

Design and Evaluation Framework for a Custom GPT Model for Teaching the Kuwaiti Arabic Dialect

Zahra A. Reda*, Sahar R. Hamzah, and Almothana M. Gasaymeh

Department of Education and Curricula, Arabian Gulf University, Manama, Kingdom of Bahrain
Email: z.reda89@hotmail.com (Z.A.R.); saharh@agu.edu.bh (S.R.H.); almothanamag@agu.edu.bh (A.M.G.)

*Corresponding author

Abstract—Due to advancements in generative Artificial Intelligence (AI), there is growing interest in using Large Language Models as conversational models in language learning. Although previous studies have shown the effectiveness of AI chatbots in accelerating second-language acquisition, most prior research focuses on English and other standard language varieties, with little attention given to Arabic dialects. This work-in-progress paper proposes the design of a custom Generative Pre-trained Transformer (GPT)-based learning system aimed at vocabulary acquisition and learner satisfaction in the Kuwaiti Arabic dialect, along with a proposed framework for evaluating the learning system. The custom GPT learning system will focus on conversational tasks centered on vocabulary learning, based on dialect-specific conversational tasks, and will provide interactive and context-specific vocabulary practice using the native language. The study adopts a design-based research approach that integrates instructional design principles with iterative system development, with the use of a mixed-methods evaluation framework. The study will be conducted on adult non-native women based in Kuwait who are learning Kuwaiti Arabic. Data collection will include vocabulary assessments, learner satisfaction surveys, semi-structured interviews, and platform usage analytics. The research aims to analyze and evaluate learning impacts, along with students' perceptions of learning environments that utilize artificial intelligence. By addressing the underrepresentation of Arabic dialects in AI-assisted language learning studies, this study aims to provide design insights and guidelines for GAI language learning that is culturally responsive.

Keywords—generative artificial intelligence, ChatGPT, Artificial Intelligence (AI) chatbots, Arabic dialect learning, Kuwaiti Arabic, non-native speaker

I. INTRODUCTION

The development and use of Generative Artificial Intelligence (GAI), especially Large Language Models (LLMs), are advancing rapidly, leading to increased interest in its application in educational contexts. Conversational Artificial Intelligence (AI) systems for learning a language have been used to produce interactive

dialogue, give immediate feedback, and provide a personalized learning experience. Research suggests that AI chatbots can help with vocabulary development, provide better learner engagement, and encourage language practice by letting learners interact with language in flexible and low-stakes ways [1–5]. As a result, AI-supported conversational systems are increasingly viewed as promising supplements to traditional language teaching.

In spite of this growing interest, existing research focuses mainly on English and other languages that are widely taught. Fewer studies have been devoted to teaching the Arabic language, and even fewer to teaching Arabic dialects. Arabic dialects differ substantially from Modern Standard Arabic (MSA) in terms of their vocabulary, structure and usage. Non-native learners need to be able to understand local dialects to integrate into society. AI-driven systems and digital learning resources for specific Arabic dialects are still quite limited.

In GAI-assisted language learning, Kuwaiti Arabic is a dialect that has received little attention. Learners of Kuwaiti Arabic are usually forced to rely on personal, informal exposure to the dialect since structured instructional materials are scarce. AI chatbots tend not to support dialect learning well because dialects are not widely included in training materials, and there is a general lack of structured instruction in dialects. This creates the need for AI systems that are intentionally designed to support learning dialect-specific vocabulary in contexts that are culturally relevant.

In response to this, this work-in-progress study focuses on the design and planned evaluation of a customized Generative Pre-trained Transformer (GPT)-based conversational learning system for teaching the Kuwaiti Arabic dialect to non-native learners. The proposed system is designed to integrate scenario-based dialogues with adaptive learning based upon individual learner responses, and with repeated usage of new vocabulary words personalized for specific learners. This study utilizes a design-based research approach that is supported by the Analysis, Design, Development, Implementation, and Evaluation (ADDIE instructional design model). This ensures the systematic development

of the system and its iterative refinement based on personalized interaction with participating learners and their personal feedback.

There are three major contributions that this paper makes. First, as previously mentioned, there is a clear gap in the literature when it comes to GAI-assisted learning for Arabic dialects versus the standard language. Second, we present a structured design framework for a custom GPT-based system that addresses this gap through guided conversation. Third, we propose an evaluation framework that combines vocabulary assessments, performance indicators for determining learner satisfaction, and analytics to measure user interaction with the GPT. These solutions provide a practical model for culturally responsive AI in language education.

II. LITERATURE REVIEW

Interest in GAI for its impact on language education is widespread. This section reviews studies that empirically and theoretically inform this study. The literature is arranged thematically and presented chronologically to show how research in GAI language learning has evolved. This includes learner perceptions of AI chatbots and the deployment of AI in Arabic language and dialect learning.

A. GAI in Language Education

Rule-based systems and tutoring systems were two of the earliest applications to benefit from AI usage. They were designed to provide practice that was structured and could provide some automated feedback to learners. As LLMs like GPT-based architectures developed, research started to focus on generative systems that produce more human-like responses that could deal with open-ended interactions. Chapelle [6] characterizes GAI as a major innovation in language education, with emphasis on its ability to support data-driven adaptive interaction in a process much more like authentic communication by humans.

Recent reviews show an increase in this type of empirical literature. Lee *et al.* [7] show that GAI tools are being widely adopted to make writing easier, enhance vocabulary development, and facilitate reflective learning. This is especially seen in higher education. Similarly, Li *et al.* [8] record an increase in experimental and quasi-experimental studies during the past two years, with results indicating a shift from exploratory implementations towards more theory-informed instructional designs.

In general, GAI tools are found to be more effective when included pedagogically within the learning curricula instead of as isolated technologies. These findings promoted more design-based research approaches that bring technology together with learning objectives and encourage iterative development [9, 10].

B. AI Chatbots and Vocabulary Acquisition

AI chatbots are perhaps the most frequent applications of GAI in language learning. Early empirical studies showed that chatbot-based interactions can support the

acquisition of vocabulary using techniques such as repetition, instant feedback and contextualized practice. Zhang and Huang [11] presented data that learners who were exposed to LLM chatbots had better vocabulary retention compared to more conventional practice techniques. This was especially true in situations where vocabulary was deployed in targeted conversational tasks. These findings are bolstered by meta-analytic evidence. Lyu *et al.* [12] conducted a meta-analysis of the implications for chatbot-supported language learning and reported that the use of chatbots makes a statistically significant improvement in vocabulary acquisition, increases learner engagement, and continues motivation in a wide range of learning contexts. The authors also note that interaction quality and task design are important predictors of learning effectiveness in recent research.

Design-based research has reinforced the relevance of using chatbots that are structured according to established learning theories. Li *et al.* [8] and Li *et al.* [13] suggest that AI chatbot designs based on activity theory encourage more active involvement of the learners by embedding vocabulary learning into meaningful communicative tasks. This concept suggests a transition in the literature from using drill-based practices to the use of more dialogue-based, context-rich learning experiences.

C. Learner Satisfaction and Acceptance of GAI Learning

Along with learning outcomes research, much literature is concerned with evaluating learners' perception and acceptance of AI chatbots. Early research focused on students' first impressions of AI tools, while more recent research has adopted formal acceptance models to investigate ongoing interaction and satisfaction with the tools. Stöhr *et al.* [14] stated that university students tend to accept AI chatbots in the process of learning as supportive learning companions that foster learner autonomy and diminish anxiety, especially in language learning situations. However, they observed that feedback quality and the relevance of chatbot responses are also very important for learner satisfaction.

Expanding on this work, Hwang *et al.* [15] applied an extended Technology Acceptance Model (TAM) to investigate learners' acceptance of ChatGPT. They found perceived usefulness and ease of use to be great predictors of learner satisfaction and commitment to continue using AI-based tools. These results emphasize learner satisfaction as a vital affective variable that directly impacts the effectiveness of AI-mediated learning environments.

Notwithstanding these advances, the majority of acceptance studies are confined to an English language learning context and the learner perceptions in a non-English and non-standard language environment continue to be underexplored.

D. AI Applications in Arabic Language and Dialect Learning

Increased research on the use of GAI for conversational Arabic demonstrates an increase in efforts

to develop chatbots that are capable of processing and generating regional dialects [16–18]. However, the extent of available literature in GAI-assisted Arabic language education still largely focuses on Modern Standard Arabic (MSA) and not regional dialects. Arabic language learning, in itself, is also limited compared to literature that focuses on English language learning.

According to Alkaabi and Almaamari [10], educators realize the pedagogical potential of GAI tools in Arabic teaching, but the practical implementation of these tools is hindered by limited contextualization, difficulties in technical adaptation to teaching and learning, and the absence of culturally relevant teaching resources. Bourhil and Younoussi [19] provide a review of Arabic chatbot technologies and point out that most existing systems are specifically tailored to MSA rather than the spoken language.

Arabic is also a diglossic language. This means that there are two distinct varieties of the same language that are used simultaneously for different specialized purposes. This causes a major challenge in learning the language, especially for AI [20]. The vocabulary, pronunciation, and everyday phrases of dialectal Arabic variations differ from MSA and from other regional dialects in varying degrees. Kuwaiti Arabic exemplifies this aspect of the language [17, 21]. The Kuwaiti dialect presents additional challenges due to the presence of phonetic variation within Kuwait. This includes geographic sub-dialects, lexical borrowing from other languages, and English code-switching among younger speakers [21].

Therefore, it is often hard for non-native students to learn dialects even though dialect competency is the basis of meaningful communication and social interactions. The lack of specialized AI systems for dialect learning is a major gap in research and practice.

E. Single-Sex Education and Language Learning Environments

Research suggests that when students learn in a single-sex education environment, it can change the way students participate and how confident they feel. Research shows that women in women-only learning environments tend to speak more, feel less nervous about making mistakes, and are more active in activities that require talking or expressing ideas [22–24]. Because these environments feel safer and more comfortable, students may be more willing to take risks without fear of judgment [23]. This is of particular relevance in the context of language learning, as practicing speaking can be perceived as a high-stakes activities that contribute to speaking anxiety [3, 25]. These environments may provide a supportive context for experimenting with language.

Affective elements, such as feelings of comfort, how motivated students feel, or how confident they feel, is essential for success in learning a new language. Studies suggest that women-only courses allow students to feel less social pressure and work together more easily. In places where cultural expectations about gender affect how students behave in class, these environments can also give learners more chances to practice the language

and help them learn better. Research on women's education in culturally specific contexts further emphasizes the importance of socially and culturally situated learning contexts [22–26].

F. Research Gap

Despite the increase in research on GAI for language learning, there are still several important limitations in the literature. As mentioned above, most research focuses on English and other languages that are globally dominant, with relatively little research on Arabic language learning. Arabic dialect acquisition is particularly underexplored. Arabic GAI research mostly addresses MSA instead of spoken dialects. The need for more dialectic research is of great cultural significance as spoken dialects occupy a central role in the everyday communication and social interactions of speakers.

Existing studies also emphasize the performance of the GAI system independently of the learners' perceptions or feelings of satisfaction. There is limited work that integrates structured vocabulary assessment, learner satisfaction, and instructional design principles into the same evaluation framework.

Additionally, research rarely examines GAI-assisted language learning for targeted learner populations within defined sociocultural contexts. Studies of Arabic conversational GAI tend to focus on developments within the GAI technology instead of pedagogy for its implementation. There are limited guidelines for creating supportive learning environments that are culturally supported. By examining the impact on GAI-assisted language learning within a women-only participant group, this study extends research on women-only education by situated the research to include GAI digital learning environments. It aims to contribute insights into how culturally responsive and supportive GAI learning conditions can influence learner satisfaction and dialectal vocabulary acquisition.

These gaps demonstrate the need for research that integrates GAI into an instructional design framework and applies it to languages that are seldom studied. There is a lack of studies that examine how customized GPT-based systems can be designed to support the acquisition of Arabic dialects while also assessing linguistic outcomes and learner satisfaction. Tackling these gaps requires a design-based approach that connects theory and system development with evaluation in an actual learning environment.

This work-in-progress study helps to fill this gap by creating a customized GPT-based learning system specifically designed for the Kuwaiti Arabic dialect. Integrating established instructional design principles we are able to evaluate its usage through key variables such as vocabulary, learners' feedback and analyzing learner interaction. This study contributes design insights and presents a novel methodology for future studies on GAI language learning that are more culturally responsive.

III. METHODS AND MATERIALS

This study uses a mixed-methods, Design-Based

Research (DBR) approach to develop and evaluate a customized GPT-based system to teach the Kuwaiti Arabic dialect. The methodology leverages qualitative and quantitative data in order to examine the acquisition of vocabulary by non-native learners and better understand their satisfaction with learning. The methods we used to achieve the research objectives are described in detail to facilitate recreation of this work.

A. Research Design

This study uses a work-in-progress DBR methodology to develop and evaluate a customized GPT-based learning system for Kuwaiti Arabic. This is well-suited to early-stage educational technology studies since it allows iterative refinements as the study is being conducted. The ADDIE instructional design model is applied during development to create synergy between the learning objectives, functionality of the system, and evaluation strategies [27, 28].

The experimental phase uses a pre-experimental, one-group, pre-test–post-test design to determine the changes in dialect vocabulary acquisition of the learners and to determine their level of satisfaction after interacting with the system. This design approach is suited to research on learning trends and user experience. The study combines quantitative and qualitative data to evaluate learning outcomes as well as learner perceptions.

B. Learning System Design and Development

The customized GPT model was designed using a prompt engineering framework that constrains the system output through predefined conversation interaction templates. These templates encourage acquisition of dialect-specific vocabulary through guided conversations that are typically short and scenario-based that depict common situations in Kuwaiti society. These situations include social events, common greetings, shopping, and family interactions.

Each interaction is structured in a particular sequence:

- (1) Words are provided within specific contexts.
- (2) Guided conversations are practiced.
- (3) Examples are given with clarification when learners need it.
- (4) The vocabulary is regularly repeated throughout various sessions.

The prompt engineering framework includes instructions that specify and define the conversational roles, vocabulary, and response formats. This ensures that there is a controlled dialogue and consistent exposure to the designed curriculum. The dialect is supported through a curated Kuwaiti Arabic lexical database. The database also addresses variations within the Kuwaiti dialect through the lexical mapping of multiple pronunciations and dialectal vocabulary. The database will also contain the most common expressions, various alternative usage patterns, and contextual prompt conditioning to better facilitate interaction with the GPT model. This addition will assist the model and the user from relying on MSA and will embed regional and colloquial differences.

The system will integrate a structured feedback mechanism that provides immediate corrective responses.

It will diagnose and reformulate the incorrect learner input and achieve the appropriate usage within the contextual scenario. A specific style of feedback will be prompted to consist of correction, explanation, and reinforcement through example responses.

The model will behave as a prototype mobile learning application created with Glide; a no-code app development platform. Glide was chosen due to its strengths in rapid prototyping and providing user authentication. Glide also allows for the automatic collection of interaction data. This makes the platform suitable for an early-stage educational application. The platform supports the ADDIE framework because of its ability to allow iterative design developments in response to analysis of the user's needs. This analysis can be used to refine the prompts and evaluation of learner performance and feedback data.

C. Participants

The target population for this study consists of adult non-native women residing in Kuwait who seek to improve their proficiency in the Kuwaiti Arabic dialect for everyday communication. Participants will be recruited based on limited exposure to formal dialect instruction and a demonstrated interest in developing spoken dialect competence for daily interaction within Kuwaiti society. Approximately 30 participants will be selected through purposive sampling in accordance with these inclusion criteria.

The focus on a women-only participant group is theoretically informed. Research indicates that single-sex learning environments may increase verbal participation and reduce communication anxiety in language-related tasks [22–24]. Since spoken interaction is central to dialect acquisition, reduced anxiety and increased confidence are particularly relevant factors in this context [3, 25].

D. Data Collection

Data will be collected through the following mixed-method instruments:

1) Vocabulary tests (pre-test and post-test)

Vocabulary acquisition tests will be given before and after the learning intervention to measure changes in learners' dialect-specific vocabulary knowledge.

2) Learner satisfaction questionnaire

A Likert-scale questionnaire will be distributed following the intervention to assess the level of satisfaction that learners have with the GPT model. Items on the Questionnaire will address the perceived usefulness, ease of interaction, level of engagement, and overall learning experience that they have.

3) Semi-structured interviews

Semi-structured interviews will be conducted with some of the participants to gather qualitative information based on item 2 in addition to better understanding their challenges using AI-supported dialect learning.

4) Platform usage analytics

System-generated analytics will be collected and

analyzed to examine user engagement, such as their interaction frequency and duration.

E. Data Analysis

Descriptive statistics will be used to identify overall trends and patterns in learning gains and perceived satisfaction. Where it is most appropriate, inferential statistical analyses will be conducted to examine any changes in vocabulary performance over time. The qualitative data gained from the semi-structured interviews will be analyzed thematically to identify patterns in learner engagement, tool usability, and system effectiveness. The analysis is expected to generate insights that inform future iterations of the system design.

F. Ethical Approval and Informed Consent

All procedures involving human participants will comply with international and institutional ethical standards. Ethical approval will be obtained from the relevant institutional ethics committee prior to data collection. Written informed consent will be obtained from all individual participants prior to participation. Participants will be informed of the purpose of the study, the voluntary nature of participation, and their right to withdraw at any time without penalty.

IV. PRELIMINARY FINDINGS AND EXPECTED OUTCOMES

Since this research is currently in progress, this study represents preliminary findings focusing on learning outcomes and system performance that are anticipated rather than representing the finalized empirical results. Based on the existing literature and the design principles that are guiding the system development, several outcomes are expected.

First, it is expected that using the customized GPT-based conversational system will result in measurable improvements in Kuwaiti Arabic vocabulary acquisition for the learners. The system's emphasis on context-specific dialogue, continued repetition of dialect-specific vocabulary, and practice of its guided situation-specific conversations is expected to encourage deeper understanding and retention of the vocabulary and improved practical use. These improvements are expected to show up in post-test vocabulary scores compared to pre-test scores.

Second, the study expects to find positive learner satisfaction and increased engagement with the system. It is expected that the conversational nature of the system, along with the immediate feedback it provides and its adaptive clarification, will enhance the learners' ease of use and its perceived usefulness, resulting overall in a better learning experience. This is particularly important in learning dialects, where confidence and the willingness to try to communicate are a necessary part of continuous language usage.

Third, system interaction analytics are expected to provide information about patterns of learner behavior, including how frequently it is used, whether the learning tasks are completed, and how much the learners engage with the conversational activities. These data are

expected to provide insight into interpreting vocabulary learning trends and learner perceptions. They can also inform future iterations of the system design.

Finally, as a DBR study, the expected outcomes are not limited to only learner performance but also include design insights into the development of GAI learning systems that are culturally responsive. Findings from this stage of the investigation are expected to inform improvements in prompt design, feedback strategies, and task sequencing, which will contribute to a scalable framework for future AI-supported Arabic dialect learning tools.

V. CONCLUSION

This work-in-progress study presents the design and planned method for the evaluation of a customized GPT-based conversational learning system for teaching the Kuwaiti Arabic dialect to non-native learners. The study focuses on the acquisition of dialect-specific vocabulary in real-world situations.

This study utilizes a design-based research methodology and is structured using the ADDIE instructional design model. The proposed system emphasizes pedagogically guided conversations and adapts to the feedback of learners in a structured manner. Rather than treating generative AI as a general conversational tool, this study demonstrates that instruction should be delivered with an acknowledgement of culturally specific elements to build a culturally appropriate and effective AI-supported language learning tool.

In the light of the ongoing process of data collection, the anticipated outcomes suggest that the proposed system will support dialect-specific vocabulary acquisition, provide learner satisfaction and encourage further engagement. The integration of vocabulary assessment, learners' feedback, and interaction analytics into the system will create a comprehensive framework that makes evaluating both learning outcomes and learner experience possible.

Further research on this project will focus on using its empirical findings to refine the system and to expand its use to other learner populations with different Arabic dialects. This study contributes to the literature by providing insights into the designs and methods that can be used in developing GAI language learning systems that are culture specific. It also provides research on the acquisition of languages that are underrepresented in the literature.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Zahra A. Reda led the conceptualization of the research, conducted the literature review, and developed the proposed framework. Dr. Sahar R. Hamzah contributed to the study's research design, methodological development, and academic co-

supervision. Dr. Almothana M. Gasaymeh provided primary supervision, offered guidance on the alignment of the study with relevant theoretical and disciplinary frameworks, and validated the overall research approach. All authors critically reviewed the manuscript and approved the final version.

ACKNOWLEDGMENT

This research was supported and created through the initiatives of His Majesty King Hamad bin Isa Al Khalifa's Digital Education Academic Chair.

REFERENCES

- [1] N. M. Meer and A. Chapman, "Assessment for confidence: Exploring the impact that low-stakes assessment design has on student retention," *Int. J. Manag. Educ.*, vol. 12, no. 2, pp. 186–192, 2014.
- [2] K. Edmunds and H. Leggett, "Active learning in the undergraduate laboratory: Giving students a safe space to experience low-stakes failure," *Open Scholarship Teach. Learn.*, vol. 3, no. 2, 2024.
- [3] I. Fauzi, R. Hartono, D. Rukmini, and H. Pratama, "AI applications for EFL learners: Enhancing speaking performance and reducing anxiety with gender-based analysis," *Forum for Linguistic Studies*, vol. 7, no. 9, pp. 282–301, 2025.
- [4] J. Fathi, M. Rahimi, and A. Derakhshan, "Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions," *System*, vol. 121, 103254, 2024.
- [5] S. M. Abdelhalim and R. A. Alsehibany, "Integrating AI-powered tools in EFL pronunciation instruction: Effects on accuracy and L2 motivation," *Comput. Assist. Lang. Learn.*, pp. 1–25, 2025.
- [6] C. A. Chapelle, "Generative AI as game changer: Implications for language education," *System*, vol. 132, 103672, 2025.
- [7] S. Lee, H. Choe, D. Zou, and J. Jeon, "Generative AI (GenAI) in the language classroom: A systematic review," *Interact. Learn. Environ.*, pp. 1–25, 2025.
- [8] B. Li, Y. L. Tan, C. Wang, and V. Lowell, "Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching," *Comput. Educ.: Artif. Intell.*, 100445, 2025.
- [9] A. A. Joubin, "Enhancing the trustworthiness of generative artificial intelligence in responsive pedagogy in the context of humanities higher education," in *Proc. General Aspects of Applying Generative AI in Higher Education: Opportunities and Challenges*. Springer, 2024, pp. 207–220.
- [10] M. H. Alkaabi and A. S. Almaamari, "Generative AI implementation and assessment in Arabic language teaching," *Int. J. Online Pedagogy Course Des.*, vol. 15, no. 1, pp. 1–18, 2025.
- [11] Z. Zhang and X. Huang, "The impact of chatbots based on large language models on second language vocabulary acquisition," *Heliyon*, vol. 10, no. 3, 2024.
- [12] B. Lyu, C. Lai, and J. Guo, "Effectiveness of chatbots in improving language learning: A meta-analysis of comparative studies," *Int. J. Appl. Linguist.*, vol. 35, no. 2, pp. 834–851, 2025.
- [13] Y. Li, X. Zhou, H. B. Yin, and T. K. Chiu, "Design language learning with Artificial Intelligence (AI) chatbots based on activity theory from a systematic review," *Smart Learn. Environ.*, vol. 12, no. 1, 24, 2025.
- [14] C. Stöhr, A. W. Ou, and H. Malmström, "Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study," *Comput. Educ.: Artif. Intell.*, vol. 7, 100259, 2024.
- [15] K. Hwang, S. Challagundla, M. Alomair, L. K. Chen, and F. S. Choa, "Towards AI-assisted multiple choice question generation and quality evaluation at scale: Aligning with Bloom's Taxonomy," in *Proc. Workshop on Generative AI for Education*, 2023.
- [16] Y. Saudi and M. M. Gammoudi, "Trends and challenges of Arabic chatbots: Literature review," *Jordanian J. Comput. Inf. Technol.*, vol. 9, no. 3, 2023.
- [17] A. Okunola and H. F. Alameir, "Assessing the role of AI chatbots in enhancing Arabic and English learning within Kuwait's bilingual education system," *Sci. Post*, vol. 2, no. 1, 2026.
- [18] M. Alhujaili and R. Abdulrahman, "Fine-tuning OpenAI GPT chatbot in Western Saudi dialect: A case study of Taibah University," *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 6, 2025.
- [19] H. Bourhil and Y. E. Younoussi, "Arabic chatbot technologies in education: An overview," arXiv Preprint, arXiv:2509.04066, 2025.
- [20] C. A. Ferguson, "Diglossia," *Word*, vol. 15, no. 2, pp. 325–340, 1959.
- [21] F. Husain, H. Alostad, and H. Omar, "Bridging the Kuwaiti dialect gap in natural language processing," *IEEE Access*, vol. 12, pp. 27709–27722, 2024.
- [22] L. J. Sax, T. A. Riggers, and M. K. Eagan, "The role of single-sex education in the academic engagement of college-bound women: A multilevel analysis," *Teach. Coll. Rec.*, vol. 115, no. 1, pp. 1–27, 2013.
- [23] S. R. Hanzah, "Pedagogy of all-women universities in AI-powered digital leadership," in *Proc. Strategic Decision-Making in Dynamic Business Environments: Systems and Control Perspectives*. Springer, 2026, pp. 895–903.
- [24] S. Kissau, L. Quach, and C. Wang, "Impact of single-sex instruction on student motivation to learn Spanish," *Can. J. Appl. Linguist.*, vol. 12, no. 2, 2009.
- [25] D. Ding and M. B. Yusof, "Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study," *Humanit. Soc. Sci. Commun.*, vol. 12, no. 1, pp. 1–16, 2025.
- [26] A. L. Booth, L. Cardona-Sosa, and P. Nolen, "Do single-sex classes affect academic achievement? An experiment in a coeducational university," *J. Public Econ.*, vol. 168, pp. 109–126, 2018. doi: 10.1016/j.jpubeco.2018.08.016
- [27] Z. Huiyu, "Unleash the power of conversational AI into the classroom with the ADDIE model," in *Proc. Perspectives of Future Learning*. CRC Press, 2025, pp. 73–111.
- [28] M. M. Rifai *et al.*, "Creating AI-powered chatbot for learning Japanese: Development and implementation," *J. Eng. Sci. Technol.*, vol. 19, pp. 1346–1364, 2024.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.