IRAQI Academic Scientific Journals
Al-Rafidain Engineering Journal (AREJ) Vol.27, No.2, September 2022,pp.178-189

A Review of Arabic Intelligent Chatbots: Developments and Challenges

Zena N. Abdulkader* zenatiq2@uomosul.edu.iq

Yousra Faisal Muhammad** yusrafaisalcs@uomosul.edu.iq

* Computer Science Departement, Collage of Education for Pure Sciences, University of Mosul, Mosul, Iraq
** Computer Science Departement, Collage of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Received: 21/12/2021 Accepted: 16/3/2022

ABSTRACT

Chatbots, or machine based conversational systems, have become popular in recent years in a variety of applications. Chatbots are intelligent technologies that employ Artificial Intelligence (AI) to converse with humans in their natural languages. Chatbot's primary function is to interpret user's questions and respond with the most convenient intelligent and natural responses. In a number of the world's most generally spoken languages, chatbots appear to be quite successful; however, Arabic chatbots have not yet attained the predicted degree of success. Many academics have recently sought to bridge the gap in the implementation of Arabic chatbots by overcoming the complicated linguistic Arabic language features.

This survey reviews the chatbots, their categories and implementation techniques. It also, presents some challenges that face the developer of Arabic chatbots due to rich complex properties of Arabic language. Then, most of published chatbot researches that deal with Arabic language are highlighted. Finally, a table of comparison among the mentioned Arabic chatbots is presented. The study concludes that, all the present AI based Arabic chatbots are developed in retrieval mode, and that Arabic chatbot literature is scarce in comparison to that in many other languages.

Keywords:
Arabic chatbot; rule based chatbots; AI based chatbots; AIML, pattern matching;

https://rengj.mosuljournals.com

1. INTRODUCTION

Modern life is not devoid of using artificial intelligence technologies due to the great improvements that aim to simulate the cognitive abilities of humans in this area [1]. One of the effective artificial intelligent technologies that have recently gained popularity is the chatbot [2]. A chatbot is a software agent that, when conversed via text or speech, reacts as if it were a smart entity that knows one or more human languages using Natural Language Processing (NLP) [3]. A chatbot is defined in the terminology as "A computer program designed to simulate conversation with human users, especially over the Internet" [4]. Terms such as smart bots, digital assistants, conversational agents, and interactive agents are all used to describe chatbots [3].

Due to dramatic improvements in artificial intelligence, machine learning and other technologies as neural networks, deep learning and natural language processing, chatbots became so popular in recent years [2]. They have served as helpful assistants in a variety of applications. Chatbots are excellent softwares for teaching, learning, and helping academia in education [5], [6], [7]. They can also be utilized in the medical field to counsel and support patients [8], [9], [10]. Chatbots have improved the e-commerce experience by acting as 24/7 customer service

employees [11], [12], [13]. Moreover, chatbots have contributed to facilitate the access to web information directly and quickly [14]. Furthermore, chatbots have expanded beyond their traditional functions to include Internet of Things (IoT) applications such as home construction automation [15].

In order to communicate with humans, chatbots rely basically on natural language. This feature puts a strain on the usefulness of chatbots, as it differs greatly from one language to another. Several researches have looked into the possibility of chatbots in a variety of international languages [1]. Arabic language suffers from shortage in NLP resources and tools; therefore, researchers appear to pay less attention to Arabic chatbots [16]. This review explores some challenges of processing Arabic language in conversational systems, the development of Arabic chatbots in term of application fields, types of dialct and their implementing technologies.

In addition to this introduction section, this review includes four sections: the second section presents Chatbot categorization according to several criteria. The third section illustrates the most common design techniques that are used to build chatbots. Arabic chatbots are explained in the fourth section, investigating the challenges facing their developers, tracing their developments since 2004, and ending with a comparison among some presented Arabic chatbots. Finally, an overall conclusion of the survey is given in the fifth section.

2. CHATBOT CATEGORIZATION

Chatbots can be categorized based on a variety of criteria. These criteria may include mode of interaction, knowledge domain, usage and the response generation methods that are typically employed in implementation of these chatbots [17]. As shown in Fig.1, the chatbot can be classified based on the following criteria:

a) Interaction mode: chatbots can be categorized into text and/or speech conversational chatbots based on how the chatbot interacts with its input and output [18].

b) Knowledge domain: It considers the knowledge that a chatbot can access or the amount of data that it has been trained on. Closed domain chatbots are focused on a certain knowledge topic and may fail to answer unrelated questions, while open domain chatbots can talk about various topics and respond effectively [3].

c) The goal: There are two categories of conversation agents, namely task-oriented and non-task-oriented chatbots. Task-oriented chatbots are designed to assist users in completing certain tasks in a specific domain, while non-task-oriented chatbots help users to participate in different domains, so they behave as informative chatbots [19].

d) The response generation approach: According to this criterion, chatbots are classified into two categories, First, *rule based chatbots* which are programmed to reply to specific questions. They essentially generate responses based on predefined rules, so the user is restricted to a limited range of questions. This chatbot is ineligible for dealing with the user's spelling and grammatical mistakes in the input as well [3], [20]. Most chatbots in this category consider the last input message, so their response selection is dedicated to single-turn conversation only. Second is the artificial intelligent (AI) based chatbots that are programmed to interact with users as real humans. They use natural language processing (NLP) and machine learning to understand the intent of the user, so they do not use keywords to generate answers. Nevertheless, they have the ability to keep track of context and word dictionary. The retrieval based model and the generative based model are two types of AI based chatbots. In the retrieval model, the response is generated by finding the best query matches in the pre-constructed conversational repository [21]. As compared to rule based model, this model provides more flexibility since it uses application programming interfaces (APIs) to query and analyze accessible resources [3]. The generative model generates answers instead of retrieving them from a predefined knowledge base; it takes into account the current user's messages as well as the previous ones. These chatbots use machine learning algorithms and deep learning techniques to achieve more human-like responses. However, there are numerous

challenges in constructing and training them [19]. All the three approaches have been used recently in designing the dialogue manager to generate appropriate responses [3].

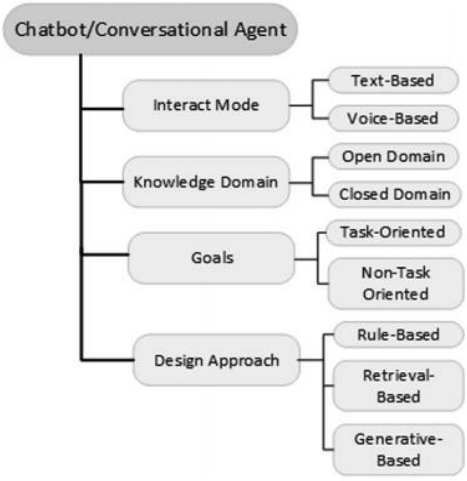

Fig. 1: Chatbot categorization.

3. TECHNOLOGY CONTEXT

The most common design techniques that used to build chatbots are:

1) Parsing: It concerns taking the user request in text form as an input, analyzing and processing it by using some NLP functions to extract meaningful information that can be easily manipulated [3], [22], [23]. To establish the grammatical structure of the text, lexical parsing coverts it into a set of simple terms [18]. Semantic parsing is an advanced paring technique that provide the machine with understandable representation of the user's input query to determine the intent of the user [17].

2) Pattern matching: Using this technique is very common in designing chatbots, specially in question-answer systems. The basic idea of this method is to take a sentence or the user intents as input while the output is the response which is created by retrieving the most consistent response. The first chatbots that use pattern recognition algorithms were Eliza and ALICE. The advantage of this technique is its flexibility in conversation creation [17]. Its disadvantage is that all possible patterns are stored manually, which results in limiting the information extraction capabilities, which are responsible for generating predictable responses. In addition, there is no way to save previous responses, which can lead to conversations looping [3].

3) Artificial Intelligence Mark-up Language (AIML): It is an XML based language that is considered as one of the scripting languages that support writing chatbots. A category is the basic unit of knowledge in AIML. Each category consists of a stimulus or input question, response or output answer, and an optional context. The stimulus is called the pattern, and the response is called the template. Each category is a rule that matches a stimulus and converts it to a response [3]. Category, pattern, and template object structure is illustrated as follows [23]:

"<category>
<pattern> User Input</pattern>
<template> Corresponding Bot Response </template>
</category>""

ALICE has been implemented using AIML. Learning and implementing AIML can be considered as an easy-flexible task for experts. So, drawbacks of AIML are that it needs some NLP and programming expertise [17]. It has weak matching patterns and is intractable to keep up with because of the large amount of content which must be manually entered to create a chatbot [24]. Fig. 2 shows Chatbot workflow using an AIML.

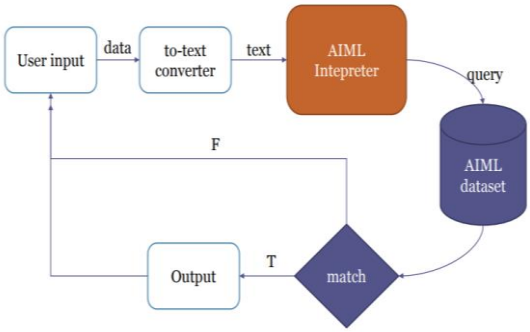

Fig. 2: Chatbot using AIML workflow

4) Chat-Script: It is a rule based chatbot development system that uses an open-source scripting language to accept user text input and generate a text response. Chat-Script begins by using replacement files to change the input words. It includes speech acts for texting, spellings, common spelling errors, contractions, abbreviations, noise, and interjections. Chat-Script aspires to be the AIML language's replacement. While AIML is simple and

straightforward to learn, it suffers from a lack of pattern matching algorithms and is difficult to maintain, as previously mentioned. Unlike AIML, which is based directly on pattern matching, Chat-Script is made up of rules associated with topics, and it begins by locating the best topic that matches the user's query string and then executing a rule in that topic. Long-term memory is provided by Chat-Script in the form of "\$" variables which are used to store some information about the user such as name or age [22], [25].

5) **Ontology:** Ontologies are AI concepts that provide formalized knowledge representations for a certain domain. They work with entities that can be categorized into hierarchical groups. Concepts, relations, functions, and instances are used to categorize such entities. The resulting ontology model might be viewed as a meta-model of a specific situation [26]. Ontologies have the general benefit of providing a structure of reusable information for a specific purpose. They allow for appropriate outputs since they set constraints and relationships between components. Furthermore, they can be expanded with additional information without changing the overall structure [27].
6) **Markov chain model:** Since this model is based on the occurrence probability of a word or a letter in the input text, it helps in the construction of replies that are more suited probabilistically and hence more correct. The postulate behind Markov Chains is that each letter or word in a data set has a fixed probability of occurrences. Simple Markov chains models are easy to program, but they do not function well with rich and sophisticated conversations [18].
7) **Natural Language Processing (NLP**): Natural Language Processing is a part of AI that deals with natural language. It investigates how computers can understand and control a natural language, whether it is text or speech. Therefore, it is an essential technology in chatbots. It also, finds a way to convert the user's speech or text into structured data to become eligible to be used to extract relevant responses for the users [28]. Currently, there are some NLP techniques in use which are: name entity recognition, dialogue planning, vector recognition, text similarity (which deals with finding the degree of similarity between two pieces of text), intent classification, lexicon and LSTM [29].
8) **Natural Language Understanding (NLU)**: In AI, this technique is a fundamental part of NLP to extract context from unstructured human input language. It's important to use NLU to reply based on the current user's intent [22].
9) **Artificial Neural Network Models**: In terms of response generation, these models can be employed for both retrieval-based and generative based chatbots. The presence of a learning algorithm in the neural network based chatbots is the main difference from the rule based ones. ANN based chatbots can be further categorized into supervised and unsupervised depending on the type of machine learning algorithms. Deep learning, which is a subset of machine learning, can work with unsupervised mode from unstructured or unlabeled data. In the field of conversational modeling, the usage of deep learning neural networks has increased, especially when the field have been dominated by: recurrent neural network (RNN), sequence to sequence (Seq2Seq), and long short term memory networks (LSTMs) [17], [30].
    a. **Recurrent Neural Network (RNN)**: is an extended feed forward network utilized in training datasets in which the network generates its current response by considering the current input and the previous output. It has the ability to remember the previous input sequence as well. So, RNN is useful when a sequential data processing is needed. It has the ability to represent the sequential nature of natural language, in which words build their semantical meaning based on the preceding words in a phrase. This enables the RNN to maintain context and create a result based on the sentence's previous words. This property makes RNN ideal for chatbots, as it is necessary to understand the conversational environment in order to comprehend the user's inputs and deliver a contextually appropriate response. The main problem with the RNN is that it cannot remember long input sequence, but this drawback is solved by the LSTM [17], [32].

b. **Long Short-Term Memory (LSTM):** is a special kind of RNN that has the ability to learn long term dependencies [25]. RNNs have a long-term dependency concern that LSTMs are essentially designed to prevent. In LSTMs, memory cells and gates are introduced, allowing cells to recall past information for long periods of time. Just like data in a computer memory, these memory cells may store information, write new information, and read it [32]. Because of their frequent ability to refer to a piece of distant information, LSTMs have proven to be quite beneficial in the design of chatbots [25].

c. **Sequence to Sequence model (Seq2seq):** is one of the most basic techniques of machine translation. The overall goal is to create a target sequence based on the source sequence [22], [34]. It is employed in language translation when the source sequence is a sentence in one language and the target sequence is the translated version in another. However, this may also be applied to chatbots, where the source sequence is the user's chat message and the target sequence is the machine's response[35].

4. ARABIC CHATBOTS

Chatbots are intelligent technologies that employ AI to converse with humans in their natural languages. Chatbot's primary function is to understand user requests and offer the most appropriate responses [36]. This is done by utilizing the NLP techniques, that's why it considered as the core of the chatbots.

Allowing users to communicate with a bot in their local language will provide more human customized interaction. In this section the Arabic language features and challenges that face the development of Arabic chatbot will be presented.

The Arabic language is both fascinating and challenging. It is fascinating because of its history, as Classical Arabic has remained unchanged for more than fifteen centuries, and its literary and cultural heritage due to its closely association with Islam, and it has prestigious group of literature. Strategically, the Arabic language is spoken by about 330 million people globally and is an official language in over twenty countries [37].

Nowadays, The Arabic language is divided into several types: The first is known as classical or Quranic Arabic language (CAL); while the second is known as Modern Standard Arabic (MSA), which is the official language of the Arab world used in formal written and spoken mediums such as journalism, education, and literature. In addition, the dialectal Arabic (DA) is the third type that utilized on a daily basis in spoken and written personal communication as well as in informal situations, yet each country and region has its own dialect [38]. Arab speakers usually use these different types of Arabic depending on the nature of conversation and sometimes they use more than one type in a single conversation which causes a diglossia [39].

4.1. Challenges Facing Arabic Chatbots

Many challenges are faced in the processing of Arabic natural language due to its features. Firstly, the presence of different types of Arabic language. In that sense, there are numerous difficulties in processing Arabic language texts, including rich morphology, orthographic deviations, and high degree of ambiguity [18]. Arabic words have a rich morphology that is inflected for a variety of characteristics like as gender, number, voice, person, and so on. This creates a problem in the context of a chatbot system. Furthermore, because verbs, adjectives, and pronouns are all gender specific, the chatbot must have two separate answer systems - one for male users and one for female users [40].

Many features in Arabic language result in high ambiguity; such as:

- The same word has different meanings.
- Lack of capitalization that makes it difficult to extract named entities [41].
- Arabic orthography includes diacritical marks which are pronounced as short vowels that are omitted in writing, and can change the meaning of the word. For example, the word "علم" becomes ambiguous without such short vowels, as illustrated in Table (1) [38].
- The rare use of consonantal doubling in the text with optional diacritical marks [40].
- It has a complex and rich grammatical structure, for example, pronouns are embedded

in the words themselves in many cases, and the pronoun can be hidden in the Arabic sentence as well [42].

The shape of Arabic letters varies depending on the position of the letter in the word. As an example, the letter "ت" has three forms. At the beginning of the word it is written as "تـ", in the middle of word it is written as "ـتـ", and at the end of word it is written as "ـت" [39]. Therefore, it is difficult to create a knowledge base that covers all the ambiguity and the diversity of the Arabic words.

Table 1:Different meanings for the same word.

Arabic	Transliteration	PoS	meaning
علم	‘alam	n	flag
عِلم	‘ilm	n	science
عُلِمَ	ulima	v	known
عَلَّمَ	‘allama	v	teach
عَلَم	‘alam	a	famous

4.2. Arabic Chatbots Development

In 2004, B. Abu Shawar and E. Atwell [43] have designed one of the earliest Arabic chatbot by using Artificial Linguistic Internet Computer Entity (ALICE) chatbot. It was an AI based chatbot since they used ML techniques to generate it by retraining ALICE on non - conversational Arabic text. They have developed a Java program to read classical Arabic, that is the Qur’anic Arabic, from corpus and convert it to AIML format to be used by ALICE. Because the Qur'an is a non-conversational Arabic text, a chatbot has been developed that can deal with that type of text. The most important word in the "Ayha" that exemplifies the category is the template for that Ayah in AIML file, hence this chatbot is a retrieval based model. The diacritical text is used to avoid sophisticated NLP. The user asks in Arabic diacritical text, and the chatbot finds "Ayahs" from the Quran which contain these words. The conversation's domain is restricted by the content of the Quran.

In 2010, M. Makatchev, *et al.* [44] published a comparative study between two receptionist chatbots: "Tank" which was located on the Pittsburgh campus of Carnegie Mellon University (CMU), and interacts in English language only; and "Hala" which was situated in Qatar's CMU reception center and interacts in both English and Arabic. They compare Hala's English dialogue corpus to that of "Tank" to report its deployment. "Hala" is supposed to be a young single Arab female chatbot who serves as a receptionist, providing information about campus directions, local events, and weather, as well as answering questions about its personal life. Therefore, it is considered as an open domain bot. Users interact with Hala by text in one of three available input modes: Arabic in MSA form, English, or Arabic written in English letters. Depending on the input mode, Hala will respond in the same mode as the input by generating a synthesized speech as well as text. Additionally, Hala has a rule based dialogue manager that consists of a knowledge base of canned mode. The purpose of Makatchev et al. work was to examine discourse patterns such as robot qualities, covered knowledge bases, and cultural variation in the community of users at CMU-Qatar to explore more about the culture of human-robot interaction[18].

In 2011, N. Mavridis, *et al.* [45] developed "IbnSina" as a multilingual conversational android robot, supporting MSA Arabic and English. The user can use speech or text for interaction, whereas "IbnSina" responds with audio. The response is in the same language as the user's request. To generate dialogue, "IbnSina" accesses the online Wikipedia and the stored Quran database, so it covers many topics. It utilizes google translate as well, and it has the basic read aloud capability. It also informs the user if there is any missing information or erroneous spelling. Pattern matching approach was used to develop this conversational robot.

In 2013, O. G. Alobaidi, *et al.* [7] developed "Abdullah" the tutor. "Abdullah" is an Arabic Conversational Intelligent Tutoring System (CITS). This online tutoring system teaches the basic topics of Islam, allowing interaction by using natural Arabic language text in either MSA or CAL form. Rather than simply presenting the answers, Abdullah can ask questions and provide problem-solving assistance. During the session, the system also uses sound effects and images to keep students' attention. Pattern matching technique was used to implement this system.

In 2014, M. Hijjawi, *et al.* [46] implemented an Arabic CA based on Pattern Matching approach to deal with the textual conversations in MSA language called "ArabChat". The user and the chatbot intercat using text mode in MSA language. The ArabChat is a rule based Conversational Agent that is developed to help

students at Jordan's Applied Science University. ArabChat's design has been divided into three main modules: scripting language, engine, and brain. The scripting language is a predefined language in which the application's domain is scripted. The scripting language is organized as a rule based language with main domain topics (contexts) and subdomain topics (rules). Each rule has a number of simulated user sentences (patterns) and related responses. The brain, on the other hand, is a structured knowledge base that stores the domain's scripts. Finally, the engine responds to user conversations by comparing them to the prescribed domain and providing an appropriate response. The conversation is turn based type, that is each of the conversation's parties (both user and ArabChat) has its turn to converse. When a user enters an utterance into ArabChat, the utterance is processed and a suitable response is provided. Till one of the parties terminates the conversation, the conversation keeps going on.

In 2015, M. Hijjawi, *et al.* [47] modified "ArabChat" [46] to a mobile based CA called Mobile ArabChat that handles the Arabic conversations between the Arab users and mobile devices. This version of ArabChat has been implemented in Android to work as an advisor for students in Applied Science University in Jordan. It interacts with users by standard Arabic texts. Mobile ArabChat is a rule based CA that is implemented by pattern matching approach. The Mobile Arabchat framework includes the same components of the ArabChat.

In 2016, M. Hijjawi, *et al.* [48] updated the ArabChat [46] producing a complement version of it called the "Enhanced ArabChat". To get better agent's performance, certain new features have been added to the previous version of *ref.*[47]. These features were "Utterance Classification" and "Hybrid Rule". To deal with keyword matching, Utterance Classification adds more keywords to the pattern of the question - based rules to recognize between question and non-question utterances. The Hybrid Rule is the second new feature, and it focuses on how to handle requests with multiple topics that necessitate the execution of many rules at once.

Again in 2016, D. A. Ali and N. Habash [40] implemented an Arabic dialect chatbot "BOTTA". BOTTA is a female companion chatbot that simulates amicable interaction with users by using the widely understood Egyptian Arabic (Cairene) dialect. By asking questions, BOTTA retains basic information about the user such as age, gender, and nationality in a temporal manner. It is an open dialogue because the chatbot can respond to a variety of topic domains. The categories comprising BOTTA's replies to user inputs are stored in AIML files in its knowledge base. It interacts with users through text, and it is constructed as a retrieval based model that relies on a pool of predetermined responses and it uses heuristics to respond with an acceptable response.

In 2017, S. Aljameel, *et al.* [49] developed an Arabic CITS called "LANA" that adapts the learning methods VAK (Visually, Auditory, Kinesthetically) for autistic children aged 10 to 16 who are qualified in Arabic writing to improve their learning. LANA engages children by delivering science lessons in modern standard Arabic. The VAK model is used to map the knowledge given to students. The proposed system's architecture uses a combination of Pattern Matching and Short Text Similarity (STS) to extract the responses.

In 2019, A. Fadhil and A. AbuRa'ed [10] introduced OlloBot which is a text based Arabic conversational agent for health care. It aids physicians and patients during the treatment process. It does not replace the physician; rather, it tracks the patients' health and assists them through conversation with care providers. OlloBot keeps track of a user's daily food intake, provides helpful advice for a healthy lifestyle, and keeps a record of their dietary habits. OlloBot creating approach relies on IBM Watson Conversation API (IBM bluemix)1 to handle the dialogue structure (conversation flow and dialogue states), and Telegram Bot Platform to build the chatbot and provide Artificial Intelligence support to catch different user intents and entities. Since some interactions of OlloBot work with Graphical UI and some others work with conversational UI, the interaction with the bot need not be conversational only.

Recently, in 2020, D. Al-Ghadhban & N. Al-Twairesh [50] developed a chatbot "Nabiha" that can support conversation with IT students at King Saud University using the Saudi Arabic dialect and it

provides entertainment as well as useful information. It is available on different platforms: Web, Android, and Twitter. "Nabiha" is developed using AIML and pattern matches approaches, and the interaction with users is text based. The limitation in "Nabiha" is due to its limited dataset, and the developers needed to deal with the limitations of Twitter's text area as well.

More recently, in 2021, AL-Hanouf et al. [13] built an Arabic flight booking dialogue system using rule based (pattern matching) and data-driven (using training data to train the system) approaches within a pipeline system architecture. As the amount of data gathered for training was not sufficiently large to depend on a data-driven approach only, the rule based approach was also used. Consequently, a merge of the two approaches was used to construct the DS. The interaction with this DS was accomplished through standard Arabic text. Real flight booking systems had been simulated because the airline companies and flight booking agents did not give permission to connect this DS to their APIs. The functionality of the system is being enhanced by self-feeding with the new user's enquiries to create new training samples. The pre-mentioned Arabic Chatbot systems are summarized in table (2).

Table (2): The presented cahtbots.

Year	Chatbot	Domain	Language	Type of Interaction		Implementation Technique	Response generation technique
				Request	Response		
2004	Quran chatbot [43]	Quran holy book	Classical Arabic	Text	Text	AIML	Retrieval based
2010	Hala[44]	Open	MSA/ English	Text	Voice and text	AIML	Rule based
2011	IbnSina[45]	Open	MSA/ English	Voice or Text	Voice	Pattern Matching	Retrieval based
2013	Abdullah[7]	Islamic fundamentals	CAL / MSA	Text	Text	Pattern Matching	Retrieval based
2014	ArabChat[46]	Applied Science for university students	MSA	Text	Text	Pattern Matching	Rule based
2015	Mobile ArabChat [47]	Applied Science for university students	MSA	Text	Text	Pattern Matching	Rule based
2016	Enhanced ArabChat[48]	Applied Science for university students	MSA	Text	Text	Pattern Matching	Rule based
2016	BOTTA[40]	open	Egyptian Arabic dialect	Text	Text	AIML	Retrieval based
2017	LANA[49]	Science for children with Autism	MSA/English	Text	Text	Pattern Matching and STS	Retrieval based
2019	OlloBot[10]	Health care	MSA	Text	Text	AIML	Retrieval based
2020	Nabiha[50]	assist the students of KingSaud University's ITdepartment	Saudi Arabic dialect	Text	Text	Pattern matching and AIML	Rule based
2021	Arabic flight booking DS[13]	airline ticket booking	MSA	Text	Text	Pattern matching	Rule based and data-driven

Table (2) reviews the evolution of those Arabic conversational agents and compares them

in several aspects. These aspects are: The application domain (whether it is open or closed to a specific field), the type of Arabic language the chat can interact with, the method of interaction (text/speech), the techniques that were used to implement the chatbot, and finally the response generation technique. Compared to English chatbots, the evolution of Arabic chatbots is very limited and slow. Also, it can be noticed from the table that the developers have determined a specific type of Arabic for interaction to eliminate the ambiguity of the language. Another point that can be extracted from the table is that all the Arabic chatbots use either rule based or retrieval based technology to generate their responses.

5. CONCLUSION

This review on Arabic chatbots developments and challenges has covered different Arabic chatbots in the period between 2004 and 2021 for different application domains. The basic principles of chatbots have been reviewed as well as their classifications according to many aspects. In addition, the most common technologies for developing chatbots have been considered. It has been shown that the developers and researchers faced difficulties in developing Arabic conversational agents because of the high ambiguity of Arabic language due to various reasons related to the Arabic language characteristics. As a result, manuscripts in this field are few compared to other languages. The developers have taken a very important step by initially determining the type of Arabic language which is suitable to their application.

Also, it has been illustrated that all the developed Arabic chatbots generate their responses using the rule based and/or retrieval based techniques.

Despite the fact that their initial application in teaching was limited to Islamic principles using classical Arabic, Arabic chatbots have been extended to many other fields, and used other types of Arabic dialects, yet this development remains slow. A compression among the pre-shown Arabic chatbots has been accomplished to prove the pre-mentioned findings, reflecting the difficulties due to Arabic language nature and the broadening of its usage.

REFERENCES

[1] A. Almurayh, “The Challenges of Using Arabic Chatbot in Saudi Universities”, *IAENG International Journal of Computer Science*, Vol. 48 Issue 1, pp. 190-201, 2021.

[2] A. Gupta, D. Hathwar and A. Vijayakumar, “Introduction to AI Chatbots”, Int J Eng Res Technol (IJERT), Vol. 9 Issue: 7, pp. 255–258, 2020.

[3] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology", *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 373-383, 2020.

[4] L. Fryer, M. Ainley, A. Thompson, A. Gibson and Z. Sherlock, “Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners”, *Computers in Human Behavior*, Vol. 75, pp. 461–468, 2017.

[5] C. Edwards, A. Beattie, A. Edwards and P. Spence, “Differences in perceptions of communication quality between a twitterbot and human agent for information seeking and learning”, *Computers in Human Behavior*, vol: 65, pp. 666-671, 2016.

[6] J. Jia, “The study of the application of a keywords-based chatbot system on the teaching of foreign languages”, *arXiv preprint cs/0310018*, 2003.

[7] O. G. Alobaidi, K. A. Crockett, J. D. O'Shea and T. M. Jarad, "Abdullah: An intelligent arabic conversational tutoring system for modern islamic education." *Proceedings of the World Congress on Engineering*. Vol. 2, 2013.

[8] R. Crutzen, G. J. Y. Peters, S. D. Portugal, E. M. Fisser, and J. J. Grolleman, “An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study”. *Journal of Adolescent Health*, Vol. *48,* Issue: 5, pp. 514-519, 2011.

[9] A. S. Miner, L. Laranjo, and A. B. Kocaballi, “Chatbots in the fight against the COVID-19 pandemic”. *NPJ digital medicine*, Vol. *3, Issue:* 1, pp.1-4, 2020.

[10] A. Fadhil, A. AbuRa’ed, “OlloBot-Towards A Text-Based Arabic Health Conversational Agent: Evaluation and Results”, In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pp. 295-303, 2019.

[11] L. Cui, S. Huang, F. Wei, C. Tan, C. Duan and M. Zhou, "Superagent: A customer service chatbot for e-commerce websites", *In Proceedings of ACL 2017, System Demonstrations*, pp. 97-102, 2017, July.

[12] N. T. THOMAS, "An e-business chatbot using AIML and LSA", In: *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, pp. 2740-2742, 2016.

[13] A. H. Al-Ajmi and N. Al-Twairesh, "Building an Arabic Flight Booking Dialogue System Using a Hybrid Rule-Based and Data Driven Approach". *IEEE Access*, Vol. *9*, pp. 7043-7053, 2021.

[14] L. Cui, S. Huang, F. Wei, C. Tan, C. Duan, and M. Zhou, "Superagent: A customer service chatbot for e-commerce websites". In *Proceedings of ACL System Demonstrations*, pp. 97-102, 2017, July.

[15] B. Kadali, N. Prasad, P. Kudav, and M. Deshpande, "Home Automation Using Chatbot and Voice Assistant". In *ITM Web of Conferences*, . EDP Sciences, Vol. 32, pp. 01002, 2020.

[16] B. A. Shawar, and E. Atwell, "Arabic question-answering via instance based learning from an FAQ corpus", In *Proceedings of the CL 2009 International Conference on Corpus Linguistics, UCREL*, Vol. 386, No. 1, pp. 1-12, 2009.

[17] S. Hussain, O. A. Sianaki, and N. Ababneh, "A survey on conversational agents/chatbots classification and design techniques". In *Workshops of the International Conference on Advanced Information Networking and Applications*, pp. 946-956, Springer, Cham, 2019.

[18] S. Al Humoud, A. Al Wazrah, and W. Aldamegh, "Arabic chatbots: a survey", *Int. J. Adv. Comp. Sci. Appl.*, Vol. 9, no. 8, pp. 535-541, 2018.

[19] E. H. Almansor, and F. K. Hussain, "Survey on intelligent chatbots: State-of-the-art and future research directions", In *Conference on Complex, Intelligent, and Software Intensive Systems*, pp. 534-543, Springer, Cham. 2019.

[20] R. Kumar, and M. M. Ali, "A Review on Chatbot Design and Implementation Techniques", *International Research Journal of Engineering and Technology IRJET*, Vol. 07 Issue: 02, pp. 2791-2800, 2020.

[21] Y. Song, R. Yan, C. T. Li, J. Y. Nie, M. Zhang, and D. Zhao, "An Ensemble of Retrieval-Based and Generation-Based Human-Computer Conversation Systems", *IJCAI-18*, pp. 4382-4388, 2018.

[22] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications", *Machine Learning with Applications*, *Elsevier Ltd*., Vol. 2, pp.100006, 2020.

[23] S. A. Abdul-Kader and J. C. Woods, "Survey on chatbot design techniques in speech conversation systems", *International Journal of Advanced Computer Science and Applications*, Vol. 6, Issue: 7, pp. 72–80, 2015.

[24] S. Arsovski, A. D. Cheok, M. Idris, and M. R. B. A. Raffur, "Analysis of the chatbot open source languages aiml and chatscript: A review", *In 9th DQM International Conference on life cycle engineering and management*, 2017, September.

[25] K. Ramesh, S. Ravishankaran, A. Joshi and K. Chandrasekaran, "A Survey of Design Techniques for Conversational Agents". *Information, Communication and Computing Technology. ICICCT 2017. Communications in Computer and Information Science CCIS, V*ol. 750, pp. 336–350, *Springer, Singapore* 2017. https://doi.org/10.1007/978-981-10-6544-6_31

[26] J. Božić, "Ontology-based metamorphic testing for chatbots", *Software Quality Journal*, DOI: 10.1007/s11219-020-09544-9 pp. 1-25, 2021. Ontology-based metamorphic testing for chatbots (springer.com)

[27] A. Quamar, C. Lei, D. Miller, F. Ozcan, J. Kreulen, R. J. Moore and V. Efthymiou, "An ontology-based conversation system for knowledge bases". In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pp. 361-376, 2020, June.

[28] S. Maher, S. Kayte and S. Nimbhore, "Chatbots & Its Techniques Using AI: an Review", *International Journal for Research in Applied Science and Engineering Technology*, Vol. *8, Issue:* 12, pp. 503-508, 2020.

[29] 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*, pp. 111-119, 2019.

[30] R. Csaky, "Deep learning based chatbot models", *arXiv preprint arXiv: 1908.08835*, 2019.

[31] J. Cahn, "CHATBOT: Architecture, design, & development". *Senior Thesis (EAS499) University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science*, 2017.

[32] H. Sak, A. Senior and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech

recognition", *arXiv preprint arXiv:1402.1128*, 2014.

[33] K. Panesar and L. Mudikanwi, "Chatterbot implementation using Transfer Learning and LSTM Encoder - Decoder Architecture", *International Journal*, Vol. *8, Issue:* 5, pp. 1709 – 1715, 2020.

[34] L. Mou, Y. Song, R. Yan, G. Li, L. Zhang and Z. Jin, "Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation", *arXiv preprint arXiv:1607.00970*, pp. 3349-3358, 2016.

[35] M. Mnasri, "Recent advances in conversational NLP: Towards the standardization of Chatbot building", *arXiv preprint arXiv:1903.09025*, 2019.

[36] A. Almurayh, "The Challenges of Using Arabic Chatbot in Saudi Universities", *IAENG International Journal of Computer Science*, Vol. 48, Issue: 1, 2021.

[37] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions", *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. *8, Issue:* 4, pp. 1-22, 2009.

[38] B. A. Shawar, "A Chatbot as a natural web Interface to Arabic web QA", *International Journal of Emerging Technologies in Learning (iJET)*, Vol. 6, Issue: 1, pp. 37-43, 2011.

[39] M. Hijjawi and Y. Elsheikh, "Arabic language challenges in text based conversational agents compared to the English language". *International Journal of Computer Science and Information Technology (IJCSIT)*, Vol. *7, Issue:* 5, pp. 1-13, 2015.

[40] 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*, pp. 208-212, 2016.

[41] W. Ahmed and P. BabuAnto, "Answer extraction for how and why questions in question answering systems", *International Journal of Computational Engineering Research (IJCER)*, Vol. *12, Issue:*6, pp.18-22, 2016.

[42] E. S. AlHagbani and M. B. Khan, "Challenges facing the development of the arabic chatbot". In *First International Workshop on Pattern Recognition*, International Society for Optics and Photonics, Vol. 10011, p. 100110Y, 2016.

[43] A. Shawar and E. S. Atwell, "An Arabic chatbot giving answers from the Qur'an", In *Proceedings of TALN04: XI Conference sur le Traitement Automatique des Langues Naturelles*, Vol. 2, pp. 197-202, ATALA. 2004.

[44] 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,)* pp. 167-168, 2010.

[45] N. Mavridis, A. AlDhaheri, L. AlDhaheri, M. Khanii and N. AlDarmaki, "Transforming IbnSina into an advanced multilingual interactive android robot", In *2011 IEEE GCC Conference and Exhibition (GCC)* IEEE, pp. 120-123, 2011.

[46] M. H ijjawi, Z. Bandar, K. Crockett and D. Mclean, "ArabChat: An arabic conversational agent", In *2014 6th International Conference on Computer Science and Information Technology (CSIT)* IEEE. pp. 227-237, 2014.

[47] M. Hijjawi, H. Qattous and O. Alsheiksalem, "Mobile Arabchat: An Arabic Mobile-Based Conversational Agent", *Int. J. Adv. Comput. Sci. Appl. IJACSA*, Vol. *6, Issue:* 10, pp. 111-119, 2015.

[48] M. Hijjawi, Z. Bandar and K. Crockett, "The Enhanced Arabchat: An Arabic Conversational Agent", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. *7*, Issue 2, 2016. http://dx.doi.org/10.14569/IJACSA.2016.070247

[49] S. S. Aljameel, J. D. O'Shea, K. A. Crockett, A. Latham and M. Kaleem, "Development of an Arabic conversational intelligent tutoring system for education of children with ASD", In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, IEEE. pp. 24-29, 2017.

[50] D. Al-Ghadhban and N. Al-Twairesh, "Nabiha: An Arabic dialect chatbot", *Int. J. Adv. Comput. Sci. Appl*, Vol. *11, Issue:* 3, pp. 1-8, 2020.

استعراض العميل الذكي للمحادثة العربية: التطورات والتحديات

زينه ناطق عبد القادر *
zenatiq2@uomosul.edu.iq

يسرى فيصل محمد **
yusrafaisalcs@uomosul.edu.iq

* جامعة الموصل - كلية التربية للعلوم الصرفة - قسم علوم الحاسوب
** جامعة الموصل - كلية علوم الحاسوب والرياضيات- قسم علوم الحاسوب

الملخص

شاع في الآونة الاخيرة استخدام روبوتات المحادثة (Chatbots)، أو أنظمة المحادثة الآلية، في مختلف التطبيقات. تعتبر روبوتات المحادثة تقنيات ذكية تستخدم الذكاء الاصطناعي (AI) للتحدث مع البشر بلغاتهم الطبيعية. تتمثل الوظيفة الأساسية لروبوتات الدردشة في تفسير أسئلة المستخدمين والرد عليها بالاستجابات الذكية الأكثر ملاءمة وبلغتهم. في عدد من اللغات الأكثر شيوعًا في العالم، تظهر روبوتات المحادثة ناجحة جد ؛ ومع ذلك، لم تحقق روبوتات المحادثة العربية (Arabic chatbots) درجة النجاح المتوقعة حتى الان. لقد سعى العديد من الأكاديميين مؤخرًا إلى سد الفجوة في تنفيذ برامج الدردشة العربية من خلال التغلب على الميزات اللغوية المعقدة للغة العربية.

تستعرض هذه الدراسة روبوتات المحادثة وطرق تصنيفها وتقنيات تنفيذها. كما أنها تطرح بعض التحديات التي تواجه مطوري روبوتات المحادثة العربية وذلك بسبب ميزات اللغة العربية الغنية والمعقدة. بعد ذلك ، تم عرض معظم روبوتات المحادثة العربية المنشورة.

أخيراً، تم تقديم جدول مقارنة بين تلك الروبوتات التي تم عرضها. وخلصت الدراسة إلى أن جميع روبوتات المحادثة العربية الحالية القائمة على الذكاء الاصطناعي تم تطويرها باستخدام تقنية الاسترجاع، وهناك فجوة كبيرة بين عدد مؤلفات روبوتات المحادثة العربية وتلك الموجودة في العديد من اللغات الأخرى.

الكلمات الدالة :

روبوت محادثة عربي ، روبوتات محادثة قائمة على القواعد ، روبوتات الدردشة الذكية، AIML ، مطابقة النمط .