ANCOLT
International Proseeding on Language Teaching
Vol. 1, No. 1, Maret 2024, pp.487-493

Beyond Boundaries in Arabic Language Teaching: A Human-Centered Framework for Integrating Generative AI and Pesantren Pedagogy

Mohammad Makinuddin
kinudd@gmail.com
Universitas Kiai Abdulah Faqih Gresik

Abstract

Arabic language teaching is entering a period in which generative artificial intelligence (GenAI), mobile learning, and longstanding local pedagogies can no longer be treated as separate domains. This paper develops a human-centered framework for future Arabic instruction by synthesizing research on AI-supported language learning, technology integration in Arabic education, and pesantren-based Arabic pedagogy. Using an integrative literature review, the study identifies four design principles: dialogic personalization, multimodal practice beyond the classroom, pedagogical continuity with authentic Arabic texts, and verification-centered AI literacy. The findings suggest that innovation is most educationally productive when AI extends rather than replaces teacher mediation, learner agency, and culturally grounded textual practices. The proposed framework offers a practical direction for institutions seeking future-ready Arabic teaching while maintaining linguistic accuracy, ethical safeguards, and the intellectual traditions that give Arabic learning its disciplinary and cultural depth.

Keywords: *Arabic language teaching; generative AI; pesantren pedagogy; technology integration; AI literacy*

INTRODUCTION

Arabic language education now operates across boundaries that were previously treated as relatively stable: classroom and home, print and digital text, teacher explanation and machine-generated feedback, formal curricula and informal learning networks, as well as global technological platforms and local intellectual traditions. The rapid diffusion of generative artificial intelligence (GenAI) intensifies this shift because learners can request explanations, translations, examples, dialogues, corrective feedback, and adapted texts at any time. Earlier generations of computer-assisted language learning already demonstrated the value of intelligent tutoring, automatic feedback, and personalization (Weng & Chiu, 2023). Chatbot research likewise shows that conversational systems can function as

interlocutors, simulations, helplines, and recommendation tools, although technological limitations, cognitive load, and novelty effects remain important concerns (Huang et al., 2022). GenAI therefore does not begin a completely new pedagogical history; rather, it expands the scale, flexibility, and linguistic reach of existing technology-supported language learning.

For Arabic, however, innovation requires more than importing tools developed primarily for English-language contexts. Arabic instruction involves a complex relationship among Modern Standard Arabic, classical texts, grammar, morphology, spoken varieties, script, vocalization, and culturally embedded meanings. In Indonesian Islamic education, this complexity is especially visible in pesantren, where Arabic literacy is often developed through close engagement with classical Islamic texts, teacher modeling, grammatical analysis, translation, and repeated guided reading. Technology-based Arabic learning in traditional Islamic boarding schools has begun to broaden access and instructional resources, yet implementation must remain attentive to pedagogical aims and institutional culture (Nasution et al., 2024). Reviews of Arabic learning also report that AI, online platforms, gamification, and related technologies can improve engagement and flexibility, while teacher readiness, infrastructure, and curricular alignment remain persistent challenges (Lubis & Nasution, 2024; Supriyanto & Toifah, 2024).

The central issue is therefore not whether Arabic teachers should choose tradition or technology. A more productive question is how technology can extend the strengths of Arabic pedagogy without displacing the human judgment required to evaluate meaning, register, grammatical accuracy, textual authority, and ethical use. Research on ChatGPT in language teaching highlights opportunities for idea generation, individualized explanations, conversational practice, and feedback, but also emphasizes the need for digital competence and critical evaluation (Kohnke et al., 2023). Broader educational reviews similarly warn that large language models can produce inaccurate information and should be integrated through redesigned assessment, teacher training, and explicit policies (Kasneci et al., 2023; Lo, 2023). UNESCO (2023) therefore recommends a human-centered approach that protects agency, inclusion, privacy, and linguistic and cultural diversity. Responding to the ANCOLT theme “Beyond Boundaries,” this paper asks: What design principles can connect GenAI-enabled innovation with culturally grounded Arabic pedagogy for future language teaching?

METHOD

This study employed an integrative literature review to build a pedagogical framework rather than to estimate a single intervention effect. The review focused on three intersecting bodies of scholarship: (1) AI, chatbots, and GenAI in language teaching; (2) technology integration in Arabic language education; and (3) digitally supported Arabic learning in Indonesian Islamic education, including pesantren. Publications from 2022–2024 were prioritized to capture the rapid expansion of GenAI while

retaining selected foundational reviews of chatbot-supported language learning. Sources were identified through targeted searches using combinations of “generative AI,” “ChatGPT,” “language learning,” “Arabic language learning,” “technology,” and “pesantren.” Peer-reviewed journal articles and authoritative international guidance were prioritized.

The analysis used thematic coding. Each source was read for its reported pedagogical affordances, constraints, teacher roles, learner roles, assessment implications, and ethical considerations. The resulting codes were compared across general language-learning research and Arabic-specific studies. Rather than treating technological affordances as sufficient evidence of pedagogical value, the synthesis asked how each affordance could be translated into an Arabic-learning task with a clear linguistic objective and a mechanism for human verification. Four recurring principles emerged and were consolidated into the framework presented below.

RESULT AND DISCUSSION

The synthesis indicates that future-ready Arabic teaching is best conceptualized as an ecology in which AI, mobile access, teachers, peers, and authoritative texts perform different but complementary functions. Four principles define this ecology.

1. Dialogic personalization without pedagogical outsourcing

GenAI expands opportunities for low-stakes, repeated interaction. Learners can request vocabulary explanations at different proficiency levels, generate short dialogues, rehearse question-and-answer sequences, compare sentence patterns, or ask for feedback on a draft. Such uses align with earlier chatbot evidence showing benefits in immediacy, personalization, and sustained interaction (Huang et al., 2022). In foreign-language contexts, learners have also reported benefits for writing, grammar, vocabulary, motivation, and engagement when ChatGPT is incorporated into structured learning experiences (Karataş et al., 2024). Arabic-specific reviews similarly identify rapid correction, conversation practice, translation support, grammar explanation, and access to learning resources as major opportunities (Supriyanto & Toifah, 2024).

The pedagogical risk is outsourcing the cognitive work that language learning requires. If students ask AI to translate every sentence, supply every i‘rab analysis, or write complete assignments, fluent output may conceal weak linguistic competence. Therefore, personalization should be designed as graduated scaffolding: AI first prompts the learner, then offers hints, examples, or contrasting forms, and only later provides a model answer. The teacher remains responsible for selecting target structures and deciding what forms of support are appropriate. In this model, AI is a rehearsal partner and feedback layer, not the final authority on Arabic usage or textual interpretation.

2. Multimodal practice beyond the classroom

The second principle is boundary extension. Arabic learning has often been constrained by limited contact hours and unequal opportunities for oral

practice. Speech-recognition chatbots and multimodal systems can create additional spaces for listening, speaking, reading, and writing. Jeon et al. (2023) show that chatbot affordances vary according to goal orientation, embodiment, and multimodality, suggesting that effective design should begin with the learning objective rather than with the novelty of the tool. For Arabic, a mobile task might combine a short audio prompt, learner speech, transcript comparison, vocabulary retrieval, and a reflective note on recurring errors. Online platforms can also distribute annotated texts, micro-quizzes, and discussion prompts before and after classroom sessions.

Moving learning beyond the classroom does not mean removing structure. Technology integration studies in Arabic education report gains in access, engagement, comprehension, and motivation, but they also identify infrastructure and readiness gaps (Lubis & Nasution, 2024). A future-oriented design should therefore use low-bandwidth, device-accessible tasks, provide downloadable resources, and avoid making advanced technology a prerequisite for participation. The boundary that should be crossed is the boundary of time and space, not the boundary of educational equity.

3. Pedagogical continuity with authentic texts and pesantren practices

Innovation becomes more sustainable when it grows from established pedagogical strengths. In pesantren, Arabic reading is connected to authentic disciplinary purposes: understanding classical Islamic texts, recognizing grammatical relations, tracing lexical meanings, and interpreting arguments under teacher guidance. Technology can enrich this ecology by providing layered glosses, morphological prompts, searchable vocabulary, audio modeling, and comparative exercises, while the teacher continues to mediate textual authority and interpretive nuance. Research on technology-supported Arabic sciences in traditional Islamic boarding schools indicates that digital tools can be integrated with established learning environments rather than treated as replacements for them (Nasution et al., 2024).

This principle also addresses a common weakness in educational AI: decontextualized fluency. A model may generate grammatically plausible Arabic but miss the register, disciplinary convention, or intended interpretation of a classical passage. Students should therefore be trained to compare machine output with dictionaries, commentaries, course texts, and teacher explanation. A useful classroom routine is “generate–verify–justify”: students obtain an AI-supported explanation, verify lexical and grammatical claims against trusted sources, and justify which elements they accept, revise, or reject. This transforms AI from an answer machine into an object of linguistic analysis and critical reading.

4. Verification-centered AI literacy and renewed teacher expertise

The final principle is that innovation must strengthen, not weaken, epistemic responsibility. Large language models can generate convincing but inaccurate content, fabricated references, oversimplified grammar explanations, and inconsistent translations. Lo (2023) identifies misinformation and academic integrity as major educational concerns, while

Kasneci et al. (2023) argue for critical thinking, fact checking, and continuous human oversight. Kohnke et al. (2023) similarly place digital competence at the center of responsible language-teaching use. UNESCO's (2023) human-centered guidance adds privacy, inclusion, cultural diversity, and institutional governance to this agenda.

For Arabic teachers, AI literacy should include prompt design, linguistic verification, recognition of model limitations, ethical attribution, assessment redesign, and protection of learner data. The teacher's role consequently becomes more expert, not less: designing tasks that require evidence of learning, diagnosing errors that AI cannot reliably interpret, curating authoritative sources, and teaching students how to challenge machine output. Assessment can emphasize oral defense, process logs, text annotation, comparative correction, and reflective justification. These forms of evidence reduce the value of uncritical AI-generated answers while preserving the benefits of assisted practice.

5. From individual tools to an institutional learning architecture

The four principles imply that future innovation should be implemented as an institutional learning architecture rather than as a collection of unrelated applications. A practical sequence begins with curriculum mapping: departments identify where students need additional exposure, feedback, conversation, textual analysis, or independent practice. Teachers then select the least complex technology capable of supporting that objective. For example, vocabulary recycling may require only a mobile quiz, while speaking rehearsal may benefit from speech-enabled dialogue, and classical-text analysis may require an annotated digital workspace plus carefully bounded AI assistance. This objective-first approach avoids the common mistake of redesigning a course around a fashionable platform. It also makes technology replacement easier when tools change.

Institutional support should combine teacher development, shared prompt and task banks, approved reference resources, privacy guidance, and common assessment rules. Pilot activities can be reviewed through evidence such as learner work, error patterns, participation data, and teacher observation rather than satisfaction alone. The most important indicator is whether students demonstrate stronger independent Arabic performance after support is reduced. This principle is consistent with the broader finding that AI-supported language learning depends on instructional design, not merely system capability (Weng & Chiu, 2023). It also responds to UNESCO's (2023) call for governance and human capacity alongside technological adoption. In pesantren and Islamic higher education, such governance can explicitly protect the role of teachers and authoritative texts while opening controlled spaces for experimentation. Innovation then becomes cumulative: successful practices are documented, verified, adapted to local levels, and integrated into the curriculum instead of remaining isolated teacher initiatives.

Table 1. Human-centered framework for future Arabic language teaching

Design principle	Primary affordance	Main risk	Pedagogical response
Dialogic personalization	Adaptive dialogue, feedback, scaffolding	Cognitive outsourcing	Use staged hints and require learner-produced responses
Multimodal extension	Practice across time, place, and mode	Access inequality; novelty-driven use	Use objective-led, low-bandwidth, accessible tasks
Textual continuity	Digital support for authentic Arabic texts	Decontextualized or inaccurate interpretation	Apply generate–verify–justify with trusted sources
Verification-centered AI literacy	Faster drafting, comparison, and feedback	Hallucination, bias, privacy, integrity	Teach verification, attribution, data awareness, and process-based assessment

CONCLUSION

Going beyond boundaries in Arabic language teaching should not be understood as moving away from established pedagogy toward technology. The more defensible direction is integration: GenAI can increase opportunities for practice, personalization, feedback, and access, while pesantren and other text-centered traditions contribute authenticity, disciplinary rigor, teacher mediation, and interpretive accountability. The proposed framework organizes this integration around four principles: dialogic personalization, multimodal learning beyond the classroom, continuity with authentic texts and local pedagogies, and verification-centered AI literacy. Its practical implication is that institutions should evaluate technology by the quality of the learning activity it enables rather than by novelty alone. Future empirical studies should test the framework across proficiency levels and institutional types, especially through designs that measure Arabic speaking, reading, writing, grammatical reasoning, and learner independence over time. The future of Arabic teaching is therefore neither machine-centered nor tradition-bound; it is human-centered, technologically extended, and culturally grounded. This orientation also creates a clear research agenda: future studies should compare teacher-guided and learner-directed AI use, document error types specific to Arabic morphology and syntax, and examine whether technology-supported practice transfers to independent reading and speaking performance without automated assistance.

REFERENCES

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. Journal of Computer Assisted Learning, 38(1), 237–257. https://doi.org/10.1111/jcal.12610

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. Computers & Education, 206, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Karataş, F., Abedi, F. Y., Gunyel, F. O., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's

effect on foreign language learners. Education and Information Technologies, 29, 19343–19366. https://doi.org/10.1007/s10639-024-12574-6

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410

Lubis, A. F., & Nasution, S. (2024). The effectiveness of technology integration in Arabic language learning: Systematic literature review. Ihya al-Arabiyah: Jurnal Pendidikan Bahasa dan Sastra Arab, 10(2). https://doi.org/10.30821/ihya.v11i3.25866

Nasution, S., Asari, H., Al-Rasyid, H., Dalimunthe, R. A., & Rahman, A. (2024). Learning Arabic language sciences based on technology in traditional Islamic boarding schools in Indonesia. Nazhruna: Jurnal Pendidikan Islam, 7(1), 77–102. https://doi.org/10.31538/nzh.v7i1.4222

Supriyanto, S., & Toifah, N. (2024). Efektifitas penggunaan artificial intelligence (AI) dalam pembelajaran bahasa Arab di era Society 5.0: Systematic literature review. Lugawiyyat, 6(2), 42–54. https://doi.org/10.18860/lg.v6i2.29713

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

Weng, X., & Chiu, T. K. F. (2023). Instructional design and learning outcomes of intelligent computer assisted language learning: Systematic review in the field. Computers and Education: Artificial Intelligence, 4, 100117. https://doi.org/10.1016/j.caeai.2022.100117