Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120437

Artificial Intelligence Chatbots: A Survey of Classical versus Deep Machine Learning Techniques

Bayan A. Alazzam, Manar Alkhatib and Khaled Shaalan*

Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

Received: 7 Jan. 2023, Revised: 10 Feb. 2023, Accepted: 12 Mar. 2023
Published online: 1 Apr. 2023

Abstract: Artificial Intelligence (AI) enables machines to be intelligent, most importantly using Machine Learning (ML) in which machines are trained to be able to make better decisions and predictions. In particular, ML-based chatbot systems have been developed to simulate chats with people using Natural Language Processing (NLP) techniques. The adoption of chatbots has increased rapidly in many sectors, including, Education, Health Care, Cultural Heritage, Supporting Systems and Marketing, and Entertainment. Chatbots have the potential to improve human interaction with machines, and NLP helps them understand human language more clearly and thus create proper and intelligent responses. In addition to classical ML techniques, Deep Learning (DL) has attracted many researchers to develop chatbots using more sophisticated and accurate techniques. However, research has paid chatbots have widely been developed for English, there is relatively less research on Arabic, which is mainly due to its complexity and lack of proper corpora compared to English. Though there have been several survey studies that reviewed the state-of-the-art of chatbot systems, these studies (a) did not give a comprehensive overview of how different the techniques used for Arabic chatbots in comparison with English chatbots; and (b) paid little attention to the application of ANN for developing chatbots. Therefore, in this paper, we conduct a literature survey of chatbot studies to highlight differences between (1) classical and deep ML techniques for chatbots; and (2) techniques employed for Arabic chatbots versus those for other languages. To this end, we propose various comparison criteria of the techniques, extract data from collected studies accordingly, and provide insights on the progress of chatbot development for Arabic and what still needs to be done in the future.

Keywords: Chatbots; Artificial Intelligence; Machine Learning; Deep Learning; Arabic Language.

1 Introduction

Nowadays, Artificial Intelligence (AI) techniques are widely used in a wide range of fields and applications. AI is often used in image processing, sounds, statistics, and texts to extract knowledge and hidden patterns and use them to make decisions or predict future events. AI has proven its efficiency in dealing with diverse types of data, especially text-related data called Natural Language Processing (NLP) [1] using Machine Learning (ML) in which machines are trained to be able to make better decisions and predictions.

NLP is an AI branch that helps machines understand texts by simulating the human skill of language understanding, thus enabling human-computer interaction. NLP a theoretically advanced technology that is used to represent and understand human languages automatically [2] by dividing the text into paragraphs, sentences, and words. Then, it learns the relationships between words, understands the meaning of text and sentences in different situations, and generates appropriate responses [3].

A chatbot, in short, is an AI system that can simulate a conversation with a user in natural language, through messaging apps, websites, mobile apps, or over the phone. Chatbots can be helpful in several customer experiences, including providing customer service and product reporting. Chatbots are one of the most exciting natural language AI applications. They are considered smart conversational agents that attempt to converse with a human partner, serving as a virtual assistant that interacts with people by emulating human dialogue and utilizing natural language [4]. Furthermore, chatbots have been developed for different fields and languages. Several related applications use the chatbot, including

* Corresponding author e-mail: 20197461@student.buid.ac.ae

customer assistance, translation, and home automation; however, the chatbot is used primarily in online business applications [5].

There have been several survey studies that reviewed the state-of-the-art of chatbot systems, for the Arabic language in particular [6,7]. However, these studies (a) did not give a comprehensive overview of how different the techniques used for Arabic chatbots in comparison with English chatbots; and (b) paid little attention to the application of deep learning techniques for developing chatbots. Therefore, the objective of this paper is to conduct a survey that analyses chatbot techniques and highlights differences between (1) classical and deep ML techniques for chatbots; and (2) techniques employed for Arabic chatbots versus those for other languages. This helps developers and researchers understand the infrastructure and select the most accurate and appropriate techniques for further development and application of chatbot systems. To this end, we propose various comparison criteria of the techniques, extract data from collected studies accordingly, and provide insights on the progress of chatbot development for Arabic and what still needs to be done in the future.

The rest of this paper is structured as follows. Section 2 gives an overview of the history of chatbots and related work. Section 3 presents the research questions that this survey aims to address. Section 4 describes our Chatbot Categorization Approach, which provides a distinction of classical versus deep learning techniques (subsection 4.1) and also Arabic chatbots versus English and other languages chatbots (subsection 4.2). Section 5 itruduced the thraets of validity. Section 6 concludes the paper and suggests directions for future work.

2 Background

2.1 Chatbot Definition

Chat is a medium of communication in which text and voice messages can be exchanged over computer networks, mobile networks, or the Internet. Chatting can be done between humans with each other, but can also involve an automated program that performs certain tasks according to given inputs, and such a program is called a chatbot. Chatbots are also known as Chatterbots, Imbots, talkbots, chat dialogue systems, and conversational agents [8].

Chatbots are one of the remarkable applications that have recently proven their efficiency and effectiveness. They are types of a simulation program that analyzes and processes human conversations, either written or spoken, and interacts with end-users through a digital device. They have significantly evolved over the past few years. They have been adopted in various fields and sectors, whether public or private. With chatbots, user's requests are responded to, analyzed, and based on that, decisions can also be made. Chatbots have the ability to learn from dialogues and conversations with users [9].

Chatbots are mainly categorized into two types: a dialogic chatbot and a rational chatbot. The dialogic chatbot is typically used in websites and social media applications and processes input text using NLP technologies and features to determine the best match for the user [10]. On the other hand, rational chatbots store the user input questions and queries; based on external knowledge, it is commonly used in the health field. Moreover, an embodied chatbot feels like the chatbot is talking with the user as if they are humans and their appearance also looks like an actual human where it responds to the questions and interacts with the users [11]. People are interested in chatbots to get assistance, retrieve information, and receive a response from a customer support service instantly, such as the Netomi application [12].

2.2 Chatbots History

Attention was drawn to machine learning in the early 1950s, when Alan Turing posed the question, "can a machine think?" [13]. Turing presented a simulation game that confronts the question of machine learning: the game is based on impersonating a human in a real-time textual discussion in which the judge cannot differentiate between the program and the actual human, now known as the Turing Test as a measure of intelligence [14].

ELIZA uses NLP techniques to recognize keywords from the input, pattern-match them against a predefined knowledge base, and generate several corresponding midfoot characters [15]. Although it allows users to communicate their issues and queries in a natural manner by speaking, writing, or pointing, user happiness is increased as a result.

After that, in 1972, Kenneth Colby built a conversational agent called PARRY that simulated a person with paranoid schizophrenia. Moreover, one of the most realistic chatbots based on NLP is ALICE (Artificial Linguistic Internet Computer Entity), which won the Loebner Prize in 2000, 2001, and 2004. It was developed by Dr. Richard Wallace using Artificial Intelligence Markup Language (AIML) in the early 1980s [16]. AIML is applied to declare pattern-matching rules that link user-submitted words and phrases with related topic categories.

In 2008, Fred Roberts created a chatbot named "Elbot," which performed conversation by understanding the meaning of synonyms. Elbot won the 18th Loebner Prize for AI and was able to convince 25% of human interrogators that he is indistinguishable from humans [17]. The conversation engine is the core of the chatbot that directs the conversation and the flow of the input and output. Right now, advanced chatbots like Amazon Echo and Alexa, Apple Siri, Microsoft Cortona, Google Assistant, and Watson take advantage of advanced machine learning technologies to process responses effectively and efficiently. Michael Mauldin coined the term "ChatterBot" in 1994 to characterize computers that could imitate human interaction and hence pass the Turing Test [18].

2.3 Arabic versus Other Languages

The Arabic language is among the five most spoken languages in the world. Arabic is spoken by more than 422 million native speakers. It is also used by more than 1.5 billion Muslim people. Arabic is different from other languages, such as English, in the sense thatt it has a classic case of diglossia where the written formal language differs substantially from the spoken vernacular. The Arabic language is a family of varieties. There are three main versions of the Arabic language: Classical Arabic (CA), Modern Standard Arabic (MSA), and Dialect Arabic (DA).

There are many challenges and limitations due to the complexity of the Arabic language, including the existence of several dialects and a scarcity of data resources, capitalisation, and the use of diacritics create ambiguity, making it difficult to distinguish between names and abbreviations. Complex morphological rules are required for tokenisation and parsing. As a result, chatbots have got less attention in Arabic than they have got in other languages.

2.4 Artificial Intelligence (AI)

Artificial Intelligence (AI) is a scientific field used for improving and developing technological solution related to the digital similarity of human intelligence. AI can "mimic human communication by learning to listen and speak, remember what has been heard, take into account the situation, time and tone of the conversation. Robots can maintain communication, develop own thoughts, and suggestions about various accidents and objects" [19]. AI can be defined and interpreted in different ways, and can be classified as narrow and general AI [20]. General AI (which does not currently exist) can be defined as a computer program that can learn and think without human assistance. Narrow AI is a program that can create patterns from data provided by human through advanced mathematical techniques and, based on that, it can predict certain events in the future, which is referred to as machine learning (ML).

2.5 Natural Language Processing (NLP)

Natural Language Processing (NLP) has significantly been employed in machine interpretation and various type of applications, such as limitation multilingual data frameworks or discourse combination and acknowledgment [21]. NLP aims to facilitate communication between machines (computers that understand machine language or programming language) and humans (who communicate and understand natural languages, such as English, Arabic, Chinese etc.). NLP is paramount as it makes a great impact on our daily lives as it has been used in many kinds of applications [21].

2.6 Deep Learning versus Classical Machine Learning

Artificial Neural Networks (ANNs) employ a number of neurons, which are interconnected together as one unit. Neural network (NN) architecture contains two types of networks: multi-layer feed-forward networks and recurrent networks [22]. Neural network-based approaches are mainly classified as retrieval-based approaches and generative approaches. Retrieval-based approaches allow to generate machine replies by identifying the most relevant response. Generative approaches allow to generate a word at a time based on the provided input. It is also possible to combine retrieval and generative approaches, for example, by feeding the retrieved replies to a generative model where the final response is provided by comparing the retrieved and generated replies based on re-ranking [23].

ANNs employ deep learning mechanisms in which multiple layers can be adopted to advance the learning process, even when datasets are relatively small. This what makes ANNs supersede classical machine learning techniques, such as Support Vector Machine (SVM), Decision Trees, Random Forests, which simple applies one short learning. Examples of Deep ANNs are described below.

2.6.1 Recurrent Neural Networks (RNNs

A recurrent neural network (RNN) is a type of neural network which, through recurrence, may accept a variable size of the sequence $x = (x_1, ..., x_n)$ as input and create a series of hidden variables $h = (h_1, ..., h_n)$. The conventional RNN solution is rarely utilized since it suffers from its vanishing gradient issue, that renders training extremely difficult.

2.6.2 Long Short-Term Memory (LSTM)

The Long short-term memories (LSTM) is a variant of RNN that has been introduced to produce more robust results by addressing the issue of long-term dependency that standard RNNs encounter [24].

2.6.3 Bidirectional Long Short-Term Memory (BiLSTM)

The Bidirectional long-short term memory(BiLSTM) is the process of making the standard RNN to have the sequence information, similar to LSTM, but in two directions: backwards (future to past) and forward (past to future). Regular LSTM makes the input flow in one direction, either forward or backwards, whereas BiLSTM makes the input flow in both directions, which enables preserving both past and future information.

2.6.4 Sequence to Sequence (Seq2Seq)

Sequence-to-sequence (Seq2Seq) models are NN models that have achieved significant success in machine translation and text summarization [25]. Seq2Seq models allow to take a sequence of elements (words, letters, etc.) and gives as an output another sequence of items. The encoder of classic Seq2Seq models is forced to send only a single vector, regardless of the input length (i.e., how many words a sentence composes). If many hidden units are used as part of the encoder to handle a larger context vector, the model can overfit with short sequences. To address this issue, Seq2Seq model with Attention [26] allows to pass not only the last hidden state to the decoder, but all the hidden states.

3 Research Questions

This survey aims to address the following research questions (RQs):

–**RQ1:** ***How different are chatbots developed using deep and classical machine learning techniques*?**
Motivation. Deep learning, or deep neural networks, techniques have been widely used in many applications across various fields. However, there is little known whether they provide an advantage to the development of chatbot systems in comparison to classical ML techniques. In this RQ, we aim to assess whether chatbots developed using deep learning achieve higher accuracies than those developed using classical ML techniques.

–**RQ2:** ***How different are chatbots developed for the Arabic language versus those developed for English and other languages*?**
Motivation. Despite the technological advances in the development of chatbot systems, much attention was paid to the English language compared to other human languages, such as Arabic. In this RQ, we aim to assess how advanced are the techniques used to develop Arabic chatbots compared to those techniques used to develop English and other languages chatbots. It has been unclear from previous survey studies how far Arabic chatbots are in terms of technological pace compared to other languages.

4 Chatbot Characterization Approach

In this section, we present our characterization approach of chatbots. Our approach distinguishes the techniques used to develop Arabic chatbots from those chatbots developed for English and other languages. To this end, we collected studies on chatbot systems using search queries performed on Google Scholar. Given the context and objective of this survey, we were only interested in identifying the methodologies (i.e., techniques, datasets, etc.) used for designing, implementing, and evaluating AI chatbots. In other words, the main goal was to identify as many methodologies as possible rather than the number of papers. Therefore, we selected a representative paper for each publication year and excluded others if they have similar methodologies.

We classify the collected studies as (1) classical versus deep learning-based chatbots; and (2) Arabic versus other languages chatbots. More details are illustrated in the following subsections.

4.1 Classical versus Deep Learning-based Chatbots

4.1.1 Classical Machine Learning Techniques

Chatbot techniques can be classified into two general methods: machine learning and neural networks chatbot [15]. Rule-based techniques were supplemented with a manually built knowledge base for bots like ALICE, hence working well for closed-domain tasks; however, they do not fare well in user engagement. An example of a rule-based chatbot is AIML, and developers use it to write rules for chatbot systems.

An AIML file must include categories that are essential units of knowledge. This category contains additional tags, such as pattern, one or more words intended to match the user's potential questions, and template, the reply that the chatbot will return to the user and natural language and that which keeps responses continue in the same context [16].

Machine learning chatbots are also referred to as decision tree bots since they use a series of defined rules, and these rules are the basis for the types of problems that chatbots will deliver solutions to it. Rule-based chatbots map out conversations depending on what a customer might ask and how the chatbot should respond. Therefore, they cannot answer any questions outside the defined rules and only work with the scenarios you train them for. Thus, they have a less flexible conversational flow, but the experience they will deliver can be guaranteed, so chatbots are more predictable than those who rely on machine learning.

Machine Learning techniques are generally faster to train (less expensive), integrate easily with legacy systems, streamline the handover to a human agent, are highly accountable and secure, can include interactive elements and media, and are not restricted to text interactions.

4.1.2 Deep Neural Network Techniques

Deep Neural Networks (Deep NNs) have been used in many aspects of our daily lives. For example, neural networks have been even used for improving food safety and quality, which includes the method of microbial growth. From here, we can predict food safety and know the functional, chemical, physical, and sensory properties of the various food products used during the processing and distribution process. Deep NNs take on many complex models and tasks by controlling the processes and simulating applications. Furthermore, many basic theories to properly apply technology to the knowledge of food safety have been discussed [27].

Each node is an artificial neuron, which is linked to the others and has a single load. Its operating principle is that if the output of any individual node exceeds the predefined limit value, the node is activated, and the data are transferred to the next layer of the neural network. Otherwise, no data will be sent to the neural network's next layer [27,28].

Deep NNs consist of a set of layers where each layer contains a set of nodes; an input layer responsible for the initial processing of input data by assigning weight factors through the activation function; one or more hidden layers responsible for providing variable strength to the input data traversing towards output layers to provide the best-fit output variables; an output layer that gathers the outputs and presents the result [28].

Neural networks rely primarily on training data to learn and improve their accuracy over time and repetition of training. However, once these learning techniques are finely tuned, they are powerful tools in AI science, allowing data to be processed, classified, and aggregated at high speed. When compared to manual work by human experts, speech and image recognition take minutes rather than hours or days. One of the most important applications of neural networks used today is the Google search algorithm [29]. Neural networks can be classified into two distinct types used for different purposes: feedforward networks and feedback networks [30].

The following are the most common neural network techniques and their use cases.

- For recurrent intralayer connections in which neurons in one layer are connected to each other, this type contains a group of branches, the most important of which is feedforward neural networks, or MLPs, which consist of an input layer, one or more hidden layers, and an output layer. In contrast, these neural networks are commonly referred to as MLPs, and it is noted that they are composed of sigmoid neurons, not receptors, because most real-world problems are nonlinear. Data are usually entered into these models to train; they are the basis for computer vision, NLP, and other neural networks [27,28].
- Convolutional Neural Networks (CNNs) are commonly used for image recognition, pattern recognition, and computer vision. These networks adapt principles from linear algebra, particularly matrix multiplication, to identify patterns in the image.
- Recurrent Neural Networks (RNNs) are defined by their feedback loops. These learning techniques are primarily used when using time series data to forecast future outcomes, such as a stock market forecast or sales forecast.

Before discussing the creation of a chatbot using neural networks, it is necessary to explain important terms: NLP and chatbot. NLP is an AI subfield that closely resembles linguistics and offers the language descriptions needed for a computer to comprehend an item. Through NLP technologies, the software can be made to analyse and simulate the understanding of natural languages, such as chatbots [31]. The chatbot architecture and process have three main components: the chat interface is the first module that interacts with the user; the knowledge engine, which consists of both tags; for example, a chatbot can be employed as helpdesk executive categories; the conversation engine is the core of the chatbot that directs the conversation and the flow of the input and output.

Below are the main studies reviewed in which neural network techniques have been used to develop chatbot systems.

Ansari et al. (2016) [32] designed a chatbot system (QAS) using a deep neural network. Their proposed system stores the various questions that are issued by users for later use and then processes the questions asked and provides the appropriate answer, that is, understanding the question and then answering it using the built deep neural network. The results indicate that the system is well learnable and gives appropriate responses.

Boyanov et al. (2017) [33] recommended to use question answering (QA) data from Web forums to train chatbots from the ground up. First, they selected pairs of question and response phrases from the generally considerably lengthier texts of forum queries and replies. They then utilize these shorter texts to more efficiently train seq2seq models. They enhance the parameter optimization even more by using a novel model selection technique based on QA measurements. Finally, they suggest that as an automated assessment approach for chatbots, they employ intrinsic evaluation in relation to a QAtask. The model obtains a Mean Average Precision (MAP) of 63.5% on the intrinsic task, according to the assessment. Furthermore, it can accurately answer 49.5% of questions that are comparable to ones answered in the forum, and 47.3% of questions that are more conversational in tone.

Singh et al. (2018) [34] suggested a chatbot program that specializes in providing answers based on the context of the conversation using a neural network model and NLP techniques. A total of 108 questions were asked, 74 of which were satisfactory, with an accuracy rate of 68.51%. The results indicate the efficiency of the proposed robot when used in the field of business to deal with customers and respond to their inquiries, which reduces the cost of operating human labor and saves time and money.

Bhartiya et al. (2019) [35] created and developed a University Counselling Auto-Reply Bot capable of answering questions about the subject of engineering at their university level. The relevant NLP methods are used to their University Data, which is created in *JSON* format, and the Feedforward neural network is employed for training this dataset, addressing the problem of overfitting. The Chat Application is then launched on Facebook Messenger, and the answer is displayed to the user on the Facebook Messenger interface, providing an effective engagement platform. The end-user testing was divided into two parts, with the probability scores of accurate replies increasing from 46% in the first phase to 72% in the second after adding more training phrases and keywords to the dataset.

Jiao (2020) [36] designed an English chatbot system using neural networks and Rasa natural language understanding (NLU) techniques. The suggested robot is designed to function in the finance sector, where it conducts comparison and verification procedures to determine its purpose, after which it may learn automatically from the questions it is asked and respond with suitable responses. They employed around 500 sentences with highlighted entities.

Jaiwai et al. (2021) [37] built a Thai educational chatbot based on deep neural network techniques and NLP techniques. The proposed application responds to students' educational questions and inquiries in the fastest time and with the best quality, which is done by segmenting Thai words and then building a neural network-based model for speech classification and training the dataset used. The application has been shown online applications to facilitate dealing with it by students. For text classification, the experimented NEWMM CutKum and DeepCut tokenized 1,400 sentences and trained a model with 500 intents.

Tsakiris et al. (2022) [38] Convolutional Neural Networks (CNNs) are utilized as the classifier, and certain particular tokenization methods are employed to create a chatbot. They use a method known as *word embedding*, which turns text into numbers in order to carry out text processing. The *Word2Vec* word embedding method was used specifically. CNN architectures such as AlexNet, LeNet5, ResNet, and VGGNet were used. The accuracy, F1-score, training time, and execution time of these designs were all compared. The findings revealed that there were considerable disparities in the performance of the architectures used. The most preferred architecture in their analysis was LeNet-5, which had the greatest accuracy relative to other designs, the shortest training time, and the fewest losses, although it was not particularly accurate on smaller datasets like their test set. Limitations and future research areas are also discussed.

Table 1 presents a summary of the most recent research that proposed the neural network chatbot techniques.

Table 1: Deep Learning-based Chatbots

Study	Finding	Strength	Weakness
A Ansari et al. (2016) [32]	Assigning deep cases to words in complicated sentences helps answers complex questions	- Gives proper answers - Useful for faster information extraction by users	Single technique
Boyanov et al. (2017) [33]	Optimizing for MAP yields the best results	Achieved a high ranking performance and BLEU score	Performed slightly worse on another second set of questions - Answers generated by the optimized model are worse
Singh et al. (2018) [34]	Useful in small industries or businesses for automating customer care	High precision as it gives responses based on the context of conversation tends to be more user friendly	High error rate, as it performs quite efficiently if queries are simple
Bhartiya et al. (2019) [35]	- Addressing overfitting helps the model to perform better. - Using softmax helps the model predict the best-matching intent tags	Successfully finds the best-matched intent tag for the query by selecting the class having the highest probability	It starts overfitting when a model learns the detail and noise in the training data
Jiao (2020) [36]	- RASA NLU has higher accuracy compared to NN when considering one sentence as a whole, whereas NN has better integrity to classify entities from segmented words	Able to identify user requests and fetch appropriate responses related to finance	Given that there is no similar expression in training data, it is expected that there is a mistake
Jaiwai et al. (2021) [37]	LaLin Chatbot in Thai with LINE applications	Fastest time and best quality	Limited dataset
Tsakiris et al. (2022) [38]	Significant differences in performance under different architectures	Multiple architectures of CNN	- Limited dataset - Most optimal architecture is still unknown

RQ1 Answer. Though classical machine learning techniques have been used to develop chatbots, deep learning techniques tend to outperform them in terms of applicability and accuracy. In particular, deep learning helps achieve good accuracy even with limited datasets. However, there is little attention in the application of the Bi-directional Long Short Term Memory (BiLSTM) technique for chatbot development, which proved better results in many other AI fields. In addition, there has been no effort to develop chatbots using multiple techniques, applied as neural network layers (e.g., Sequence-to-Sequence (Seq2Seq) with LSTM), to allow achieving much higher accuracies, thus providing more reliable chatbots that satisfy users' needs.

4.2 Arabic versus Other Languages Chatbots

4.2.1 Arabic Chatbots

Despite the abundance of research on English chatbots, there is some scarcity of Arabic chatbots due to challenges in the Arabic language. In this article, we present a survey on Arabic chatbots. The growing number of online Arabic users has prompted the development of Arabic chatbots.

Brini et al. (2009) [39] proposed a textual conversation chatbot using the AIML technique, and Saud Aljameel (2018) [40] proposed a textual conversation Arabic chatbot uses pattern matching. The speech conversation chatbot is based on the voice as an input, output, or both, such as Makatchev et al. (2010) [41]. They discovered that all Arabic chatbots employ a retrieval-based dataset model and that pattern matching and AIML are used to create Arabic conversational agents. One of the most challenging to achieve in Arabic is generative-based and deep learning models due to the lack of resources to train the learning model compared to those available in other languages.

Ali and Habash (2016) [42] presented the first BOTTA chatbot in Arabic (Egyptian dialect) using AIML, also described the challenges facing the task of creating an Arabic chatbot, and presented some proposed solutions to meet these challenges. BOTTA provides future Arab bot owners with a basic chatbot that contains basic greetings, general knowledge groups, and other useful features. BOTTA keeps track of categories for various issues, such as gender and nationality, and the set files include lists of words and phrases organized by theme. Finally, to overcome the inconsistent spelling variations of certain characters in Arabic, they used orthographic normalization.

AlHumoud et al. (2018) [6] presented a state-of-the-art review of research on Arabic chatbots. They classified Arabic chatbots based on the chatbot conversation interaction type into two groups, text and speech.

Bashir et al. (2018) [43] proposed a method for using deep learning methods to text categorization and *Named Entity Recognition* in the area of home automation in Arabic. To identify the users' intent, they employed cutting-edge neural network text classification algorithms such as CNN and LSTM. Both intent classification implementations were benchmarked, and the results showed that the LSTM performance (F1-score of 92%) was somewhat better than the CNN performance. They employed a hybrid representation of word embeddings and character-based word embeddings to extract the user aims and goals from the input, which was then fed into a BiLSTM network. The BiLSTM with the char embeddings model earned a high F1-score of 94%, indicating that its performance is extremely close to existing English Named Entity Recognition standards.

Joukhadar et al. (2019) [44] identified the dialogue act of users in a *Textual Dialogue* system employing Levantine Arabic dialect in this article. Our dialogue actions are divided into eight categories: *greeting, goodby, thanks, confirm, negate, ask repeat, ask for alt,* and *apology*. To determine the right speech act categories, the following ML methods were used: Logistic Regression, SVM, Multinomial NB, Extra Trees Classifier, and Random Forest Classifier. To make the best forecast from each classifier, they employed a Voting Ensemble approach. They compared the suggested models' outcomes on a hand-crafted corpus in the restaurant ordering and airline tickets domains. With 86% accuracy, the SVM model with 2-gram produced the best results.

Alshareef and Siddiqui (2020) [45] designed an Arabic chatbot using the Seq2Seq neural network technique. They collected a total of 5.2k pairs of post replies from Twitter. Their application is characterized by its immediate response to any inquiry and its assistance to an unlimited number of users. They employed Human Evaluation and Bilingual Evaluation Understudy (BLEU) to examine the accuracy of their suggested application, with a score of 25.1% for Arabic Seq2Seq transformer-based CA with pretrained word embedding. The results indicate that their program outperforms many chat programs that use deep learning techniques.

Khan and Yassin (2021) [46] suggested an Arabic chatbot (SeerahBot) specialized in the biography of the Prophet, based on machine learning and NLP techniques. The proposed application is concerned with users answering their inquiries about the hadiths of the Prophet and provides them with essential information related to the Prophet's Biography. The application was made available on Telegram to facilitate access and use for users. The results indicate the accuracy and superiority of the application in answering inquiries related to the Prophet's Biography. To answer the Prophet's Biography, a set of 200 questions was used. Furthermore, they organized the data and saved it in text files. As a very precise result, the quality and accuracy of information provided by SeerahBot is 35.71%.

Boussakssou et al. (2022) [47] proposed an Arabic chatbot application (midoBot) using the Seq2Seq deep learning technique. The proposed application is characterized by being able to talk with humans on several traditional topics, as it has been trained on approximately 81,659 conversations. The results show the superiority of the application in answering most questions and producing new answers. They achieved satisfactory results by combining LSTM and GRU. In training, they discovered that GRU outperforms LSTM. GRU took about 140 seconds per epoch, while LSTM took 470.78 seconds. LSTM provides more proper answers with higher accuracy than GRU. LSTM's final accuracy is 0.89, whereas that of GRU is 0.85.

AlHumoud et al. (2022) [48] This contribution introduces an Arabic chatbot that is helpful to leisure visitors traveling in Saudi Arabia. The chatbot was constructed using IBM Watson, and the first implementation is written in Java. Both chatbots may be seen operating on the Telegram platform. The testing of the chatbots shown that they are very usable and useful in providing assistance to tourists. The consumers that tested the usability of our chatbots gave them an average rating of 99% simple, which indicates that they were generally pleased with the experience overall. Also, it demonstrates that users favor the interface that provides buttons as opposed to the interface that requires them to type in requests, with a difference of at least 6% in the evaluation of conversation to the favor of the interface that provides buttons. This can be seen by comparing the two.

Summary. Overall, Arabic chatbots face many challenges regarding technology, application, and the Internet. These challenges are that the representation style of the language is unique and different, which makes reading and understanding the sentence's meaning difficult. Second, the Arabic language has a huge number of unique characters, decreasing the chances of finding a match for the user's needs. Third, the unique structure of the Arabic language makes the translation from English sentences into Arabic and vice versa difficult [49]. Last, dialect variants in Arabic are quite different from each other. The limitations of the most previous model are the small size of the dataset created.

Table 2 summarizes the most recent research about chatbots in the Arabic language.

Table 2: Arabic Chatbots

Study	Summary	Strength	Weakness
Ali and Habash (2016) [42]	To overcome the inconsistent spelling differences of certain letters in the Arabic language, they used orthographic normalization	Addressed Arabic challenges	- BOTTA stores categories for various topics such as gender and nationality and the selected files contain lists of words and phrases under one topic - No exploiting existing tools to perform morphological analysis on the input and to - experiment with lemma-based pattern matching
Bashir et al. (2018) [43]	CNNs performed better than LSTMs. For entity extraction, the model obtained comparable results to the named entity recognition benchmarks in English	Used LSTMs and CNNs, BiLSTM, and character-based word embeddings. The study used data collected via an online survey and the AQMAR dataset. The data was filtered and labelled according to the Conll-2003 NER format	
Joukhadar et al. (2019) [44]	Comprehensive evaluation of dialog act systems	Multiple ML models	Limited dataset
Alshareef and Siddiqui (2020) [45]	Can understand and speak Gulf Arabic	Able to learn and understand deep languages	- Difficulty in analyzing and analyzing Arabic language data - Small sample size - Single comparison reference - BLEU was calculated by just replying to one sentence, not multi-turn conversations
Khan and Yassin (2021) [46]	Answering inquiries related to the Prophet's biography	Accurate answers	- Unable to respond to certain inquiries - Inquiries were not present in Seerah Bot's knowledge base - Seerah Bot must be improved by expanding the dataset
Boussakssou et al. (2022) [47]	Answering most questions and producing new answers	Good and formulated answers to questions that people can understand	Complex and requires constant training by the human being to get used to the idea
AlHumoud et al. (2022) [48]	Usability test of the chatbot show a relatively high satisfaction of the users, as they evaluated them as 99% easy on average	Both Arabic chatbots were evaluated as 100% easy to use	- Many of the users who tested the English chatbot speak English as a second language - The difference in satisfaction between the typing interfaces provided by IBM Watson and the button interface in the Java-based chatbot was up to 7%.

4.3 English and Other Language Chatbots

The English language is considered an essential element in the creation and development of these applications and the development of the ability to communicate between cultures [50]. Therefore, many English chatbots have been developed in various sectors (commerce, education, finance, entertainment, economy, health, news, etc.) [51,52].

All related works indicate the importance of creating chatbots and benefiting from them in various fields, whether educational, medical, economic, entertainment, etc. We noticed diversity in the languages in which chatbots are created, including English, Turkish, Latin, Arabic, and Chinese. A wide range of machine and deep learning techniques have been used in creating all the proposed chatbots due to the proof of their efficiency and high ability to achieve the highest rates of accuracy and performance.

Most of the studies achieved high results in performance and accuracy due to the use of the latest deep learning networks and their training on enough datasets after performing an appropriate set of preprocessing procedures on it. We can benefit from related works using some of the datasets used in them, using part of the preprocessing techniques, in addition to more than one technology that can be combined to obtain better results.

Serban et al. (2017) [53] proposed an educational chatbot (MILABOT) using a set of deep learning techniques: bag-of-words, template-based models, latent variable neural network, and Seq2Seq network. MILABOT interacts with users and answers their inquiries and questions through text and speech, as she has been trained on a wide range of suggested questions so that she chooses the appropriate answer to the question posed. A portion of the research sample's thoughts was gathered, and it was discovered that they were all favourable and happy with the product. On the other hand, negative samples indicate a lack of consumer satisfaction with the product. The model's precision is 72.96%. These search engines used are *Google*, *Bing*, and i. Because of that, *Google* was used as the primary search engine in this strategy to provide clear and relevant results. Aside from computational, material, and technological resources, there was also assistance. The Amazon program was used to accomplish this. Additionally, 20,000 stickers were included in the Amazon app. Tesla K80 graphics and more than 32 processor units have been used dedicated to the program's operation. The results indicate that it is superior to many competing programs due to its combination of more than one deep learning technology.

Su et al. (2017) [54] proposed a chatbot system for the elderly using an LSTM-based multilayer embedding model. The proposed system was set up and trained using a group of normal conversations among the elderly in Chinese and extracted patterns and semantic sentences using the LSTM technique. They collected daily discussions with the elderly in order to develop the MHMC chitchat database. In total, 2239 message-response pairs make up the MHMC chitchat database (MR pairs). The results showed the superiority of the proposed system with an accuracy of 79.96% in choosing the appropriate answer.

Muangkammuen et al. (2018) [55] proposed a chatbot system in the e-commerce field based on customer service via the Internet. The proposed system is based on RNN in the form of LSTM, in which the system responds to usual questions in the field. There are 2,636 pairs of questions and responses in the dataset. Then they manually classified such pairs into 80 groups (based on the number of FAQs (frequently asked questions) categories) and assigned them to an integer. Then the questions and answers are separated. Questions were utilized for training AI, whereas answers were produced to respond to clients. The results showed that the proposed robot could recognize 86.36% of the questions asked.

Palasundram et al. (2019) [56] proposed an educational chat robot to support the work of teachers and answer and clarify the questions of students who are absent from class, using an RNN-based Seq2Seq. The proposed model contributes to generating new answers corresponding to the questions asked of all kinds because it can learn and train on a renewable dataset and cannot rely on limited databases.

Bhartiya et al. (2019) [35] suggested a chatbot for university counselling in the engineering department using the feedforward neural model to train the dataset used. The application responds to all engineering-related inquiries and questions; it has been published on the Messenger application so that the inquiry-and-response procedure may be undertaken; the findings suggest that the probability of the right replies techniques is 0.72.

Dhyani and Kumar (2019) [57] proposed an English chatbot based on bidirectional recurrent neural networks (BRNN). The proposed system is characterized by being able to respond to long sentences consisting of 40 words as a maximum. The TensorFlow framework is used to implement the proposed system.

Muangnak et al. (2020) [5] designed an educational chatbot (KUSE-ChatBOT) using ANN based on deep learning techniques. The proposed chatbot was built and used within the famous LINE application so that students can access and use it to provide valuable advice and information to students regarding academic advising. In total, 75 commonly requested Thai language questions from two main groups were used in the experiments. The data for this investigation were gathered by Kasetsart University students at the Chalermphrakiat Sakon Nakhon Province Campus and from the FAQs website (KU.CSC). Students at the Faculty of Science and Engineering had questions about services; thus, the FAQs were used to obtain the answers. A total of 375 questions were created, each with its own unique word pattern. Among the themes addressed are enrollment, academic report, graduation, examination, tuition charges, library services, dormitory information, scholarships, and other fundamental information. Each question was updated by the faculty members who work in the registration department. Thai NLP and binary encryption have been used to process the FAQs data. The results indicate the superiority of the proposed chatbot in answering students' inquiries with an accuracy of over 75%. Moreover, it helped employees and academics reduce workload in answering FAQs and answering in the fastest time.

Lee et al. (2020) [58] created a voice chatbot in the Chinese Language that recognizes the emotional state of a speaker using CNN and RNN models. The suggested application analyses the emotional state of the speaker, whether they are pleased, sad, fearful, furious, or tranquil, using the CNN GoogLeNet technique and then the RNN technique to develop a dialogue and response application based on the classified positive or negative sentiments. The results indicate the superiority of the proposed robot in classifying the five emotional states of the user with an accuracy of 79.81%, and the robot also showed its effectiveness in working in the economic field.

Khan et al. (2021) [59] suggested creating a university chatbot using machine learning techniques: decision tree, random forest, and SVM. The proposed robot provides answers to various students' questions regarding the university, in addition to its answers to people's inquiries regarding university procedures or available jobs. The results show the superiority of the random forest technique in the accuracy rate over other techniques; the proposed robot showed superiority in answering all inquiries and its positive impact on university communications.

Segni et al. (2021) [60] proposed a chatbot system in the field of agriculture using RNN and DNN techniques. The proposed robot helps farmers support corn productivity by answering their inquiries and providing them with solutions to the agricultural problems they face. The robot supports the Turkish language. The results indicate that the proposed robot has an accuracy of 84.4% using DNN, while it has achieved 80% accuracy when using RNN-based "Seq2Seq".

Setiawan et al. (2021) [61] built a Banjaras chatbot based on LSTM technology on a dataset of 5000 lines. The proposed system responds to various inquiries after being adequately trained on the different conversations in the Banjaras language. In total, 400 epochs were considered ideal in this study, and the results indicate the superiority of the proposed system in terms of precision, with 85.20%.

Bagwan et al. (2021) [62] proposed a medical chatbot based on deep learning techniques and deep neural networks. The suggested application is trained on various health-related queries and WebMD database responses. The results show a high accuracy rate. While training the deep learning model, the system achieved an accuracy of 0.97 and a loss of 0.02 in terms of correct response prediction for a given query.

Prasetyo and Santoso (2021) [63] proposed a chatbot to classify conversations according to the appropriate topic for the user, based on deep learning techniques to classify conversations based on different phrases of sentences. The data source is a university guest book, which can be accessed via the website and contains numerous visitor remarks. There are 120 instances of comment data from visitors and website managers in the training set, with five different sorts of categories. The results indicate that the model achieved a high accuracy rate of 81%.

Table 3 summarizes the most recent research about chatbots using various machine and deep learning techniques.

RQ1 Answer. Despite the pace in chatbot development for the English language, research on Arabic chatbot development seems on track as more sophisticated techniques have been used. However, there is a lack of datasets in Arabic NLP research in general, and conversational and questions/answers data in particular, which still poses some challenges in Arabic chatbot performance.

4.4 Summary of Deep Neural Network and Arabic Chatbots

Given that the focus of this survey is on deep neural network techniques and Arabic chatbots, we give a summarized overview of these research studies on these two topics in Table 4. We use a set of characterization attributes, namely performance, dataset, chatbot type, technique type, additional techniques employed, domain, training time, and targeted language.

Comparison Attributes

We describe the attributes we used to compare the studies below.

- **Study.** The reference of the chatbot study
- **Performance.** The performance results of the proposed chatbot system
- **Dataset.** The amount and type of data used to train and test the chatbot model
- **Chatbot Type.** The type of the chatbot (either conversational (dialogue) or question answering
- **Technique Type.** The type of machine learning technique(s) used to develop the chatbot, either classical learning (CL) or deep learning (DL)
- **Technique.** The name of the technique(s) used to develop the chatbot
- **Algorithm.** The name(s) of the additional algorithm(s) used to optimize the chatbot
- **Domain.** The domain in which the chatbot is applied
- **Training time.** The time it takes to train the chatbot model
- **Language.** The language(s) that the chatbot supports

Observations

Looking at Table 4, we observe about half of the studies on Arabic chatbots used deep learning, whereas the other half used classical machine learning. We also observe that, despite being more sophisticated, deep learning techniques tend to achieve comparable results to classical machine learning techniques. However, we should note that the performance results we report in the Table are obtained from the studied papers, i.e., we did not evaluate the techniques on a unified

Table 3: Chatbots for Other Languages

Study	Finding	Strength	Weakness
Behera (2016) [64]	Semi-automatic chatbot	Uses NLP to analyse chats and extracts intent of the user, then uses such information and AIML to make a conversation with the user. This is the marked difference between Chappie and existing chatbots work on AIML	- Not evaluated - Small amount of data - Needs more sophisticated techniques to extract intent and classify chats more accurately
Serban et al. (2017) [53]	Interacts with users and answers their inquiries and questions	combination of more than one deep learning technology	
Su et al. (2017) [54]	Useful for elderly care, which can accompany and chat with the elderly people to reduce their loneliness	This chatbot will increase the wellness of the elderly by chatting with the system	Need more topics to be provided in the dialog
Muangkammuen et al. (2018) [55]	Recognizes the questions asked	Using modern technology	Limited dataset
Palasundram et al. (2019) [56]	Generating new answers	not rely on limited databases	High error rate
Bhartiya et al. (2019) [35]	Answers all inquiries and questions related to engineering disciplines		- High error rate - Low accuracy
Dhyani and Kumar (2019) [57]	Able to respond to long sentences	quick response	
Muangnak et al. (2020) [5]	Reduces workload in answering frequently asked questions	Quick response	Supports one app (LINE)
Lee et al. (2020) [58]	Classifies the five emotional states of the user	Creative	Low accuracy
Khan et al. (2021) [59]	Its positive impact on university communications	Quick response	No deep learning models used
Segni (2021) [60]	A chatbot system in the field of agriculture using RNN and DNN techniques.	The proposed robot helps farmers to support corn productivity by answering their inquiries and providing them with solutions to the agricultural problems they face - Support the Turkish language	Limited dataset
Setiawan et al. (2021) [61]	Achieves high accuracies at the 50th epoch	Testing on the optimization of the two conversation models resulted in a somewhat high metric score and a low loss in the 50th epoch	- Exceedingly long training process. - Limited RAM and GPU - Limited dataset
Jaiwai et al. (2021) [62]	LaLin Chatbot in Thai with LINE Applications	High accuracy rate	No AI techniques used
Prasetyo and Santoso (2021) [63]	Sentence expressions can be understood by chatbots more accurately		Limited dataset

benchmark. Therefore, achieving high or low quality is dependent on what data is being used for training/testing. While having more data is expected to produce higher results, we observe that the technique used by Alshareef and Siddiqui (2020) [45] achieved relative a lower BLEU score while being evaluated on a relatively larger dataset compared to other techniques.

In terms of the techniques used, the most common techniques is Seq2Seq, which was used by four studies. RNN, CNN, and AIML come next as there were used by two studies each. When it comes to deep learning techniques, we observe that almost all of them were applied to question answering chatbots. In addition, despite the promising performance of BiLSTM in understanding user inputs, it has only be used in one study by Bashir et al. [43]. In addition, we observe that the performance of BiLSTM was evaluated on a conversational dataset. This encourages future research to apply BiLSTM for question answering chatbots on rather large datasets of questions/answers.

Table 4: Summary of deep neural network and Arabic chatbots

Study	Performance	Dataset	Chatbot Type	Technique Type	Technique used	Algorithm	Domain	Training time	Language
Ansari et al. (2016) [32]	Not evaluated	a document which contains the details about Mahatma Gandhi	QA	DL	CNN	Net beans IDE	History		English
Boyanov et al. (2017) [33]	MAP=63% BLEU=63% Loss=63%	SemEval	QA	DL	Seq2Seq	TF-IDF			English
Nguyen et al. (2018) [65]	BLEU=1.4%	1,331 questions	QA	DL	Seq2Seq + Attention			4.8 hours	Vietnamese
Bhartiya et al. (2019) [35]	Prob.=72%	62 intents & 1059 query patterns	QA	DL	FFNN	ADAM optimizer	Education		English
Jiao (2020) [36]	Accuracy=95%	500 sentences with marked entities	QA	DL	RNN	RASA NLU	Stocks		English
Jaiwai et al. (2021) [37]	Accuracy=73-86%	1,400 sentences + 500 intents	QA	DL	RNN	NEWMM Deepcut Cutkum	Education		Thai
Tsakiris et al. (2022) [38]	Accuracy=21% F1-score=12% Loss=7%	826 questions	QA	DL	CNN				English
Ali and Habash (2016) [42]	Not evaluated	Multiple datasets*	QA	CL	AIML		Entertaining		Arabic
Bashir et al. (2018) [43]	Accuracy=94-97% F-Score=92-94% Precision=92-94% Recall=92%	768 entries	CV	DL	CNN, LSTM, BiLSTM		Restaurant & ticket ordering		Arabic
Joukhadar et al. (2019) [44]	Accuracy=86%	873 sentences	CV	CL	SVM		Restaurant & ticket ordering		Arabic (Levantine)
Alshareef and Siddiqui (2020) [45]	BLEU=25%	49,000 tweets	QA	DL	Seq2Seq	CA as machine translation	Traditional topics		Arabic (Gulf)
Zubair et al. (2021) [46]	Accuracy=93%	200 questions and answers	QA	CL	Retrieval		Religious		Arabic
Boussakssou et al. (2022) [47]	Accuracy=85-89%	81,659 conversations	CV	DL	Seq2Seq		Traditional topics	12 hours	Arabic
AlHumoud et al. (2022) [48]	User Satisfaction=79-87%	845 data entries	CV	CL	AIML		Tourism		Arabic English

CV: Conversational, **QA**: Question Answering, **CL**: Classical learning, **DL**: Deep Learning,
Seq2Seq: Sequence to Sequence, **FFNN**: Feedforward Neural Network
* http://camel.abudhabi.nyu.edu/resources

With respect to the datasets, we observe that almost all techniques were evaluated on completely different datasets from other techniques. This alerts researchers in the chatbots areas to work on a standardized benchmark on which all techniques should be evaluated to ensure robustness and validity of the chatbots' performance. Studies on Arabic chatbots seem competing with those in on other languages, as we can see relatively larger datasets were used to evaluate Arabic chatbots.

Finally, in terms of applicability, we observe that the majority of the studied chatbots were designed and evaluated for certain domain, thus raising generalizability issues. Looking at studies where no domain was specified, we observe that the chatbot performance is relatively lower that those domain-specific chatbots. This suggests future research to pay more attention to building more general chatbots to expand their applicability to serve as many sectors as possible.

5 Threats to Validity

5.1 Internal Validity

Papers selected in this survey were collected based on search queries performed on Google Scholar. While many journals and conferences are indexed by Google Scholar, our survey might have missed some relevant studies. However, given the context and objective of this survey, we were only interested in identifying the methodologies (i.e., techniques, datasets, etc.) used for designing, implementing, and evaluating AI chatbots. In other words, the main goal was to identify as many methodologies as possible rather than the number of papers. Therefore, we restricted the number of papers per year and, to that end, we excluded some of relevant papers that have similar methodologies to other selected papers. We also excluded low-quality studies for which identifying the methodology was not possible. Nevertheless, future research should extend this survey to cover other objectives and thus systematically collect additional studies to allow generating more comprehensive findings.

5.2 External Validity

The results and findings reported in this survey are based on the studies we reviewed. Therefore, we cannot generalize our conclusions to AI chatbots that do not use machine or deep learning algorithms. We encourage future research to further investigate whether our conclusions hold for other chatbot studies.

6 Conclusion

Nowadays, chatbots are widely spread in many fields and languages, as they are known as programs that have been dedicated to assisting in performing certain tasks during conversations, usually in messaging applications, as they are smart tools that use the concept of AI to communicate with computers using the natural language of humans. These robots interact directly with human beings and answer their questions and inquiries in an intelligent and fast manner based on the field in which they specialize. Chatbots are divided into two main types: the conversational chatbot and the logical chatbot. Chatbot methods can be categorized into two general methods: rule-based and AI techniques. Over the years, various chatbots have been developed for many languages, such as English and French; however, Arabic chatbots are rare due to the nature, complexity, and terminology of the Arabic language. The most famous Arabic-based chatbots are BOTTA and Nabiha. This paper discussed the proposed applications in several languages, such as Arabic, English, Turkish, and Chinese, and the deep and machine learning techniques used in building these robots. Moreover, we discussed the most important robots used in the educational field and explained their importance in supporting the educational process and saving time and effort. All previous studies prove the efficiency of the machine and deep learning techniques in achieving the best results and the highest accuracy rates when used in the process of building chatbots. Both classical and deep learning techniques have been used, such as DNN, CNN, RNN, LSTM, decision tree, random forest, and SVM, where we notice a lack of studies that use the Bidirectional LSTM (BiLSTM) model in developing chatbots for different languages. According to our analysis of chatbot studies, there is only one study that used BiLSTM for developing an Arabic chatbot, which achieved promising results, thus encouraging future research to use it to develop more accurate chatbots for Arabic language on large corpora.

Conflict of Interest

The authors declare that there is no conflict regarding the publication of this paper.

References

[1] V. Galaz, M. A. Centeno, P. W. Callahan, A. Causevic, T. Patterson, I. Brass, S. Baum, D. Farber, J. Fischer, D. Garcia *et al.*, Artificial intelligence, systemic risks, and sustainability, *Technology in Society* **67** (2021) p. 101741.

[2] E. Cambria and B. White, Jumping nlp curves: A review of natural language processing research, *IEEE Computational intelligence magazine* **9**(2) (2014) 48–57.

[3] S. Tembhekar and M. Kanojiya, A survey paper on approaches of natural language processing (nlp) j, *International Journal of Advance Research, Ideas and Innovations in Technology* **3**(3) (2017) 1496–98.

[4] B. Setiaji and F. W. Wibowo, Chatbot using a knowledge in database: human-to-machine conversation modeling, in *2016 7th international conference on intelligent systems, modelling and simulation (ISMS)*, IEEE2016, pp. 72–77.
[5] N. Muangnak, N. Thasnas, T. Hengsanunkul and J. Yotapakdee, The neural network conversation model enables the commonly asked student query agents, *Int. J. Adv. Comput. Sci. Appl* **11** (2020) 154–164.
[6] S. AlHumoud, A. Al Wazrah and W. Aldamegh, Arabic chatbots: a survey, *International Journal of Advanced Computer Science and Applications* **9**(8) (2018).
[7] A. Ahmed, N. Ali, M. Alzubaidi, W. Zaghouani, A. Abd-alrazaq and M. Househ, Arabic chatbot technologies: A scoping review, *Computer Methods and Programs in Biomedicine Update* (2022) p. 100057.
[8] E. O. Oduntan and A. James, Enhancing communication technology through an intelligent chatbot system, in *National Media Communication & Information Technology Conference for Tertiary Institutions*, 2017.
[9] M. Mohasses, How ai-chatbots can make dubai smarter?, in *2019 Amity international conference on artificial intelligence (AICAI)*, IEEE2019, pp. 439–446.
[10] A. Rahman, A. Al Mamun and A. Islam, Programming challenges of chatbot: Current and future prospective, in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, IEEE2017, pp. 75–78.
[11] B. R. Ranoliya, N. Raghuwanshi and S. Singh, Chatbot for university related faqs, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE2017, pp. 1525–1530.
[12] A. Følstad and M. Skjuve, Chatbots for customer service: user experience and motivation, in *Proceedings of the 1st international conference on conversational user interfaces*, 2019, pp. 1–9.
[13] A. M. Turing, Computing machinery and intelligence, *MIT Press* (1950) 11–35.
[14] K. M. Colby, F. D. Hilf, S. Weber and H. C. Kraemer, Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes, *Artificial Intelligence* **3** (1972) 199–221.
[15] W. Maroengsit, T. Piyakulpinyo, K. Phonyiam, S. Pongnumkul, P. Chaovalit and T. Theeramunkong, A survey on evaluation methods for chatbots, in *Proceedings of the 2019 7th International conference on information and education technology*, 2019, pp. 111–119.
[16] M. S. Satu, M. H. Parvez *et al.*, Review of integrated applications with aiml based chatbot, in *2015 International Conference on Computer and Information Engineering (ICCIE)*, IEEE2015, pp. 87–90.
[17] M. Amilon, Chatbot with common-sense database (2015).
[18] M. L. Mauldin, Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition, in *AAAI*, **94**1994, pp. 16–21.
[19] E. Bryndin, Collaboration robots as digital doubles of person for communication in public life and space, *American Journal of Mechanical and Industrial Engineering* **4**(2) (2019) 35–39.
[20] K. Kerasovitis, Artificial unintelligence: how computers misunderstand the world/the message is murder: substrates of computational capital: by meredith broussard, cambridge and london, mit press, 2018, 248 pp.,£ 19.95 (hardback), isbn 9780262038003/by jonathan beller, london, pluto press, 2018, 224 pp.,£ 75.00 (hardback), isbn 9780745337319,£ 18.99 (paperback), isbn 9780745337302, *Information, Communication & Society* **22**(10) (2019) 1514–1516.
[21] K. Shaalan, S. Siddiqui, M. Alkhatib and A. A. Monem, Computational linguistics, speech and image processing for arabic language (2018).
[22] K. Polachowska and M. Trojnarski, 10 use cases of AI in manufacturing – Neoteric https://neoteric.eu/blog/10-use-cases-of-ai-in-manufacturing, (2019), Accessed: 2022-09-01.
[23] S. Bird, E. Klein and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit* ("O'Reilly Media, Inc.", 2009).
[24] J. Eisenstein, *Introduction to natural language processing* (MIT press, 2019).
[25] I. Sutskever, O. Vinyals and Q. V. Le, Sequence to sequence learning with neural networks, *Advances in neural information processing systems* **27** (2014).
[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* **30** (2017).
[27] Y. Huang, L. J. Kangas and B. A. Rasco, Applications of artificial neural networks (anns) in food science, *Critical reviews in food science and nutrition* **47**(2) (2007) 113–126.
[28] K.-T. Yang, Artificial neural networks (anns): a new paradigm for thermal science and engineering, *Journal of heat transfer* **130**(9) (2008).
[29] R. M. Balabin and E. I. Lomakina, Support vector machine regression (ls-svm)—an alternative to artificial neural networks (anns) for the analysis of quantum chemistry data?, *Physical Chemistry Chemical Physics* **13**(24) (2011) 11710–11718.
[30] J. Zou, Y. Han and S.-S. So, Overview of artificial neural networks, *Artificial Neural Networks* (2008) 14–22.
[31] P. M. Nadkarni, L. Ohno-Machado and W. W. Chapman, Natural language processing: an introduction, *Journal of the American Medical Informatics Association* **18**(5) (2011) 544–551.
[32] A. Ansari, M. Maknojia and A. Shaikh, Intelligent question answering system based on artificial neural network, in *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, IEEE2016, pp. 758–763.
[33] M. Boyanov, I. Koychev, P. Nakov, A. Moschitti and G. D. S. Martino, Building chatbots from forum data: Model selection using question answering metrics, *arXiv preprint arXiv:1710.00689* (2017).
[34] R. Singh, M. Paste, N. Shinde, H. Patel and N. Mishra, Chatbot using tensorflow for small businesses, in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, IEEE2018, pp. 1614–1619.

[35] N. Bhartiya, N. Jangid, S. Jannu, P. Shukla and R. Chapaneri, Artificial neural network based university chatbot system, in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, IEEE2019, pp. 1–6.
[36] A. Jiao, An intelligent chatbot system based on entity extraction using rasa nlu and neural network, *Journal of Physics: Conference Series* **1487**(1) (2020) p. 012014.
[37] M. Jaiwai, K. Shiangjen, S. Rawangyot, S. Dangmanee, T. Kunsuree and A. Sa-nguanthong, Automatized educational chatbot using deep neural network, in *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, IEEE2021, pp. 5–8.
[38] G. Tsakiris, C. Papadopoulos, G. Patrikalos, K.-F. Kollias, N. Asimopoulos and G. F. Fragulis, The development of a chatbot using convolutional neural networks, in *SHS Web of Conferences*, **139**, EDP Sciences2022, p. 03009.
[39] W. Brini, M. Ellouze, S. Mesfar and L. H. Belguith, An arabic question-answering system for factoid questions, in *2009 International Conference on Natural Language Processing and Knowledge Engineering*, IEEE2009, pp. 1–7.
[40] S. S. Aljameel, *Development of an Arabic conversational intelligent tutoring system for education of children with autism spectrum disorder*, PhD thesis, Manchester Metropolitan University2018.
[41] M. Makatchev, I. Fanaswala, A. Abdulsalam, B. Browning, W. Ghazzawi, M. Sakr and R. Simmons, Dialogue patterns of an arabic robot receptionist, in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE2010, pp. 167–168.
[42] D. A. Ali and N. Habash, Botta: An arabic dialect chatbot, in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, 2016, pp. 208–212.
[43] A. M. Bashir, A. Hassan, B. Rosman, D. Duma and M. Ahmed, Implementation of a neural natural language understanding component for arabic dialogue systems, *Procedia computer science* **142** (2018) 222–229.
[44] A. Joukhadar, H. Saghergy, L. Kweider and N. Ghneim, Arabic dialogue act recognition for textual chatbot systems, in *Proceedings of The First International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) co-located with ICNLSP 2019-Short Papers*, 2019, pp. 43–49.
[45] T. Alshareef and M. A. Siddiqui, A seq2seq neural network based conversational agent for gulf arabic dialect, in *2020 21st International Arab Conference on Information Technology (ACIT)*, IEEE2020, pp. 1–7.
[46] M. Z. Khan and S. M. Yassin, Seerahbot: An arabic chatbot about prophet's biography, *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)* (2021).
[47] M. Boussakssou, H. Ezzikouri and M. Erritali, Chatbot in arabic language using seq to seq model, *Multimedia Tools and Applications* **81**(2) (2022) 2859–2871.
[48] S. AlHumoud, A. Diab, D. AlDukhai, A. AlShalhoub, R. AlAbdullatif, D. AlQahtany, M. AlAlyani and F. Bin-Aqeel, Rahhal: A tourist arabic chatbot, in *2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, IEEE2022, pp. 66–73.
[49] N. Albalooshi, N. Mohamed and J. Al-Jaroodi, The challenges of arabic language use on the internet, in *2011 International Conference for Internet Technology and Secured Transactions*, IEEE2011, pp. 378–382.
[50] S. Gupta, D. Borkar, C. De Mello and S. Patil, An e-commerce website based chatbot, *International Journal of Computer Science and Information Technologies* **6**(2) (2015) 1483–1485.
[51] M. McTear, Z. Callejas and D. Griol, Conversational interfaces: Past and present, in *The Conversational Interface*, (Springer, 2016) pp. 51–72.
[52] J. Bozic, O. A. Tazl and F. Wotawa, Chatbot testing using ai planning, in *2019 IEEE International Conference On Artificial Intelligence Testing (AITest)*, IEEE2019, pp. 37–44.
[53] I. V. Serban, C. Sankar, M. Germain, S. Zhang, Z. Lin, S. Subramanian, T. Kim, M. Pieper, S. Chandar, N. R. Ke *et al.*, A deep reinforcement learning chatbot, *arXiv preprint arXiv:1709.02349* (2017).
[54] M.-H. Su, C.-H. Wu, K.-Y. Huang, Q.-B. Hong and H.-M. Wang, A chatbot using lstm-based multi-layer embedding for elderly care, in *2017 International Conference on Orange Technologies (ICOT)*, IEEE2017, pp. 70–74.
[55] P. Muangkammuen, N. Intiruk and K. R. Saikaew, Automated thai-faq chatbot using rnn-lstm, in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*, IEEE2018, pp. 1–4.
[56] K. Palasundram, N. M. Sharef, N. Nasharuddin, K. Kasmiran and A. Azman, Sequence to sequence model performance for education chatbot, *International journal of emerging Technologies in Learning (iJET)* **14**(24) (2019) 56–68.
[57] M. Dhyani and R. Kumar, An intelligent chatbot using deep learning with bidirectional rnn and attention model, *Materials today: proceedings* **34** (2021) 817–824.
[58] M.-C. Lee, S.-Y. Chiang, S.-C. Yeh and T.-F. Wen, Study on emotion recognition and companion chatbot using deep neural network, *Multimedia Tools and Applications* **79**(27) (2020) 19629–19657.
[59] Z. M. Khan, H.-u. Rehman, M. Maqsood and K. Mehmood, Artificial intelligence based university chatbot using machine learning, *Pakistan Journal of Engineering and Technology* **4**(2) (2021) 108–112.
[60] B. Segni, *Developing Afaan Oromoo Text based Chatbot for Enhancing Maize Productivity*, PhD thesis, ASTU2021.
[61] B. A. Setiawan, E. Utami and A. D. Hartanto, Banjarese chatbot using seq2seq model, in *2021 4th International Conference on Information and Communications Technology (ICOIACT)*, IEEE2021, pp. 233–238.
[62] F. Bagwan, R. Phalnikar and S. Desai, Artificially intelligent health chatbot using deep learning, in *2021 2nd International Conference for Emerging Technology (INCET)*, IEEE2021, pp. 1–5.
[63] A. Prasetyo and H. A. Santoso, Intents categorization for chatbot development using recurrent neural network (rnn) learning, in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, **1**, IEEE2021, pp. 51–55.
[64] B. Behera, Chappie-a semi-automatic intelligent chatbot, *cse. iitb. ac. in* (2016).

[65] T. Nguyen and M. Shcherbakov, A neural network based vietnamese chatbot, in *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*, IEEE2018, pp. 147–149.