Bilingual AI-Driven Chatbot for Academic Advising

Ghazala Bilquise[1]
Computer Information Science
Department, Higher Colleges of
Technology, Dubai
United Arab Emirates

Samar Ibrahim[2]
School of Arts and Sciences
American University in Dubai
Dubai, United Arab Emirates

Khaled Shaalan[3]
The British University in Dubai
Informatics Department
The British University in Dubai

Abstract—Conversational technologies are revolutionizing how organizations communicate with people, thereby raising quick responses and constant availability expectations. Students often have queries about the institutional and academic policies and procedures, academic progression, activities, and more in an academic environment. In reality, the student services team and the academic advisors are overwhelmed with several queries that they cannot provide instant responses to, resulting in dissatisfaction with services. Our study leverages Artificial Intelligence and Natural Language processing technologies to build a bilingual chatbot that interacts with students in the English and Arabic languages. The conversational agent is built in Python and designed for students to support advising-related queries. We use a purpose-built domain-specific corpus consisting of the common questions advisors receive from students and their responses as the chatbots knowledge base. The chatbot engine determines the user intent by processing the input and retrieves the most appropriate response that matches the intent with an accuracy of 80% in English and 75% in Arabic. We also evaluated the chatbot interface by conducting field experiments with students to test the accuracy of the chatbot with real-time input and test the application interface.

Keywords—*Chatbot; conversational agent; academic advising; natural language processing; deep learning; bilingual English Arabic*

I. INTRODUCTION

Conversational technologies are transforming the interaction landscape between organizations and people, causing digital communication to be propelled by technology rather than humans. A chatbot, also known as a conversational agent, is a software system that processes and simulates human conversation to provide digital assistance in real-time [1]. The constant availability of chatbots and the ability to respond immediately and communicate in a natural language have escalated their popularity across all domains [2], [3]. Chatbots are being entrusted with various tasks previously handled by human agents, such as providing customer service, healthcare advice, e-shopping, and answering queries. Organizations pervasively rely on chatbots to support customers' needs and increase customer satisfaction with services. Therefore in this digital era, chatbots have the potential to support student queries and assist in the academic advising process in the education domain.

Academic advising is an integral function of Higher Education Institutions (HEIs) and has been widely acknowledged as a principal strategy for confronting the challenges of persistence and retention [4]–[6]. While advising encompasses several tasks, one of the crucial tasks of advising is to provide students with the essential information required for navigating their academic journey successfully. This task involves a high degree of interaction between advisors and students and often leads to dissatisfaction with advising services when students cannot get timely and accurate information. The large number of students assigned to each advisor makes it impossible for the advisor to respond to all students in a satisfactory amount of time [7]. Moreover, students' expectations and information requirements for their daily tasks have intensified with today's technological advancement. Providing adequate channels for student communication is vital for their academic progression and integration with the academic environment. Therefore, a chatbot can provide numerous benefits to the students and the academic institution by providing instant responses to students, thereby enhancing student satisfaction.

This study aims at building a chatbot for the students at an academic institution in the UAE. The institution offers four undergraduate programs of study. There are nearly 3000 students of Arab origin and almost 100 faculty members employed at the institution. Each faculty member serves as an advisor to nearly 25-30 students per semester. This large ratio makes it challenging for the advisor to contribute quality time to advisees and answer all their queries or make them aware of the college policies related to registration, courses, pre-requisites, and more. A chatbot would assist in reducing the workload of the advisor so they may focus on more cognitive tasks such as creating an ideal study plan for their advisees.

Considering the aforementioned challenges of advising at the institution of study, the study aims to develop a chatbot that supports students in answering queries on college and academic-related matters and thereby improve student satisfaction with college services. The chatbot will be bilingual and provide an interface in both English and Arabic. Moreover, the chatbot will be developed using a neural network and Natural Language (NLP) technologies. Thus, our study is novel in its context with bilingual conversational support.

The rest of the paper is organized as follows. Section II provides a literature review on the background of chatbots and related studies of chatbot use in the education sector and bi-lingual chatbots. Next, Section III describes our research methodology, while Section IV presents the evaluation and results of the study. Finally, the study concludes with Section V, which summarizes the paper, significance of the study, limitations, and new directions for future research.

II. Literature Review

A. Overview of Chatbots

Chatbots, are dialog systems that mimic human conversations in text, voice, or multimodal form [1]. A chatbot, also known as a conversational agent, processes user input to discover the query's intent and respond appropriately. In the last few years, there has been a tremendous rise in chatbot applications worldwide [3]. Organizations rely on chatbots to respond to customer service queries and automate tasks [8]. Chatbots are also being used in the healthcare sector for psychiatric and medical diagnosis, raising awareness of health and safety issues [9], [10]. In the educational sector, chatbots are used for teaching and learning activities, student advising, and administrative tasks [11]. Chatbots offer a cost-effective means of delivering services to consumers eliminating repetitive and time-consuming human-agent communication, enabling them to focus on high-end complex tasks [2].

Several classifications exist in literature to categorize chatbots. A chatbot may be rule-based or driven by Artificial Intelligence (AI). A rule-based chatbot provides predefined responses based on keywords and a defined set of rules. ELIZA and PARRY were among the earliest rule-based chatbots developed in the 1960s, built using pattern recognition technology [3]. Artificial Intelligence Markup Language (AIML) [12] was used to develop the ALICE chatbot in 1995. The markup language is based on an XML structure. Chatbots developed with AIML use a rule-based approach to respond to user queries based on inputs that match a pattern.

On the other hand, an AI-driven chatbot is powered by NLP techniques to recognize the intents of the user input and generate an appropriate response based on the intent. AI-driven chatbots are technologically superior and can meet consumers' language and conversational expectations [3]. Several AI techniques have been employed in the literature to develop chatbots, such as machine learning, neural network [13], deep learning with sequence to sequence model [14], and CVAE Models [15].

Chatbots have been classified as task-oriented or non-task-oriented based on their functionality [16]. A task-oriented chatbot responds to domain-specific user queries and performs tasks such as making a reservation, placing an order, or answering queries. On the other hand, a non-task-oriented chatbot responds to open-ended queries that are not domain-specific, also called an open-domain. The main purpose of these chatbots is to act as digital assistants using an open-ended dialog. Siri and Alexa are an example of non-task-oriented virtual assistants.

Chatbots have also been classified based on their response generation method as retrieval-based and generative chatbots [17]. A retrieval-based chatbot retrieves responses from a knowledge base using machine learning algorithms, and NLP techniques process the user input, allowing users to communicate in natural language. However, the responses generated in a retrieval-based chatbot are fixed. On the other hand, a generative chatbot is trained on a conversational corpus to generate new and diverse responses that do not exist in the dataset. A limitation of the generative model is that it requires massive training data and may provide unpredictable responses not stored in the corpus.

This study uses a domain-specific knowledge base to develop a task-oriented chatbot that responds to student queries. The students ask questions in a natural language, yet the responses provided by the chatbot must be precise and accurate. Hence we use an AI-driven retrieval-based chatbot that uses NLP techniques to process user input and retrieve precise responses from a corpus of advising queries. The chatbot determines the user intent by processing the input and retrieving the response that matches the intent.

B. Chatbots in Education

Some studies used NLP techniques with a rule-based approach for developing chatbots in the educational setting to answer student queries [18], [19]. Reference [18] developed a rule-based conversational agent using PHP and NLP to respond to student queries with an accuracy of 80%. While reference [19] developed a chatbot using a social conversation dataset between students and advisors. The chatbot was developed using a frequent intent pattern by discovering rules from the dataset.

Several studies develop retrieval-based chatbots to answer student queries using AI and NLP techniques. Refrence [20] developed a chatbot based on pattern matching using AIML and Latent Semantic Analysis (LSA). The chatbot answers informational queries on college and academics. In a similar study, [21] proposed a chatbot that answers frequently asked questions. The knowledge base of the chatbot consisted of 300 questions. Both studies did not evaluate the performance of the chatbot.

Reference [22] developed an AI-driven chatbot that allows students to enquire about college admission rules and policies. The chatbot is developed using the RASA framework. The performance of the chatbot was evaluated using the confidence of the responses. However, the confidence does not indicate the accuracy of the response. Moreover, the study did not specify how they handled spelling errors in the user input.

In another study, [13] developed a chatbot using machine learning and NLP techniques that answer campus-related queries published as FAQs on the website. The study compares two chatbot models, RNN based Seq2Seq model and a semantic similarity model. The results show that the semantic similarity model performs better in cases where the dataset size is small. Furthermore, this study uses a deep learning model to process the input patterns and retrieve the most accurate response rather than constructing responses, similar to our study. However, the chatbot is developed in one language only.

Several studies have developed chatbots to answer students' admissions, policies, or academic advising queries. However, only a few have used neural networks with NLP techniques to process the user input.

C. Chatbots in Arabic and other Languages

Due to its complexity, the Arabic language is underrepresented in NLP and chatbot development and is not given enough attention by researchers. Few studies have

examined Arabic chatbots in general and education, some of which were bilingual or multilingual.

BOTTA is an Arabic Egyptian dialect female public chatbot proposed by [23] that simulates friendly conversations with users. It is a retrieval-based model designed for open domain conversations responds. Arabchat and enhanced ArabChat are conversational agents designed for students at Applied Science University in Jordan [24]. Both are interactive chatbots that use Arabic MSA textual language. The study [25] proposed a conversational social chatbot "Nabiha" for Information Technology (IT) students at King Saud University using the Saudi Arabic dialect. Nabiha is a retrieval-based chatbot that uses AIML. It serves as an academic counselor to interact with students about their courses and academic progress inquiries.

A bilingual chatbot called "Jooka" was designed by [26] to improve the admissions process at the German University in Cairo (GUC). It understands queries written in English and Arabic and responds based on the query language. Google Cloud, Translation API was used to translate Arabic to English. However, in our study we found that translation of APIs for Arabic language are still not mature and result in an unnatural response.

Reference [27] proposed a voice-interactive chatbot that adopts a multilingual interface specifically to detect and respond to exam stress of university students. The chatbot application analyzes the tone of the user's voice to determine their feelings towards their exams with an accracy of 76.5%.

Multilingual chatbots have been developed in domains other than education. For example, [28] proposed a multilingual health chatbot application that can diagnose disease based on user symptoms and supports three languages: English, Hindi, and Gujarati. Reference [29] presents a bilingual retail chatbot that can handle Filipino-Tagalog and English languages that employs k-fold cross-validation on a dataset generated using a bilingual automatic corpus engine.

Supporting users with chatbot conversations in English as well as local dialects is gaining importance and is highlighted in the literature. The above studies show that there are two ways of creating a bilingual chatbot. The first method is using translation services to perform translation between both languages while maintaining a corpus in the primary language only. And the other method is to create and maintain a two corpus files, one for each language. Our study adopts the second method as experimentation with the first method resulted in unnatural translation between English and Arabic languages.

III. METHODOLOGY

This section presents the methodology adopted for planning, designing, and developing the chatbot system. This system provides bilingual advice through a chatbot with an easy-to-use interface to communicate in either Arabic or English. Our chatbot is equipped with sufficient information to provide students with answers to their specific advising inquiries. The advice chatbot adopts a bilingual corpus as the knowledge source type used to generate responses adopting the retrieval-based model. As part of the retrieval-based model, a

chatbot uses heuristics to select the most appropriate response from a predefined pool of responses. This retrieval-based model is selected due to the need for precise and accurate responses to a specific task and domain. The following sub-sections present the three phases of the methodology - data collection, building the chatbot model, and the chatbot GUI development.

A. Data Collection

The conversational data required for the chatbot was collected through interaction with students, advisors, and referrals to university policy documents. The data consists of the most commonly asked queries that advisors usually receive from students and responses to those questions. We followed four main steps in collecting the conversational data required for the chatbot. First, we identified eight primary contexts to classify each query. The context is the domain of the user's request, such as attendance, course delivery, and more. Second, we added queries to the contents and tagged each query with a unique intent tag that identifies the main purpose of the query. Third, we created patterns for each query to depict the variety of ways the question may be presented to the chatbot. Last, we added a variety of responses for each intent to incorporate diversity in the response.

In summary, each intent reflects what students would like to accomplish when interacting with the chatbot. Table I illustrates the different contexts and the number of intents in each context. For the purpose of this study, we developed 152 English and Arabic intents, with a total of 356 patterns.

TABLE I. DISTRIBUTION OF INTENTS AND PATTERNS

Context	Number of intents	Patterns	Description
Greeting	8	28	That greet, welcome, and thank the user
Academic Standing	22	50	Students' academic status/probation
Registration	32	76	inquiries related to registration/ scheduling and retaking courses
Summer	6	24	Inquiries related to Summer Courses /credits
WP	26	56	Inquiries related to work placement, schedule/registering.
COVID	14	28	Inquiries related to requirements related to COVID on campus
Final Exam	20	48	Inquiries related to materials scheduling /attending / missing / to final exams
Attendance	6	12	Inquiries related to attendance
Course Delivery	18	34	Inquiries about online, Hybrid courses
	152	356	

The corpus, consisting of the conversational data, was stored in JSON format. We use two corpus files to store the English intents and the other to store the Arabic intents as the initial experiments revealed that translation services from Arabic to English and vice versa are still very weak and result in unnatural statements. For example, when translating the Arabic statement "ما هو المعدل المناسب للنجاح", the resulting translation is "what is the appropriate rate of success," which is

not a natural way of phrasing the statement in the English language.

The complexity and NLP challenges inherent in the Arabic language, such as dialectal differences, orthographic ambiguity, and inconsistencies, are more prevalent while translating [23]. Moreover, the existing translation functions are inaccurate and do not reflect the correct English statements.

Furthermore, using a separate Arabic corpus allows us to integrate English words that are typically used by students when they write in the Arabic language, such as "probation", "covid," "GPA," and more. Also, the Arabic corpus uses Arabic words written in local dialects. Table II shows sample intents from the Arabic and English corpora, with patterns, and responses.

After building the chatbot, we conducted a pilot implementation with eight students and three faculty members to augment the corpus with additional queries. Students and advisors were asked to type questions in natural language (English and Arabic) within the contexts identified earlier. The purpose of the pilot was to re-examine the initial corpus and augment it with additional patterns in which a query may be composed by the user. Furthermore, the pilot was also meant to identify any gaps in data collection within the contexts identified. After conducting the pilot, we examined the results and added new intents or patterns to existing intents. In addition, we identified queries that were not addressed in the initial corpus development; for example, questions, such as blackboard password, arriving before the final exam, and materials needed for the final exam were not included in the initial corpus development. Therefore, the pilot implementation was crucial to extend the corpus.

B. Chatbot Modelling with Deep Learning

The chatbot model was developed in Python using a supervised deep learning algorithm. Deep learning is a subset of machine learning based on an artificial neural network, in which layers of nodes simulate the neurons of a human brain. Input neurons are interconnected with multiple hidden layers to produce output by automatically adjusting the weights of the nodes in each layer [30]. We used the keras library in Python to build our deep learning network to build two chatbot models, each trained on the English and Arabic corpus, respectively. Fig. 1 shows the steps involved in developing the English chatbot model. Similar steps were also applied for developing the Arabic chatbot model.

First, to train our chatbot model, we pre-processed the training data and encoded each intent to make it suitable for the neural network algorithm. Pre-processing is crucial to transform the corpus data in an appropriate form for the neural network algorithm. Pre-processing the data enhances the efficiency and performance of the model. The pre-processing phase includes transforming input to lower case, removing punctuations and special characters, tokenization, and vectorization of the words. We used the NLTK library in Python to perform all the pre-processing steps.

Tokenization is the process of extracting words from sentences. We tokenized all the patterns in the corpus to extract individual words. The words were then simplified to their base

forms using the process of lemmatization and stemming for English and Arabic words, respectively. Lemmatization converts the words to mean their original form based on the context, while stemming reduces the words to their base by removing the last characters without preserving the meaning. We used the NLTK WordNetLemmatizer library to lemmatize the English words using the parts of speech tag. For the lack of a sound library in Python for lemmatizing Arabic words, we used the ISRIStemmer to stem the words. However, some Arabic words did not retain their meaning when stemmed, such as "يمك" does not have any meaning originally "يمكنني," also "ادة" did not preserve its meaning originally "مادة." Table III shows input patterns in Arabic and English and the extracted words. Word extraction and reducing to its base form resulted in 247 unique words in English and 250 words in Arabic.

TABLE II. SAMPLE ARABIC AND ENGLISH INTENTS

Description	Sample Intent
Arabic intent where there is the use of a UAE dialect such as "أقدر"	{"tag": "numberofcurses-ar," "patterns": [" كم عدد المواد أقدر اسجل " , " كم مادة ", "كم مادة ممكن اسجل"," أقدر اسجل "], "responses": ["هذا يعتمد على مستواك الاكاديمي"], "context_set": "academic-standing", }
Arabic intent that includes English word "Probation" and written in Arabic "بروبيشن" and also using a dialect UAE "شتقصد" and using	{"tag": "Probation-ar," "patterns": [" Probation ما معنى "," بروبيشن", "ما معنى تحت الاختبار","شتقصد"], "responses": ["هذا يعتمد على مستواك الاكاديمي"], "context_set": "academic-standing", }
English intent for the registration context student clarifying of a section that is full in a course	{"tag": "sectionfull-en", "patterns": ["If all sections are full", "There are no seats available"], "responses": ["Contact your advisor to change your plan"], "context_set": "registration",}
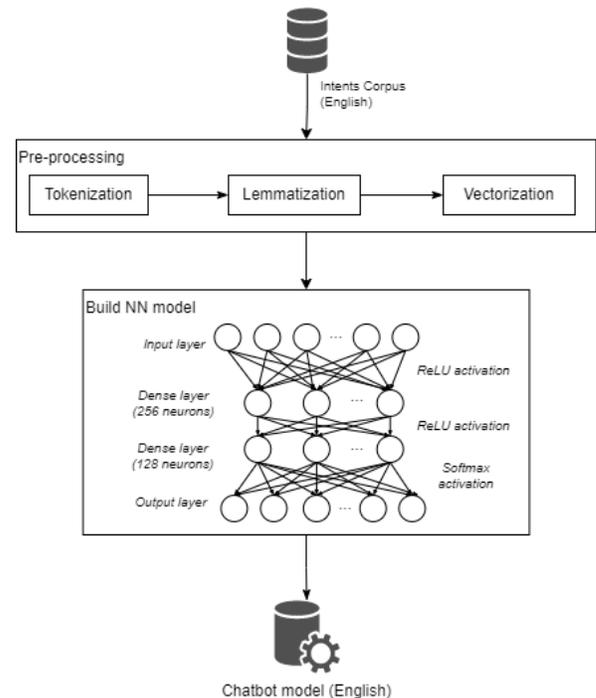
Fig. 1.   English Chatbot Model Development Process.

TABLE III.   SAMPLE OF LEMMATIZATION AND STEMMING OF PATTERNS

Pattern	Extracted words
'What is going to happen if I do not raise my GPA	['what', 'be', 'go', 'to', 'happen', 'if', 'I', 'do', 'not', 'raise', 'my', 'gpa']
'What is the time for adding and dropping courses'	['what', 'be', 'the', 'time', 'for', 'add', 'and', 'drop', 'course']
كيف يمكنني معرفة ما إذا كانت المادة عبر الإنترنت أم لا	['كيف', 'يمك', 'عرف', 'ما', 'اذا', 'كانت', 'ادة', 'عبر', 'نرن', 'ام', 'لا']
ماذا أفعل بحضوري إذا لازم أذهب للخدمة العسكرية	['اذا', 'فعل', 'حضر', 'اذا', 'لزم', 'ذهب', 'خدم', 'عسكر']

The next step of pre-processing is the process of vectorization. In this step, the words were converted to numerical form by creating a list of word vectors, which is a two-dimensional representation of each unique word and its frequency of occurrence. These word vectors are used as features of the neural network input layer.

After the pre-processing phase, we build two Neural Network (NN) models with deep learning for English and Arabic, respectively. The keras library in Python was used to build the NN model. The network consists of an input layer, two hidden layers, also known as the dense layers, and the output layer. The input layer comprises of all the unique features extracted from the respective corpus and has approximately 250 neurons in each model. The output layer represents the classes or the intents that should be predicted.

The first dense layer has 256 neurons, and the second has 128 neurons with a dropout rate of 0.5. The number of neurons in the layers is considered ideal since a smaller number would lead to underfitting, and a larger number would result in overfitting. Therefore, we selected the number of neurons in the dense layers between the input and output neurons. We configured the neural network with the following settings:

Optimizer – Stochastic Gradient Descent (SGD). The SGD estimates the expected risk gradient based on a single randomly selected sample instead of computing the precise value. Thus it is an optimization algorithm because the samples are randomly selected from the distribution [31].

Activation Function – Rectified Linear Unit (ReLU), was used as an activation function in the hidden layers. ReLU is a piecewise function in which if the input value is zero or less, then the output value will also be zero; otherwise, the output value will equal the input value. When data value is forced to be zero, a sparse characteristic is created, making the function fast and efficient. In addition to providing a faster computer rate, the ReLU function does not cause gradient diffusion problems, i.e., minor errors. However, because it always returns 0 for negative values, it can kill some neurons permanently and affect the final results or the output, i.e., generate exploding gradients [32].

Learning Rate – 0.01. The learning rate is a configurable hyperparameter used in training neural networks. Typically, between 0.0 and 1.0, it has a small positive value that must be carefully selected. That value determines how quickly the models are adapted to the problem. Lower learning rates result in more training cycles, and the process can get stuck, whereas larger learning rates lead to rapid changes and require fewer training cycles [33].

Classification function – Softmax. In artificial neural networks, the classification function, also known as the activation function, identifies a node's output given an input or set of inputs. The activation function allows neural networks to recognize complex relationships and patterns in data. This refers to the activated neurons features that can be retained and mapped out by nonlinear functions and employed to solve complex nonlinear problems. Furthermore, the activation function increases the neural network's ability in which the nonlinear ability of the activation function makes the deep neural network have real artificial intelligence [32].

Epoch – 200. The epoch determines how many cycles are used to train the model. Since the dataset size is small, we set the epoch size to 200.

C. Chatbot Engine and GUI

The chatbot engine interacts with the Graphical User (GUI) to get the user query as an input and returns the most suitable response. Fig. 2 shows the architecture of our chatbot engine. There are three logical components of our chatbot engine – Natural Language Understanding (NLU), Natural Language Processing (NLP), and Natural Language Generation (NLG).

1) NLP: In this component, the user submits a query; the chatbot application first determines the language used for communication and accordingly uses the appropriate chatbot model for getting the response. The input query is first corrected for spelling mistakes using a spell check function implemented from the TextBlob Python library. In the case of Arabic input, however, there is no spell check function performed due to the complexity of the Arabic language and the inconsistency of the spellchecking function on the Arabic language, which led to many errors while applying it. This is considered a limitation of the study and a potential area for further development, research, and analysis. Also, in the NLP component, the input query is pre-processed using the same methods used in the training phase: tokenization, lemmatization/stemming, and vectorization of the words. In addition, all intents features are extracted from the input query in this component.
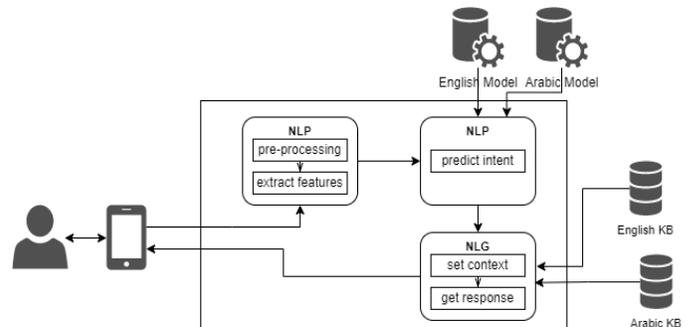
Fig. 2.   Chatbot Engine.

2) NLU: This component bridges the gap between what computers understand and how people speak. The appropriate chatbot model is used for prediction by providing the word vectors to the two different models, Arabic or English, for classification. The prediction returns all the matching intents along with the probability of prediction. We set an error threshold of 25% to accept all predictions that have a probability above this threshold. Thus, in this component, if the model is not confident of the intent it detects, the user is requested to rephrase or restate their intent because of missing vocabulary.

3) NLG: In this component, the user's intent context is set based on the user query and language selected. The prediction is performed according to the training model discussed in the previous section. The function matches the intents tags and generates the response from either the Arabic or English knowledge base. If the model is unable to generate the response, a message will be displayed in English or Arabic, "contact your advisor," "اتصل بمشرفك".

The advising chatbots Graphical User Interface (GUI) was developed using Python's tkinter library. Our chatbot application employs a simple natural language user interface similar to an instant messaging application, which has a text box to type the input, a button to submit the message, and a display to show the input and response of the chat conversation. In addition, our interface consists of a language button that allows the user to toggle between the English or Arabic language mode to communicate with the chatbot. Fig. 3 shows three screenshots of an English and Arabic conversation, respectively. The screenshot (a) shows that the chatbot accepts spelling errors as the spellings are corrected in the pre-processing phase. For example, despite the spelling mistake of the word "available," the chatbot retrieves the correct response. In screenshots (b), the chatbot appends an additional message to rephrase the question when the response retrieval has a low confidence rate (below 0.75).

IV. EVALUATION AND RESULTS

Evaluation metrics are essential to determine the machine learning algorithm's performance and assess the chatbot application. Since there are no standard evaluation methods of a chatbot application [34], the evaluation measure should be adapted to the chatbot type of service. Some studies used both automatic and human evaluations to measure the performance of chatbots [35].

Automatic evaluation measures the machine learning model's performance using known metrics such as accuracy, F1-Score, BLEU, and more, while human evaluation measures the quality of the responses using people as evaluators. Hence, human evaluation is suitable for generative chatbots that generate diverse responses, which do not exist in the corpus. However, since our chatbot is retrieval-based, we use only automatic evaluation to assess the chatbot performance.

We used two methods of evaluation. First, we used a test set consisting of queries with labeled intents, and second, we used human input to test the chatbot application and performance using ad-hoc queries. Finally, we used accuracy

as a metric to evaluate the chatbot model for both methods. Accuracy measures the ratio of correct responses over the total responses that are predicted on an unlabeled set of inputs.

We developed two new test sets in the English and Arabic languages in the first evaluation method, which were not used to train the model. Each set is approximately 30% of the corpus size. The test set is populated with queries labeled with the actual intent tags. The label is hidden from the prediction algorithm when the test is performed. Table IV shows a few sample queries in English and Arabic from the test set. The queries are phrased differently than the patterns that exist in the corpus. The evaluation aims to determine the percentage of accurate responses retrieved by the chatbot. After running the prediction on the test set, predicted intents were compared to the actual intents to determine the number of correct responses. The accuracy of the English model was 80%, while the accuracy of the Arabic model was 75%.

(a)

(b)
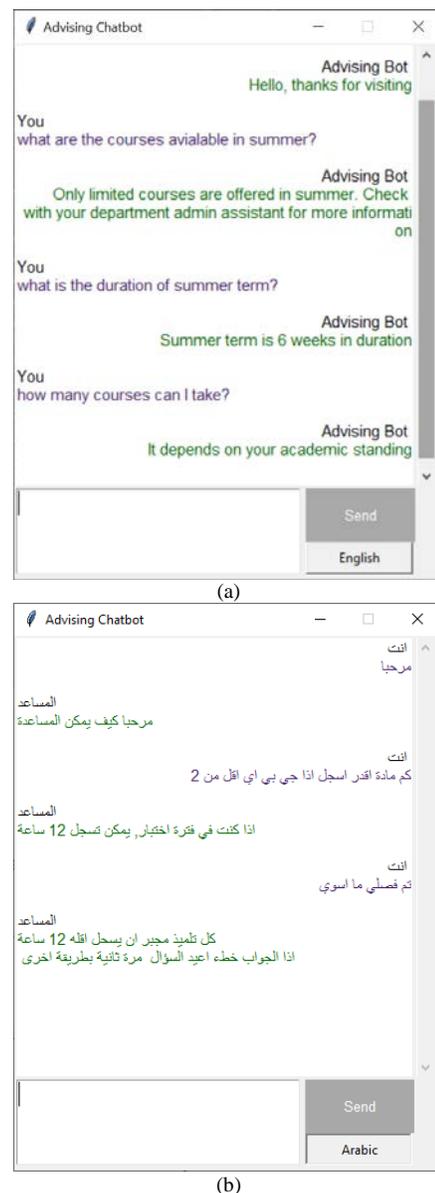
Fig. 3.   Figure 1 Screenshot of an English and Arabic Conversation.

TABLE IV.   SAMPLE QUERIES FROM THE TEST SET

Query	actual_intent (label)
If I'm working do I still need to take work placement course?	wp-working-en
how long is a summer semester?	sum-duration-en
امت اضافة حذف	add-drop-ar
كيف اعرف اذا الدورة هيبرد او وجها لوجه	course-how-ar

In the second evaluation method, we involved the end-users, students, and advisors, to test the chatbot GUI application and performance of the prediction model. The objective of this evaluation was threefold, to test the chatbot interface, the effectiveness of the conversational system, and the accuracy of responses based on context.

Thirty students and three advisors evaluated the chatbot both in English and Arabic language. The participants were briefed on the context of the chatbot corpus and asked to provide random queries. The interactions were recorded in a CSV file along with the response's predicted intent, context, and confidence. When the response confidence was below 0.75, the chatbot requested the user to rephrase the question if they thought the response was not accurate. In nearly 20% of the cases, the chatbot engine could not determine the intent due to out of vocabulary words or out-of-context queries, so the standard response "Contact your advisor" was displayed. This result shows that it is essential for the chatbot corpus to be extended to include a wider domain of queries. From all the captured test inputs, we considered only the intents within the context specified to determine the accuracy of the response.

Our study does not evaluate the user satisfaction of the chatbot application. This type of study involves gathering empirical feedback from end-users from the Human-Computer Interaction perspective, which is outside the scope of our paper. However, during the testing phase of the chatbot application, several students commented that they found the chatbot useful and would prefer to use it instead of going to their advisor. In addition, they appreciated the quick response and constant availability of a chatbot application. Another observation we made from this evaluation is that students preferred to use English rather than Arabic when writing their queries as it was faster for them to type. There were several words that they did not know how to write in Arabic, such as "probation" or "covid.".

V.   CONCLUSION

In today's world, conversational agents are proving to be one of the most innovative forms of user interaction. This paper presents a new bilingual task-oriented, domain-specific Arabic English chatbot explicitly designed to advise university students to ease their academic journey. The chatbot uses NLP and neural network algorithms to retrieve English or Arabic responses. Through the bot, students may communicate and receive responses to their inquiries. Two chatbot models have been created in Python using a supervised deep learning algorithm, trained on English and Arabic corpora, respectively. An Arabic and English corpus of roughly 152 intents in both English and Arabic has been developed, with 356 patterns. In order to train the model, we pre-processed the training data and

encoded each intent using the Python library so that it is suitable for the neural network algorithm. In the absence of a good library in Python for lemmatizing Arabic words, ISRIStemmer was used to stem the words. We use three logical components (NLP), (NLU), and (NLG) in our chatbot engine in order to pre-process the input query and to predict and generate a response based on the user's request. The prediction error threshold was set at 25%, and all predictions with probabilities above this threshold were accepted.

Moreover, the chatbots graphical user interface was developed using the Python tkinter library to interact with the user and display the most appropriate response. Two types of evaluations were performed to measure the performance of the system; the confidence score and another automated evaluation performed by the system users. The first provides 80% accuracy in English and 75% in Arabic. The second evaluation performed by the user also has similar results.

A. Limitations and Future Work

The bilingual chatbot system has some limitations. It was challenging to spellcheck Arabic, and many errors were produced when the results did not match the input inquiry after the check was performed. There was another issue with lemmatizing in Arabic. Some of the words did not retain their meaning, so the response was incorrect. There were also challenges with getting a response when the model confidence level was low, and the model did not understand the user's intent.

Both Arabic and English corpora should be expanded to include more vocabulary in each intent tag. Additionally, adding more intents with new context will broaden the scope of the corpora used in English and Arabic and expand advisory areas. Finally, the Arabic spellchecker needs further study and analysis to be used in the system.

Another limitation of the study is that the developed chatbot does not provide personalized assistance to students. Future work would enhance the chatbot with intelligent capabilities that allow personalized responses containing information such as students' GPA, academic standing, and courses required for graduation. Such a chatbot could assist advisors in developing study plans and communicating with the students. Another enhancement to the chatbot that can add value to the communication is to send push notifications to remind students of upcoming deadlines for registration, add and drop periods, and more.

REFERENCES

[1] M. Allouch, A. Azaria, and R. Azoulay, "Conversational Agents: Goals Technologies and Challenges," Sensors (Switzerland), pp. 1–48, 2021.

[2] M. Adam, M. Wessel, and A. Benlian, "AI-based chatbots in customer service and their effects on user compliance," Electron. Mark., vol. 31, no. 2, pp. 427–445, 2021, doi: 10.1007/s12525-020-00414-7.

[3] J. Grudin and R. Jacques, "Chatbots, humbots, and the quest for artificial general intelligence," Conf. Hum. Factors Comput. Syst. - Proc., pp. 1–11, 2019, doi: 10.1145/3290605.3300439.

[4] S. Campbell and C. Nutt, "Academic Advising in the New Global Century: Supporting Student Engagement and Learning Outcomes Achievement," Peer Rev., vol. 10, no. 1, p. 4, 2008.

[5] J. K. Drake, "The Role of Academic Advising in Student Retention and Persistence," About Campus Enrich. Student Learn. Exp., vol. 16, no. 3, pp. 8–12, 2011, doi: 10.1002/abc.20062.

[6] T. Fricker, "The Relationship between Academic Advising and Student Success in Canadian Colleges: A Review of the Literature.," Coll. Q., vol. 18, no. 4, p. n4, 2015.

[7] O. Iatrellis, A. Kameas, and P. Fitsilis, "Academic advising systems: A systematic literature review of empirical evidence," Educ. Sci., vol. 7, no. 4, 2017, doi: 10.3390/educsci7040090.

[8] A. Miklosik, N. Evans, A. Mahmood, and A. Qureshi, "The Use of Chatbots in Digital Business Transformation : A Systematic Literature Review," IEEE Access, vol. 9, pp. 106530–106539, 2021, doi: 10.1109/ACCESS.2021.3100885.

[9] M. R. Pacheco-Lorenzo, S. M. Valladares-Rodríguez, L. E. Anido-Rifón, and M. J. Fernández-Iglesias, "Smart conversational agents for the detection of neuropsychiatric disorders: A systematic review," J. Biomed. Inform., vol. 113, p. 103632, 2021, doi: https://doi.org/10.1016/j.jbi.2020.103632.

[10] M. Jovanovic, M. Baez, and F. Casati, "Chatbots as Conversational Healthcare Services," IEEE Internet Comput., vol. 25, no. 3, pp. 44–51, 2021, doi: 10.1109/MIC.2020.3037151.

[11] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," Comput. Educ. Artif. Intell., vol. 2, p. 100033, 2021, doi: 10.1016/j.caeai.2021.100033.

[12] R. S. Wallace, "The anatomy of ALICE," in Parsing the turing test, Springer, 2009, pp. 181–210.

[13] M. Daswani, K. Desai, M. Patel, R. Vani, and M. Eirinaki, "CollegeBot: A Conversational AI Approach to Help Students Navigate College," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12424 LNCS, no. October 2020, pp. 44–63, 2020, doi: 10.1007/978-3-030-60117-1_4.

[14] N. Asghar, P. Poupart, J. Hoey, X. Jiang, and L. Mou, "Affective neural response generation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10772 LNCS, pp. 154–166, 2018, doi: 10.1007/978-3-319-76941-7_12.

[15] X. Zhou and W. Y. Wang, "MOJITALK: Generating Emotional Responses at Scale," Proc. 56th Annu. Meet. Assoc. Comput. Linguist., pp. 1–2, 2018.

[16] S. Hussain, O. Ameri Sianaki, and N. Ababneh, A Survey on Conversational Agents/Chatbots Classification and Design Techniques, vol. 927, no. March. Springer International Publishing, 2019.

[17] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology BT - Artificial Intelligence Applications and Innovations," 2020, pp. 373–383.

[18] E. M. Latorre-Navarro and J. G. Harris, "An Intelligent Natural Language Conversational Agent System for Academic Advising," Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 1, 2015.

[19] S. Alias, M. S. Sainin, T. S. Fun, and N. Daut, "Intent Pattern Discovery for Academic Chatbot-A Comparison between N-gram model and Frequent Pattern-Growth method," ICETAS 2019 - 2019 6th IEEE Int. Conf. Eng. Technol. Appl. Sci., 2019, doi: 10.1109/ICETAS48360.2019.9117315.

[20] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, "Chatbot for university related FAQs," 2017 Int. Conf. Adv. Comput. Commun. Informatics,

ICACCI 2017, vol. 2017-Janua, pp. 1525–1530, 2017, doi: 10.1109/ICACCI.2017.8126057.

[21] C. Asakiewicz, E. A. Stohr, and S. Mahajan, "Building a Cognitive Application Using Watson DeepQA," IT Prof., vol. 19, no. 4, pp. 36–44, 2017.

[22] S. Meshram, N. Naik, V. R. Megha, T. More, and S. Kharche, "College Enquiry Chatbot using Rasa Framework," in IEEE Asian Conference on Innovation in Technology (ASIANCON), 2021, pp. 1–8, doi: 10.1109/ASIANCON51346.2021.9544650.

[23] D. A. Ali and N. Habash, "Botta: An Arabic dialect chatbot," COLING 2016 - 26th Int. Conf. Comput. Linguist. Proc. COLING 2016 Syst. Demonstr., pp. 208–212, 2016.

[24] S. AlHumoud, A. Al Wazrah, and W. Aldamegh, "Arabic Chatbots: A survey," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 8, pp. 535–541, 2018, doi: 10.14569/ijacsa.2018.090867.

[25] D. Al-Ghadhban and N. Al-Twairesh, "Nabiha: An Arabic dialect chatbot," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 3, pp. 452–459, 2020, doi: 10.14569/ijacsa.2020.0110357.

[26] W. El Hefny, Y. Mansy, M. Abdallah, and S. Abdennadher, "Jooka: A Bilingual Chatbot for University Admission," World Conf. Inf. Syst. Technol., vol. 1367 AISC, no. March, pp. 671–681, 2021, doi: 10.1007/978-3-030-72660-7_64.

[27] K. Ralston, Y. Chen, H. Isah, and F. Zulkernine, "A voice interactive multilingual student support system using IBM watson," in 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, 2019, pp. 1924–1929, doi: 10.1109/ICMLA.2019.00309.

[28] S. Badlani, T. Aditya, M. Dave, and S. Chaudhari, "Multilingual healthcare chatbot using machine learning," 2021 2nd Int. Conf. Emerg. Technol. INCET 2021, pp. 1–6, 2021, doi: 10.1109/INCET51464.2021.9456304.

[29] J. K. Catapang, G. A. Solano, and N. Oco, "A Bilingual Chatbot Using Support Vector Classifier on an Automatic Corpus Engine Dataset," 2020 Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2020, pp. 187–192, 2020, doi: 10.1109/ICAIIC48513.2020.9065208.

[30] A. R. Martinez, "A Survey of the Usages of Deep Learning for Natural Language Processing," IEEE Trans. Neural Networks Learn. Syst., vol. 32, no. 3, pp. 352–357, 2021, doi: 10.1002/wics.76.

[31] L. Bottou, "18 Stochastic Gradient Descent Tricks," pp. 421–422, 2012.

[32] Y. Wang, Y. Li, Y. Song, and X. Rong, "The influence of the activation function in a convolution neural network model of facial expression recognition," Applied Sciences (Switzerland), vol. 10, no. 5. 2020, doi: 10.3390/app10051897.

[33] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," 2012, [Online]. Available: http://arxiv.org/abs/1212.5701.

[34] P. Huo, Y. Yang, J. Zhou, C. Chen, and L. He, "TERG: Topic-Aware Emotional Response Generation for Chatbot," 2020, doi: 10.1109/IJCNN48605.2020.9206719.

[35] P. Huo, Y. Yang, J. Zhou, C. Chen, and L. He, "TERG: Topic-Aware Emotional Response Generation for Chatbot," 2020, doi: 10.1109/IJCNN48605.2020.9206719.