attention based sequence to sequence (seq2seq) recurrent neural network. DeepProbe has an advantage over the ontbot designed by Al-Zubaide and Issa (2011) as it can rephrase, evaluate, and as well actively ask questions, leveraging the generative ability and likelihood estimation made possible by seq2seq models.

DeepProbe makes decisions based on a derived uncertainty measure conditioned on user inputs, which may involve multiple rounds of interactions that can help to maximize information gain, allowing for a more efficient user intention idenfication process. It selection of the best response is done through sequence to sequence estimates conditional entropy, having its generative response produced with the sequence to sequence model. The limitation of this approach is that it is domain specific and the number of rounds of interaction that will occur for the user to acquire the needed information (Yin *et al*, 2017).

Serban *et al.* (2017) presented MILABOT: a deep reinforcement learning chatbot developed with the motivation to have a system that is capable of conversing with humans on popular small talk topics through both speech and text. It knowledge domain is open which includes general knowledge from recent news topics, entertainment and basic human knowledge unlike the Deepprobe system which is domain specific (Serban *et al.*, 2017). The system consists of an ensemble of natural language generation and retrieval models, including template-based models, bag-of-words models, sequence-to-sequence neural network and latent variable neural network models for retrieval and generation of output response in word-by-word sequence (Serban *et al.*, 2017).

According to Serban *et al. (*2017) it was stated that by harnessing the power of reinforcement learning techniques and utilizing crowd-sourced data along with real-world user interactions, this system underwent training to intelligently discern and select suitable responses from the models present in its ensemble.

The system was evaluated through testing with real users and it was said to perform significantly than some other competing system. The system is likely to improve with more additional data as the robustness of the data determines how well it can relate with intents of a question for the right response (Serban *et al.*, 2017).

Cui et al. (2017) presented SuperAgent, a customer service chatbot that leverages large-scale and publicly available e-commerce data. It is an add-on extension engine that provides the customers with the best answer among a huge existing data sources within a webpage. This Chatbot is a great way to supplement customer service offerings since a Chatbot is more economical and indefatigable than the traditional customer service (Cui et al, 2017).

This innovative system leverages data derived from in-page product descriptions and user-generated content reviews sourced from e-commerce websites. By employing a combination of natural language processing and machine learning techniques, including fact-based question-answering (QA), FAQ search, opinion-oriented text QA, and chit-chat conversation modelling, it drives its engaging and interactive chats;

I. The fact QA engine is utilized for product information (PI)
II. FAQ search engine for customer QA pairs (QA)
III. Opinion-Oriented text QA for customer reviews (CR)

Effortlessly tapping into vast amounts of large-scale data, crowd-sourcing styles, and publicly available e-commerce data, it gains a significant advantage in its capabilities. It has a set of state-of-the-art Natural Language Processing (NLP) and machine learning techniques, which are used in the Chabot’s sub-engines (Cui et *al*, 2017). It does not need to deploy web crawlers for the websites since it is associated with each product webpage as an add-on extension, which is a strength over the MILABOT.

However, there is the need for the integration of a customer’s query intent detection module, so as to better leverage individuals engines. Also, advancement could be made to have a deeper delve on multi-turn queries, where the concept of context modeling could be worked on.

Ranoliya et al. (2017) designed a chatbot for university related FAQs. The objective is to provide an efficient and accurate answer for any query based on the dataset of FAQs, the developed system uses the Artificial Intelligence Markup Language (AIML). User post query on chatbot, and processing is done on the users query to match the predefined format by the developer. Pattern matching is performed between user query and knowledge pattern (Ranoliya et al., 2017).

And answer to user query is presented to the user. The limitation of this system is that knowledge is presented as an instance of AIML files. And if knowledge is created based on data that is collected from the internet, it will not be automatically updated and would have to be updated periodically. Original AIML has no extension possible. Hence, a need for improvement.

Muangkagmmuen et al. (2018), developed a Frequently Asked Questions (FAQs) Chatbot which was motivated on the basis of manual response by system administrator to customers concern for an e-commerce customer service through live chat, taking extensive time. And in advancing over this concern, this FAQs chatbot was built to automatically respond to customers by using a Neural Network (RNN) in the form of feed forward for text classification.

The experimental results shown that the chatbot could recognize 86.36% of the questions and answer with 93.2% accuracy (Muangkagmmuen et al, 2018). However, the system wasn't considered in it's designed to learn from users inputs or questions and couldn't retrieve resource for giving answers from other source which could be a web page or other external data centre.

Heo and Lee (2019) built an Inclusive Chatbot Service for International Students and Academics which was termed ‘CISA’. It was developed to enable international students and academics to effectively acquire important information as touching their academic and campus life. The researchers conducted interviews and surveys to be able to identify the needs of the target group.

In regards to the key findings from qualitative analysis, the concept was carried out. The design was carried out using Google’s DialogFlow, and implemented in Facebook Messenger (Heo and Lee, 2019). It is designed with fixed information, which offers limited help. It is basically advantageous to individual with limited access to the institution and handles more of repetitive questions.

Chandra and Suyanto (2019) in their work used a sequence-to-sequence model with attention mechanism to create a question answering chatbot. With a bidirectional recurrent neural network at its core, the chatbot produced a 44.68 BLEU score. The method used however is limited because it only searches for similar questions to the current question asked by the user and therefore has no memory for continued conversations.

Samyukth and Supriya. (2020) developed a chatbot to provide a human interaction for admission enquiry system. The researcher deploy the Latent Semantic Analysis algorithm (LSA) and cosine similarity in choosing right sentences for decision making, and considers maintaining data of questions not answered to be used for future analysis. However, these stored data of questions should have been utilized to train the system to improve its robustness and adequacy.

Tham (2020) developed a conversational question and answer chatbot for university. The motivation for this work is based on operational inefficiencies and FAQ site ineffectiveness when several individuals seek to make enquiries from the university at same period of time. This system is developed to also reduce manual work done by University marketing and promotion teams in handling information querying (Tham, 2020). The Naïve Bayes approach was chosen as the NLP Engine for the chatbot development implementation. The deployment of the system demonstrated that the chatbot was able to handle most conversations in a limited domain. And further stabilization is required to improve the NLP Engine for the chatbot so that all possible user interactions could be handled.

Lala *et al* (2020) developed an Improved Rapid Response Model for University Admission Enquiry System Using Chatbot. The objective of this system is to allow for real-time response on admission related enquiries and it is developed with the aim of bridging the straggle usually experienced through the conventional approach of phone call and email. IBM Watson was used to implement the design of the Chatbot for rapid response to admission enquiries (Lala *et al*, 2020). Botium was used to evaluate the performance of the Chatbot which gave an accuracy of 95.9% with instance of 212 successful test cases and 9 failed test cases. This approach introduces users to new and emerging technological solutions for optimal and rapid response in the educational sector. However, the underlying functionality and usability level could not be determined.

Khin and Soe (2020) implemented a Questioning Answering based University Chatbots using sequence to sequence model which leveraged on RNN. This chatbot was designed to converse in the Myanmar language to assist the human requirements in their daily routine. The result shows that it can generate the required information on the user question about the university information. The developed chatbot was noted to help students, teachers and others who want to know information about the university. The conversation model was trained on both CPU, GPU and implemented by Python language. However, there are some limitations associated with this system in terms of low resource information in regards to the language used in training the Seq2Seq model with attention mechanism. Further developments will lead to improve the system for greater accuracy and solution.

Daswani et al. (2020) developed a CollegeBot: A conversational AI Approach to Help Students Navigate College. The motivation of the development of this system was based on the limitation experienced when inquiry is being made by an individual as regards a college. Assistance over the phone or in person is often restricted to office hours and the information online is scattered through numerous web pages, which is often independently administered and maintained. In order to address this, the CollegeBot, is built, a conversational AI agent that uses natural language processing and machine learning to assist visitors of a university's web site in easily locating information related to their queries. The knowledge base is created by collecting and appropriately preprocessing information that is used to train the conversational agent for answering domain-specific questions. Two basic algorithms were utilized in training the conversational model for the chatbot, namely a semantic similarity model and deep learning, leveraging Sequence-to-Sequence learning model. The system is able to capture the user's intent and switch context appropriately. It also leverages the open source AIML chatbot ALICE to answer any generic (non domain-specific) questions. However, single intents that can be determined by the system from a sentence. This could be improved on, where multiple intent in a statement could be determined and processed.

Nguyen et. al. (2021) in their research developed a chatbot for inquiries using deep learning as provided in the Rasa platform. The said solution involved Detection of entities and intent from the question, retrieval of matching answer from the database and performance of specified action. The solution produced a 97.1% accuracy, after training on a training set of 1500 examples consisting of 50 intents. The major limitation noted was the fact that whenever admission requirements or scholarship amount change, the bot can't figure it out by itself, it is the admission officer that need to manually update the database yearly to match with the most recent information. Moreover, if there is a standard place on the University site or database where the bot can extract this information every time it needs it, it would work over this limitation which means integration with those other platforms to extract recent information may affect the responsiveness of the bot (Nguyen et. al, 2021).

One potential solution to address this limitation is for the bot to maintain its database of information. Whenever the external source of information undergoes editing or updates, the bot can receive notifications and subsequently update its own database accordingly. By doing so, the chatbot eliminates the need to frequently request information from the external source, as it keeps a personal and up-to-date copy of the data. The external source proactively keeps the chatbot informed of any changes, ensuring the chatbot's information remains accurate and current.

2.1.1 Existing Chatbot Approaches

In chatbot development, various approaches can be pursued, such as opting for Finite State Machines Route (FSMR), Deep Learning Route (DLR), or exploring alternative directions. The choice of development methodology significantly impacts the chatbot's success (Santos et al., 2022). Some of the existing chatbot approach will be highlighted here;

2.1.1.1 Rule-Based Chatbots

Rule-based chatbots adhere to a predetermined collection of rules and patterns. Their design revolves around providing responses to particular keywords or phrases, making their implementation relatively straightforward (Adamopoulou and Moussiades, 2020). Nonetheless, these chatbots face limitations in handling intricate or unforeseen queries, given their dependence on predefined rules (Singh et al, 2019).

The 'APU Admin Bot,' a chatbot, developed by Singh et al in 2019, which leverages this rule-based approach aims to offer students faster solutions to their queries, reducing their reliance on administrative offices. Utilizing a rule-based approach, the chatbot recognizes specific words, phrases, and actions, triggering predefined responses (Singh et al, 2019). Developed entirely on the Chatfuel platform and hosted on Facebook Messenger, this chatbot relies on a code-less authoring tool and a messaging platform, eliminating the need for traditional programming and architectural structures (Singh et al, 2019).

2.1.1.2 Retrieval-Based Chatbots

This approach relies on a database of predefined responses and matches the user's input to existing data using similarity measures. These chatbots are more flexible compared to rule-based ones, capable of handling variations in phrases and providing appropriate responses (Park *et al*, 2018). However, their limitation lies in being confined to the available data, and they may struggle to comprehend entirely new or novel queries (Akkineni *et al*, 2022).

A retrieval-based chatbot was developed by Akkineni *et al*, in 2022. The web-based software application was specifically created for Prasad V Potluri Siddhartha Institute of Technology to address diverse queries related to the college, encompassing facilities, procedures, policies, and more (Akkineni *et al*, 2022). Implemented using the Flask framework, the application captures text inputs from users through a console and employs machine learning concepts to generate text-format responses. The application follows a retrieval approach, utilizing logic adapters to process the input and provide suitable answers (Akkineni *et al*, 2022).

This approach implementation of a chatbot basically leads to a reduction in the organization's workload. However, the chatbot's limitations include its lack of domain knowledge and the absence of a distinct personality (Akkineni *et al*, 2022).

2.1.1.3 Generative Chatbots

Generative chatbots utilize natural language processing (NLP) and machine learning techniques to construct responses. By analyzing patterns and context in the input they receive, they have the ability to generate appropriate and contextually relevant responses (Sheth et al, 2019).

These chatbots are more flexible and can handle a wider range of queries, making them more human-like in conversation. However, they require large amounts of data and complex algorithms, and their responses may not always be coherent or accurate (Sheth et al, 2019).

Esfandiari *et al* (2023) developed a conditioner generative chatbot, which learns answer distribution using the generative models, such as sequence generative adversarial net which uses the reinforcement learning techniques. And the reinforcement learning rewards comes from the discriminator, which is judged on a complete sequence and fed back to the intermediate state (Wu *et al*, 2020).

The Generative Adversarial Network (GAN) utilized was evaluated step by step, which allows the discriminator to be adjusted to automatically assign scores and assess the suitability of each generated subsequence (Esfandiari *et al*, 2023).

However, the limitation considered for the generative chatbot model was based on the fact that the chatbot might not have the ability to understand the context of a conversation or user intent in relation to videos response. And this may not allow us be able to guarantee the quality of answer it provides in a customer service environment (Esfandiari *et al*, 2023).

2.1.1.4 Contextual Chatbots

Contextual chatbots take into account the conversation history to provide more relevant and personalized responses. They use memory or context to understand the user's intent and maintain continuity in the conversation (Jain et al, 2018). Contextual chatbots can be more engaging and efficient in understanding user needs (Jain et al, 2018).

Google Assistant as a contextual chatbot was developed by Google, which is AI-powered virtual assistant designed to provide users with personalized and contextually relevant information and services (Bhosale and Ravekar, 2019). As a contextual chatbot, Google Assistant utilizes advanced natural language processing (NLP) and machine learning algorithms to understand user queries and provide appropriate responses (Bhosale and Ravekar, 2019).

One of the standout features of Google Assistant is its contextual awareness. It can retain information from previous interactions within a conversation, allowing for more coherent and personalized responses over time. This context retention significantly enhances the user experience (Brachten *et al*, 2021).

While Google Assistant is a robust contextual chatbot, it still faces certain limitations. In complex or ambiguous queries, it may struggle to provide accurate responses, occasionally resorting to generic answers (Bhosale and Ravekar, 2019).

2.1.1.5 AI Virtual Assistants

A virtual assistant is a software application capable of executing tasks or providing services in response to commands or inquiries (Janssen et al, 2020). While not all virtual assistants are contextual, certain ones can achieve a significant level of contextual comprehension by retaining and recalling information from past interactions with the user, in relation to this based on the google assistant previously discussed (Janssen *et al*, 2020).

Other examples of virtual assistants are Siri and Alexa, which combines various technologies such as NLP, machine learning, and voice recognition to provide a wide range of services beyond basic chatbot functionality. They can perform tasks, answer questions, control smart devices, and more (Brill *et a*l, 2022).

2.1.1.6 Hybrid Approaches

Certain chatbots integrate various methods, utilizing rule-based systems to address straightforward inquiries and relying on retrieval or generative techniques for more intricate interactions. This blended approach seeks to capitalize on the respective strengths of these different techniques.

Maeng and Lee (2021) utilized a hybrid approach to develop a Chatbot for Survivors of Sexual Violence, this approached leveraged on the rule-based and generative approach in achieving its course.The hybrid chatbot overcomes the limitations of each by combining a rule-based and an ML-based chatbot. To explore its potential in supporting survivors of sexual violence, researchers conducted a study (Maeng and Lee, 2021).

A total of 349 questions asked from the perspective of survivors was collected and analysed. Among the questions asked from survivors, the most frequent inquiries revolved around the punishment for sexual violence, the process of filing a police report, and the availability of support centers. Also, 30% of the questions omitted contextual information about sexual violence, and 10% of questions consisted of only keywords, not complete sentences (Maeng and Lee, 2021).

Furthermore, the choice of the approach depends on the specific use case, available resources, and the desired level of complexity. While some businesses may find that rule-based or retrieval-based chatbots meet their needs sufficiently, others may opt for more advanced generative or contextual chatbots to provide users with a more personalized and human-like experience (Folstad *et al,* 2021).

3.0 Chatbot Model With Existing Technologies

Chatbot model leverages artificial intelligence and natural language processing to comprehend and generate responses to human questions or messages (Agarwal and Wadhwa, 2020). Chatbot models are also specialized in comprehending and generating human language. Their purpose is to grasp statistical patterns, semantic connections, and syntactic arrangements within language (Tejasri, 2023). These models serve as the foundation for numerous natural language processing (NLP) techniques, a field in computer science dedicated to empowering computers with the capacity to comprehend written and spoken language in a manner akin to human understanding (Tejasri, 2023).

The progress in chatbot models, notably the introduction of transformer-based models like GPT (Generative Pre-trained Transformers), has brought about a remarkable transformation in NLP systems. These advancements have sparked a revolution by empowering NLP systems to achieve greater precision and contextually meaningful language comprehension and generation (Yenduri, 2023). Other related model and existing technologies includes;

3.1 GPT-4

Generative Pre-trained Transformer 4 (GPT-4) stands as the latest iteration in the GPT series, crafted by OpenAI, with the potential to usher in substantial progress within the realm of natural language processing (NLP) (Baktash and Dawodi, 2023).

GPT-4 stands as a cutting-edge language model, harnessing the power of deep learning algorithms to craft natural language text. It boasts a formidable parts of NLP capabilities, encompassing tasks like language translation, text summarization, question-answering, and dialogue generation. Its emergence ushers in a new era of AI-driven language processing, amplified by a host of advanced features that distinctly set it apart from its predecessor (Baktash and Dawodi, 2023).

Some of the amazing features of the GPT-4 are;

i. Enhanced Multilingual Proficiency: GPT-4 surpasses its predecessors in multilingual capabilities, enabling seamless comprehension and generation of text in various languages.
ii. Enhanced model dimensions: GPT-4 boasts a notably larger model size compared to GPT-3, which had 175 billion parameters. This augmentation in model size results in improved accuracy and performance of the GPT-4 model.
iii. Improved Contextual Comprehension: GPT-4 exhibits superior proficiency in understanding the context in which text is generated. This enhancement empowers it to produce text that is more precise and relevant, fostering a heightened level of accuracy.
iv. Enhanced Logical Reasoning: GPT-4's improved reasoning capabilities equip it to excel at complex logical reasoning tasks, setting the stage for even more sophisticated problem-solving abilities.

3.2 GPT-3

Generative Pre-trained Transformer 3 (GPT-3) is a state-of-the-art language model developed by OpenAI. GPT-3 emerges as a potent chatbot, proficient in a diverse array of natural language processing tasks like language translation, question-answering, and text generation, among others. Renowned for its substantial size and remarkable abilities, GPT-3 is accessible solely through specialized partnerships with OpenAI or API access (Katar *et al*, 2022).

GPT-3 demonstrates robust performance across a multitude of NLP datasets, encompassing translation, question-answering, cloze tasks, and other on-the-fly reasoning or domain adaptation challenges. These include tasks like unscrambling words, incorporating novel words into sentences, and executing 3-digit arithmetic (Brown *et al*, 2020).

Various users of this model have achieved astonishing outcomes across diverse applications and enthusiastically shared their experiences on multiple social media platforms (Katar *et al*, 2022).

The performance of this model was also evaluated and was considered advantageous for the performance academic article writing (Katar *et al*, 2022).

3.3 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a noteworthy language model crafted by Google in 2018 (Devlin *et al*, 2018). Though it was not initially created as a chatbot, BERT has been successfully adapted to cater to diverse NLP tasks, ranging from question-answering to chatbot applications (Devlin *et al*, 2018).

It revolutionized the field of natural language processing (NLP) by introducing the concept of bidirectional training in transformer-based architectures (Kenton *et al*, 2019). Unlike previous models that processed text in a unidirectional manner, BERT employs a bidirectional approach, allowing it to consider the entire context of a sentence or paragraph (Kenton *et al*, 2019).

The primary objective of BERT is to create powerful word embedding's that capture rich semantic meanings by considering both the left and right context of each word. This bidirectional context allows BERT to better understand the nuances and dependencies between words, enabling it to perform remarkably well on various NLP tasks (Liu *et al*, 2019).

BERT is pre-trained on massive amounts of text data, including Wikipedia articles and BooksCorpus, to learn contextual representations of words. After pre-training, it can be fine-tuned on specific downstream tasks like question-answering, sentiment analysis, named entity recognition, and chatbot applications (Devlin *et al*, 2018).

BERT stands out from other models due to its exceptional skills in handling long-range connections within text and its remarkable ability to excel in tasks that demand a profound comprehension of context (Gillioz *et al*, 2020). This has resulted in its outstanding performance in numerous NLP benchmarks, making it the foundation for various advanced NLP models (Gillioz *et al*, 2020).

3.4 LaMDA's

LaMDA is an AI language model designed specifically for dialogue applications. It falls under the category of artificial intelligence systems capable of producing natural and coherent responses during conversations. Developed by Google, LaMDA represents a family of large language models with the remarkable ability to engage in fluid and authentic dialogues across a wide range of topics (Thoppilan *et al*, 2022).

LaMDA, similar to modern language models like BERT and GPT-3, is built upon the Transformer neural network architecture, which was first introduced by Google Research and made available as open-source in 2017 (Thoppilan *et al*, 2022).. Leveraging this architecture, LaMDA undergoes training to comprehend vast text segments, such as sentences or paragraphs, enabling it to grasp the relationships between words and predict subsequent words accurately (Thoppilan *et al*, 2022).

LaMDA distinguishes itself from numerous other language models due to its distinct training methodology (O'Leary, 2023).. Specifically tailored to dialogues, LaMDA's training allows it to master the complexities that differentiate open-ended conversations from other linguistic forms. As a result of this specialized training, LaMDA has acquired a noteworthy nuance known as "sensibleness." (O'Leary, 2023).

4.1 Chatbot Model Based On Non-English Existing Approaches

A Chatbot model can draw inspiration from non-English existing methodologies, with one crucial aspect being the consideration of a heterogeneous culture. This concept pertains to the variations in cultural identity arising from factors like social class, ethnicity, language, traditions, religion, and a sense of belonging to a specific location, among numerous other factors. (Goga et al, 2021).

According to a study conducted by Chen et al. (2020), it was found that a specific chatbot, which is the L2 chatbot, had a significant impact on increasing learners' cultural awareness and their level of engagement in learning a new language while being in a foreign country, as opposed to their home country.

Hence, the authors noted that this particular design of L2 chatbot functions as a valuable social integration tool. The L2 chatbot can play a crucial role in assisting migrants in adapting to a new culture, social norms, and the traditions of the host country, drawing from fields like social science and classic anthropology (Chen et al, 2020).

Goga et al. (2021) introduced a chatbot called Rosa, designed specifically to address the main objective of decreasing the dropout rate among L2 learners studying abroad. The authors stressed the significance of an L2 chatbot's ability to improve learners' second language proficiency while considering their socio-economic status and providing support for their cultural differences.

Intercultural language learning is an approach to L2 (second language) acquisition that focuses on fostering awareness of one's own culture and encouraging recognition and understanding of other cultures among L2 learners. This approach takes into consideration cognitive characteristics related to how learners process information, perceive it, and organize it in their minds. Mehrotra and Yilmaz (2015) argue that the cognitive characteristics coexist in language proficiency development.

One is an analytical procedural system, rule-based; the other is a declarative system, example-based. Adult learners are more characterized on the former, whereas children are defined by the latter (Mehrotra and Yilmaz, 2015).

The way a learner processes information, perceives it, and organizes it in their mind is influenced by these cognitive characteristics. As a result, the incorporation of cultural dimensions such as cultural awareness and a diverse cultural context into conversational chatbots can have a positive effect on the learning outcomes of L2 learners (Zhu and Luo, 2022).

Research on L2 chatbots equipped with cultural awareness revealed that the interactions between the chatbot and L2 learners were not only enjoyable but also productive, leading to enhanced cognitive abilities among the learners (Shin *et al*, 2021). Additionally, the study demonstrated that integrating heterogenous cultural elements into L2 chatbots can further enhance the second language learning experience for learners, taking into account their socio-economic status and providing support for their unique cultural differences (Shin *et al*, 2021).

Another profound approach to non-English existing chatbot is known as the multi-lingual chatbot which can engage in conversations with users using multiple languages, achieved through the utilization of language detection and translation services (Ralston *et al*, 2019).

Badlani et al (2021) developed a Multilingual healthcare chatbot using machine learning. The chatbot application has the ability to diagnose diseases based on user symptoms. Moreover, it addresses user queries by employing Term Frequency Inverse Document Frequency of records (TF-IDF) and Cosine Similarity techniques to calculate sentence similarity, selecting the most relevant response from its knowledge database. With its multilingual capabilities, the chatbot system proves to be highly suitable for implementation in rural India. Currently, it supports three languages: English, Hindi, and Gujarati. Additionally, the chatbot utilizes Natural Language Processing concepts to converse with users, and it also supports speech-to-text and text-to-speech conversion, enabling users to communicate using voice inputs.

Conclusion

In conclusion, the systematic review of chatbot models and technologies provides a comprehensive overview of the current state of this rapidly evolving field. The diverse range of chatbot models, from rule-based to AI-driven, reflects the dynamism in technological advancements. It also shows the importance of understanding the strengths and limitations of various chatbot applications from several authors, considering factors such as scalability, user experience, and adaptability to different domains. As chatbots continue to play a pivotal role in enhancing human-computer interactions, future research should focus on refining existing models, exploring innovative technologies, and addressing ethical considerations to ensure the responsible development and deployment of chatbot systems.

References

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. International Journal of Advanced Computer Science and Applications, 6(7).

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427-445.

Agarwal, R., & Wadhwa, M. (2020). Review of state-of-the-art design techniques for chatbots. SN Computer Science, 1(5), 246.

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006.

Al-Zubaide, H., & Issa, A. A. (2011). Ontbot: Ontology based chatbot. In International Symposium on Innovations in Information and Communications Technology (pp. 7-12). IEEE.

Akkineni, H., Lakshmi, P. V. S., & Sarada, L. (2022). Design and development of retrieval-based chatbot using sentence similarity. In IoT and Analytics for Sensor Networks: Proceedings of ICWSNUCA 2021 (pp. 477-487). Springer Singapore.

Badlani, S., Aditya, T., Dave, M., & Chaudhari, S. (2021). Multilingual healthcare chatbot using machine learning. In 2021 2nd International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.

Bani, B. S., & Singh, A. P. (2017). College enquiry Chatbot using ALICE. International Journal of New Technology and Research, 3(1), 263368.

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., & Kaplan, J. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862.

Baktash, J. A., & Dawodi, M. (2023). Gpt-4: A Review on Advancements and Opportunities in

Natural Language Processing. arXiv preprint arXiv:2305.03195.

Brachten, F., Kissmer, T., & Stieglitz, S. (2021). The acceptance of chatbots in an enterprise context–A survey study. International Journal of Information Management, 60, 102375.

Brill, T. M., Munoz, L., & Miller, R. J. (2022). Siri, Alexa, and other digital assistants: a study of satisfaction with artificial intelligence applications. In The Role of Smart Technologies in Decision Making (pp. 35-70). Routledge.

Bhosale, V., & Ravekar, D. (2019). Google Assistant: Need Of The Hour. Dostupno na: https://www. researchgate. net/publication/352478016_Google_Assistant,[pristupljeno: 01.12. 2021].

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

Bui, T. H., & Nguyen, V. P. (2023). The impact of artificial intelligence and digital economy on Vietnam's legal system. International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 36(2), 969-989.

Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. Information, 13(1), 41.

Chandra, Y. W., & Suyanto, S. (2019). Indonesian chatbot of university admission using a question answering system based on sequence-to-sequence model. Procedia Computer Science, 157, 367-374.

Chen, Y., & Luo, Z. (2023). Pre-Trained Joint Model for Intent Classification and Slot Filling Semantic Feature Fusion. Sensors, 23(5), 2848.

Chen, X., & Cardie, C. (2018). Efficient Supervised Self-training of Named Entity Recognizers with Arbitrary Lexicons. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2.

Chen, Q., Gong, Y., Lu, Y., & Tang, J. (2022). Classifying and measuring the service quality of AI chatbot in frontline service. Journal of Business Research, 145, 552-568.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). Superagent: A customer service chatbot for e-commerce websites. In Proceedings of ACL 2017, system demonstrations (pp. 97-102).

Curran, J. M., & Meuter, M. L. (2005). Self-service technology adoption: comparing three technologies. Journal of services marketing, 19(2), 103-113.

Crunchbase.com (2019) https://www.crunchbase.com/organization/manychat/people. Accessed and Retrieved, 2023.

Daswani, M., Desai, K., Patel, M., Vani, R., & Eirinaki, M. (2020). CollegeBot: A Conversational AI Approach to Help Students Navigate College. In International Conference on Human-Computer Interaction (pp. 44-63).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

Esfandiari, N., Kiani, K., & Rastgoo, R. (2023). A Conditional Generative Chatbot using Transformer Model. arXiv preprint arXiv:2306.02074.

Gibbons (2018) https://manychat.com/blog/the-future-of-messenger-marketing. Accessed and Retrieved 2023.

Gnewuch, U., Morana, S., Adam, M. T., & Maedche, A. (2022). Opposing Effects of Response Time in Human–Chatbot Interaction: The Moderating Role of Prior Experience. Business & Information Systems Engineering, 64(6), 773-791.

Folstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. Computing, 103(12), 2915-2942.

Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. (2020). Overview of the Transformer-based Models for NLP Tasks. In 2020 15th Conference on Computer Science and Information Systems (FedCSIS) (pp. 179-183). IEEE.

Goga, N., Radu, M. D., Vasilățeanu, A., Dragomir, R., Buligan, R. M., Dinu, B. G., & Scurtu, D. (2021). Rose-An intelligent system for student counseling and tutoring. In 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-5). IEEE.

Fischer, M., Menon, S., & Khatib, O. (2016). From bot to bot: Using a chat bot to synthesize robot motion. In 2016 AAAI Fall Symposium Series.

Heo, J., & Lee, J. (2019). CiSA: An inclusive chatbot service for international students and academics. In International Conference on Human-Computer Interaction (pp. 153-167). Springer, Cham.

Janssen, A., Passlick, J., Rodríguez Cardona, D., & Breitner, M. H. (2020). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. Business & Information Systems Engineering, 62, 211-225.

Joren, W., (2023) https://chatimize.com/reviews/manychat. Access and Retrieved 2023.

Katar, O., Ozkan, D., Yildirim, Ö., & Acharya, U. R. (2022). Evaluation of GPT-3 AI language model in research paper writing. Turkish Journal of Science and Technology, 18(2), 311-318.

Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of naacL-HLT (Vol. 1, p. 2).

Lala O. G., Okedigba T.O. & Aworinde H. O. (2020). "An Improved Rapid Response Model for University Admission Enquiry System Using Chatbot". International Journal of Computer (IJC) (2020) Volume 38, No 1, pp 123-131.

Liu, C. Z., Sheng, Y. X., Wei, Z. Q., & Yang, Y. Q. (2018). Research of text classification based on improved TF-IDF algorithm. In 2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE) (pp. 218-222). IEEE.

Marietto, M., Varago de Aguiar, R., Barbosa, G., Botelho, W., Pimentel, E., Franca, R., & Silva, V. (2013). Artificial intelligence markup language: A brief tutorial. International Journal of Computer Science and Engineering Survey, 04, http://dx.doi. org/10.5121/ijcses.2013.4301

Maeng, W., & Lee, J. (2021). Designing a chatbot for survivors of sexual violence: Exploratory study for hybrid approach combining rule-based chatbot and ml-based chatbot. In Asian CHI Symposium 2021 (pp. 160-166).

Mehenni, T., & Boudjellal, N (2018). A new deep learning-based chatbot system for the Customer service companies.

Mehrotra, R., & Yilmaz, E. (2015). Terms, topics & tasks: Enhanced user modelling for better personalization. In Proceedings of the 2015 international conference on the theory of information retrieval (pp. 131-140).

Misischia, C. V., Poecze, F., & Strauss, C. (2022). Chatbots in customer service: Their relevance and impact on service quality. Procedia Computer Science, 201, 421-428.

O'Leary, D. E. (2023). An analysis of three chatbots: BlenderBot, ChatGPT and LaMDA. Intelligent Systems in Accounting, Finance and Management, 30(1), 41-54.

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018). Convey: Exploring the use of a context view for chatbots. In Proceedings of the 2018 chi conference on human factors in computing systems (pp. 1-6).

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.

Kumari, V., Sharma, A. A., Sharma, Y., & Goel, L. (2023). Scalability and Sustainability in Chatbot and Mobile Application Development. In 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 397-403). IEEE.

Kumar, R., & Ali, M. M. (2020). A review on chatbot design and implementation techniques. Int. J. Eng. Technol, 7(11).

Kumar, A., Tejaswini, P., Nayak, O., Kujur, A. D., Gupta, R., Rajanand, A., & Sahu, M. (2022). A survey on IBM watson and its services. In Journal of Physics: Conference Series (Vol. 2273, No. 1, p. 012022). IOP Publishing.

Liu, J., Symons, C., & Vatsavai, R. R. (2022). Persona-Based Conversational AI: State of the Art and Challenges. In 2022 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 993-1001). IEEE.

Mero, J. (2018). The effects of two-way communication and chat service usage on consumer attitudes in the e-commerce retailing sector. Electronic Markets, 28, 205-217.

Mnasri, M. (2019). Recent advances in conversational NLP: Towards the standardization of Chatbot building. arXiv preprint arXiv:1903.09025.

Park, M., Aiken, M., & Salvador, L. (2018). How do humans interact with chatbots?: An analysis of transcripts. International Journal of Management and Information Technology, 14, 3338-3350.

Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related FAQs. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1525-1530). IEEE.

Ralston, K., Chen, Y., Isah, H., & Zulkernine, F. (2019). A voice interactive multilingual student support system using IBM Watson. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 1924-1929). IEEE.

Reddy, T. (2017). Chatbots for customer service will help businesses save $8 billion per year. Watson Blog. Retrieved April, 19, 2021.

Samyuktha M., Supriya M. (2020) Automation of Admission Enquiry Process Through Chatbot-A Feedback-Enabled Learning System. In: Bindhu V., Chen J., Tavares J. (eds) International Conference on Communication, Computing and Electronics Systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/10.1007/978-981-15-2612-1_18.

Santos, G. A., de Andrade, G. G., Silva, G. R. S., Duarte, F. C. M., Da Costa, J. P. J., & de Sousa, R. T. (2022). A conversation-driven approach for chatbot management. IEEE Access, 10, 8474-8486.

Sheth, A., Yip, H. Y., Iyengar, A., & Tepper, P. (2019). Cognitive services and intelligent chatbots: current perspectives and special issue introduction. IEEE internet computing, 23(2), 6-12.

Senthilkumar, M., & Chowdhary, C. L. (2019). An AI-Based Chatbot Using Deep Learning. In Intelligent Systems (pp. 231-242). Apple Academic Press.

Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., & Bengio, Y. (2017). A deep reinforcement learning chatbot. arXiv preprint arXiv:1709.02349.

Scherer, A., Wünderlich, N. V., & Von Wangenheim, F. (2015). The value of self-service: Long-term effects of technology-based self-service usage on customer retention. MIS Quarterly, 39(1), 177–200.

Shin, D., Kim, H., Lee, J. H., & Yang, H. (2021). Exploring the use of an artificial intelligence chatbot as second language conversation partners, 21, 375-391.

Singh, J., Joesph, M. H., & Jabbar, K. B. A. (2019). Rule-based chabot for student enquiries. In Journal of Physics: Conference Series (Vol. 1228, No. 1, p. 012060). IOP Publishing.

Tejasri G. (2023). ChatGPT: Unveiling the power of advanced language models. https://interestingengineering.com/innovation/ai-openai-chatgpt-chatbots-language-models

Tham, Z. J. (2020). A conversational question and answer chatbot for university (Doctoral dissertation, UTAR).

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H. T., & Le, Q. (2022). Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239.

Wallace, R. S. (2009). The anatomy of ALICE (pp. 181-210). Springer Netherlands.

Weizenbaum, J. (1966). ELIZA-A computer program for the study of natural language communication between man and machine. Commun. ACM, 9(1), 36–45. http: //dx.doi.org/10.1145/365153.365168.

Wu, E. H. K., Lin, C. H., Ou, Y. Y., Liu, C. Z., Wang, W. K., & Chao, C. Y. (2020). Advantages and constraints of a hybrid model K-12 E-Learning assistant chatbot. Ieee Access, 8, 77788-77801.

Yin, Z., Chang, K. H., & Zhang, R. (2017). Deepprobe: Information directed sequence understanding and chatbot design via recurrent neural networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 2131-2139).

Yenduri, G., Srivastava, G., Maddikunta, P. K. R., Jhaveri, R. H., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. arXiv preprint arXiv:2305.10435.

Zeifman, I. (2017). Bot traffic report 2016. Imperva Incapsula Blog. Accessed and Retrieved, 2023.

Zhu, Q., & Luo, J. (2022). Generative pre-trained transformer for design concept generation: an exploration. Proceedings of the Design Society, 2, 1825-1834.