IJAIT (International Journal of Applied Information Technology)
http://journals.telkomuniversity.ac.id/ijait/

Conversational AI for Enhancing English Speaking Proficiency: A Mobile App

Areeg Fahad Rasheed [a, *], M. Zarkoosh [b]

[a] *Network Dept, Al-Nahrain University, Baghdad, Iraq*
[b] *Independent Researcher, Baghdad, Iraq*
areeg.fahad@coie-nahrain.edu.iq, mzarkoosh@gmail.com

ARTICLE INFO

Received December 19th, 2023
Revised May 6th, 2024
Accepted May 6th, 2024
Available online March 11th, 2025

Keywords
Conversational AI, Chat GPT, DialoGPT Blenderbot, machine learning

ABSTRACT

As one of the most widely spoken languages globally, English plays a pivotal role in academia, digital communication, and numerous professional domains. English proficiency is vital for academic pursuits, international communication, and careers. The profound impact of artificial intelligence on our daily lives is exemplified by techniques such as chatbots and speech recognition. Leveraging these techniques (i.e., Chatbot and speech recognition), the paper aims to create a user-friendly mobile app to enhance English-speaking proficiency at minimal cost. This study employs three conversational AI models- ChatGPT, Blenderbot, and DialoGPT as chatbots, and uses Google's speech-to-text/text-to-speech models for voice recognition and generation. The developed app and the used models are evaluated through questionnaires and response time measurements, highlighting user satisfaction with the system's realism, accuracy, and user-friendliness. ChatGPT and Blenderbot emerge as top performers in response quality, while DialoGPT excels in response speed.

* Corresponding author at:
Al-Nahrain University, College of Information Engineering, Network department
Country. Baghdad, Iraq
E-mail address: areeg.fahad@coie-nahrain.edu.iq

ORCID ID:
First Author: 0000-0002-7078-6200

http://dx.doi.org/10.25124/ijait.v8i1.6960
Paper_reg_number IJAIT000080105 2024

1. Introduction

English stands as the world's most widely spoken language and holds the status of being the official language in numerous nations. It serves as the primary language in virtually all digital resources, dominates academic discourse, and is integral to university applications [1], [2]. Furthermore, English is the chosen medium for cross-country communication and finds applications in various professional domains. Mastering the art of learning and speaking English is crucial for obtaining scholastic degrees and pursuing higher education opportunities abroad. Proficiency in the language is equally vital for international travel and conducting global business affairs [3].

Proficiency in English speaking constitutes one of the four essential language proficiencies, along with writing, reading, and listening. It is essential for any job application and professional engagement. It is the first impression you make on potential employers and colleagues, and it can have a significant impact on your chances of success [4] Strong speaking skills can also give you a competitive edge in other areas of your career, such as networking, negotiating, and giving presentations. Additionally, they can help you develop your leadership skills and become more influential in your workplace [5][6][7]

In recent years, we have witnessed a revolution in Artificial Intelligence (AI) that has significantly impacted various fields of our lives [8]-[11]. This transformation has been particularly pronounced in areas such as healthcare, cybersecurity, etc [12]–[14]. One notable application of AI in this revolution is the development of chatbots. Chatbots have emerged as powerful tools with the ability to assist and answer a wide range of questions [15]. This innovation has had a profound effect on the lives of many individuals, making access to information easier and more efficient than ever before. ChatGPT is one of the most versatile chatbot systems, capable of performing a wide range of tasks and providing responses [16]. Additionally, over the years, several other chatbot models have been introduced such as blenderbot from Facebook [17] and DialoGPT from Microsft [18], further expanding the capabilities of AI-driven conversations.

Numerous programs are available for users to enhance their English language skills, and most of the ones that offer a high-quality learning experience come with a price tag [19]. Therefore, the main objective of this paper is to develop a comprehensive system that combines the power of conversational AI and speech recognition technologies by building a user-friendly mobile app that helps people improve their English-speaking skills almost free. In the following the main contributions of the works. Contributions of the works are as follows.

- Integration of Conversational AI Models: The paper introduces a mobile app that seamlessly integrates three state-of-the-art conversational AI models: ChatGPT, Blenderbot, and DialoGPT. This integration allows users to interact with these models to improve their English-speaking skills.
- User-Friendly Mobile Interface: The authors have designed a user-friendly mobile app interface that simplifies the user experience. The interface allows users to select their preferred AI model, adjust voice preferences (including gender and speed), and engage in real-time conversations. The simplicity of the interface enhances the app's accessibility for users seeking to improve their English proficiency.
- User Evaluation and Performance Metrics: The paper conducts user evaluations and performance measurements to assess the effectiveness

of the mobile app. User feedback is gathered through questionnaires, which provide insights into user satisfaction, interface usability, and the perceived benefits of using the app to enhance language proficiency.

The rest of the paper is organized as follows: In Section 2, we explain the conversational AI models used in this study. Section 3 details the system architecture and the workflow, from sending the voice to receiving and playing the response, with the main components of the mobile app. Section 4 covers the evaluation of the mobile app, and Section 5 concludes of the paper.

2. Conversational Al Models Description

We have employed three conversational AI models to develop the app: ChatGPT from OpenAI, BlenderBot from Facebook, and DialoGPT from Microsoft. In the following sections, we will briefly explain each model.

1. ChatGPT: ChatGPT is a natural language model that has been instructed to provide human-like responses [15], [20]. It accepts text input and generates a variety of answers based on the instructions in the input text. These responses can range from answering inquiries and supplying code to solving arithmetic problems and delivering various other services. ChatGPT is available in several versions, each trained on a dataset of a different size, providing a wide range of response capabilities. In this work, we utilized a free trial of the GPT-3.5-turbo version of ChatGPT, which contains 175 billion parameters.
2. Blenderbot-400M-distill: It is the distilled version of Blenderbot from Facebook, boasting 400 million parameters. This model falls under the category of Retrieval-Augmented Generation, where 'retrieval' refers to the model providing a response based on existing results. At the same time, 'generation' refers to the model's capability to create responses from scratch, leveraging its language modelling capabilities. BlenderBot is trained on a massive dataset of text and code and can be used for various tasks [17].
3. DialoGPT is a conversational language model based on the GPT architecture. DialoGPT is trained on a massive dataset of 147M multi-turn dialogues extracted from Reddit discussion threads. This dataset includes various topics, from casual conversations to debates on controversial issues. DialoGPT can generate text that is both informative and engaging [18].

3. System Overview and App Components

This section is divided into two parts. The first illustrates the model architecture and outlines the flow of data between the client and the server. In the second, we delve into the design of the mobile app, providing an in-depth overview of its key components.

3.1. System Overview

As mentioned in the preceding section, the primary objective of this work is to design and implement an app based on conversational AI with the explicit aim of enhancing English language proficiency. The architectural framework for this system is illustrated in Figure 1. The first step in our model involves converting

user-generated voice input into text format. We leverage Google's advanced voice-to-text conversion service [21]. Subsequently, the recognized text undergoes further processing and is dispatched via an API to the server. Alongside this text, we send the name of the selected model. The text is added to the conversation history on the server side and tokenized using the appropriate model. It is then fed to the model to generate the output. The resulting output is consolidated, sent to the client side, rejoined, and then passed to the recognition model for conversion into voice. This voice is played, and the corresponding text is displayed on the screen.

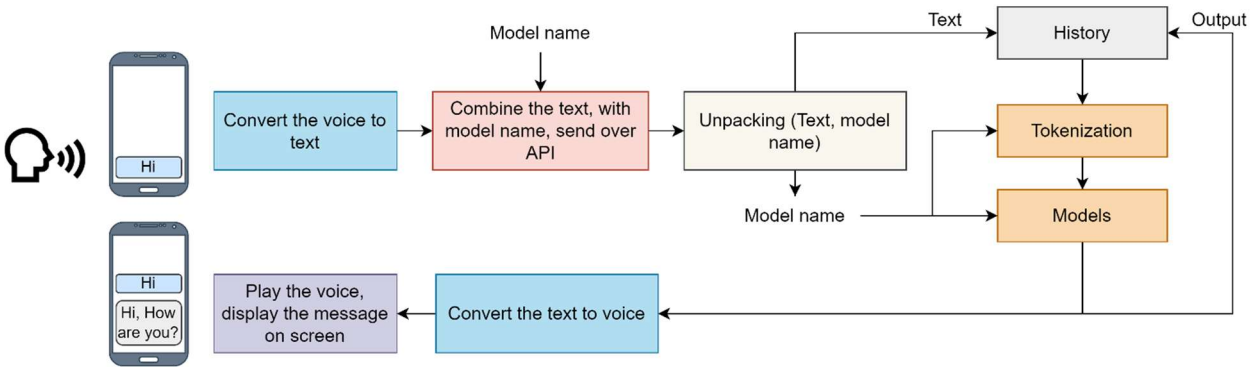

Figure 1 System Architecture

3.2. Apps' Components

We have developed a straightforward mobile app to facilitate the use of our proposed system. When designing the mobile app, we strongly emphasised user-friendliness and simplicity. As a result, we have kept the interface clean and intuitive. The mobile app, depicted in Figure 2, consists of two screens. The first screen (Figure 2a) displays user and model messages and a microphone icon for voice recording. The second screen (Figure 2b) presents multiple choices for the user. On the second screen, users can choose their preferred model. We use the Chat GPT model by default, but users can choose any of the three available models. The system can seamlessly produce responses for the selected model without any issues. Another option available to users is voice output. By default, Android Studio supports a female voice for text-to-speech, but we have also incorporated a male voice option. Additionally, users can adjust their talking speed. This feature is essential, as some individuals may need a slower pace to comprehend spoken words better. We have included a higher-speed talking option for those who prefer a faster pace.

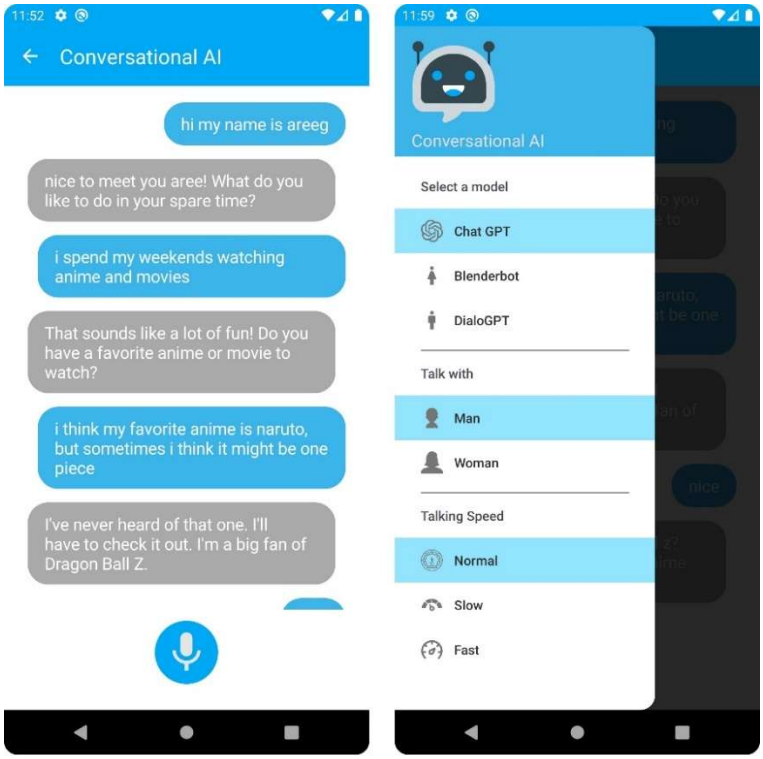

Figure 2 Mobile App Screens: (a) The main screen displays messages sent and received as text. (b) The options screen allows users to select their preferred model, talking speed, and voice-over gender.

4. Mobile app and model evaluation

We have created a form with multiple questions to evaluate the proposed system and determine its effectiveness. These questions cover various aspects, including system performance (accuracy, realism, speed), mobile usability, user interface friendliness, and how the system improves English skills. Collect questions based on the most popular questions used to evaluate mobile apps [22]. All the questions are listed along with the corresponding evaluation percentages results in Table 1. 15 users are involved in evaluating this app. Concerning accuracy and appropriate responses, the chatGPT model shows high performance and the highest accuracy results, while the lowest one is DialogGPT. Considering technical issues, user-friendliness, and recommendations to other users, the results show high performance and positive feedback concerning all questions.

Table 1 User Feedback on Conversational Models and Mobile App Performance

Questions related to model performance	DialoGPT (%)	Blenderbot (%)	ChatGPT (%)
Rate the model in terms of realism and accuracy in responses.	48	82.5	91
How satisfied are you with the speed at which the models respond?	90	72	84
Have you encountered instances where the models provided inaccurate or inappropriate responses?	40	23	10
Questions related to the app's performance			
How user-friendly is the app's interface?		94	
Did you encounter any technical issues or glitches while using the app?		7	
Is the app a good choice for learning and improving speaking proficiency?		81	
How likely are you to recommend this app to others based on your overall experience?		79	

In addition to evaluating user feedback, we have also measured the time it takes to receive responses from the models. This time is calculated from when the text message is tokenized and fed to the model until the response is received. Figure 3 illustrates the results of this analysis. The performance of the DialogGPT model is the lowest. Its response time is approximately one and a half times faster than that of the ChatGPT model and three times faster than that of the BlenderBot model. In contrast, the BlenderBot model has demonstrated the highest response time compared to the other models. While it may offer certain advantages regarding its capabilities, it is essential to consider its relatively longer response time, which may impact real-time interactions.

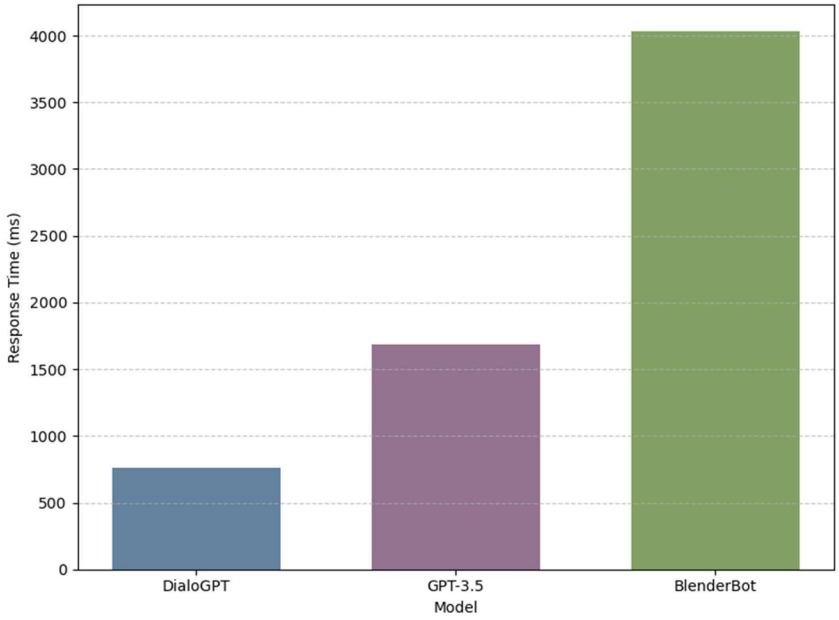

Figure 3 Comparison of Response Times for Different Models (msec)

5. Conclusions

This paper introduces a mobile app that employs advanced AI techniques to significantly enhance English-speaking skills. Our system leverages Google's voice recognition system to convert recorded speech into text. Additionally, we use three state-of-the-art conversational AI chatbot models: ChatGPT-3.5 Turbo, Blenderbot, and DiloadGPT. These models generate responses based on the received text input. Our proposed system is implemented within a user-friendly mobile application developed using Android Studio. We conducted two separate assessments to evaluate the proposed system and mobile application. The first involved administering questionnaires to users, while the second focused on measuring response times. Our results show that ChatGPT and Blenderbot consistently outperformed DiloadGPT regarding response realism and accuracy. Regarding response times, DiloadGPT exhibited the fastest response generation, followed by ChatGPT and then Blenderbot. DiloadGPT is approximately twice as fast as ChatGPT and three times faster than Blenderbot. Our user feedback indicates high-performance levels in terms of user friendliness user friendliness, and technical robustness. Overall, our proposed mobile app can significantly enhance English-speaking skills, and our results demonstrate the effectiveness of our AI techniques.

6. Limitations and Future Work

We utilized a free trial of ChatGPT, which offers a limited period for free usage. On the other hand, if they are willing to tolerate slower responses, they can use Blenderbot, which is entirely free with no usage limits, as mentioned earlier. For future work, we also recommend exploring the utilization of different models, such as ChatGPT-4 or BART, and comparing them with the models used in this paper. To enhance the mobile experience, consider adding features like login functionality and a history feature.

Bibliography

[1] M. F. Patel and P. M. Jain, *English language teaching: (methods, tools & techniques)*. Jaipur, India: Sunrise Publishers & Distributors, 2008.

[2] Abdulqader Talib Naeem and Juma'a Q. Hussein, "Difficulties of Understanding English Breaking News Headlines by Iraqi EFL Learners at the University Level: عبد القادر طالب نعيم, و جمعة قادر حسين," *JCEW*, vol. 32, no. 4, pp. 1–21, Dec. 2021, doi: 10.36231/coedw.v32i4.1546.

[3] L. T. Schächinger Tenés, J. C. Weiner-Bühler, L. Volpin, A. Grob, K. Skoruppa, and R. K. Segerer, "Language proficiency predictors of code-switching behaviour in dual-language-learning children," *Bilingualism*, vol. 26, no. 5, pp. 942–958, Nov. 2023, doi: 10.1017/S1366728923000081.

[4] J.-M. Dewaele and L. M. Pavelescu, "The relationship between incommensurable emotions and willingness to communicate in English as a foreign language: a multiple case study," *Innovation in Language Learning and Teaching*, vol. 15, no. 1, pp. 66–80, Jan. 2021, doi: 10.1080/17501229.2019.1675667.

[5] J. Currie *et al.*, "A scoping review of clinical skill development of preregistration registered nurses in Australia and five other English-speaking countries," *Journal of Clinical Nursing*, vol. 32, no. 1–2, pp. 283–297, Jan. 2023, doi: 10.1111/jocn.16239.

[6] A. Syaepul Uyun, "Teaching English Speaking Strategies," *JELL*, vol. 6, no. 1, pp. 14–23, Jun. 2022, doi: 10.31949/jell.v6i1.2475.

[7] H. Pallathadka, E. H. Ramirez-Asis, T. P. Loli-Poma, K. Kaliyaperumal, R. J. M. Ventayen, and M. Naved, "Applications of artificial intelligence in business management, e-commerce and finance," *Materials Today: Proceedings*, vol. 80, pp. 2610–2613, 2023, doi: 10.1016/j.matpr.2021.06.419.

[8] H. Al-Sahaf *et al.*, "A survey on evolutionary machine learning," *Journal of the Royal Society of New Zealand*, vol. 49, no. 2, pp. 205–228, Apr. 2019, doi: 10.1080/03036758.2019.1609052.

[9] Areeg Fahad Rasheed, M. Zarkoosh, Safa F. Abbas, and Sana Sabah Al-Azzawi. "Arabic offensive language classification: Leveraging transformer, lstm, and svm." In 2023 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), pp. 1-6. IEEE, 2023.

[10] Areeg Fahad Rasheed, M. Zarkoosh. "Mashee at SemEval-2024 Task 8: The Impact of Samples Quality on the Performance of In-Context Learning for Machine Text Classification." *arXiv e-prints* (2024): arXiv-2406.

[11] Areeg Fahad Rasheed, M. Zarkoosh, and F. R. Elia. "Enhancing graphical password authentication system with deep learning-based arabic digit recognition." *International Journal of Information Technology* 16.3 (2024): 1419-1427.

[12] A. R. Talib and H. M. Ali, "Deep Learning-based CAD System for Predicting the COVID-19 X-ray Images," *Karbala International Journal of Modern Science*, vol. 9, no. 3, Aug. 2023, doi: 10.33640/2405-609X.3316.

[13] Areeg Fahad Rasheed, M. Zarkoosh, and S. S. Al-Azzawi, "Multi-CNN Voting Method for Improved Arabic Handwritten Digits Classification," in *2023 9th International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia: IEEE, Aug. 2023, pp. 205–210. doi: 10.1109/ICCCE58854.2023.10246065.

[14] Areeg Fahad Rasheed, M. Zarkoosh, and S. S. Al-Azzawi, "The Impact of Feature Selection on Malware Classification Using Chi-Square and Machine Learning," in *2023 9th International Conference on Computer and Communication Engineering*

(ICCCE), Kuala Lumpur, Malaysia: IEEE, Aug. 2023, pp. 211–216. doi: 10.1109/ICCCE58854.2023.10246084.

[15] J. Deng and Y. Lin, "The Benefits and Challenges of ChatGPT: An Overview," *FCIS*, vol. 2, no. 2, pp. 81–83, Jan. 2023, doi: 10.54097/fcis.v2i2.4465.

[16] A. L. Alkhaqani, "Potential Benefits and Challenges of ChatGPT in Future Nursing Education," *Maaen Journal for Medical Sciences*, vol. 2, no. 2, Aug. 2023, doi: 10.55810/2789-9136.1020.

[17] S. Roller *et al.*, "Recipes for building an open-domain chatbot." arXiv, Apr. 30, 2020. Accessed: Dec. 19, 2023. [Online]. Available: http://arxiv.org/abs/2004.13637

[18] Y. Zhang *et al.*, "DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Online: Association for Computational Linguistics, 2020, pp. 270–278. doi: 10.18653/v1/2020.acl-demos.30.

[19] M. N. Hasnine and J. Wu, "Wordhyve: A context-aware language learning app for vocabulary enhancement through images and learning contexts," *Procedia Computer Science*, vol. 192, pp. 3432–3439, 2021, doi: 10.1016/j.procs.2021.09.116.

[20] M. Aljanabi, M. Ghazi, A. H. Ali, and A. Abed, "ChatGpt: Open Possibilities," vol. 4, no. 1, 2023.

[21] J. Choi, H. Gill, S. Ou, Y. Song, and J. Lee, "Design of Voice to Text Conversion and Management Program Based on Google Cloud Speech API," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA: IEEE, Dec. 2018, pp. 1452–1453. doi: 10.1109/CSCI46756.2018.00286.

[22] Pragadeesh60+ Mobile App Survey Questions to Ask," Surveysparrow.com, Available at: https://surveysparrow.com/blog/mobile-app-survey-questions/, Accessed on: 2024-2-14