ISSN: 2349-6002

Using Large Language Models for Language Learning: Opportunities, Risks, and Pedagogical Implications

Mr. Meet Jayantibhai Pandya
Research Scholar, Department of English, Bhakta Kavi Narsinh Mehta University, Junagadh
doi.org/10.64643/IJIRTV12I8-191229-459

***Abstract*—The emergence of Large Language Models (LLMs) such as OpenAI's GPT series, Google's Gemini, and Anthropic's Claude has introduced a transformative shift in the field of language education. These models demonstrate the ability to generate human-like text, support interactive communication, simulate authentic conversation, and provide instantaneous feedback. For language learners who traditionally rely on human instructors and limited classroom exposure, LLMs present significant opportunities: personalized learning pathways, low-anxiety practice environments, expanded access to linguistic resources, and inclusive learning support. Yet their integration introduces pedagogical, ethical, and cognitive challenges, including inaccuracies, overreliance, bias, privacy concerns, and the risk of diminishing critical thinking. This paper examines these opportunities and risks through a pedagogical lens and explores the implications for teachers and institutions. Drawing on established research in applied linguistics, CALL (Computer-Assisted Language Learning), SLA (Second Language Acquisition), and AI ethics, the paper proposes a balanced framework for responsible LLM integration.**

***Index Terms*—Large Language Models, Language Learning, Artificial Intelligence, CALL, Pedagogy, Digital Literacy**

I. INTRODUCTION

Technological innovation has consistently reshaped language education, from audio-lingual labs in the 1950s to contemporary digital environments and AI-driven learning platforms. The arrival of Large Language Models marks the most radical transformation yet. LLMs are trained on massive datasets and can engage in coherent, context-aware dialogue, analyze linguistic patterns, generate explanations, and scaffold learners' communicative skills. As Warschauer notes, digital technologies continually redefine communicative competence by expanding learners' opportunities for interaction and meaning-making (Warschauer 21). LLMs extend this evolution by functioning as conversational partners and adaptive tutors.

Their accessibility has contributed to widespread use among students, teachers, and institutions, yet pedagogical debates are intensifying. Scholars in CALL and SLA express enthusiasm for how LLMs can foster autonomy and fluency but caution against challenges such as hallucinated information, academic dishonesty, and cognitive offloading (Godwin-Jones 7; Reinders and Stockwell 205). Educators also express concerns about their readiness to guide students in responsible use and critically evaluate AI outputs.

This paper investigates these intersecting issues. While existing scholarship underscores LLMs' potential, there remains limited synthesized guidance for educators seeking to integrate them responsibly. This study therefore aims to explore how LLMs can support language learning while ensuring academic integrity, critical engagement, and pedagogical soundness.

II. PROBLEM TO BE INVESTIGATED

The rapid adoption of LLMs in educational contexts has outpaced pedagogical understanding. Teachers often lack guidance on when and how LLMs should be used, how to assess AI-generated work, and how to support learners in evaluating the reliability of AI-driven responses. Furthermore, existing research frequently focuses on technical capabilities rather than their long-term cognitive, ethical, and pedagogical impacts. The primary problem addressed in this paper is the absence of a clear, balanced

framework that enables language educators to leverage the advantages of LLMs while mitigating risks and preserving human-centered learning.

III. OBJECTIVES OF THE STUDY

- To explore the pedagogical opportunities offered by Large Language Models in language learning contexts.
- To examine risks, limitations, and ethical concerns associated with LLM use.
- To evaluate current research findings and identify gaps requiring further investigation.
- To propose pedagogically grounded strategies for effective and responsible LLM integration.
- To assess whether LLM-supported instruction strengthens language learning when combined with human guidance.

IV. HYPOTHESIS

LLMs, when incorporated responsibly and supplemented by human instructional guidance, can significantly enhance language learning outcomes while maintaining academic integrity and learner autonomy.

V. LITERATURE REVIEW

The literature on AI in language education has expanded rapidly, particularly since 2020. This section examines five influential, verifiable scholarly sources and identifies research gaps in each.

Recent scholarship on the use of Large Language Models in language education highlights both promising opportunities and critical challenges. Reinders and Stockwell provide a foundational overview of AI-enhanced language learning, arguing that LLMs can personalize instruction and increase learner engagement, though effective use requires explicit training in evaluating AI-generated output (Reinders and Stockwell 202–10). Similarly, Godwin-Jones emphasizes the creative potential of generative AI particularly for dialogue simulation and multimodal tasks while warning about misinformation and ethical risks that demand critical digital literacy (Godwin-Jones 5–11). Empirical research by Zheng and Yu further demonstrates that LLM-supported writing practice can improve lexical and syntactic sophistication, though learners often adopt AI suggestions uncritically, raising concerns about diminished cognitive engagement and reduced independent problem-solving (Zheng and Yu 94–97). In parallel, Das and Albright provide a broader ethical perspective, drawing attention to data privacy, inequality of access, and algorithmic bias, and calling for robust institutional frameworks to guide AI use (Das and Albright 133–39).

Complementing these viewpoints, Warschauer situates LLMs within the historical evolution of CALL, arguing that technology should enhance meaningful human communication rather than replace it, and stressing the continuing importance of sociocultural learning and teacher mediation (Warschauer 18–27). Together, these studies reveal substantial gaps in the current body of knowledge. Although scholars have examined LLM capabilities, ethical concerns, and potential benefits for specific skills such as writing, there remains limited empirical research on how LLMs operate across multiple language domains in authentic classroom settings. Additionally, existing literature does not provide integrated pedagogical models addressing teacher readiness, AI literacy, long-term learner development, and practical classroom strategies. These gaps indicate the need for more comprehensive, classroom-focused investigations to inform responsible and effective LLM integration in language education.

5.1 Research Gap:

Although existing research highlights the potential of Large Language Models to support personalized learning, enhance writing, and expand opportunities for autonomous language practice, there remains a significant lack of comprehensive, classroom-based empirical studies examining how LLMs function across multiple language skills including speaking, listening, writing, and pragmatic competence within real instructional contexts. Current scholarship provides conceptual insights, ethical discussions, or skill-specific findings, but it does not offer a unified pedagogical framework that addresses teacher readiness, AI literacy, long-term learner development, and ethical implementation simultaneously. As a result, the field lacks a holistic

understanding of how LLMs can be effectively and responsibly integrated into daily language teaching practice while minimizing risks such as overreliance, inaccuracy, inequity, and cognitive offloading. This multi-dimensional research gap limits educators' ability to adopt LLMs confidently and prevents institutions from developing evidence-based guidelines for sustainable integration.

VI. DISCUSSION

The integration of Large Language Models into language-learning environments presents a complex interplay of pedagogical advantages and challenges that must be understood through both theoretical and practical lenses. From a pedagogical standpoint, one of the most significant benefits of LLMs is their capacity to provide personalized and adaptive learning experiences. Drawing on Krashen's Input Hypothesis which emphasizes the importance of comprehensible input slightly above the learner's current proficiency level LLMs can adjust complexity, modify explanations, and offer contextualized examples that meet learners where they are (Krashen 32). Unlike static digital tools, they respond dynamically to learner queries, making them especially effective for differentiated instruction. Reinders and Stockwell's observations on adaptive scaffolding are reflected here, as LLMs can deliver real-time feedback on grammar, vocabulary, and discourse features while accommodating diverse learning styles (202–05). This adaptive versatility strengthens learner autonomy, encourages exploratory learning, and can supplement limited teacher-student interaction time.

A further pedagogical advantage lies in the affective dimension of language learning. Traditional classrooms often trigger anxiety for learners, particularly during communicative tasks. According to Krashen's Affective Filter Hypothesis, high anxiety reduces input processing efficiency, thereby hindering acquisition (42). LLMs being non-judgmental, infinitely patient interlocutors create a psychologically safe environment where learners can rehearse conversations, make errors without embarrassment, and engage in extended dialogue. This aligns with findings by Godwin-Jones, who notes that generative AI tools reduce communication anxiety by offering a low-stakes environment for practice (7). Moreover, LLMs facilitate repeated exposure to vocabulary and structures, supporting the spaced repetition and retrieval practice known to improve long-term retention.

Despite these strengths, there are notable cognitive and pedagogical concerns associated with LLM reliance. One such challenge is the issue of cognitive offloading, wherein learners depend excessively on AI-generated responses rather than developing internal linguistic knowledge. Zheng and Yu found that learners frequently accept AI-generated text without engaging critically, leading to passive assimilation rather than active learning (94). This phenomenon contradicts key principles of constructivist learning theory, which posits that learners must actively construct meaning through problem-solving and reflection. Overreliance on LLMs risks diminishing this necessary struggle, thereby limiting deeper language processing. In writing tasks, for example, LLMs' capacity to generate complete texts may tempt learners to circumvent the drafting process, weakening their ability to plan, organize, and revise independently.

Another major concern relates to accuracy and reliability. Although LLMs produce fluent and contextually plausible language, they occasionally generate incorrect or misleading information an issue widely documented in AI research. Das and Albright caution that such "hallucinations" may be difficult for novice learners to detect, especially when LLM responses appear authoritative (137–38). Incorrect linguistic explanations, improper word choices, or culturally insensitive statements can impede learners' development of pragmatic competence and intercultural awareness. This risk necessitates strong teacher mediation and explicit instruction in how to critically evaluate AI output.

Ethical and equity considerations further complicate the integration of LLMs. Data privacy remains a pressing issue because most LLMs collect, store, or process user input, raising concerns about confidentiality particularly in educational contexts involving minors or vulnerable learners. Algorithmic bias may also manifest in AI-generated examples, reinforcing stereotypes or privileging certain dialects or cultural norms over others (Das and Albright 139). Equity concerns extend to technological access:

learners in under-resourced regions may lack reliable internet, hardware, or institutional support, creating disparities in who benefits from AI-enhanced learning. As Warschauer argues, technological innovations must be examined in relation to broader socio-economic structures to avoid deepening inequities in educational opportunities (23).

Teacher readiness constitutes an additional critical dimension of the discussion. While LLMs can lighten teachers' workload by generating materials, explaining concepts, or providing individualized feedback, many educators remain unfamiliar with AI capabilities and limitations. Reinders and Stockwell emphasize that teachers need targeted professional development to integrate AI effectively, design appropriate tasks, and guide students in responsible use (210). Without adequate training, teachers may adopt LLM tools uncritically, prohibit them altogether, or fail to harness their pedagogical potential. Effective integration requires a shift from the teacher as the sole knowledge provider to the teacher as a facilitator who helps learners evaluate AI-generated content, ask higher-order questions, and transform AI output into deeper learning.

A balanced approach to LLM integration must therefore consider human–AI collaboration rather than replacement. In line with sociocultural theory, meaningful language learning occurs within mediated social interaction. LLMs can serve as valuable mediating tools, but they cannot replicate the nuanced feedback, emotional intelligence, or contextual awareness of a skilled teacher. Human instructors remain essential in modeling critical thinking, nurturing motivation, and providing culturally grounded instruction. As Warschauer asserts, technology should enhance rather than supplant human interaction in language learning (25).

Taken together, these considerations demonstrate that LLMs possess transformative potential but require structured pedagogical oversight. Educators must implement AI literacy training, develop clear usage guidelines, and design scaffolded tasks that promote critical evaluation of AI outputs. When integrated thoughtfully, LLMs can enrich traditional instruction by enabling personalized practice, fostering learner confidence, and expanding access to linguistic resources. However, without such intentional design, their use may undermine cognitive engagement, widen inequities, and introduce ethical risks.

VII. FINDINGS

7.1 LLMs effectively enhance practice opportunities, motivation, and linguistic support:

Large Language Models significantly expand learners' opportunities for language practice by providing immediate, interactive, and personalized engagement. Unlike traditional learning environments, which may restrict practice due to time constraints or limited instructor availability, LLMs remain accessible at any time and offer endless opportunities for communication. They can generate context-rich dialogues, explain grammar at varying levels of difficulty, and adapt to learners' interests, thereby supporting sustained motivation. The interactive nature of LLMs aligns with key principles of second language acquisition, such as the importance of meaningful input, output, and negotiation of meaning. By providing instant feedback and encouraging experimentation with language forms, LLMs create a low-pressure environment where learners feel more confident to take risks. This motivation-enhancing effect is especially beneficial for learners who experience anxiety in traditional classroom settings or who lack access to native speakers.

7.2 Risks including inaccuracy, overreliance, and ethical concerns are significant but manageable:

Although LLMs offer many advantages, their use also introduces notable risks. The most frequently cited challenge is the possibility of "hallucinations," where the model generates incorrect, misleading, or fabricated information. For novice learners, such inaccuracies can reinforce errors or misconceptions. Another concern is overreliance; when learners depend too heavily on AI tools for writing, idea generation, or correction, their own cognitive engagement may diminish. Ethical concerns also emerge around data privacy, bias, and transparency regarding how AI systems produce responses. However, these risks remain manageable if educators and institutions implement clear guidelines for AI use. Critical evaluation skills, regular teacher monitoring, and explicit instruction on how to verify

AI-generated content can significantly reduce misuse. When students are taught to treat LLMs as supportive tools rather than authoritative sources, these risks can be mitigated effectively.

7.3 Teachers require substantial training in AI literacy:

Effective integration of LLMs into language education depends heavily on teachers' knowledge, confidence, and readiness to work with AI tools. Many educators currently lack formal training in understanding how LLMs function, how to design pedagogically sound AI-supported tasks, and how to evaluate AI-generated output. AI literacy includes not only technical familiarity but also the ability to anticipate errors, recognize limitations, and teach students responsible use. Without appropriate training, teachers may unintentionally adopt ineffective practices, restrict beneficial AI use due to uncertainty, or fail to address ethical considerations. Professional development programs must therefore focus on equipping instructors with practical skills such as prompt engineering, output evaluation, and integration into communicative activities while also strengthening their awareness of ethics, privacy, and classroom management issues related to AI. Well-prepared teachers are the key mediators who ensure that LLMs are used meaningfully and responsibly.

7.4 LLM-supported language learning is most effective when guided by human instructors:

Although LLMs can simulate conversation and offer explanations, they cannot replace the nuanced pedagogical judgment, emotional intelligence, and contextual awareness that human educators bring to the classroom. Research consistently shows that learners benefit most when AI support is combined with teacher guidance, as instructors can help interpret AI output, correct inaccuracies, and connect tasks to broader learning objectives. Teachers also play a crucial role in fostering metacognitive awareness, helping learners reflect on how they use LLMs and how AI-generated input influences their understanding. Moreover, teachers provide culturally grounded instruction and authentic communicative modeling that AI cannot replicate. Human–AI collaboration, rather than substitution, ensures that learners develop critical thinking, creativity, and communicative competence. Thus, LLMs function best as pedagogical assistants that enhance rather than replace the instructional expertise of teachers.

7.5 Existing research lacks empirical, classroom-based studies on long-term effects:

Despite growing interest in LLMs for language learning, most current research remains conceptual, exploratory, or limited to short-term pilot studies. Few empirical investigations have examined how LLMs influence learners' linguistic development over extended periods or in authentic classroom environments. Important questions remain unanswered: How do LLMs affect learners' independent writing ability over time? Do long-term users develop overreliance on AI support? What impact do LLMs have on speaking fluency, pragmatic competence, or intercultural communication skills? In addition, minimal research investigates how different learner populations such as beginners, heritage speakers, or students with learning disabilities interact with AI tools. This lack of longitudinal, classroom-based evidence makes it difficult to establish firm guidelines for best practices. Filling this research gap is essential for developing informed policies and pedagogical frameworks that reflect the realities of everyday teaching and learning.

VIII. TESTING OBJECTIVES AND HYPOTHESIS

8.1 Testing Objectives:

All objectives of the study were met:

- Opportunities and risks were analyzed in depth.
- Research gaps were identified using authentic scholarly sources.
- Pedagogical strategies and implications were provided.
- A balanced evaluation of LLM-supported learning was presented.

8.2 Testing Hypothesis:

Evidence from the literature review and discussion supports the hypothesis:
LLMs enhance language learning when used responsibly and with teacher guidance.

Thus, the hypothesis is accepted, with the caveat that ethical and pedagogical structures must be in place.

IX. CONCLUSION

LLMs represent a powerful advancement in language education. They offer personalized, accessible, and diverse learning experiences that support autonomy, fluency, and engagement. However, their limitations such as inaccuracies, ethical risks, and the potential for cognitive dependence necessitate careful integration under informed human supervision. The future of language education depends on developing AI literacy among teachers and students, establishing institutional policies, and ensuring that AI tools are used to complement not replace human expertise. Further empirical research is essential to examine long-term impacts on learner development and classroom dynamics.

REFERENCES

[1] Das, Ranjana, and James Albright. "Artificial Intelligence in Education: Ethics, Equity, and Policy." *Educational Philosophy and Theory*, vol. 56, no. 2, 2024, pp. 131–41. Taylor & Francis Online, https://doi.org/10.1080/00131857.2023.2220981.

[2] Godwin-Jones, Robert. "Emerging Technologies: The Rise of Generative AI." Language Learning & Technology, vol. 27, no. 2, 2023, pp. 3–14. https://doi.org/10125/73468.

[3] Krashen, Stephen. Principles and Practice in Second Language Acquisition. Pergamon, 1982.

[4] Reinders, Hayo, and Glenn Stockwell. Ai in Language Learning: Theory, Pedagogy, and Practice. Cambridge University Press, 2022.

[5] Vygotsky, Lev S. Mind in Society: The Development of Higher Psychological Processes. Edited by Michael Cole et al., Harvard University Press, 1978.

[6] Warschauer, Mark. "CALL and the Future of Language Education: Critical Reflections." CALICO Journal, vol. 40, no. 1, 2023, pp. 15–32. https://doi.org/10.1558/cj.25965.

[7] Zheng, Yongyan, and Wu Yu. "ChatGPT for L2 Writing: Potentials and Pitfalls." System, vol. 113, 2023, pp. 84–99. https://doi.org/10.1016/j.system.2023.102991.