

Artificial Intelligence in Improving the Domain of English Language Learning

Manal Ali

AlShaab University, Baghdad, Iraq
Email: manal.ali@alshaab.edu.iq

Received : 30 June - 2026

Accepted : 25 August - 2026

Published online : 28 August - 2026

Abstract

Artificial Intelligence (AI) has been widely promoted as a transformative force in English language teaching, yet claims about its pedagogical value have often outpaced the available evidence. This paper appraised the methodological strengths and limitations of recent research in this area and identified recurring patterns across that work. A narrative critical review was conducted, in which seven peer-reviewed studies were purposively selected from databases such as Scopus and Google Scholar to represent variation in methodology, systematic review, survey, interview, and bibliometric analysis, and in geographical context. Each study was appraised against its stated aims, design, sample, findings, and acknowledged limitations; PRISMA guidelines were not followed and no standardized appraisal instrument was applied. The studies converged on adaptive feedback and personalized learning as the mechanisms through which AI most plausibly supported language acquisition, and on teacher training as a precondition for adoption. Reported benefits rested largely on self-reported perception data gathered through closed questionnaires. Samples were frequently small and single-site, four of the seven studies were situated in Middle Eastern or Central Asian contexts, and none tracked learners longitudinally or compared outcomes against a control condition. The reviewed literature therefore established the plausibility of AI's contribution to English language learning rather than confirming its effectiveness, indicating a need for controlled, outcome-based research across more diverse settings.

Keywords: Artificial Intelligence, Education, English Language Learning, Linguistics, Natural Language Processing (NLP).

1. Introduction

Artificial intelligence (AI) has moved rapidly from the periphery of language education to its center. English, sustained by demand across business, science, and culture (Crystal, 2003), has been among the first domains to absorb it. AI refers to a set of technologies designed to emulate aspects of human intelligence, principally machine learning and natural language processing (Russell & Norvig, 2020), and what distinguishes these from earlier classroom technologies is their capacity to respond. Systems can process learner interaction data at scale to adjust content against individual level and pace (Luckin et al., 2016; Woolf et al., 2013), construct customized learning paths from those interactions (Pane et al., 2014), and support interactive strategies aligned to learner interest rather than fixed sequence (Heffernan & Heffernan, 2014). In language learning specifically, conversational applications now allow learners to rehearse speaking with systems that simulate human interlocutors, offering real-time correction of pronunciation and fluency (Divekar et al., 2021), and such interactive practice has been associated with gains in retention and critical thinking (Carter, 2009). Parallel developments in learning analytics allow teachers to identify individual strengths and weaknesses from performance data (Baker & Inventado, 2014) and adapt instruction



accordingly (Means et al., 2013), while generative tools produce videos, games, and quizzes that sustain engagement (Clark & Mayer, 2008; Plass et al., 2020).

Research has followed this uptake closely, and the limitations are as well documented as the affordances. Unequal access to technology risks widening disparities between learners from different economic backgrounds (Warschauer, 2003), and systems remain imperfect at interpreting cultural and grammatical nuance (Bender, 2013). Direct engagement with teachers and peers continues to underpin the development of critical thinking and social communication (Vygotsky, 1978), which is why AI is widely positioned as a supplement to rather than a replacement for human instruction (Alam, 2021). Several reviews have synthesized this body of work, mapping tool categories, application areas, and research distribution (Zawacki-Richter et al., 2019).

What this literature has not examined is its own evidential foundation. Existing reviews are largely descriptive: they establish what has been studied rather than how well, and they aggregate findings without interrogating the designs that produced them. This matters because the field's claims and its evidence appear to be diverging. Assertions that AI can transform language education circulate widely (Luckin, 2018), yet much of the underlying research relies on perception surveys, single-institution samples, and short observation windows, designs that measure attitudes towards a technology rather than its effects on learning. Where such limitations are acknowledged, they are typically listed rather than examined for their consequences.

This paper addresses that gap. It appraises a purposively selected set of recent studies on AI in English language teaching and asks three questions: what research designs currently dominate the field; what those designs can and cannot support in the way of conclusions; and which methodological gaps most constrain confidence in the reported benefits. The aim is not to establish whether AI improves English language learning, since the available evidence does not settle that, but to clarify what would be required to answer the question and to give researchers and practitioners a more accurate account of where the field presently stands.

2. Literature Review

2.1. The Scope of AI Adoption in ELT

Research on AI in English language teaching has grown quickly, though it has not settled into agreement about how far the technology has actually reached classrooms. Tobing et al. (2023) found that university instruction still underuses chatbots and virtual reality, and that mobile applications, where adopted at all, concentrate narrowly on vocabulary rather than integrated skills. Sharma and Suvin (2026) and Crompton et al. (2024) describe a rather different field, one already reshaped by adaptive systems delivering personalized instruction. The accounts may be measuring different things, since the first documents classroom practice while the others draw largely on tool capability and pilot work (Cahyono & Rosita, 2023). That gap between capability and uptake is itself worth noting, and the more optimistic literature tends to pass over it.

2.2. Personalization and Its Evidence Base

Personalization is the benefit most consistently claimed. Sharma and Suvin (2026) attribute engagement gains to adaptive feedback and individually tailored content, while Al-Midlij and Alotaibi (2023) report improved speaking fluency and cultural understanding among 280 Kuwaiti learners. The second claim is stronger than the first, though it rests on survey data, which captures learners' reported sense of improvement rather than measured

gains, and cannot separate the platform's contribution from novelty or from the extra practice that any new tool tends to prompt.

2.3. Method and the Shape of Findings

Method appears to shape findings in ways the literature rarely acknowledges. Alharbi (2025) reports strong support among Saudi teachers and students, framed by Vision 2030 modernization goals, and Al-Midlij and Alotaibi (2023) report similarly favorable results. Both used closed-response instruments with participants already using the technology. Dilzhan (2024), working through open-ended interviews with Kazakhstani teachers, surfaced concerns about over-reliance and academic integrity that questionnaires seem not to elicit. Some of the positive consensus may therefore reflect instrument design as much as the technology itself.

2.4. An Untested Consensus and a Narrow Evidence Base

A further consensus runs through nearly all of this work: that AI should complement teachers rather than replace them. Alharbi (2025), Al-Midlij and Alotaibi (2023), and Dilzhan (2024) all reach this conclusion, and Crompton et al. (2024) build it into their proposed frameworks. No study in this set tests it. Geographic concentration compounds the difficulty. Liu et al. (2024) identify uneven distribution in AI-ELT research, an observation confirmed by the studies reviewed here, several of which are situated in Gulf and Central Asian contexts where national policy actively promotes educational technology. Whether findings from these settings transfer elsewhere remains an open question, and one that Crompton et al. (2024) raise conceptually without testing empirically.

3. Methods

This paper is a narrative critical review rather than a systematic review. It does not follow PRISMA guidelines, and no claim is made that the studies examined constitute an exhaustive or representative sample of the literature on artificial intelligence in English language teaching (ELT). The aim is interpretive: to appraise the strengths and limitations of a small set of recent studies and to identify recurring methodological patterns across them.

3.1. Identification of Sources

Sources were identified through searches of the databases such as Scopus and Google Scholar, using combinations of the terms artificial intelligence, English language teaching, EFL, and language learning. Searches were limited to peer-reviewed articles published in English. The search was not exhaustive, and no formal record was kept of the number of records screened or excluded; consequently, the selection process is not replicable in the sense required of a systematic review.

3.2. Selection of Studies

Seven studies were selected purposively. Selection was guided by three considerations: (a) explicit focus on AI applications in English language teaching or learning; (b) variation in methodological approach, so that systematic reviews, survey research, interview-based studies, and bibliometric analysis were each represented; and (c) variation in geographical context, including studies situated in Saudi Arabia, Kuwait, Kazakhstan, and international settings. Because selection was purposive rather than criterion-driven, the resulting set reflects the authors' judgment about illustrative value and should not be read as a comprehensive mapping of the field.

3.3. Analytical Procedure

Each study was read in full and summarized with attention to its stated aims, methodology, sample, principal findings, and acknowledged limitations. Studies were then appraised individually along two dimensions; methodological strengths and methodological limitations, with appraisal grounded in what each study reported about its own design rather than in an external quality-assessment instrument. No standardized critical appraisal tool (such as CASP or MMAT) was applied, and no formal inter-rater reliability procedure was used; appraisals reflect the authors' reading and were reconciled through discussion. Recurring patterns across the appraisals were subsequently grouped into the thematic categories reported.

4. Results and Discussion

4.1. Methodological Characteristics of the Reviewed Studies

The seven studies differ considerably in design. Two are reviews: Tobing et al. (2023) applies PRISMA guidelines, giving a transparent and replicable selection procedure, while Liu et al. (2024) uses bibliometric methods with documented search parameters. Both are constrained by what these approaches can deliver, since a systematic review inherits the limitations of its constituent studies and bibliometric analysis maps publication activity rather than pedagogical effect. Liu et al.'s case for AI as a classroom tool therefore rests on research volume rather than research findings.

Crompton et al. (2024) proceeds largely by conceptual analysis, offering a framework for professional development and cross-context implementation that is not empirically validated. The remaining studies are empirical. Alharbi (2025) as well as Al-Midlij and Alotaibi (2023) use closed-response questionnaires, the former with a small single-institution sample that limits inference, the latter with 280 Kuwaiti learners. Dilzhan (2024) is the only study using open-ended interviews. Sharma and Suvin (2026) reports on adaptive systems adjusting content in real time, drawing chiefly on tool capability rather than classroom data.

4.2. Patterns Across the Reported Findings

Four of the seven studies rely substantially on participants' reports of their own experience, and this appears to shape their conclusions. Alharbi (2025) and Al-Midlij and Alotaibi (2023) both administer closed instruments to participants already using AI tools, and both report favorable results. The convergence is weaker evidence than it appears, since closed questions constrain the available answers and existing users are a poor test of whether a technology works. Al-Midlij and Alotaibi (2023) finding on speaking fluency illustrates the difficulty: fluency is measurable, but the study measures perception of it. Dilzhan (2024) provides a contrast. Open-ended interviews surfaced concerns about over-reliance and academic integrity that the questionnaire studies do not report, suggesting these concerns exist in survey populations but are not elicited by the instruments used.

A second pattern concerns the distance between capability and practice. Tobing et al. (2023) find that available technologies, including neural machine translation, are not reaching classrooms systematically, while Sharma and Suvin (2026) describe adaptive systems already improving access for disadvantaged learners. Both may be accurate, describing uptake and capability respectively, but neither examines the space between them. Tobing et al. (2023) identify training needs without investigating why teachers resist adoption, and Sharma and Suvin underestimate the preparation required for pedagogically coherent use.

Ethical concerns recur throughout and are consistently raised rather than resolved. Sharma and Suvin (2026) note data privacy without proposing safeguards; Crompton et al. (2024) propose ethical frameworks without engaging algorithmic bias as a mechanism affecting particular learner groups; Dilzhan (2024) documents integrity concerns without examining institutional responses.

The empirical studies are also geographically concentrated. Al-Midlij and Alotaibi (2023), Alharbi (2025), and Dilzhan (2024) are situated in Kuwait, Saudi Arabia, and Kazakhstan respectively, all settings where national policy actively promotes educational technology. The favorable reception reported may reflect the tools, the policy environment, or both. Liu et al. (2024) identify this concentration as a field-level problem and Crompton et al. (2024) argue for local adaptation, though neither tests how outcomes vary across settings.

5. Conclusion

Taken together, the studies examined here suggest that AI has begun to reshape practice in English language teaching, though the supporting evidence is thinner and more geographically concentrated than the enthusiasm surrounding it. The clearest contribution of this body of work is descriptive: the reviews map which tools are in use and where research is concentrated, while the studies of adaptive platforms point to feedback and personalized pacing as the mechanisms through which AI most plausibly affects learning. Studies grounded in particular institutions add what the broader reviews cannot, showing that adoption depends on training, workload, and concerns about academic integrity rather than on technical capability alone. The limitations, however, are substantial and largely shared. Empirical grounding is weak, with several contributions proceeding by conceptual analysis or drawing on small single-site samples. Where data are collected, the instruments are narrow: closed questionnaires capture perception rather than performance, and the richer interview evidence is confined to one national context. Outcomes are predominantly self-reported and short-term, with no longitudinal tracking and no comparison against control conditions. What emerges is not a settled finding that AI improves English language learning, but a body of work that has established the plausibility of that claim and shown where it should be tested.

Several directions follow. The most pressing is experimental and quasi-experimental research measuring language outcomes directly rather than learner perceptions, with follow-up intervals long enough to separate durable gains from novelty effects. Mixed-methods designs would help, since the perception surveys dominating this literature capture attitudes but not the classroom processes producing them. Geographical breadth is a second priority: findings from settings with differing infrastructure, class sizes, and institutional policies are needed before any claim about AI in ELT can be called general. Third, the conditions under which teachers adopt these tools, and the reasons some decline to warrant direct study rather than inference. Finally, the ethical questions raised in passing across this literature, particularly around student data and assessment integrity, would be better served by empirical investigation of current practice than by further calls for frameworks in the abstract.

6. References

- Al-Midlij, N., & Alotaibi, N. (2023). Advantages of Using Artificial Intelligence in Teaching English as a Second Language Public Authority for Applied Education and Training Language Centre. *Multi-Knowledge Electronic Comprehensive Journal For Education & Science Publications (MECSJ)*, 68, 1–13. https://mecsaj.com/uplode/images/photo/musaedalioo7@gmail.com_1.pdf
- Alam, A. (2021). Should Robots Replace Teachers? Mobilisation of AI and Learning Analytics in Education. *2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)*, 1–12. <https://doi.org/10.1109/ICAC353642.2021.9697300>
- Alharbi, J. M. (2025). Adoption of Artificial Intelligence Tools for English Language Learning Among Saudi EFL University Students: The Moderating Role of Faculty. *Journal of Ecohumanism*, 4(2), 804–819. <https://doi.org/10.62754/joe.v4i2.6349>
- Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In *Learning Analytics* (pp. 61–75). Springer New York. https://doi.org/10.1007/978-1-4614-3305-7_4
- Bender, E. M. (2013). Linguistic Fundamentals for Natural Language Processing. In *Synthesis Lectures on Human Language Technologies*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-02150-3>
- Cahyono, A. E., & Rosita, R. (2023). The Impact of Using AI-Based Language Learning Platforms on English Speaking Skills of College Students. *Transformational Language, Literature, and Technology Overview in Learning*, 2(2), 1–8. <https://doi.org/10.55047/transtool.v2i2.1352>
- Carter, M. (2009). Visible learning: a synthesis of over 800 meta-analyses relating to achievement. *Educational Psychology*, 29(7), 867–869. <https://doi.org/10.1080/01443410903415150>
- Clark, R. C., & Mayer, R. E. (2008). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.
- Crompton, H., Jones, M. V., & Burke, D. (2024). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education*, 56(3), 248–268. <https://doi.org/10.1080/15391523.2022.2121344>
- Crystal, D. (2003). *English as a global language*. Cambridge university press.
- Dilzhan, B. (2024). *Teaching English and Artificial Intelligence: EFL Teachers' Perceptions and Use of ChatGPT*. SDU University. <https://doi.org/10.35542/osf.io/fwy92>
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2021). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning*, 35(9), 2332–2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Liu, H., Li, Y., & Luo, W. (2024). Artificial Intelligence in English Language Teaching. *3rd International Conference on Education, Language and Art (ICELA)*, 120–131. https://doi.org/10.2991/978-2-38476-214-9_16
- Luckin, R. (2018). *Machine Learning and Human Intelligence. The future of education for the 21st century*. UCL institute of education press.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An argument for AI in Education*. Pearson. <https://static.googleusercontent.com/media/edu.google.com/en//pdfs/Intelligence-Unleashed-Publication.pdf>
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. *Teachers College Record: The Voice of Scholarship in Education*, 115(3), 1–47. <https://doi.org/10.1177/016146811311500307>

- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at Scale. *Educational Evaluation and Policy Analysis*, 36(2), 127–144. <https://doi.org/10.3102/0162373713507480>
- Plass, J. L., Mayer, R. E., & Homer, B. D. (2020). *Handbook of game-based learning*. Mit Press.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Sharma, B., & Suvin, V. M. (2026). A Descriptive Bibliometric Study of Artificial Intelligence in Education in India. *Indian Journal of Educational Technology*, 8(1), 13–29. <https://doi.org/10.47191/ijsshr/v6-i6-78>
- Tobing, M. B., SNA, F. I., Hayrunnisa, N. R., Haswuri, N. I. T., Sutarsyah, C., & Munifatullah, F. (2023). An Exploration of Artificial Intelligence in English Language Teaching As a Foreign Language. *International Journal of Social Science and Human Research*, 6(6), 3837–3843. <https://doi.org/10.47191/ijsshr/v6-i6-78>
- Vygotsky, L. S. (1978). *Mind In Society: The Development Of Higher Psychological Processes*. Harvard University Press.
- Warschauer, M. (2003). *Technology and Social Inclusion*. MIT Press.
- Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI Grand Challenges for Education. *AI Magazine*, 34(4), 66–84. <https://doi.org/10.1609/aimag.v34i4.2490>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>