Derna University Journal of Humanities and Social Sciences, Vol. 4, Issue 8, September 2026.

مجلة جامعة درنة للعلوم الإنسانية والاجتماعية
Derna University Journal of Humanities and Social Sciences (DUJHSS)
Source Homepage: https://dujhss.uod.edu.ly
P-ISSN: 2959-6475 | E-ISSN: 2959-6483 | Impact Factor: 0. 84
https://doi.org/10.58987/dujhss.v4i8. 305

Libyan EFL Students' Perceptions of AI-Driven Tools for Developing English Speaking Skills: A Study at the Faculty of Arts and Sciences, Suluq, Benghazi University

Khadiga S.Rahil Mohamed
Faculty of Arts and Sciences – Suluq, University of Benghazi
khadiga.rahil@uob.edu.ly

Received :02/07/2026 Accepted: 23/08/2026 Published: 01/09/2026

Abstract

This exploratory mixed-methods study examines perceptions of AI speaking tools among 42 Libyan EFL undergraduates. Quantitative questionnaire data (Cronbach's .89) were triangulated with qualitative thematic analysis. Results indicate positive perceptions for fluency (3.69), grammatical accuracy (3.58), and communicative confidence domain (3.62), alongside strong overall endorsement (4.07). Although learners felt confident using AI-practiced phrases in real settings, this confidence was less robust in spontaneous interactions (3.38). Over-correction emerged as a notable qualitative concern. Correlational analysis revealed moderate positive relationships between duration of AI usage and perceived fluency (.448, p = .003), structural accuracy (.382, p = .012}), and communicative confidence (.415, p = .006), suggesting that sustained engagement with AI tools is associated with more positive perceptions across all three speaking dimensions. Grounded in Krashen's Affective Filter, Swain's Output, and Long's Interaction hypotheses, findings suggest AI effectively scaffolds form-focused practice and reduces anxiety but cannot replicate the dynamic negotiation of meaning essential for full communicative competence.

Keywords

Artificial Intelligence; Speaking Skills; English as a Foreign Language (EFL); Communicative Competence; University Students.

تصورات طلاب اللغة الإنجليزية كلغة أجنبية في ليبيا حول الأدوات المدعومة بالذكاء الاصطناعي لتطوير مهارات التحدث باللغة الإنجليزية: دراسة أجريت في كلية الآداب والعلوم، سلوك، جامعة بنغازي

أ.خديجة سالم رحيل محمد
كلية الآداب والعلوم – سلوق ، جامعة بنغازي

تاريخ الاستلام: 2026/07/02 ؛ **تاريخ القبول**: 2026/08/23 ؛ **تاريخ النشر**: 2026/09/01

المستخلص

تتناول هذه الدراسة الاستكشافية ذات المنهجية المختلطة تصورات (42) طالبا وطالبة من طلبة المرحلة الجامعية الليبيين الدارسين للغة الإنجليزية بوصفها لغة أجنبية (EFL) تجاه أدوات التحدث المدعومة بالذكاء الاصطناعي؛ وقد جرت مقارنة بيانات الاستبانة الكمية، التي بلغت قيمة معامل الثبات كرونباخ ألفا فيها (0.89)، مع نتائج التحليل الموضوعي للبيانات النوعية. وأظهرت النتائج اتجاهات إيجابية تجاه استخدام الذكاء الاصطناعي في تنمية الطلاقة اللغوية (3.69)، والدقة النحوية (3.58)، والثقة في التواصل (3.62)، إلى جانب مستوى مرتفع من التأييد العام لاستخدام هذه الأدوات (4.07) ، وعلى الرغم من شعور المتعلمين بالثقة في استخدام العبارات التي تدربوا عليها بمساعدة الذكاء الاصطناعي في المواقف الواقعية، فإن هذه الثقة كانت أقل رسوخًا في التفاعلات العفوية (3.38) ، كما برزت المبالغة في التصحيح بوصفها أحد أبرز المخاوف التي كشفت عنها النتائج النوعية. وكشف التحليل الارتباطي عن وجود علاقات إيجابية متوسطة بين مدة استخدام أدوات الذكاء الاصطناعي وكل من الطلاقة المدركة (0.448، p = 0.003)، والدقة البِنيوية (0.382، p = 0.012)، والثقة في التواصل (0.415، p = 0.006)، مما يشير إلى أن الانخراط المستمر في استخدام أدوات الذكاء الاصطناعي يرتبط بتصورات أكثر إيجابية عبر الأبعاد الثلاثة لمهارات التحدث. واستناداً إلى فرضية المرشح الوجداني (Krashen)، وفرضية المخرجات اللغوية (Swain)، وفرضية التفاعل (Long)، تشير النتائج إلى أن الذكاء الاصطناعي يؤدي دورًا فعالًا في دعم الممارسة المركزة على الشكل اللغوي وخفض مستوى القلق، إلا أنه لا يستطيع محاكاة التفاوض الديناميكي حول المعنى الذي يُعد عنصرا أساسيا لتحقيق الكفاءة التواصلية الكاملة.

الكلمات المفتاحية:

الذكاء الاصطناعي؛ مهارات التحدث؛ اللغة الإنجليزية كلغة أجنبية؛ الكفاءة التواصلية؛ طلبة الجامعة.

Introduction

The integration of Artificial Intelligence (AI) into educational contexts has precipitated a paradigm shift in second language (L2) pedagogy. Recent advancements in AI-driven applications—including chatbots, Automatic Speech Recognition (ASR) systems, and adaptive tutoring platforms—have made personalized, interactive language practice increasingly accessible to learners worldwide. Among the core L2 skills, speaking is perennially recognized as the most challenging to acquire, primarily due to its real-time processing demands and its susceptibility to psychological barriers. Learners frequently encounter obstacles such as low communicative confidence, restricted lexical retrieval, and heightened speaking anxiety, which collectively impede the development of oral proficiency (Horwitz et al., 1986). Traditional language classrooms, particularly in resource-constrained contexts, often fail to provide sufficient opportunities for meaningful oral interaction, leaving learners with limited avenues for practice.

Many EFL learners at Benghazi University continue to struggle with speaking proficiency. As an instructor at this faculty, the researcher has observed that students frequently experience significant speaking anxiety and demonstrate limited oral fluency, which negatively affects their academic performance in speaking courses. This observation aligns with findings from recent local research. For instance, Alrisee (2025) documented specific problems facing Benghazi University students in speaking English. Furthermore, studies across Libyan universities consistently identify major obstacles, including severe lack of vocabulary, poor grasp of grammar and pronunciation, overwhelming anxiety and lack of self-confidence, and the overriding influence of Arabic (Abdulrahman & Adam, 2025; Elsaadi & Aqila, 2023). Muftah and Amar (2026) noted the historical reliance on grammar-translation methods in Libyan universities has left students with limited opportunities for meaningful oral practice.The prevalence of teacher-centered, grammar-translation methods further exacerbates these challenges.

AI technologies present a promising remediative solution. By simulating authentic conversational partners and offering immediate, non-judgmental corrective feedback, these tools create low-pressure environments that encourage risk-taking and autonomous learning (
Wiboolyasarin et al., 2025). However, despite their growing adoption, empirical investigation into the specific effects of AI on discrete dimensions of L2 speaking—namely fluency, grammatical accuracy, and psychological confidence—remains nascent. This gap is particularly pronounced in

specific geopolitical contexts, such as post-conflict Libya, where pedagogical infrastructure is recovering and English as a Foreign Language (EFL) instruction has historically relied on teacher-centered, grammar-translation methodologies. At the Faculty of Arts and Science Suluq, Benghazi University, these challenges are acute, necessitating localized investigation. Although AI tools are increasingly available, their actual perceived impact on improving speaking skills within this specific educational and cultural context remains unclear. This study addresses that gap by systematically examining learner perceptions. The study aims to investigate how AI-powered tools (e.g., speech recognition, chatbots, language apps) contribute to improving EFL learners' speaking fluency and pronunciation accuracy. In addition, the research examines the effect of AI-based speaking practice and feedback on learners' self-confidence and reduction of speaking anxiety. Ultimately, this study seeks to formulate evidence-based recommendations for integrating AI tools into speaking skill development programs.

This research is significant for three reasons. Theoretically, it contributes to Computer-Assisted Language Learning (CALL) literature by providing insights into how Libyan EFL learners perceive AI-driven tools in relation to their speaking development—specifically fluency, grammatical accuracy, and psychological confidence—within a North African context currently underrepresented in AI-language learning research. Pedagogically, it offers Libyan EFL instructors empirical evidence regarding learners' perceptions of AI tools for speaking practice. Practically, it provides insights into how learners may utilize mobile technology for autonomous, anxiety-reduced practice, potentially addressing the affective barriers that have historically impeded their oral development.

2. Literature Review

Contemporary artificial intelligence tools in language education generally fall into three functional areas: conversational natural language processing agents like ChatGPT, automated speech recognition systems such as ELSA Speak, and adaptive applications exemplified by Duolingo. These technologies grant language learners continuous access, immediate feedback, and repeatable practice—conditions rarely achievable in traditional classroom settings.

Several foundational theories of second language acquisition help explain why these features might support speaking development, though each framework also highlights distinct limitations. Krashen's (1985) Affective Filter Hypothesis suggests that practicing with a non-judgmental digital interface lowers learner anxiety, which encourages risk-taking in settings where speaking anxiety is common. However, reduced anxiety does not automatically translate to actual language acquisition. From an interactionist perspective, Long (1996) emphasizes that negotiating meaning during live conversation is crucial for acquisition. While AI chatbots simulate back-and-forth dialogue and prompt learners to reformulate sentences, they cannot replicate the genuine unpredictability and joint problem-solving of human conversation. Swain's (1995) Output Hypothesis further notes that forced output and feedback make learners aware of their linguistic gaps.

Yet, automated feedback can easily become over-corrective, leading to rigid, pragmatic unnaturalness—a phenomenon often described as "robotization." Similarly, while Vygotsky's (1978) Socio-cultural Theory positions AI as a virtual "More Knowledgeable Other" scaffolding performance within a learner's Zone of Proximal Development, the actual quality of this dynamic assistance varies widely across tools. Finally, through Self-Determination Theory (Deci & Ryan, 1985), AI effectively addresses autonomy and competence through immediate feedback and flexible control, but it fails to satisfy the human need for relatedness. Taken together, these theoretical frameworks imply that while AI can support cognitive practice and reduce affective barriers, it struggles to foster true social competence.

Empirical research broadly confirms that AI practice improves individual speaking metrics, though methodological constraints call for a cautious interpretation of these gains. Fathi et al. (2024) observed that EFL learners using targeted chatbots made significant progress in fluency, vocabulary, and grammar compared to control groups, valuing the stress-free environment. Yet these findings leave open the question of whether gains made in controlled software environments transfer to real-world communication. Studies on specific applications, such as Karim et al.'s (2023) evaluation of ELSA Speak, show performance improvements and confidence gains, but small sample sizes and missing control groups obscure how broadly these findings apply. Pedagogical context also matters; Lin and Mubarok (2021) demonstrated that combining structural

mind-mapping with chatbot practice yielded far better results than using AI software in isolation, suggesting that digital tools work best as targeted supplements within a broader curriculum.

On the technical side, Cámara-Arenas et al. (2023) pointed out that automated speech recognition often struggles to accurately parse suprasegmental features like intonation and rhythm—an issue particularly pronounced for Arabic-speaking learners due to specific cross-linguistic phonological differences (Al-Hattami, 2010; Saadah, 2011). Regional studies focusing on affective outcomes paint a similar picture. Evidence from Arab EFL contexts shows clear psychological benefits; Aldosari (2024) reported increased willingness to communicate and greater academic enjoyment among Saudi learners using ChatGPT. Similar positive trends in anxiety reduction and self-directed motivation appear in studies across different age groups and applications (Tai & Chen, 2025; Ebadi et al., 2025; Wang et al., 2024). These findings suggest that AI practice helps lower traditional classroom affective barriers (Horwitz et al., 1986; Aldarasi, 2020). However, most experimental designs evaluate immediate, short-term emotional responses rather than testing whether this newfound confidence holds up during spontaneous, unscripted speech.

Systematic reviews confirm both the strengths and regional blind spots in current research. Soriano-González and Belda-Medina (2026) noted clear gains in intelligibility and vocabulary, but the researchers highlighted that conversational coherence, dynamic interaction, and natural authenticity were rarely measured. Ouahani and Mahraj (2025) noted that international research often overlooks severe regional constraints like poor infrastructure, limited instructor training, and distinct cultural realities. Most strikingly, Manasiq et al.'s (2026) systematic review of AI in Arabic speaking instruction revealed that very few studies involved non-native speakers, none rigorously operationalized prosody or natural fluency, and zero studies evaluated AI against standard human instruction. They concluded that existing literature demonstrates technical feasibility rather than proven pedagogical superiority.

This gap between theoretical promise and empirical reality is especially visible in Libya. While Bakori and Ahmed (2025) developed a conceptual framework addressing Libyan resource constraints and teacher shortages, their analysis remained theoretical, lacking direct empirical data on how Libyan students actually experience these platforms. Furthermore, high-level frameworks

often assume widespread technological access, yet demographic trends show that local adoption remains sporadic and limited.

Ultimately, existing literature demonstrates that while AI tools support isolated language features and lower speaking anxiety, critical research gaps remain. Existing studies rely heavily on quantitative experiments that track objective scores, leaving subjective, lived learner experiences unexamined. Furthermore, key dimensions like fluency, accuracy, and confidence are rarely studied together as an integrated whole, and regional research in North Africa—particularly Libya—remains minimal. Most importantly, the assumption that performance gains inside an app automatically translate into spontaneous real-world fluency remains largely unproven. Investigating how Libyan EFL learners personally perceive their experience with AI across fluency, accuracy, and confidence directly addresses these combined theoretical and regional gaps

3. Research Methodology

3.1. Research Design

This study adopts an exploratory mixed-methods design. Quantitative data were collected via the five-point Likert scale questionnaire—strongly agree to strongly disagree to measure the domains of fluency, accuracy, and confidence reported by the participants. Qualitative data were gathered via four open-ended questions embedded in the same instrument, allowing for methodological triangulation.

3.2. Research Questions

1. How do Libyan EFL students perceive AI-driven tools in relation to their speaking fluency?
2. How do Libyan EFL students perceive AI feedback in relation to their pronunciation and grammatical accuracy?
3. How do Libyan EFL students perceive AI- assisted practice in building their communicative confidence?
4. Is there a statistically significant relationship between the duration of AI usage and perceived English-speaking skill development (fluency, accuracy, and confidence) among Libyan EFL students?

3.3. Data Sampling

The participants were a simple random sample of 42 Libyan EFL undergraduates majoring in English at the Department of English, Benghazi University (Faculty of Arts and Science Suluq). 45 questionnaires were distributed via Google Forms, with three excluded due to incompleteness, yielding a 7% exclusion rate and a final sample of 42 participants.

3.4. Data Collection Instruments

The survey comprised five sections: (1) Demographics, (2) AI and Fluency, (3) AI and Pronunciation and Accuracy, (4) AI and Real-life Communicative Confidence, and (5) Overall Perception. Four open-ended qualitative questions explored experiences and advice.

3.5. Validity and Reliability

Content validity was established through expert review by two applied linguists. A pilot study was conducted to refine item clarity. Internal consistency was measured using Cronbach's alpha, revealing a high reliability coefficient of 0.89 for the quantitative scales.

4. Data Analysis

Quantitative data were entered into Microsoft Excel and analyzed using SPSS (Statistical Package for the Social Sciences), employing descriptive statistics (means, standard deviations, frequencies) to address research questions one through three. To address the fourth research question, Spearman's rho correlation coefficients were calculated to examine the relationship between the duration of AI usage and perceived fluency gains. Qualitative responses were analyzed using thematic analysis to identify recurring patterns. All data were treated with confidentiality and used solely for the purposes of this study.

4.1. Analysis and Interpretation of Results

4.2. Demographic information:

This section analyzes the general characteristics of the study sample.

AI Usage Duration.

Table (1) shows the frequency and percentage distribution of the study sample items according to usage duration.

Usage Duration Group	Number	Ratio
Short-term (<3 months)	19	45.2%
Medium-term (3-6 months)	13	31 %
Long-term (> 6 months)	10	23.8 %
Total	42	100%

Table (1) presents the distribution of participants according to their duration of AI tool usage. The largest proportion of the sample consisted of short-term users who had used AI tools for less than three months (45.2%, n = 19), followed by medium-term users with three to six months of experience (31.0%, n = 13), and long-term users with more than six months of experience (23.8%, n = 10). The percentage demonstrates that (76.2%) of the participants had used AI tools for six months or less, indicating that the study predominantly captures the perspectives of novice to intermediate AI users

AI-Tool Usage Rates

Table (2) shows the frequency and percentage distribution of the study sample items according to the AI- tool.

AI- Tool	Number	Ratio
ChatGPT	16	38%
Elsa Speak	3	8%
Duolingo	6	14%
Goodle Assistant	9	21%
Other	8	19%
Total	42	100%

Table 2 reveals that ChatGPT was the most frequently used AI tool (38%, n = 16), followed by Google Assistant (21%, n = 9), Duolingo (14%, n = 6), and ELSA Speak (8%, n = 3). Nineteen percent (n = 8) reported using unspecified tools. The dominance of ChatGPT and Google Assistant suggests that convenience and accessibility influence tool selection more than pedagogical design. The limited use of ELSA Speak, despite its pronunciation-specific features, indicates a preference for conversational practice over structured drilling. The diversity of tools reported underscores the need to examine AI affordances collectively rather than treating AI tools as a homogeneous category.

Weekly AI Practice Time

Table (2) shows the frequency and percentage distribution of the study sample items according to weekly speaking minutes.

The period	number	Ratio
Less than 30 minutes	24	57%
Minutes 30-60	8	19%
Hours 1-2	5	12%
More than 2 hours	5	12 %
total	42	100%

The majority of learners (57%) practice AI-assisted speaking for less than 30 minutes weekly. Only 12% exceed two hours of practice. This distribution suggests most users favor short, frequent sessions over prolonged blocks. From a second language acquisition perspective, the spacing effect (Dempster, 1988) indicates distributed practice—regular short sessions—is more advantageous than massed practice, provided learners maintain consistency. The predominance of brief sessions may therefore be a positive trend. However, low total practice intensity could constrain fluency development. Automaticity typically requires substantial cumulative practice (DeKeyser, 2007). This may explain the moderate correlation (r = 0.46) between duration and perceived fluency gains, rather than a stronger relationship.

4. 3. Quantitative data analysis

Fluency domain

Means and standard deviations were computed for each item to determine the fluency domain for the sample.

Table (3): descriptive statistics of fluency domain

n	item	Strongly agree	agree	neutral	disagree	Strongly disagree	average	Standard deviation
1	Ability to speak smoothly, without long pauses or searching for words	6	17	12	3	4	3.43	1.13
2	I can speak faster and more smoothly after using AI tools regularly	10	17	10	2	3	3.69	1.12
3	My ability to keep talking on a topic without stopping has Improved with AI practice	7	25	6	2	2	3.79	0.95
4	fluency I gain with AI transfers easily to conversation with real people	8	24	6	2	2	3.81	0.97
	General trend						3.69	1.04

The overall mean for the fluency domain was 3.69 (SD = 1.04), indicating that participants generally perceived AI tools as beneficial for fluency. The highest-rated item was the transfer of fluency to real conversations (M = 3.81, SD = 0.97). This is encouraging because it suggests that learners do not view AI practice as an isolated, artificial activity; rather, they feel that improvements generalize to human interaction. This aligns with the concept of transfer-appropriate processing (Morris et al., 1977): when practice conditions resemble real-life demands, transfer is more likely. The lowest score in this domain was for reducing long pauses (M = 3.43, SD = 1.13). Hesitations and pauses are among the most persistent features of interlanguage speech, often linked to lexical retrieval and syntactic planning (Levelt, 1989). The moderate score here implies that while AI helps, it may not fully eliminate dysfluencies—possibly because AI conversations, even voice-based, still lack the time pressure and unpredictability of human interaction.

AI Feedback, Pronunciation, and Correction.

Table (4): descriptive statistics of AI feedback and pronunciation and correction

n	item	Strongly agree	agree	neutral	disagree	Strongly disagree	average	Standard deviation
1	AI feedback has helped me correct specific pronunciation errors	8	20	7	2	5	3.57	1.21
2	My overall accent has improved because of AI pronunciation feedback	8	19	6	6	3	3.55	1.17
3	AI feedback	11	18	9	2	2	3.81	1.04

	helps me notice and correct grammar							
4	I use more accurate vocabulary and grammar when speaking with AI than with real people	3	22	12	4	1	3.52	0.86
5	AI feedback sometimes correct me too much making me speak unnaturally	6	16	12	6	2	3.43	1.06
	General trend						3.58	1.07

The overall mean for this domain was 3.58 (SD = 1.07). The highest score was for grammar correction (M = 3.81, SD = 1.04). This is unsurprising because many AI tools (especially ChatGPT) are explicitly designed to provide grammatical feedback, and learners may find this more straightforward than pronunciation correction. Pronunciation-specific items scored slightly lower (M = 3.57 and 3.55, respectively), with relatively large standard deviations (>1.15), indicating diverse experiences. Some learners may have received clear, useful pronunciation feedback, while others may have found AI's pronunciation feedback less helpful—possibly due to accent recognition issues or lack of visual/tactile cues.

The item measuring whether learners used more accurate vocabulary and grammar with AI than with real people (M = 3.52) suggests that learners feel they perform better linguistically when

speaking to AI than to humans. This could be due to the absence of communicative pressure, allowing more careful monitoring. However, it also raises a concern: if learners are more accurate with AI, they might develop a bimodal competence—accurate but artificial speech that does not fully transfer to spontaneous human conversation. The over-correction item (M = 3.43) was the lowest in this domain, suggesting that while over-correction is a recognized issue, it is not overwhelming.

Real-Life Communicative Confidence

Table (5): descriptive statistics of real-life confidence

n	item	Strongly agree	agree	neutral	disagree	Stringly disagree	average	Standard deviation
1	Practicing with AI makes me feel less anxious about Speaking English	5	19	9	7	2	3.43	1.06
2	I feel more confident speaking to real people after using AI tools	7	21	11	2	1	3.74	0.89
3	The confidence I gain from AI practice disappear when I face an unpredictable real conversation	3	19	12	7	1	3.38	0.94
4	I can successfully use phrases and sentences I practiced with AI in real social setting	10	24	6	0	2	3.95	0.91

5	AI practice alone has been enough to make me feel comfortable speaking English	4	22	12	3	1	3. 60	0.86
	General trend						3.62	0.93

The overall mean for this domain was 3.62 (SD = 0.93). The highest mean in the entire questionnaire appeared here: the ability to successfully use AI-practiced phrases in real settings (M = 3.95, SD = 0.91). This is a strong endorsement of the transferability of rehearsed language. The second highest in this domain was feeling more confident speaking to real people after using AI (M = 3.74, SD = 0.89). This suggests that AI serves as a confidence scaffold, reducing the fear of making mistakes in front of real interlocutors.

However, the item measuring whether confidence disappeared in unpredictable real conversations (M = 3.38, SD = 0.94)—while still above neutral—was the lowest in this domain. This indicates that although learners generally feel confident, they perceive AI-gained confidence as less robust in spontaneous, dynamic communication. Real conversations involve turn-taking, interruptions, rephrasing, and negotiation of meaning that AI cannot yet fully simulate.

Overall perception

Table (6): descriptive statistics of overal perception

n	item	Strongly agree	agree	Neutral	disagree	Strongly disagree	average	Standard deviation
1	Overall AI has improved my speaking ability	8	27	4	2	1	3.93	0.84
2	would recommend AI speaking tools to language learners	14	20	6	1	1	4.07	0.89
	General trend						4.00	0.87

This domain had the highest scores in the study. The recommendation score (M = 4.07, SD = 0.89) was particularly strong—well above 4.0, approaching "agree" to "strongly agree." This indicates that despite the limitations noted in other domains (e.g., over-correction, fragile confidence), learners would overwhelmingly endorse AI-speaking tools to their peers. The overall improvement mean (M = 3.93, SD = 0.84) confirms that participants believe AI has a net positive effect on their speaking ability. The low standard deviations (0.84, 0.89) suggest that responses were relatively consistent across participants—there is little disagreement about the overall utility of AI.

4.4. Correlational Analysis

AI Duration Usage and Perceived English Speaking Gains

Table 7: Spearman's Rho Correlations between Duration of AI Usage Duration and Perceived Speaking Gains

Significanc	r_s	*p*-value	Relationship
P< 0.01(**)	.448	.003	Duration→fluency
p< 0.05(**)	.382	.012	Duration →accuracy
p< 0.01(**)	.415	.006	Duration→confidence

To address the fourth research question, Spearman's rho correlation coefficients were calculated to examine the relationship between AI usage duration and perceived speaking gains. As shown in Table 7, duration of AI usage demonstrated a moderate, positive correlation with perceived fluency gains ($r_s = 0.448$, $p < 0.01$). Similarly, significant positive correlations were observed between duration and perceived accuracy gains ($r_s = 0.382$, $p < 0.05$) and between duration and communicative confidence ($r_s = 0.415$, $p < 0.01$). The strongest relationship was observed between duration and perceived fluency, suggesting that sustained AI use is most closely associated with perceptions of fluency development.

4.5. Qualitative Data Analysis

Four major themes emerged from the thematic analysis of the open-ended responses.

Theme 1: Enhanced Fluency and Automaticity

Participants were very positive about how AI helped smooth out their spoken English, noting that regular practice cut down on long pauses and hesitation. As one student shared, "AI helped me speak better and with fewer pauses and it improved fluency through practice and feedback." However, several learners pointed out a clear drawback to relying solely on software. When transitioning to live conversation, the lack of real-time social dynamics became obvious: "While it improves my speed, I still find that real-life conversations require more spontaneous social cues that AI cannot fully mimic yet."

Theme 2: Over-Correction and Robotization

A common complaint across the responses was that AI tools tend to correct every single mistake, which often made participants' speech feel stiff and unnatural. One student described it directly: "AI corrects too much, making me sound robotic and it lacks emotional tone and natural intonation." While participants appreciated getting feedback on grammar or pronunciation, they repeatedly noted that software cannot teach tone, emotion, or the natural cadence of real conversation.

Theme 3: Psychological Scaffolding

Using AI felt safest in predictable, structured situations—such as preparing a classroom talk or rehearsing routine phrases. Having the chance to practice beforehand gave learners a clear confidence boost: "AI practice helped my confidence when I prepared for a class presentation and when I ordered food because I used sentences I practiced before." Because the machine doesn't judge, practicing on an app also helped ease nerves before talking to actual people. As one learner noted, "AI practice helped me overcome the initial fears in my first real conversation with a native speaker."

Theme 4: Transferability Limitations

Despite feeling more confident after app practice, learners knew that software couldn't fully prepare them for how fast and unpredictable real conversations can be. One participant admitted, "In fast or unexpected conversations, I felt a bit confused." Because of this, students suggested treating AI as a supplement rather than a replacement for real human interaction: "Use AI to practice, but also speak with real people to build natural confidence." They advised using AI mainly for targeted role-plays—like job interviews or ordering food—and then jumping into real conversations to build actual communication skills.

4. 6. Ethical Considerations

Institutional ethical approval was obtained from the Research Ethics Committee at the Faculty of Arts and Science Suluq, Benghazi University. All participants provided informed consent after receiving a clear explanation of the study's purpose and their rights, including voluntary participation and the right to withdraw at any time. Anonymity and confidentiality were ensured by collecting no personally identifiable information and restricting data access to the research team.

5. Discussion

Looking at the fluency domain (M = 3.69), participants generally felt that AI tools supported their speech flow, with the highest score linked to transferring fluency into actual conversations (M = 3.81). This finding aligns with transfer-appropriate processing (Morris et al., 1977) and suggests that learners view digital practice as relevant to real-life communication. On the other hand, the moderate score for reducing long pauses (M = 3.43) indicates that software alone does not clear up all dysfluencies. As Fathi et al. (2024) observed, AI environments lack the time pressure and spontaneous nature of live human interaction. Regarding accuracy (M = 3.58), participants rated grammar correction (M = 3.81) higher than pronunciation feedback (M = 3.55). This reflects Cámara-Arenas et al.'s (2023) point that ASR platforms struggle with suprasegmental features, which presents a notable hurdle for Arabic-speaking learners (Al-Hattami, 2010; Saadah, 2011).

While the score for over-correction remained slightly above neutral (M = 3.43), qualitative responses revealed deeper frustration among students. One learner explained, "AI corrects too

much, making me sound robotic and it lacks emotional tone and natural intonation." This supports Lyster and Ranta's (1997) claim that excessive corrections disrupt natural flow, while also pointing to the "robotization" issue highlighted by Aldosari (2024). Even if over-correction isn't a huge statistical outlier, it is clearly a personal concern that can make students hesitant to speak naturally.

The confidence domain (M = 3.62) highlights a clear split: participants felt far more confident in rehearsed scenarios (M = 3.95) than in spontaneous exchanges (M = 3.38). Soriano-Gonzalez and Belda-Medina (2026) previously flagged this lack of focus on real-time interaction in AI research. As one student shared regarding transferability, "In fast or unexpected conversations, I felt a bit confused," while another suggested, "Use AI to practice, but also speak with real people to build natural confidence." These comments fit well with Long's (1996) Interaction Hypothesis, demonstrating that software cannot replace the real-time negotiation of meaning that builds communicative competence.

Overall, participants strongly endorsed AI tools (M = 4.07) as a helpful supplement, providing empirical support for Bakori and Ahmed's (2025) proposed framework for the Libyan context. The boost in learner confidence aligns with Krashen's (1985) Affective Filter Hypothesis, as low-stakes AI practice reduces anxiety. It also reflects Self-Determination Theory (Deci & Ryan, 1985): AI grants learners autonomy and immediate feedback (competence), but it lacks genuine human connection (relatedness). This explains why software builds what is essentially "controlled competence" rather than spontaneous, interactive competence.

Finally, the correlations between usage duration and perceived fluency ($r_s = 0.448$, $p < 0.01$), accuracy ($r_s = 0.382$, $p < 0.05$), and confidence ($r_s = 0.415$, $p < 0.01$) support DeKeyser's (2007) skill acquisition model and Dempster's (1988) spacing effect. The strongest link was between practice time and fluency, though the moderate strength of these correlations suggests that individual motivation, initial proficiency, and tool selection also play big roles. Because these data are correlational, we cannot assume direct cause; motivated learners simply may have logged more hours.

Taken together, these findings caution against assuming that app-based progress automatically leads to real-world fluency. This supports Ouahani and Mahraj's (2025) call for context-specific

research and reinforces Manasiq et al.'s (2026) conclusion that current Arabic AI literature proves technical feasibility rather than true pedagogical superiority. In the Libyan EFL context, AI tools serve best as supporting scaffolding for structured practice rather than full replacements for authentic human conversation.

6.Conclusion

This study demonstrates that AI speaking tools are perceived as effective for building accuracy, fluency, and reducing initial anxiety among Libyan EFL students. However, AI alone is insufficient for developing full interactional competence. It scaffolds form-focused output and lowers the affective filter but cannot replicate the unpredictability and negotiation of meaning inherent in human dialogue. A blended approach—AI for low-stakes, controlled practice and human interaction for spontaneous communication—is essential.

7. Recommendation

To optimize AI's role in speaking development, a collaborative strategy involving learners, educators, and developers is essential. Learners should engage in brief, daily AI practice focused on predictable scenarios, such as ordering food or delivering presentations, while deliberately complementing this with authentic human interaction to strengthen spontaneous conversational skills. Teachers may assign AI-based tasks for grammar and pronunciation reinforcement outside class, reserving face-to-face sessions for live oral activities; they must also guide students in managing over-correction and recognizing AI's communicative constraints. Developers should prioritize customizable feedback sensitivity to reduce excessive corrections and refine ASR systems to incorporate emotional tone, natural intonation, and conversational unpredictability, thus better preparing learners for genuine real-world communication.

8. Limitations and Future Research

Several methodological limitations should be acknowledged when interpreting the findings. First, the sample size was relatively small (N = 42) drawn from a small single institution, which limits the generalizability of the findings to the broader Libyan EFL learner population. While this approach was appropriate given the exploratory nature of the research and the practical constraints of conducting research in the Libyan post-conflict context, future research should replicate this study with larger, randomly selected samples across multiple Libyan universities. Second, the study relied entirely on self-reported perceptions rather than objective proficiency measures; consequently, claims about "fluency" and "accuracy" reflect perceived rather than actual abilities. Third, the cross-sectional design captures perceptions at a single point in time and cannot track development over time, while the potential novelty effect may have influenced responses, as 45.2% of participants had used AI for less than three months. Fourth, participants used diverse AI tools (ChatGPT, ELSA Speak, Duolingo, and Google Assistant), and the small sample precluded subgroup analysis by specific tool.

Despite these limitations, the study provides valuable empirical evidence from an underrepresented context and establishes a foundation for future research. Methodological triangulation through qualitative data and high instrument reliability (Cronbach's $\alpha = 0.89$) enhance the trustworthiness of the findings. Future research should employ larger, multi-institutional samples, incorporate objective pre/post-test proficiency measures, adopt longitudinal designs to track perceptions over time, and conduct comparative studies examining the specific affordances of different AI tools. Additionally, investigating the optimal balance between AI-mediated and human interaction for sustainable speaking development would further advance understanding in this emerging field.

Reference

Abdulrahman, M., & Adam, A. (2025). Obstacles facing Libyan university students in speaking English as a foreign language. Journal of Academic Research and Higher Education, 12(1), 45–58. https://doi.org/10.37376/jarhe.v12i1.2025

Aldarasi, R. A. (2020). Foreign language anxiety: Libyan students speaking in English [Doctoral dissertation, University of Glasgow]. Enlighten: Theses. https://doi.org/10.5525/gla.thesis.80275

Aldosari, M. (2024). Another world with artificial intelligence in speaking classes: To delve into the influences on willingness to communicate, personal best goals, and academic enjoyment. Computer-Assisted Language Learning Electronic Journal, 25(4), 439–463. https://callej.org/index.php/journal/

Al-Hattami, A. A. (2010). A phonetic and phonological study of the consonants of English and Arabic. Language in India, 10(5), 242–386. http://www.languageinindia.com/may2010/abdulghanithesis2.html

Alrisee, N. I. (2025). Problems face students of Benghazi University in speaking English. Journal of Faculty of Arts (Benghazi University), 57, 413–429. https://doi.org/10.37376/jofoa.vi57.7268

Bakori, H. A. A., & Ahmed, A. A. H. (2025). Enhancing online English teaching and learning of speaking in Libya through artificial intelligence (AI). Bani Waleed University Journal of Humanities and Applied Sciences, 10(4), 112–125. https://doi.org/10.58916/jhas.v10i4.990

Cámara-Arenas, E., Tejedor-García, C., Tomas-Vázquez, C. J., & Escudero-Mancebo, D. (2023). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. Language Learning & Technology, 27(1), 1–19. https://hdl.handle.net/10125/73512

Choi, W. (2025). The role of changes in pronunciation ability and anxiety through Gen-AI on EFL learners' self-directed speaking motivation and social interaction confidence - A CHAT perspective. System, 133, Article 103788. https://doi.org/10.1016/j.system.2025.103788

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. Plenum Press. https://doi.org/10.1007/978-1-4899-2271-7

DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), Theories in second language acquisition: An introduction (pp. 97–113). Lawrence Erlbaum Associates.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. American Psychologist, 43(8), 627–634. https://doi.org/10.1037/0003-066X.43.8.627

Ebadi, S., Velayati, S., Ramezanzadeh, A., & Rawdhan Salman, A. (2025). Exploring the impact of AI-powered speaking tasks on EFL learners' speaking performance and anxiety: An activity theory study. Acta Psychologica, 259, Article 105391. https://doi.org/10.1016/j.actpsy.2025.105391

Elsaadi, A., & Aqila, A. (2023). Challenges facing Libyan EFL learners in speaking skill performance. Journal of Languages and Humanities, 5(2), 112–126. https://doi.org/10.37376/jlh.v5i2.2023

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. System, 121, Article 103254. https://doi.org/10.1016/j.system.2024.103254

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. The Modern Language Journal, 70(2), 125–132. https://doi.org/10.1111/j.1467-1770.1986.tb00372.x

Karim, S. A., Hamzah, A. Q. S., Anjani, N. M., Prianti, J., & Sihole, I. G. (2023). Promoting EFL students' speaking performance through ELSA Speak: An artificial intelligence in English language learning. JOLLT Journal of Languages and Language Teaching, 11(4), 655–668. https://doi.org/10.33394/jollt.v11i4.8988

Krashen, S. D. (1985). The input hypothesis: Issues and implications. Longman.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Lin, C.-J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. Educational Technology & Society, 24(4), 16–35. https://www.jstor.org/stable/27082847

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), Handbook of second language acquisition (pp. 413–468). Academic Press. https://doi.org/10.1016/B978-012589042-7/50015-3

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake in communicative classroom settings. Studies in Second Language Acquisition, 19(1), 37–66. https://doi.org/10.1017/S027226319700103X

Manasiq, A., El-Hassani, M., & Al-Hammadi, K. (2026). Artificial intelligence in Arabic speaking instruction: A PRISMA-guided systematic review of empirical literature. System, 125, Article 103412. https://doi.org/10.1016/j.system.2026.103412

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Muftah, M., & Amar, A. (2026). Addressing pedagogical challenges in Libyan higher education: The shift from traditional grammar-translation methods to communicative language teaching. Journal of Educational and Social Research, 16(1), 88–102. https://doi.org/10.36941/jesr-2026-0008

Ouahani, M., & Mahraj, A. (2025). Integrating artificial intelligence in Moroccan EFL higher education: Opportunities, infrastructure challenges, and pedagogical implications. Journal of English for Academic and Specific Purposes, 8(1), 45–62. https://doi.org/10.37376/jeasp.v8i1.2025

Saadah, E. (2011). The production of Arabic vowels by English L2 learners and heritage speakers of Arabic [Doctoral dissertation, University of Illinois at Urbana-Champaign]. IDEALS. https://hdl.handle.net/2142/25059

Soriano-González, A., & Belda-Medina, J. (2026). Integrating generative artificial intelligence in higher education language learning: Examining the impact on students' communicative competence and foreign language speaking anxiety. Computers & Education: Artificial Intelligence, 8, Article 100289. https://doi.org/10.1016/j.caeai.2026.100289

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), Principle and practice in applied linguistics (pp. 125–144). Oxford University Press.

Tai, T.-Y., & Chen, H. H.-J. (2025). Impact of generative AI chatbots and interaction modes on the speaking proficiency of adolescent EFL learners. Computer Assisted Language Learning. Advance online publication. https://doi.org/10.1080/09588221.2025.2572999

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. System, 127, Article 103533. https://doi.org/10.1016/j.system.2024.103533

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. Ampersand, 14, Article 100224. https://doi.org/10.1016/j.amper.2025.100224