AirLogiBot: A Bilingual Chatbot for Customer Service in Airports

Taha M. Mohamed
College of Business, University of Jeddah,
Jeddah, Saudi Arabia
tmmohamed@uj.edu.sa

Alaa A. Qaffas
College of Business, University of Jeddah,
Jeddah, Saudi Arabia
aaqaffas@uj.edu.sa

***Abstract*—Most airways companies have designed chatbots for answering questions related to their services. However, there is a lack of chatbots dedicated to customer service in airports. This paper proposes AirLogiBot, a transformer based chatbot intended to provide customer service in airports. The proposed chatbot answers customer inquiries and questions. Five chatbot models are simulated and compared: RoBERTa, DistilBERT, Rule-based for both English and Arabic languages, and AraBERT. Simulation results show that the performance of transformer based chatbots exceeds other models in accuracy, f-score, user intents, and many other metrics. RoBERTa excels in accuracy, precision and recall, while DistilBERT excels on the average response time. The performance of the English chatbot exceeds the Arabic chatbot. The response time of the rule based chatbots is generally less than that of the transformer based chatbots. AraBERT is very efficient in handling Arabic questions. However, Arabic based chatbots are still suffering from several limitations that affect their performance.**

Keywords— Arabic Logistics Chatbots; BERT chatbots; Airport Chatbots; and Customer Service

I. Introduction

Companies are seeking to offer friendly customer services and offer great advancements for business in digital revolution age and digitalization of services. These services can be carried out through intelligent chatbots [12]. Chatbots are available on 24/7/365 basis as virtual assistants to serve enterprise customers. They can understand user intent, manage context, and execute many complex tasks. So, the overall costs, human resources demand, and response times are reduced. These achievements let human experts focus on more important tasks in the enterprise [1]. Although Internet is more mature than chatbots in searching, customers often prefer using conversational chatbots to get additional information about products and services [19]. Additionally, chatbots could be used by different interfaces such as mobile phones, touch screens and physical robots in the service locations [12]. Statistics show that research on chatbots is growing significantly from 2020 to 2024 in many different countries worldwide, indicating the growing importance of this topic [14]. Chatbots are utilized in various industries such as e-commerce, education, and many business fields [11]. In business, customer service team is often suffering from crowding phone calls, WhatsApp messages, and emails from different customers and clients. Also, employees are bound by work hours only. However, customers often seek rapid and real-time updates about different logistics issues. So, the responses provided by human resources often decrease user satisfaction. Alternatively, conversational chatbots can participate in solving many customer service issues such as answering routine questions and performing routine tasks. Therefore, chatbots can enhance customer satisfaction by giving quick answers to customers' inquiries. The operational efficiency is also increased by responding to many customers simultaneously [11]. Consequently, the answer times of the human resources could be saved to answer sophisticated questions and perform other sophisticated tasks.

In airports, chatbots could be considered as the first-level customer support agents. They can significantly reduce passengers' waiting times and efforts by providing the necessary assistance to passengers [11]. They can also handle large amounts of customer queries related to baggage, tickets, gates, boardings, and many other airport logistics. This leads to improving customer satisfaction, decreasing headaches, increasing operational efficiency, and decreasing costs. However, there is limited research on using automated agents and chatbots in airports [11]. Straightforward customer service chatbots use deterministic rules to answer the questions. They are called rule-based chatbots. More advanced chatbots can use large language models (LLM) to answer more advanced questions [1]. These chatbots use the bidirectional encoder representations from transformers (BERT) models exploiting huge advancements in artificial intelligence (AI) and deep learning. These chatbots are called retrieval based chatbots [14]. In the two types of chatbots, enterprise databases and digital content could be combined with the models to answer more sophisticated questions.

However, although the great benefits of using chatbots in customer service, there are still some challenges. One of the challenges is the multilingual capability of the chatbot. Also, there are some challenges related to using chatbots in more than one language. For example, although the Arabic language is spoken by over 400 million native speakers, chatbot solutions for Arabic language are still in its infancy [3]. The Arabic language has many complexities and challenges [2]. Arabic sentences are not simple translation of English-language sentences. However, an understanding of Arabic's linguistic structure is crucial [5]. Another challenge is the hallucinations of some LLMs models such as generative pre-trained transformers (GPT). However, research is growing in this area for decreasing hallucinations and biasness of LLMs.

This paper proposes a new chatbot for logistics operations in airports called AirLogiBot. The proposed chatbot uses BERT based models and supports two languages (English and Arabic). The proposed chatbot saves customers' time in airport and perform many important tasks. The chatbot can foster the tourism and logistics industry and advances the automation in airports. To perform this task, we compare the performance of five chatbots: RoBERTa, DistilBERT, Rule-based for both English and Arabic languages, and AraBERT. The performance of each chatbot is evaluated and compared. Simulation results show that each of the simulated chatbot has its advantages and drawbacks. This rest of this paper is organized as follows: section II presents the necessary background of the paper including chatbots design, performance metrics, and challenges. Section III presents the previous related work of chatbots. Section IV explains the proposed AirLogiBot. Section V presents the simulation results and comparisons. Finally, the paper is concluded in section VI.

II. Background

In this section, we will overview some important concepts about chatbot design and the techniques used for getting responses from chatbots. Some important challenges of chatbot design are also discussed. Additionally, the section reviews some important metrics to assess the performance of chatbots.

A. Chatbot Design

Chatbots could be categorized into three classes. These classes include rule based, retrieval based, and generative based chatbots [13], [20]. Rule based methods searches for exact matching of questions and responses that may use algorithms such as the longest common subsequence techniques [7]. These algorithms use "if then" rules and decision trees for pattern matching to answer user questions. The chatbots are 100% safe as they do not generate any new text. Questions, keywords, and responses must be previously identified to the system. These chatbots are suitable for simple FAQ bots, menu-driven chatbots, and customer support with predefined paths. The major drawbacks of these chatbots are that the responses are often repetitive and predictable [20]. Also, these chatbots do not learn from data and fail when there is no matching for question.

Both retrieval-based and generative based chatbots use transformers and large language models (LLM) [13], [20]. Retrieval based chatbots use machine learning (ML) and natural language processing (NLP) to select the most relevant response from a predefined database. These chatbots do not generate any new text. Alternatively, they select the best response from a database based on classification results. These chatbots are suitable for customer support with semi-flexible questions and large knowledge bases chatbots [20]. They use word embeddings, sentence embeddings, and cosine similarity for prediction. The addition of new questions to the database is straightforward. They are suitable for Customer service FAQ chatbots even if the question form is changed with different synonyms [10]. BERT models are encoder only transformer models that understand text, not generating it which is very important for understanding user's intent detection [8].

In retrieval based chatbot, the user enters a question, and BERT tries to understand user intent. The database is searched for best answer based on prediction [23]. There are many examples of BERT models. Two common BERT models are RoBERTa and DistilBERT. RoBERTa is a configured BERT model for robustly optimized approach. It is trained using masked language modeling and next sentence prediction objective. Massive amount of text is used to train RoBERTa from Internet contents and data sources. The model proves superiority in many fields such as emotion detection tasks [23]. Knowledge distillation is exploited in DistilBERT. DistilBERT is trained by using very few features. This gives fast and good performance of the model. DistilBERT is smaller in size. So, it is preferred by many researchers that are seeking faster and accurate responses [23]. Both models are well implemented in the Hugging Face Transformers library [22]. However, the two above BERT models are English models. Their performances in Arabic are not excellent. However, AraBERT is a BERT model designed for Arabic language tasks. The model is optimized for Arabic vocabulary, tokenizer, dialects, morphology, and language representations [20]. Generally, all BERT models should be fine-tuned to the specific language tasks such as question answering, NER, and classification. The fine-tuning process is much faster than training the model from scratch.

The third chatbots class is the generative based chatbots (GPT) which give broader question answering and are more suitable for longer conversations and nonspecific domain questions [13]. They can understand context using self-attention and generate completely new sentences with multilingual support [10]. They have decoders only transformers instead on encoders transformers. However, these models are computationally expensive. Also, their responses are harder to control, and there is a risk of hallucinations and biasness [13]. They are designed to generate language and storytelling. So, chatbots based on GPT are not suitable for customer service that have predetermined answers.

There are some challenges still associated with chatbots. One of them is related to the language itself, especially Arabic. Arabic is a morphologically rich language having few resources compared to English. So, many NLP tasks in Arabic language are still more difficult than those of English. Such tasks include named entity recognition (NER), question answering, and sentiment analysis [8]. Additionally, task-oriented dialog systems are very efficient in English language. However, this is not the case for Arabic language due to the limited existence of Arabic datasets [9]. Another challenge is related to answering many users at the same time. Chatbots that can answer many users simultaneously are called polyadic chatbots. Moreover, generating human-like responses and personalization is another challenge [21]. Also, AI models may have bias towards certain dialects. Additionally, data privacy is also an important issue that the chatbot should comply with local laws and regulations.

B. Chatbot Performance Evaluation

The performance of chatbots is often measured by many performance metrics. These metrics may include usefulness, ease of use, intention prediction, and trust of answers [11]. Another metric is the perceived feeling of users about generated responses whether these responses are like human answers or not [17]. Accuracy, precision, recall, and F-Score are also other important metrics. The response time measures the time between sending a question and receiving a response. The chatbot throughput includes successfully processed conversations. Herein, we will show the computation of some of these metrics. The Throughput T could be computed from equation 1 as:

$$T = S / P \qquad (1)$$

where S is the total number of successful conversations. P is the total measurement period. For example, if the chatbot successfully handles 2000 complete conversations in 60 minutes, its throughput is throughput = 2000 conversations / 1 hours = 2000 conversations per hour. The overall accuracy of the chatbot could be expressed using equation 2 as:

$$Accu = (C / Q) . 100 \qquad (2)$$

Where $Accu$ is the Accuracy. C is the number of correct responses; Q is the total number of questions asked. The Precision, Recall, F-Score is given by equations 3, 4, and 5 as:

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

$$Recall = \frac{TP}{TP+FN} \qquad (4)$$

$$Fscore = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \qquad (5)$$

Where TP is the number of true negatives. FN is the number of false negatives. The intent accuracy is given by Equation 6 as:

$$I = (C_i / U) . 100 \qquad (6)$$

Where I is the intent accuracy. C_i is the number of correctly identified intents. U represents the total number of user utterances. The Task completion rate is given by Equation 7 as:

$$TC = (SC / N_i) . 100 \qquad (7)$$

Where T_C is the Task completion rate. S_C is the number of successfully completed tasks. N_i represents the total number of initiated tasks. The error rate is given by equation 8 as:

$$E_r = (Ne / N_r) . 100 \qquad (8)$$

Where E_r is the error rate, N_e is the number of errors responses, whereas N_r represents the total number of user requests. Finally, the average response time is given by Equation 9 as:

$$Avg_r = S_r / T_r \qquad (9)$$

Where Avg_r is the average response time. S_r is the sum of all response times. T_r is the total number of responses.

III. LITERATURE REVIEW

In this section we will review the related work of adopting chatbots in airports. The authors in [4] present chatbots framework that integrates data from different enterprise systems such as ERP and other systems. Their chatbot can provide answers and execute actions. The authors in [6] discuss in detail the opportunities and challenges of using Industry 4.0 technologies on SCM. They show that chatbots can be used extensively in customer services and other logistics domains. The authors in [16] address the challenges of SCM chatbots. They show the promising landscape of AI integration in SCM and logistics domain. The authors in [13], explore the principles and technologies of chatbot design. They also show chatbot applications in many fields such as interviews, education, and healthcare. The authors also discuss some challenges of chatbot design including privacy and hallucinations issues in responses [13].

The authors in [15] develop AI-driven chatbot for real-time news automation. The chatbot can improve news summarization. The average $F1$ score of the chatbot is 0.94 for summarization. Also, for correlation analysis, the chatbot achieves $F1$ score of 0.92. The average response time of the query is 9 seconds. The authors in [10] evaluate the performance of five major AI chatbots. The performance is evaluated by using a set of essays and multiple-choice questions. Their results show that GPT-4 achieves the best performance. The authors in [14] conduct a literature review of chatbot research classified by various domains. They show the growing concentration of shifting chatbot research towards AI chatbots. The progress is shifted from rule-based to advanced retrieval and generative chatbots [14].

The authors in [3] and [8] address the problems and challenges related to Arabic chatbots. These problems are related to language issues such as orthography, morphological complexity, diacritics, and diglossia. Additionally, the authors present a systematic literature review on Arabic chatbot challenges. They survey academic papers from different database sources from 2000 to 2023. They show that the complexity of Arabic led many researchers to adopt rule-based chatbots approaches. Human-based evaluation metrics are used to assess chatbot performance due to the Arabic complexities

[3]. In [8], the authors compare the performance of AraBERT to multilingual BERT. Their results show that AraBERT achieved high performance on many Arabic tasks. The authors in [9] introduce the AraConv which is the Arabic task dialog system. They used multi-lingual transformers. Their results indicate that the performance is compared to other languages including English and Chinese.
The authors in [11] address the problem of using chatbots in airports by evaluating passengers' acceptance to this technology by using technology acceptance model. They used some evaluation metrics such as perceived trust, ease of use, perceived usefulness, and perceived enjoyment. The authors in [12], discuss design and implementation of a customer support chatbot in Venice Airport. They also develop a working prototype of the intended chatbot. The authors in [17], use machine learning techniques to identify whether a text is generated by human or chatbots. They use random forest prediction that achieves 88% accuracy [17]. Interestingly, their results indicate that the generated text by chatbots is still comparable to the generated text by human when answering user questions and inquiries.

The authors in [18] compare the responses generated by AI chatbots versus the responses generated by search engines in product purchasing domain. So, the authors present some insights for fostering purchasing decisions through AI chatbots [18]. The authors in [19] compare perceptions of conversational chatbots with the Internet searching from consumer perspective. Their findings show that the internet is still perceived as better than chatbots from many dimensions. In [21], the authors review some chatbot responses challenges including personalization and human-like responses. In [22], the authors compare the performance of some BERT models. Their results show the superiority of RoBERTa compared to other BERT models in many accuracy metrics. The superiority of RoBERTa is also shown in [23].

IV. Proposed AirLogiBot Architecture

In this section, we will discuss the proposed AirLogiBot chatbot. AirLogiBot uses BERT to understand user intent and generate the appropriate response to airport customers from the database. To recommend the best BERT model, we assess the performance of three models: RoBERTa, DistilBERT, and AraBERT. Also, rule-based from both English and Arabic languages are also evaluated and compared. Meaning that, in total, five models are compared. In the beginning, the algorithm starts by detecting the language of the inquiry (Arabic / English) asked by the customer. If English is detected, then either RoBERTa or DistilBERT is chosen for determining user intents. Alternatively, AraBERT is chosen when Arabic is detected. For performance comparison, the same inquiry is forwarded to a rule-based system (either Arabic or English).
BERT based chatbots generate embeddings from the text. All database texts are converted into embeddings using BERT and the embeddings are stored in a database. When a customer asks a query, the query is also converted to embeddings by BERT. Then, cosine similarity is computed with the database to find the most similar vector. Then, the appropriate response is retrieved. However, when rule based chatbots are used, the classifier tries to extract the entity and intent from customer inquiry. Then, the appropriate response is generated by matching.
To illustrate the details of AirLogiBot, in the beginning, a customer in an airport needs to answer a query. For example, the customer may ask about the times of certain flights, gate direction, baggage allowance, or any other service in the airport. Also, the customer may need to perform an action by chatbot such as making or modifying bookings. So, this inquiry passes through several steps before generating a response or action. These steps are shown in Figure 1.

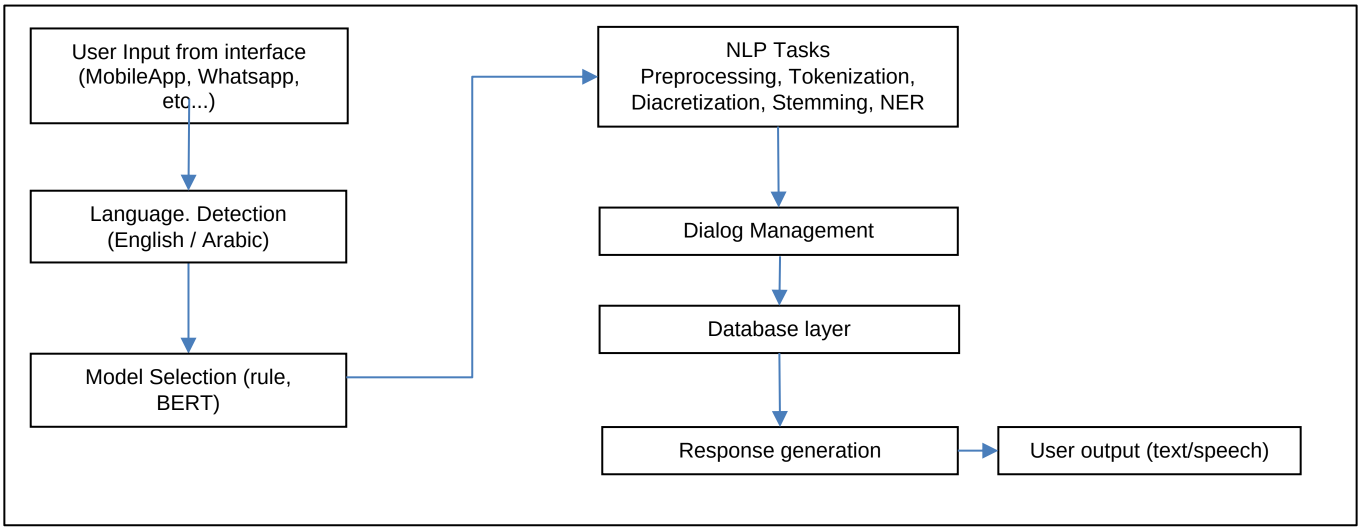

Fig. 1: Flow of the proposed chatbot (AirLogiBot)

First, the inquiry is asked through an interface such as mobile application, WhatsApp message, or any other types of user interfaces. The query may be asked in the form of text or audio. Moreover, the queries may be asked in bilingual (either English or Arabic). Next, this inquiry is passed on to the NLP layer. The main target of NLP processing is to understand user intents. This layer includes many NLP tasks such as preprocessing, tokenization, diacretization, stemming, and named entity recognition (NER). Text preprocessing may also include text normalization, especially for Arabic language. That is, in Arabic, there are different forms of the letter *Alef* and *Hamza*. These different forms should be normalized to ensure that they are referring to the same letter. Tokenization is the splitting of the text into words to simplify handling of words. Diacritization is an optional task which is the process of restoring the diacritical marks. Also, lemmatization (stemming) is used to reduce words to their lemma or root. The root of a word is a simplified form of the word that helps to identify user intent. NER is a custom-trained model to identify and classify the intent domain from Arabic text. For example, if a customer asks, "where is my bag?", then, the system identifies the intent as the baggage entity (Baggage Table).

Next, the dialog manager delivers the identified intent to the appropriate database tables. Such tables may include information about flights, navigation, gates, shopping, or other entities. It manages the conversation's flow and context. It determines the next response, or action, based on the intent classified from the previous stage. The dialog manager should also care about the current state of the dialog. For example, if the user asks about his/her bag, and he/she doesn't enter the bag number from the previous NER state, then the chatbot should ask about bag number to complete the conversation. So, the model is selected to answer user's inquiry.

Although rule-based models are simple, fast, and straightforward, the model doesn't learn from data. It just checks if user messages match known templates, words, or sentences. For example, if the user searches for something related to "flight" and "status" then the rule-based model classifies this query as flight status. The search in the database is on the flight table in this case. Also, there is no context attention of the query. However, the model fails when the question is asked with different words. Multi intent queries are not handled accurately. This occurs when a query has more than one intent. For example, a query may ask about baggage for certain flight.

BERT based models are efficient. The responses are generated based on the predicted intents and entities. The model can handle ambiguous queries better than rule-based chatbots. BERT based chatbots integrate reinforcement learning and context tracking to optimize dialogs. Also, the model connects the context of the current dialog with the previous dialogue turns by using the attention property of the transformer. In this case, the prediction of intents and entities is perfect. So, the generated response of the model seems to be natural response like human responses. Moreover, the generated responses are personalized according to the context of the dialog. The ability of the model to perform clarification and reasoning is a large advantage. Due to these advantages, we propose the AirLogiBot to be based on BERT rather than rule based chatbots. This choice is supported by the proposed simulation results as shown in the following section.

V. Simulation and Performance Comparison

To simulate AirLogiBot performance, we conduct a simulated logistics environment for airport data. In the simulation, we simulate 3000 sessions for 3000 users. Each session has about 4.1 turns on average. The turn is a part of the dialog between customer and chatbot. So, each user asks about 4.1 queries. Therefore, the total number of turns equals 12300 turns for all customers. These queries are stored in a database table. Each row of the table contains a user query that is also called user utterance. For each query, the true goal intent is stored for evaluating the performance of the chatbot. Additionally, relevant entities are stored for each intent. For example, if a user asks about the status of a certain flight (intent), then this intent may be associated with a relevant flight number and flight date. The database is provided by bilingual (Arabic / English). Depending on the language of the query (Arabic / English), the chatbot model is chosen from the five tested models mentioned above. These models are rule based and BERT based. Moreover, we generate another 2500 user queries for testing the performance of the models.

Table 1: Examples of possible questions categorized by domain in airports

Entity	Sample Question
Flights	Show me details for Saudi Arabia Airlines SV 303
Gates	Why was the gate of my flight changed?
Baggage	One of my bags didn't arrive, what is the correct procedure to get my bag?
Shopping	What is the best place to buy souvenirs in the airport?
Car Rent	Could you tell me how to rent comfortable car?
Airport Navigation	Please give me a description of how to reach terminal 3

The predictions results are compared to the previously stored intents to assess the performance of the chosen model. If the intent is obtained successfully, then the prediction succeeds. Otherwise, the prediction fails if the predicted intent is incorrect. In our simulations, we simulate asking questions that are often asked in airports that include different entities. These entities may include flights, gates, baggage, facilities, car rent, and other facilities. We use some evaluation metrics to assess the performance of the proposed AirLogiBot. The Python Scikit-learn library is used for computing the metrics. Table 1 shows some of the simulated questions related to each entity.

Figure 2 shows the average response time of the tested models. From the figure, DistilBERT achieves the least response time in BERT models. RoBERTa achieves average response time of 45 milliseconds. AraBERT achieves the maximum response time due to Arabic complexities. Rule based chatbots achieve the least response times compared to BERT classes because inference is straightforward. Again, Arabic rule based chatbots have greater response times compared to English.

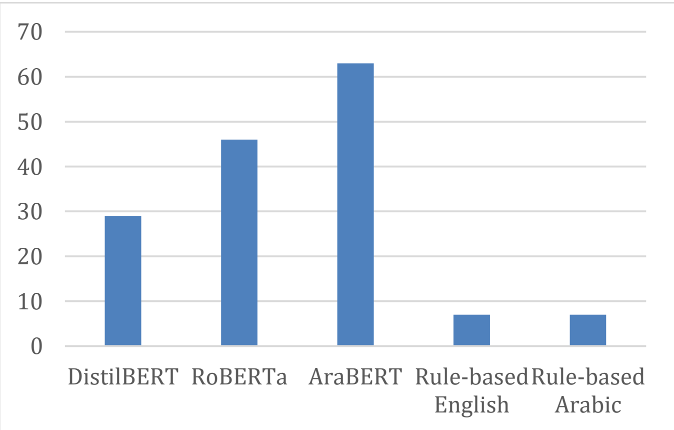

Fig. 2: Average response times of the tested models (milliseconds)

TABLE 2: MODELS PERFORMANCE COMPARISON

Model	Accu.	Prec.	Rec.	F-Sc.
RoBERTa	93.6%	93.2%	93.4%	93.7%
DistilBERT	89.1%	89.0%	87.7%	88.5%
AraBERT	88.4%	88.1%	86.4%	87.2%
Rule-based (English)	72.3%	69.0%	67.9%	68.7%
Rule-based (Arabic)	63.3%	61.1%	59%	59.7%

Table 2 shows the performance of the tested models. The first column represents accuracy. The second column represents precision. The third column represents the recall. The last column represents the f-score. It is clear from the table that the best accuracy is achieved by RoBERTa that exceeds 93%. This is the case for the rest of metrics. DistilBERT achieves less performance compared to RoBERTa. The performance is degraded by about 5% in f-score. When using AraBERT on Arabic inquires, the metrics also decreased compared to both RoBERTa and DistilBERT. However, the performance of the rule based chatbots is degraded on both English and Arabic chatbots. This may be explained as that any variation of the morphology or context of the inquiry doesn't match with the true intents. The accuracy of the rule based Arabic chatbots is about 63% which is not good for customer service airport chatbots. A graphical comparison of the performance of the five tested chatbot models is shown in figure 3. From the figure it is shown that the best performance is achieved by using RoBERTa model. Both DistilBERT and AraBERT are comparable in performance. Rule based chatbot models achieve the least accuracy and f-score. Rule based English chatbots achieve better performance compared to Arabic rule based chatbots.

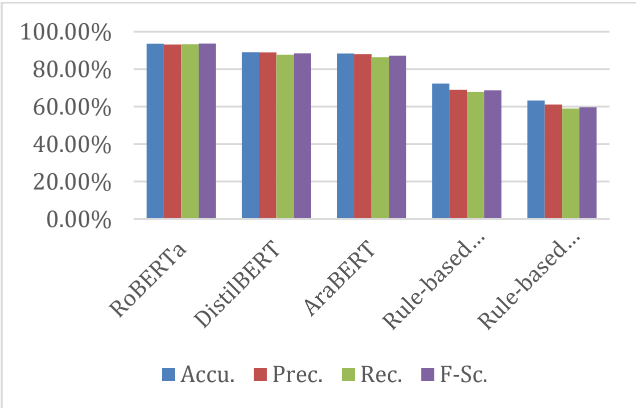

Fig. 3: Graphical comparison of the five tested chatbot models

The previous results show some conclusions. Performance results of Arabic chatbots show some challenges. Arabic has many dialects that lead to more ambiguity. Also, tokenization in Arabic is more complex because Arabic words have many morphologies and variations. Also, when Arabic words are tokenized, many sub words resulted which leads to increased response time. RoBERTa has more computational load compared to DistilBERT. Therefore, RoBERTa is less suitable for real-time chatbot applications. Rule-based chatbots are more predictable and deterministic. They are suitable for predefined patterns and responses. However, rule based chatbots are not efficient in synonyms and paraphrasing of words. In fact, they are not suitable for customer service in airports with different queries variations. RoBERTa has the highest accuracy and F score. So, it is the best for customer support systems. DistilBERT is quicker and faster. AraBERT is much better than Arabic rule-based and can handle different word morphologies. Rule-based Arabic chatbots are the worst in performance. Also, rule-based English chatbots are the worst English chatbot in performance. Overall, using chatbots is expected to reduce calls significantly as the achieved accuracy is about 93% in RoBERTa. So, customer service is enhanced significantly by using chatbots in airports. Additionally, many of the customer efforts, delivery times, and movements are decreased or totally cancelled. This leads to increased customer satisfaction. Finally, the efforts of the customer service team could be saved for other critical tasks instead of just responding to customer inquiries.

VI. Conclusion

This paper proposes a bilingual chatbot, called AirLogiBot based on BERT, that is intended for customer service in airports. The chatbot can answer customer inquiries from a database and uses RoBERTa model for answering user inquiries in English language. The chatbot also uses AraBERT model for answering user inquiries in Arabic. After simulating the performance of five chatbot models, the simulation results show the superiority of BERT based chatbots compared to rule based chatbots in accuracy and f-score. However, rule-based models excel in response time compared to BERT based models. Additionally, simulation results show a lot of challenging issues that faces Arabic chatbots compared to English chatbots. Overall, using chatbots in airports can minimize passenger delays, save efforts, and optimize logistic resources. Also, the efforts of customer service team could be directed for other critical tasks instead of just responding to simple customer inquiries.

Acknowledgment

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-23-SHR-47). Therefore, the authors thank the University of Jeddah for its technical and financial support.

References

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

[2] Alhuwaydi, A. M., "Challenges and Opportunities for Arabic Natural Language Processing", International Journal of Advanced Computer Science and Applications, 12(7), 2021.

[3] Soufiyane Ouali, Said El Garouani El Garouani, "Arabic Chatbots Challenges and Solutions: A Systematic Literature Review", Iraqi Journal for Computer Science and Mathematics, 5(3), pp.:128-169, 2024.

[4] Winkelmann, S., Kärkkäinen, H., & Lahtinen, T., "Chatbots in the supply chain: A framework for implementation", Proceedings of the European Conference on Information Systems (ECIS), 2020.

[5] Wang, L., Ma, C., Feng, X. et al. "A survey on large language model based autonomous agents", Front. Comput. Sci. 18, 186345, 2024.

[6] Taha M. Mohamed, Alharbi, A., Alhassan, I., Kholeif, S, "Applications of Industry 4.0 on Saudi Supply Chain Management: Technologies, Opportunities, and Challenges", In Machine Learning and Data Analytics for Solving Business Problems", Book Chapter, Springer, 2022, Cham. https://doi.org/10.1007/978-3-031-18483-3_10

[7] T. Mahdy, H. El-Mahdy, and H. Onsi, "Efficient Watermark Detection by Using the Longest Common Substring Technique", the Egyptian Informatics Journal, Elsevier, Vol. 12, Issue 2, July 2011. http://www.sciencedirect.com/science/article/pii/S1110866511000193

[8] Antoun, W., Baly, F., and Hajj, H. “AraBERT: Transformer-based Model for Arabic Language Understanding.”, Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection", Marseille, France, May 2020.

[9] Fuad, A., & Al-Yahya, M., "AraConv: Developing an Arabic Task-Oriented Dialogue System Using Multi-Lingual Transformer Model", Applied Sciences, 12(4), 2022.

[10] Firuz Kamalov, David Santandreu Calonge, Patrik T. Hultberg, Linda Smail, "Comparative analysis of leading artificial intelligence chatbots in the context of entrepreneurship", Journal of Innovation and Entrepreneurship, 14(1), June 2025

[11] Auer, I., Schlögl, S., & Glowka, G. "Chatbots in Airport Customer Service—Exploring Use Cases and Technology Acceptance", Future Internet, 16(5), 2024

[12] Carisi, M., Albarelli, A., Luccio, F.L. "Design and implementation of an airport chatbot", In Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good, Valencia, Spain, September 2019, pp. 49–54.

[13] Yigit, G., Bayraktar, R., "Chatbot development strategies: a review of current studies and applications", Knowl Inf Syst, 67, 2025, pp. 7319–7354.

[14] Nazruddin Safaat Harahap, Aslina Saad Author, Nor Hasbiah Ubaidullah, "Comprehensive Bibliometric Literature Review of Chatbot Research: Trends, Frameworks, and Emerging Applications", International Journal of Advanced Computer Science and Applications (IJACSA), Volume 16 Issue 1, 2025, DOI: 10.14569/IJACSA.2025.0160185

[15] Sufi, F., & Alsulami, M. "AI-Driven Chatbot for Real-Time News Automation", Mathematics, 13 (5), 2025

[16] Robert Glenn Richey Jr., Soumyadeb Chowdhury, Beth Davis-Sramek, Mihalis Giannakis, Yogesh K. Dwivedi, "Artificial intelligence in logistics and supply chain management: A primer and roadmap for research", Journal of Business Logistics,

44(4), September 2023, https://doi.org/10.1111/jbl.12364

[17] Mohammed Salah Ibrahim, Jabbar Abed Eleiwy, Hassan Mohamed Muhi-Aldeen, Yusra Al-Yasiri, Ahmed Adil Nafea, "Human to Chatbot Text Classification Using Multi-Source AI Chatbots and Machine Learning Models", Journal of Intelligent Systems and Internet of Things Vol. 16, No. 01, PP. 152-165, 2025, DOI: https://doi.org/10.54216/JISIoT.160113

[18] Kim, S., & Priluck, R., "Consumer Responses to Generative AI Chatbots Versus Search Engines for Product Evaluation", Journal of Theoretical and Applied Electronic Commerce Research, 20(2), 2025, https://doi.org/10.3390/jtaer20020093

[19] Tafesse, W., & Mamo, Y. "A comparison of conversational chatbots and the internet for consumer information search", Behaviour & Information Technology, 1–18, 2025.

[20] Shafquat Hussain, Omid Ameri Sianaki, and Nedal Ababneh, "A Survey on Conversational Agents/Chatbots Classification and Design Techniques", Advances in Intelligent Systems and Computing, Springer, March 2019

[21] Shurooq Almutairi, Sana A. Khan, Mohammad Amin Kuhail, Imran Taj, "Chatbot Design Challenges and the Effect on User Behavior", In book: Trends, Applications, and Challenges of Chatbot Technology, pp.24-35, , February 2023, DOI:10.4018/978-1-6684-6234-8.ch002

[22] Very Sugiarto, Alvionitha Sari Agstriningtyas, Yus Mochammad Cholily, "Performance Comparison of BERT, ALBERT and RoBERTa for Sentiment Analysis in Critical Pilot Communication Prior to Aviation Accident", International Journal of Artificial Intelligence and Information Technology (IJAIIT), Vol. 1, No. 1, 2025.

[23] Arslan Ahmed, Barkha, Tushar Kanjwani, Shahzad Nasim, "Comparative Analysis of DistilBert, XLNET, RoBERTa & BigBird for Emotion Detection in Conversational Text", ILMA Journal of Technology & Software Management (IJTSM), Vol. 6 Issue 1, 2025.