



Generative Artificial Intelligence in English as a Foreign Language (EFL) Higher Education: A Critical Review of Opportunities and Challenges

Hager Mohamed Ben Ali *

Department of English Language, Faculty of Arts, University of Benghazi, Benghazi, Libya

الذكاء الاصطناعي التوليدي في تعليم اللغة الإنجليزية كلغة أجنبية في التعليم العالي: مراجعة نقدية
للفرص والتحديات

هاجر محمد بن علي *

قسم اللغة الإنجليزية، كلية الآداب، جامعة بنغازي، بنغازي، ليبيا

*Corresponding author: Hager.benali@uob.edu.ly

Received: May 28, 2026

Accepted: June 29, 2026

Published: August 03 2026

Abstract:

Generative Artificial Intelligence (GenAI) is rapidly changing higher education, offering opportunities for English as a Foreign Language (EFL) teaching and learning while also posing pedagogical, ethical, and institutional challenges. This critical literature review examines recent research on GenAI implementation in EFL higher education. The review includes eighteen peer-reviewed publications, including empirical studies and systematic literature reviews, retrieved from Scopus, Google Scholar, and ScienceDirect between 2024 and 2026. The review indicates that GenAI has potential to support language learning through personalised feedback, adaptive learning, formative assessment, and increased learner engagement, with the most consistent evidence found for writing, reading, grammar, vocabulary development and formative feedback specifically. Evidence for other areas, including speaking and listening proficiency as well as benefits for learners at lower proficiency levels, is more limited and comes mainly from individual studies. The findings also suggest that AI-assisted learning can promote learner autonomy and self-regulated learning, and can support instructors in developing more student-centred instructional approaches. At the same time, the review identifies recurring challenges, including academic integrity, overreliance on AI-produced materials, data security, AI literacy, teacher preparation, and the need for clear institutional policy. These challenges appear to depend more on how GenAI is implemented than on the technology itself. The evidence suggests that GenAI is most effective as a tool that supports teachers rather than replacing them, provided it operates within appropriate pedagogical, ethical, and institutional frameworks. The paper concludes by discussing practical implications for EFL higher education, with particular reference to Benghazi University as an illustrative case, and by identifying priorities for further research, including longitudinal studies, broader geographical representation, and more direct investigation of learners at different English proficiency levels.

Keywords: Generative Artificial Intelligence (GenAI), English as a Foreign Language (EFL), Higher Education, Language Learning, English Language Proficiency, Learner Autonomy, Teacher Readiness, Artificial Intelligence in Education.

المخلص

يشهد الذكاء الاصطناعي التوليدي تطورًا سريعًا في مجال التعليم العالي، وأصبح يوفر فرصًا جديدة لتعليم اللغة الإنجليزية كلغة أجنبية وتعلمها، إلى جانب ما يرافق ذلك من تحديات تربوية وأخلاقية ومؤسسية. تهدف هذه المراجعة النقدية إلى تحليل الدراسات الحديثة التي تناولت استخدام الذكاء الاصطناعي التوليدي في تعليم اللغة الإنجليزية كلغة أجنبية في مؤسسات التعليم العالي. واعتمدت الدراسة على مراجعة ثماني عشرة دراسة علمية محكمة، شملت دراسات تجريبية ومراجعات منهجية، نُشرت خلال الفترة من 2024 إلى 2026، وتم الحصول عليها من قواعد بيانات Scopus وGoogle Scholar وScienceDirect. أظهرت نتائج المراجعة أن الذكاء الاصطناعي التوليدي يمكن أن يدعم تعلم اللغة من خلال تقديم تقييم للطالب لمعرفة أخطاءه وتحسين ادائه، وتصميم أنشطة تعليمية تتناسب مع احتياجات كل طالب ومستواه، ودعم التقويم التكويني، وزيادة مشاركة الطلاب في التعلم. وكانت أقوى الأدلة على فاعليته في تنمية مهارات الكتابة والقراءة، وتحسين القواعد والمفردات، وتقديم ملاحظات تقييم. أما الأدلة المتعلقة بتنمية مهارات التحدث والاستماع، أو بدعم الطلاب ذوي المستويات اللغوية المنخفضة، فما زالت محدودة وتعتمد في الغالب على عدد محدود من الدراسات. كما أشارت النتائج إلى أن استخدام الذكاء الاصطناعي يمكن أن يعزز استقلالية المتعلم والتعلم المنظم ذاتيًا، ويساعد المعلمين على تبني أساليب تدريس أكثر تركيزًا على الطالب. وفي المقابل، أبرزت المراجعة عددًا من التحديات، مثل النزاهة الأكاديمية، والاعتماد المفرط على مخرجات الذكاء الاصطناعي، وأمن البيانات، والكفاءة في استخدام الذكاء الاصطناعي، وتأهيل المعلمين، والحاجة إلى سياسات مؤسسية واضحة. وتشير الأدلة إلى أن هذه التحديات ترتبط بطريقة توظيف الذكاء الاصطناعي في العملية التعليمية أكثر من ارتباطها بالتقنية نفسها، وأن أفضل النتائج تتحقق عندما يُستخدم بوصفه أداة تدعم دور المعلم ولا تحل محله. شريطة أن يستخدم ضمن أطر تربوية ومؤسسية مناسبة. وتختتم الدراسة بمناقشة الآثار التطبيقية لهذه النتائج في مؤسسات التعليم العالي، مع الإشارة إلى جامعة بنغازي بوصفها مثالًا توضيحيًا، كما تقدم مجموعة من المقترحات للبحوث المستقبلية، من أهمها إجراء دراسات طويلة المدى، وتوسيع نطاق البحث ليشمل مناطق جغرافية أكثر تنوعًا، ودراسة أثر الذكاء الاصطناعي على المتعلمين باختلاف مستوياتهم في اللغة الإنجليزية.

الكلمات المفتاحية: الذكاء الاصطناعي التوليدي، اللغة الإنجليزية كلغة أجنبية، التعليم العالي، تعلم اللغة، الكفاءة في اللغة الإنجليزية، استقلالية المتعلم، جاهزية المعلمين، الذكاء الاصطناعي في التعليم.

1. Introduction

The rapid development of Generative Artificial Intelligence (GenAI) has created new opportunities for teaching and learning in higher education. Unlike traditional educational technologies, GenAI can generate content, adapt to learner needs, and provide real-time support. The availability of large language models, such as ChatGPT and other AI applications, has increased their use in education, encouraging educators and researchers to reconsider approaches to teaching, assessment, and student learning in English as a Foreign Language (EFL) education.

In EFL higher education, GenAI can help address challenges associated with teaching students with different levels of English proficiency and learning needs. Research suggests that AI tools can provide personalized learning experiences, adaptive feedback, independent practice opportunities, and flexible learning pathways beyond the classroom. Studies reported improvements in writing, reading, vocabulary development, grammar, learner engagement, motivation, and self-regulated learning. However, as this review will show, the evidence is not equally strong across all of these areas. The strongest evidence is for writing, reading, grammar, vocabulary development and formative feedback, while evidence for speaking, listening, and language proficiency remains more limited.

Despite these benefits, the growing use of GenAI in language education has also raised several concerns. These include academic integrity, overreliance on AI-generated content, privacy, algorithmic bias, assessment validity, unequal access to technology, and the reliability of AI-generated responses. Researchers also highlighted the changing role of teachers, the importance of AI literacy, and the need for clear institutional policies. This suggests that the value of GenAI depends largely on how it is used in education.

Previous studies have explored different aspects of GenAI in EFL higher education, including pedagogical practices, learner autonomy, assessment, ethical issues, teacher readiness, and institutional implementation. However, these topics have often been examined separately. Bringing them together provides a more comprehensive understanding of GenAI in EFL higher education, particularly in contexts where learners have different English proficiency levels and require differentiated instructional support.

Accordingly, this review brings together findings from recent studies to provide a comprehensive overview of GenAI in EFL higher education. It aims to present a balanced discussion of its opportunities and challenges while identifying where the evidence is strong, where it is still limited and where further research is needed.

Although this review is based on international studies, its practical implications are discussed with reference to Benghazi University, a Libyan public university where English is taught as a foreign language, and where students have different levels of English proficiency. Benghazi University is used as an example to consider how these findings may inform EFL higher education in similar contexts.

2. Methodology

2.1 Research Design

This study uses a critical literature review approach to examine the opportunities and challenges of GenAI in EFL higher education. Unlike a systematic review, which aims to identify all available evidence through a comprehensive search, a critical review brings together and evaluates existing research to provide a broader understanding of the topic. This approach is suitable because GenAI in education is a rapidly developing field that includes educational, technological and pedagogical perspectives.

The review identifies common themes and compares findings across studies. It focuses on English language proficiency, teaching practices, learner autonomy, teacher readiness, ethical issues, and institutional implementation.

2.2 Literature Search Strategy

The literature search was conducted using three major academic databases: Scopus, Google Scholar, and ScienceDirect. These databases were selected because they include a wide range of peer-reviewed research in education, applied linguistics, educational technology, and artificial intelligence.

Searches were performed using combinations of keywords related to Generative Artificial Intelligence and English language education. The main search terms included "Generative AI," "ChatGPT," "Large Language Models," "English as a Foreign Language," "EFL," "higher

education," "language learning," "English proficiency," "teacher readiness," "AI literacy," "formative assessment," and "artificial intelligence in language education." The search terms were combined using AND and OR to make the search more accurate and find relevant studies on different aspects of GenAI in EFL higher education.

The search focused primarily on studies published between 2024 and 2026 to include the most recent studies on the use of large models in education.

2.3 Study Selection Process

The initial search of Scopus, Google Scholar, and ScienceDirect identified a number of relevant studies. The title and abstracts were screened using the inclusion and exclusion criteria. Studies that appeared relevant were then reviewed in full. Studies were excluded if they did not focus on education, examined AI applications unrelated to English language education, or did not meet the study criteria. After the screening process, 18 studies met the inclusion criteria and were used in the review.

2.4 Inclusion and Exclusion Criteria

Studies were selected according to predefined inclusion and exclusion criteria to ensure that the reviewed literature was relevant to the objectives of this study.

The inclusion criteria were as follows:

- Peer-reviewed journal articles and systematic literature reviews.
- Quantitative, qualitative, and mixed-methods empirical studies.
- Studies examining the educational applications of Generative Artificial Intelligence in English language teaching and learning within higher education.
- Studies focusing on one or more of the following themes: English language proficiency, pedagogical opportunities, learner autonomy, assessment, teacher readiness, ethical issues, or institutional implementation.
- Studies published in English between 2024 and 2026.

Studies were excluded if they:

- Focused primarily on the technical development of artificial intelligence without an educational focus.
- Examined AI applications unrelated to English language education.
- Were editorials, opinion papers, conference abstracts, dissertations, or unpublished manuscripts without sufficient methodological detail.
- Were duplicate publications reporting the same findings.

2.5 Approach to Synthesis

The eighteen studies included in this review used different research designs, including systematic literature reviews, randomised controlled trials, and mixed-methods studies. Because of this variation, a thematic synthesis approach was used. Each study was first summarised and then compared with the others to identify common themes, similarities, and differences. Six main themes were identified: general opportunities of GenAI in the EFL context, English language proficiency, pedagogical opportunities, ethical and educational challenges, teacher readiness, and institutional implementation. These themes form the basis of section 3. When several studies reported similar findings, this was highlighted. Where evidence came from only one or a few studies, this is clearly indicated.

3. Literature Review

3.1 General Overview of Generative AI in EFL Higher Education

GenAI has become an important topic in EFL higher education because of its increasing use in teaching and learning. Recent studies show that it is being used for instructional support, feedback, assessment, lesson planning, material development, and personalised learning. These developments suggest that GenAI can improve language learning skills and proficiency, but

they also raise pedagogical, ethical and institutional challenges (Deng & Jamaludin, 2026; Griche & Bennis, 2026).

Across the reviewed studies, GenAI is described as a useful tool for language learning. It can provide immediate feedback, individual support, and opportunities for learner autonomy. Griche and Bennis (2026), who reviewed 152 empirical studies, found that the main benefits of GenAI in higher education include personalised learning, affective support, language proficiency development, creativity, critical thinking, and teacher support. On the other hand, Deng and Jamaludin (2026) explored 50 empirical studies concentrated mainly in East Asia and the Middle East, which focused more on function: their findings show that GenAI can act as a feedback provider, teaching assistant, instructional designer, and learning partner, supporting both students and teachers.

Individual studies also support these findings. Alhitty (2026) found that EFL university students believed GenAI improved learner engagement, motivation, writing, vocabulary, grammar, and self-regulated learning. Many students also continued using GenAI outside the classroom, suggesting it can encourage independent language learning. Al Mahmud (2026) also found that teachers viewed GenAI as a useful tool for language assessment, while emphasising that human judgement remains essential.

Although these studies highlight the benefits of GenAI, they also emphasise that it should not replace teachers. The main concerns include academic integrity, privacy, algorithmic bias, inaccurate responses, overreliance on AI, and limited teacher preparation. These issues are discussed in more detail in section 3.4. To sum up, the studies suggest that GenAI can improve EFL teaching and learning when it is supported by well-prepared teachers, clear institutional policies, and responsible use (Al Mahmud, 2026; Alhitty, 2026; Deng & Jamaludin, 2026; Griche & Bennis, 2026).

3.2 Generative AI and English Proficiency Development

Both empirical research and systematic reviews show that GenAI can improve English language development among university students learning EFL. The strongest evidence is for writing, reading, vocabulary, and grammar, although studies also report improvements in listening, speaking, pronunciation, text analysis, and summarisation (Deng & Jamaludin, 2026; Griche & Bennis, 2026; Limiya & Kumar, 2025).

Writing is the most studied language skill in the reviewed literature. Three systematic reviews found that writing showed the greatest improvement with the use of GenAI. These improvements include generation, grammar, vocabulary, text organisation, coherence, sentence structure, and revision through immediate and personalised feedback (Deng & Jamaludin, 2026; Griche & Bennis, 2026; Limiya & Kumar, 2025). These findings are reinforced by Chan et al. (2024). In their randomised controlled trial of 918 university students in Hong Kong, students who received GPT-3.5 feedback produced better revised essays than those who did not. They also reported higher levels of motivation and engagement.

Some other studies also report benefits for other language skills. Unlike some reviews that focus mainly on writing, Limiya and Kumar (2025) demonstrate that AI benefits the entire spectrum of English language learning. In their systematic review of 55 articles, they indicated that GenAI can significantly enhance all four English language skills by creating personalised and adaptive learning environments through tailoring activities to learners' proficiency levels, interests, and needs.

Evidence on the use of AI to improve speaking skills is less consistent than writing. Although systematic reviews report benefits for speaking and pronunciation, relatively few empirical studies have examined these outcomes directly. Cheon et al. (2025), through a randomised field experiment that included 248 South Korean university participants, found that students who practised with AI conversational agents used a wider range of vocabulary than those who practised with human conversation partners. The study also found that students with below-

average English proficiency benefited the most. However, because these findings come from a few studies, more research is needed to confirm them. Similarly, Griche and Bennis (2026) demonstrated that GenAI facilitates improvements in speaking and pronunciation, as well as willingness to communicate, by creating a supportive, non-judgemental learning environment. In addition, their study also indicated that GenAI can develop reading, listening, vocabulary, grammar, summarisation, and text analysis by delivering real-time feedback, supporting interactive learning and providing personalised learning pathways. Alhitty (2026) also found that university EFL students believed GenAI helped improve their vocabulary, grammar, writing, and confidence in using English.

In summary, the reviewed literature suggests that GenAI can be a valuable tool for English language learning in higher education. The strongest evidence supports improvements in writing, grammar, and vocabulary, while evidence for other skills is more limited. Some studies also suggest that GenAI may be particularly beneficial for students with lower levels of English proficiency. However, this conclusion was reported only in one study (see Section 4.4).

3.3 Pedagogical Opportunities of Generative AI in EFL Higher Education

Besides supporting language proficiency, GenAI also offers several benefits for EFL teaching. It supports more learner-centred and flexible teaching by providing adaptive tasks for different language skills, promoting active classroom learning, encouraging a supportive learning environment, supporting independent learning, and reducing teachers' workload. (Yeşil, 2025; Lukešová & Jennings, 2026; Deng & Jamaludin, 2026).

One important pedagogical advantage of GenAI is its ability to provide personalised learning experiences. Yeşil (2025) states that AI shifts language teaching from teacher-centred instruction toward learner-centred learning, where students become more active participants in their own learning process. Limiya and Kumar (2025) highlighted that GenAI can adapt learning activities and feedback to students' individual needs, learning pace, and English proficiency level.

GenAI can also help students become more independent learners. Yeşil (2025) showed that AI-supported language learning promotes learner autonomy by enabling students to take greater responsibility for planning, monitoring, and evaluating their own learning. Liu and Zhou (2026) investigated how AI-assisted task-based language learning influences English proficiency among 412 Chinese university undergraduates by examining the mediating roles of cognitive engagement and self-regulated learning, as well as the moderating role of AI trust. Their study shows that AI-based language learning tasks significantly improved students' English proficiency. They also concluded that students who trusted AI were more likely to engage effectively with learning activities using AI.

Cheon et al. (2025) also highlighted that AI conversational agents can reduce students' anxiety caused by fear of being negatively judged during language learning. Moreover, Self-paced learning increased learner autonomy and enhanced language development. This was linked to better learning outcomes, especially for students with lower levels of English proficiency. It should be noted that this result comes from a single study included in this review.

GenAI can also support more interactive and student-centred teaching. Amuseghan and Emmanuel (2026), in their Quasi-experimental study, examined a ChatGPT-supported classroom modelling in which students used ChatGPT before class to prepare for lessons, while class time was used for discussion, group work, and problem solving. Their study found that these students achieved better academic results and showed a higher level of independent learning than those taught using traditional methods. In addition, Lukešová and Jennings (2026), in their Empirical classroom study, used ChatGPT to enhance 58 students' writing revision through indirect prompts instead of direct corrections. This approach encouraged students to reflect on their work, revise independently, and become active learners.

To sum up, these studies suggest that the pedagogical benefits of GenAI go beyond improving language skills. They also promote personalised learning, learner autonomy, and formative feedback. These teaching approaches may explain the improvement in English language proficiency discussed in Section 3.2. Nevertheless, the reviewed studies did not directly examine this relationship.

3.4 Challenges and Ethical Concerns of Generative AI in EFL Higher Education

Despite the benefits discussed above, the reviewed studies also highlight several ethical and educational challenges related to the use of GenAI in EFL higher education. These include academic integrity, overreliance on AI, reduced critical thinking, limited AI literacy, data privacy, equality, and assessment validity (Chuang & Yan, 2025; Maleki, 2026).

Academic integrity is one of the main concerns discussed in the literature. Chuang and Yan (2025), in their systematic review of 77 studies, reported that AI-assisted academic dishonesty is a major challenge in language assessment. Educators are increasingly concerned that AI-generated work may compromise authentic assessment and make it difficult to determine students' actual language ability. They also indicated that AI detection tools, such as Turnitin and GPTZero, can identify fully AI-generated texts with reasonable accuracy, but are less reliable when students combine AI-generated and their own writing. As a result, the authors argued that such tools should not be used on their own to make decisions about academic dishonesty.

The way students use GenAI may be more important than whether they use it. Alsadoon (2026) identified three patterns of AI use among EFL female students at Imam Mohammad Ibn Saud Islamic University: Constructive, hybrid, and substitutive. The study indicated that the students who used ChatGPT for brainstorming, language support, and revision generally showed stronger academic integrity than those who relied on AI to complete large parts of their work. This suggests that the main concern is not in the use of AI itself, but how it is used to support learning rather than replacing it. Furthermore, Interview data showed that many students were unsure what their university actually permitted. The authors argue that clearer institutional guidance and policy awareness are needed to support responsible AI use.

Overreliance on AI is another concern discussed in several studies. Relying too much on AI-generated responses may reduce students' opportunities to think independently and develop their language skills if they accept AI-generated content without evaluating it (Deng & Jamaludin, 2026; Griche & Bennis, 2026; Limiya & Kumar, 2025; Chuang & Yan, 2025). In addition, Chan et al. (2024) reported that although students valued AI-generated feedback, some still preferred feedback from teachers because they consider it more personalised and emotionally supportive.

AI literacy is another common theme in the reviewed studies. For instance, Zhang et al. (2024) investigated how digital literacy, AI literacy, and trust influence different types of ChatGPT use among 947 U.S. college students and how these patterns relate to academic self-efficacy and perceived academic creativity through a digital divide perspective. They reported that AI literacy was more important than general digital literacy in helping students use ChatGPT effectively. The study also shows that students who used ChatGPT for productive educational uses such as brainstorming, feedback, language learning, writing support, and study planning reported higher levels of self-efficacy and creativity.

Finally, privacy, fairness, and equal access remain important issues. Chuang and Yan (2025) highlighted concerns about data privacy, algorithmic bias, and unequal access to AI tools. AI systems may reflect biases embedded in their training data, potentially disadvantaging learners from different linguistic or cultural backgrounds. In addition, unequal access to AI tools may create assessment inequalities because some students receive more effective AI support than others. Using AI systems often requires uploading students' written or spoken work. The authors note that this raises concerns regarding data privacy, emphasising the need for

responsible AI implementation. Likewise, Maleki (2026) found that privacy concerns, limited access to technology, and cultural bias can make it more difficult to use GenAI for online formative assessment. Teachers in the study also questioned the security of student data uploaded to commercial AI platforms.

Across these studies, the challenges of GenAI depend more on how it is used. Academic integrity, overreliance, limited AI literacy, and privacy issues require responsible use, clear institutional policies, and human guidance.

3.5 Teacher Readiness and Institutional Implementation

Addressing the challenges identified in Section 3.4 depends mainly on well-prepared teachers, institutional support, and clear implementation strategies. Teacher readiness involves more than basic digital skills.

Wang et al. (2026) investigate how 187 university EFL teachers' affective and cognitive responses influence their willingness to adopt ChatGPT for teaching using an extended Technology Acceptance Model (TAM). They indicate that teachers' confidence in using ChatGPT and institutional support increased their acceptance of the technology. However, teachers also reported concerns about reliability, digital equity, and how ChatGPT fits with teaching practices. The authors also emphasise that AI should complement, not replace, teachers because it lacks human qualities such as emotional intelligence and meaningful interaction. Instead, teachers viewed AI as a complementary tool capable of supporting lesson planning, routine tasks, and personalised learning while educators retained responsibility for teaching and learner support. Therefore, the authors stress that universities should provide targeted training and practical support to increase teachers' confidence and effective AI integration.

Al Mahmud (2026) found that EFL teachers generally had a high level of GenAI assessment literacy and were confident in using AI to design assessments. However, they emphasised the need for human judgement when evaluating AI-generated assessment materials while maintaining academic integrity, student privacy, and ethical guidelines.

Institutional support is also important for the successful use of GenAI. Maleki (2026) showed that teachers faced several challenges, including technical problems, heavy workloads, limited preparation time, limited AI literacy, few opportunities for professional development, and lack of institutional guidance. The study further emphasised that successful GenAI integration requires AI literacy, teacher training, clear policies, and technical resources. Alsadoon (2026) also found differences between students' and teachers' understanding of acceptable AI use, highlighting the need for clearer institutional policies. Liang et al. (2026) also reported that successful GenAI integration depends on appropriate teaching practices, AI literacy, ethical governance, and institutional support. Similarly, Yeşil (2025) and Amuseghan and Emmanuel (2026) emphasised the significance of clear guidance on the appropriate use of AI, while Lukešová and Jennings (2026) recommended continuous professional development instead of one-time technology training.

In summary, these findings show that successful GenAI implementation requires collaboration between teachers and institutions. Teacher preparation, clear policies, and ongoing professional development all play an important role in ensuring that GenAI is used effectively in EFL higher education.

4. Discussion

4.1 Synthesis of the Main Findings

The reviewed research in this study highlights that GenAI has become a valuable tool in EFL higher education, with the strongest evidence for writing, grammar, reading and vocabulary development. These areas are supported by both systematic reviews and empirical studies. Evidence for other outcomes, such as speaking and listening proficiency and support for lower

proficiency learners, is more limited and often based on a single study rather than several studies reporting similar findings (see Section 3.2 and 3.3). One possible explanation is that GenAI is still a new area of research, with writing receiving more attention than other language skills.

Another important point is the type of evidence used in the reviewed studies. Many studies relied on students' and teachers' opinions or short-term classroom activities instead of measuring actual improvement in English proficiency over a longer period. Therefore, findings such as higher motivation, engagement, or confidence reflect participants' experience rather than objective learning outcomes. In contrast, the few studies that used objective measures such as essay-quality outcomes or lexical diversity measures provide stronger evidence. However, they focused on specific language skills rather than overall English proficiency.

A related issue concerns the relationship between the teaching approaches supported by GenAI and improvement in English language proficiency discussed in Section 3.2. The reviewed studies report both teaching and language learning benefits, but they do not directly examine whether one leads to the other. Although there seems to be a connection, no single study has demonstrated it directly. This is a gap that has not yet been addressed in the literature, and any link should be interpreted with caution until further evidence is available.

Finally, across the literature on ethical issues, teacher readiness, and institutional implementation (Sections 3.4–3.5), a consistent pattern emerges. The risks associated with GenAI, including overreliance, academic integrity issues, and unclear institutional expectations, are repeatedly linked to limited AI literacy, insufficient teacher training, and unclear policies, indicating that these challenges are shaped mainly by implementation. These findings emphasise the need for clear guidance, teacher preparation, and responsible AI use in higher education.

4.2 Implications for EFL Higher Education

The findings of this review have several implications for EFL higher education. However, these implications should be interpreted with caution because of the limitations discussed in Section 4.4.

Based on the available evidence, EFL higher education institutions may benefit from introducing GenAI first in writing instruction, grammar, reading, vocabulary, revision support, and formative feedback, where the current evidence is strongest. Applications to speaking, listening, and support for learners with lower levels of English proficiency should be introduced more cautiously until stronger evidence becomes available.

Another important implication concerns the role of teachers. Rather than replacing teachers, GenAI should be used to support teaching while teachers continue to guide learning, evaluate AI-generated content, and make decisions about assessment. This suggests that teacher training should focus not only on how to use AI tools but also on how to integrate them effectively into teaching and learning.

Providing access to AI tools is not enough on its own. Students also need AI literacy to use these tools effectively. EFL higher education institutions may therefore need to include AI literacy as part of their curriculum so that students learn to use GenAI responsibly and productively. This is consistent with the reviewed evidence, which indicates that AI literacy is more important than general digital literacy in helping students use GenAI effectively for learning rather than relying on it to complete their work.

Finally, ethical considerations should be included from the beginning when GenAI is introduced into EFL higher education. Instead of treating issues such as academic integrity, privacy, and equal access as separate concerns, institutions should consider them when planning AI-supported learning activities and assessment. This approach reflects one of the main conclusions of this review: many of the challenges are largely influenced by how GenAI is integrated into teaching and assessment.

4.3 Practical Implications for Benghazi University and Similar Contexts

As explained in the Introduction, Benghazi University is used as an example to illustrate how the findings of this review may apply to an EFL higher education context. However, this review did not investigate Benghazi University directly. Therefore, the following points are offered as evidence-based considerations rather than conclusions about the university itself.

Based on the findings of this review, Benghazi University and similar EFL institutions may benefit from introducing GenAI gradually, beginning with areas where the evidence is strongest, particularly writing instruction, feedback, and revision support, rather than applying it across all language skills at the same time. Any wider use of GenAI should be accompanied by training for both teachers and students, so that it can be used effectively and responsibly.

Given that students at Benghazi University enter with different levels of English proficiency, the introduction of GenAI may also need to take these differences into account. Since the available evidence for learners with lower levels of English proficiency is still limited, the use of GenAI should be evaluated carefully across students with different proficiency levels before it is introduced more widely.

Teacher training, clear institutional policies, and adequate technical resources are essential for the successful use of GenAI. Since most of the available evidence comes from East Asia and the Middle East, and research on GenAI is still developing (see Section 4.4), any implementation should include regular evaluation to monitor its outcomes in the local context rather than assuming that findings from other institutions will apply in the same way.

4.4 Limitations of the Reviewed Literature and Future Research

The findings of this review should be interpreted considering several limitations in the available evidence. Most of the reviewed studies were published between 2024 and 2026, so current knowledge is based mainly on short-term research. As a result, the long-term effects of GenAI on English proficiency, learner autonomy, and academic achievement remain unclear. In addition, most studies were conducted in mainly East Asia and the Middle East, with relatively little research from Africa, Europe, and other regions. This limits how far the findings can be applied to other contexts, including Benghazi University.

Another limitation is the uneven focus of the existing research. For example, writing has received much more attention than other language skills, while evidence on speaking and listening, and on learners with lower levels of English proficiency, remains limited. In addition, many studies relied on participants' perceptions and short-term classroom interventions rather than direct measures of English language proficiency or long-term outcomes. Only a small number of studies used rigorous experimental designs, making it difficult to draw strong conclusions about the effects of GenAI on English language learning.

Future research should focus on the main gaps identified in this review. In particular, more evidence is needed on the long-term effects of GenAI, its impact on speaking and listening, its use with learners who have lower levels of English proficiency, and its application in under-researched contexts such as Libya. Research should also examine more directly how AI-supported teaching approaches influence English language proficiency and how institutions develop effective policies for the use of GenAI.

5. Conclusion

Generative Artificial Intelligence (GenAI) is becoming an important part of EFL higher education, with growing applications in language teaching, learning, and assessment. This critical review examined evidence from eighteen recent empirical and systematic reviews to explore the opportunities, challenges, and implications of using GenAI in EFL higher education.

The review shows that GenAI can support language learning through personalised feedback, adaptive learning, and opportunities for independent practice. The strongest evidence is for writing development, reading, grammatical awareness, vocabulary and formative feedback, while evidence for speaking, listening and learners with lower English language proficiency is still developing. The review also highlights that GenAI can support teachers in assessment, lesson planning, and learner-centred teaching when it is used within appropriate pedagogical, ethical, and institutional frameworks. At the same time, challenges such as academic integrity, AI literacy, overreliance, data privacy, and institutional guidance remain important. The findings indicate that these challenges depend mainly on educational practice and institutional support.

Several areas for further research are noted. In particular, more evidence is required in speaking and listening development as well as for learners with lower levels of English proficiency, and the long-term effects of GenAI across different educational contexts, especially in under-researched African EFL settings.

In conclusion, GenAI offers valuable opportunities to support both teaching and learning in EFL higher education, but its success depends on how it is integrated into educational practice. Careful implementation, teacher preparation, clear institutional policies, and continued research will be essential to ensure that GenAI supports meaningful language learning while addressing the challenges identified in this review.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] Al Mahmud, F. (2026). *EFL instructors' assessment literacy for generative AI in Saudi higher education: Readiness, ethical concerns, and policy needs*. *Education Sciences*, 16, 1081. <https://doi.org/10.3390/educsci16071081>
- [2] Alhitty, A. A. (2026). *Learner engagement & motivation in EFL higher education: The pedagogical role of generative AI tools in educational ecosystems* (Doctoral dissertation, Lancaster University). <https://doi.org/10.17635/lancaster/thesis/3157>
- [3] Alsadoon, R. (2026). *Perceptions vs. practices: Academic integrity and actual ChatGPT use among EFL students*. *Frontiers in Psychology*, 17, 1796737. <https://doi.org/10.3389/fpsyg.2026.1796737>
- [4] Amuseghan, A. S., & Emmanuel, I. F. (2026). *Effects of ChatGPT-flipped instruction on the academic performance and self-regulated skill of L2 pre-service teachers in Ondo State, Nigeria*. *Hungarian Educational Research Journal*, 16(1), 1–17. <https://doi.org/10.1556/063.2025.00389>
- [5] Chan, S. T. S., Lo, N. P. K., & Wong, A. M. H. (2024). *Enhancing university level English proficiency with generative AI: Empirical insights into automated feedback and learning outcomes*. *Contemporary Educational Technology*, 16(4), ep541. <https://doi.org/10.30935/cedtech/15607>
- [6] Cheon, G., Choi, Y., Lee, D., & Baek, J. (2025). *Generative AI agents in language learning: A randomized field experiment*. In *Proceedings of the 58th Hawaii International Conference on System Sciences* (pp. 4949–4958). University of Hawai'i at Mānoa. <https://hdl.handle.net/10125/109443>
- [7] Chuang, P.-L., & Yan, X. (2025). *Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions*. *System*, 134, 103846. <https://doi.org/10.1016/j.system.2025.103846>

- [8] Deng, L., & Jamaludin, K. A. (2026). *Roles of generative artificial intelligence (GenAI) in English as a foreign language (EFL) instruction: A systematic literature review*. SAGE Open. <https://doi.org/10.1177/21582440261418315>
- [9] Griche, M. S., & Bennis, M. (2026). *Generative Artificial Intelligence Integration in Foreign Language Education in Higher Education: A Systematic Literature Review*. *Technology in Language Teaching & Learning*, 8, 102827. <https://doi.org/10.29140/tltl.2026.102827>
- [10] Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). *Generative AI (GenAI) in the language classroom: A systematic review*. *Interactive Learning Environments*, 34(1), 335–359. <https://doi.org/10.1080/10494820.2025.2498537>
- [11] Liang, Z., Yang, K., Sha, L., Gašević, D., Yan, L., & Chen, G. (2026). *A systematic review of generative AI in education: Empirical insights from a human–AI interaction perspective*. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.70055>
- [12] Limiya V. G., A., & Kumar N. S., P. (2025). *Generative artificial intelligence in enhancing English language skills: A systematic review*. *Studies in Media and Communication*, 13(4), 256–267. <https://doi.org/10.11114/smc.v13i4.7763>
- [13] Liu, J., & Zhou, J. (2026). *AI-assisted task-based language learning and English proficiency: A dual-pathway model of cognitive engagement and self-regulated learning with AI trust as a boundary condition*. *Frontiers in Psychology*, 17, 1842773. <https://doi.org/10.3389/fpsyg.2026.1842773>
- [14] Lukešová, A., & Jennings, P. J. (2026). *Clue before correction: ChatGPT-enhanced strategy for promoting autonomous and reflective language learning*. *Innovation in Language Learning and Teaching*. Advance online publication. <https://doi.org/10.1080/17501229.2025.2612532>
- [15] Maleki, A. (2026). *"We cannot incorporate them because ...": Challenges of integrating generative AI into online formative assessments in EFL contexts*. *Computers in Human Behavior Reports*, 21, 100934. <https://doi.org/10.1016/j.chbr.2026.100934>
- [16] Wang, C., Zou, B., Zhang, W., Du, Y., & Hu, W. (2026). *Understanding EFL teachers' affective and cognitive responses to ChatGPT in higher education*. *Humanities and Social Sciences Communications*, 13, 822. <https://doi.org/10.1057/s41599-026-07360-3>
- [17] Yeşil, Ş. (2025). *Fostering learner autonomy in AI-mediated language learning environments: A review of literature*. *ELT Research Journal*, 14(2), 243–259. <https://doi.org/10.71362/eltrj.1817405>
- [18] Zhang, C. (X.), Rice, R. E., & Wang, L. H. (2024). *College students' literacy, ChatGPT activities, educational outcomes, and trust from a digital divide perspective*. *New Media & Society*, 28(2), 673–695. <https://doi.org/10.1177/14614448241301741>

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of CJHES and/or the editor(s). CJHES and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.