Arabiyatuna
Jurnal Bahasa Arab

Vol. 10. No. 2, November 2026, 719-734
P-ISSN: 2580-5045, E-ISSN: 2580-5053
DOI: http://dx.doi.org/10.29240/jba.v10i2.17624

Responsible AI in Arabic Language Learning: Students' Perceptions and Engagement

Iis Susiawati[1*], Wiwin[2], Habsah Bte Musliman[3]
Institut Agama Islam Al-Zaytun, Indonesia[1,2,3]

[*]Corresponding Author: iis.susiawati@iai-alzaytun.ac.id

Cite this article:
Susiawati, I., Wiwin, & Bte Musliman, H. (2026). Responsible AI in Arabic Language Learning: Students' Perceptions and Engagement. Arabiyatuna: Jurnal Bahasa Arab, 10(2), 719–734. https://doi.org/10.29240/jba.v10i2.17624

Received: 20-06-2026	Revised: 10-09-2026	Accepted: 02-11-2026

Abstract

Generative AI provides immediate language assistance, but questions remain about students' engagement and academic integrity in Arabic learning. Evidence linking learners' perceptions, engagement, and ethical awareness is limited. This exploratory study compares the independent statistical associations of perceptions and ethical awareness with engagement and describes how these constructs relate to self-reported responsible AI use. A cross-sectional survey involved 40 purposively recruited students with experience using AI to learn Arabic. Responses on a five-point Likert scale were analyzed using descriptive statistics, internal-consistency coefficients, Pearson correlations, and two-predictor ordinary least squares regression; diagnostic and heteroskedasticity-robust checks supplemented the primary model. ChatGPT (87.5%) and translation tasks (80.0%) predominated. Perceptions correlated strongly with engagement ($r = .795$, $p < .001$). The engagement model explained 63.7% of sample variance ($R^2 = .637$): perceptions were a significant independent predictor ($B = 0.748$, 95% CI [0.519, 0.976], $p < .001$), whereas ethical awareness was not ($B = 0.071$, 95% CI [−0.118, 0.261], $p = .452$). Ethical awareness was nevertheless correlated with reported responsible AI use ($r = .501$, $p = .001$). In this exploratory Arabic-learning sample, positive AI perceptions were associated with engagement after adjustment, whereas ethical awareness was not an independently significant correlate. This difference between bivariate and adjusted associations is the study's principal contribution; it does not imply that ethical guidance is unnecessary. Teachers may pair useful AI

Copyright © 2026 **Author'(s)**
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

tasks with explicit verification and academic-integrity expectations. Larger, independently validated studies are needed to establish whether these practices improve Arabic learning outcomes.

Keywords: Arabic language pedagogy; generative AI; ChatGPT; academic integrity; higher education.

Introduction

Generative AI now offers language learners instant translations, feedback, grammatical explanations, and dialogue simulations, while also creating risks of inaccurate output, overreliance, and unclear authorship. Reviews of educational AI emphasize that its benefits depend on instructional design, governance, and critical evaluation (Holmes & Tuomi, 2022; Chiu et al., 2023; Kasneci et al., 2023; Li et al., 2024; S. Wang et al., 2024). These tensions make it necessary to distinguish students' willingness to use AI from the quality and integrity of their learning practices.

Arabic presents specific challenges for automated language assistance: derivational morphology, agreement, case-related patterns, diglossia, dialect variation, and culturally or religiously marked registers. Research on AI-assisted Arabic learning identifies opportunities for feedback and accessible materials but also warns about linguistic reliability and contextual appropriateness (Al-Shaboul et al., 2025; Moustafa et al., 2026). Arabic pedagogy also prioritizes communicative practice, integrated skills, and language-sensitive feedback (Ryding, 2013). Here, 'responsible AI use' is operationalized as students' self-reported practices of verifying outputs, acknowledging assistance, protecting information, and resisting uncritical substitution of machine-generated work for their own effort—not as an independently observed measure of Arabic proficiency.

The longer pedagogical emphasis on acquired language competence through repeated practice also appears in Ibn Khaldun's discussion of linguistic habit (Khaldun, 1958). This historical perspective is used only to motivate independent practice, not as empirical evidence about AI or as a construct measured in the survey.

Recent Arabic-focused research examines mobile AI, curriculum integration, standardization, teacher implementation, and generated reading materials (Alasmari, 2025; Alkaabi & Almaamari, 2025; Allaithy & Zaki, 2026; Sapawi & Yusoff, 2025; Sabilar et al., 2026). These contributions are valuable, yet they seldom consider, within one student dataset, how perceived usefulness, learning engagement, ethical awareness, and reported responsible use relate to one another. The present study addresses this narrower descriptive and correlational gap rather than proposing a new technology or measurement theory.

The technology acceptance model links perceived usefulness and ease of use with the acceptance of digital systems (Davis, 1989; Venkatesh et al., 2003). Applied to Arabic learning, perceptions concern whether AI helps students understand text, practice vocabulary, or obtain feedback. However, positive perceptions alone cannot verify achievement or ethical behaviour, and students may value speed even when learning remains superficial (Chan & Hu, 2023).

Engagement theory differentiates behavioural, emotional, and cognitive participation (Fredricks et al., 2004), while technology-enhanced learning research emphasizes sustained interaction and purposeful feedback use (Bond et al., 2020). This distinction is important for Arabic because asking AI to translate a passage is not equivalent to independently interpreting, revising, or producing Arabic. Students' reported engagement therefore reflects perceived activity, not directly assessed language development.

Ethical AI scholarship highlights transparency, authorship, privacy, fairness, reliability, and learner agency (Akgun & Greenhow, 2022; Schiff, 2022; Nguyen et al., 2023; Tlili et al., 2023). AI literacy also includes recognizing limitations and evaluating generated content rather than merely operating a tool (Long & Magerko, 2020). Accordingly, ethical awareness is conceptually distinct from engagement: a student can recognize risks without using AI more frequently, and active use does not prove responsible practice.

Existing work rarely estimates these relationships jointly among students learning Arabic, while discussion of language-specific risks often exceeds what survey instruments actually measure. This study does not assess errors in AI-generated Arabic, dialect competence, or learning gains. Its distinctive contribution is instead a bounded, student-cantered comparison of perceived usefulness-related evaluations and ethical awareness as correlates of engagement, together with descriptive evidence about reported responsible practices. Results are framed as preliminary because of the small, nonprobability sample. This contribution concerns the contrast between unadjusted ethical-awareness correlations and its nonsignificant adjusted association with engagement; it does not establish a new theory or prove that ethical awareness affects learning behaviour.

Four questions guide the analysis: (1) Which AI tools and activities do participating Arabic learners report using? (2) What are the levels of perceived usefulness-related evaluations, engagement, ethical awareness, and responsible AI use? (3) How are these four self-reported constructs correlated? (4) Do perceptions and ethical awareness show distinct associations with engagement in a two-predictor regression? The study contributes an initial comparison that can inform Arabic-specific task design and later validation research, without inferring causal effects from cross-sectional responses.

Method

A nonexperimental, cross-sectional quantitative survey was used to describe AI-use patterns and associations among four self-reported constructs. The primary explanatory model examined engagement as the outcome, with perceptions and ethical awareness as the two predictors. Responsible AI use was analyzed descriptively and correlationally as a separate reported-practice construct; a supplementary regression treated it as an exploratory outcome, not as a preregistered confirmatory model. None of these analyses establishes a causal effect (Ponto, 2015).

The sample comprised 40 volunteers who were enrolled in Arabic-related programs and reported prior use of at least one AI tool for studying Arabic. Recruitment used purposive eligibility criteria, not probability sampling. Of these students, 38 studied Arabic Language Education and two studied Arabic Language and Literature; 28 were in semester eight. The analytic file does not document the institutional distribution, recruitment denominator, or response rate, so representativeness cannot be assessed. The small and concentrated sample restricts external validity, particularly for regression-based inference.

No a priori sample-size or statistical-power calculation was documented before data collection. Accordingly, the sample of 40 respondents is not presented as an adequately powered confirmatory target. All regression estimates are interpreted as exploratory, with uncertainty reported using 95% confidence intervals, heteroscedasticity-robust standard errors, influence diagnostics, and leave-one-out sensitivity analyses. Future confirmatory studies should prospectively determine sample size from a prespecified smallest effect of substantive interest and recruit a larger and more heterogeneous sample.

The questionnaire comprised background questions and 42 five-point agreement items (1 = strongly disagree; 5 = strongly agree). Composite scores were calculated as the arithmetic mean of each respondent's item responses: perceptions (10 items), learner engagement (12), ethical awareness (12), and responsible AI use (8). These scores capture reported experience or judgment rather than observed AI practices or Arabic competence (Koo & Yang, 2025).

Construct domains were informed by technology acceptance theory for perceptions (Davis, 1989), multidimensional engagement theory for engagement (Fredricks et al., 2004), and educational AI ethics and literacy frameworks for ethical awareness and reported responsible use (Long & Magerko, 2020; Nguyen et al., 2023). The survey includes references to learning Arabic, but it does not operationally test morphological errors, diglossia, or Qur'anic registers; these are pedagogical considerations rather than measured Arabic-specific subscales.

The questionnaire items were organized around the specified theoretical domains, and internal consistency was assessed using responses from the 40 participating students. No retrievable expert-validation forms, expert-rating data, or separately identifiable pilot-test dataset are available for this study. Consequently, the article does not claim that a formal expert-validation study or an independent pilot test was completed, and it does not report content-validity indices, pilot statistics, or expert-agreement coefficients. The instrument should therefore be regarded as exploratory. Cronbach's alpha provides within-sample internal-consistency evidence only and cannot establish content validity, construct validity, or discriminant validity (Tavakol & Dennick, 2011; Boateng et al., 2018). This limitation is acknowledged rather than replaced with retrospectively invented validation results. The absence of these records means that expert-based content validity and pilot-test performance cannot be asserted retrospectively; the authors report internal consistency only, not confirmed construct validity.

The survey was completed online by volunteers meeting the stated eligibility criteria. According to the study procedures, respondents were informed of its purpose and confidentiality arrangements, and participation was voluntary. The de-identified 40-case response matrix was used for the present analyses. Dates and site-level recruitment details were not retained in the available analytic records. Nor are there verifiable records of a formal expert review or separate pilot test; neither procedure is claimed as completed.

Counts and percentages summarized background responses, while composite means, standard deviations, and Cronbach's alpha described the four scales. Pearson correlations quantified pairwise associations. The primary ordinary least squares model predicted engagement from perceptions and ethical awareness; estimates are reported with two-sided tests and 95% confidence intervals. We checked residual normality (Shapiro–Wilk), heteroskedasticity (Breusch–Pagan), predictor collinearity (variance inflation factors), and potentially influential cases (Cook's distance). HC3 heteroskedasticity-robust intervals and leave-one-out estimates assessed coefficient stability. An exploratory second model examined responsible AI use as the outcome. Given N = 40, the secondary estimates are hypothesis-generating rather than confirmatory.

Findings and Discussion

Students' AI Use Background in Arabic Language Learning

All 40 participants reported prior experience with AI in Arabic learning. Twenty-five were male (62.5%) and 15 female (37.5%); 38 were undergraduate students (95.0%). Arabic Language Education accounted for 38 participants (95.0%), and semester eight accounted for 28 (70.0%). This concentration is

relevant to interpretation: the dataset reflects a narrow group of respondents rather than the wider population of Arabic learners (Table 1).

Table 1. Demographic Profile of Respondents

Variable	Category	n	%
Gender	Male	25	62.5
	Female	15	37.5
Education level	Undergraduate	38	95.0
	Doctoral	2	5.0
Study program	Arabic Language Education	38	95.0
	Arabic Language and Literature	2	5.0
Semester	1st semester	1	2.5
	2nd semester	3	7.5
	4th semester	4	10.0
	6th semester	3	7.5
	8th semester	28	70.0
	9th semester or above	1	2.5

Data Source: Online questionnaire, 2026.

Most participants reported occasional (47.5%) or frequent (42.5%) AI use; the detailed distribution is shown in Table 2.

ChatGPT (87.5%) and Google Translate (70.0%) were the most common tools. Translation (80.0%) and assignment completion (57.5%) dominated the reported purposes; speaking practice was selected by only 10.0% (Table 2). As participants could choose multiple tools and purposes, these percentages are not mutually exclusive.

Table 2. Frequency, Tools, and Purposes of AI Use in Arabic Language Learning

Variable	Category	n	%
Frequency of AI use	Rarely	1	2.5
	Occasionally	19	47.5
	Frequently	17	42.5
	Very frequently	3	7.5
AI tools used	ChatGPT	35	87.5
	Google Translate	28	70.0
	Gemini	18	45.0
	Copilot	11	27.5
	Speech recognition application	4	10.0

	Other tools	3	7.5
Purpose of AI use	Translation	32	80.0
	Completing assignments	23	57.5
	Vocabulary learning	20	50.0
	Grammar learning	17	42.5
	Writing practice	12	30.0
	Reading practice	7	17.5
	Pronunciation practice	5	12.5
	Speaking practice	4	10.0
	Other purposes	2	5.0

Data Source: Online questionnaire, 2026. Respondents could select more than one AI tool and more than one purpose of use; therefore, percentages may exceed 100%.

The predominance of translation and task completion suggests that students mainly seek rapid access to meaning and support for course assignments, a pattern consistent with generative AI's language-support affordances (Kasneci et al., 2023; Kohnke et al., 2023). Yet this survey cannot determine whether students checked the translations or improved their Arabic as a result. Unlike general-purpose adoption studies, Arabic pedagogy must additionally address agreement, morphology, register, and culturally embedded meaning (Moustafa et al., 2026; Ryding, 2013).

A practical implication is to require students to explain which Arabic expressions they verified and why a translation was accepted or revised. Such tasks distinguish reflective reading and interpretation from automatic copying. The study did not directly observe these practices; they are recommendations informed by the pattern of reported tool use.

Students' Perceptions, Engagement, Ethical Awareness, and Responsible AI Use

Across the 40 responses, Cronbach's alpha was .870 for perceptions (10 items), .912 for engagement (12 items), .942 for ethical awareness (12 items), and .950 for responsible AI use (8 items). These high coefficients indicate internal consistency but may also reflect item redundancy, especially in the latter two scales; additional construct-validation evidence is needed (Table 3).

On the five-point scale, participants reported means of 3.71 for perceptions, 3.73 for engagement, 4.05 for ethical awareness, and 3.90 for responsible AI use (Table 3). The label 'high' follows prespecified descriptive score bands; it should not be taken as evidence that actual ethical conduct or language performance was independently observed.

Table 3. Reliability and Descriptive Statistics of the Main Constructs

Construct	Items	N	M	SD	Cronbach's α	Interpretation
Students' perceptions	10	40	3.71	0.66	.870	High
Learner engagement	12	40	3.73	0.65	.912	High
Ethical awareness	12	40	4.05	0.79	.942	High
Responsible AI use	8	40	3.90	0.91	.950	High

Data Source: Online questionnaire, 2026. Scale range: 1 = strongly disagree to 5 = strongly agree. Mean interpretation: 1.00–1.80 = very low, 1.81–2.60 = low, 2.61–3.40 = moderate, 3.41–4.20 = high, and 4.21–5.00 = very high.

Perceptions and engagement had similarly elevated averages, suggesting a favorable self-reported learning experience among these respondents. The results are consistent with acceptance frameworks (Davis, 1989; Venkatesh et al., 2003), but the cross-sectional instrument cannot distinguish whether favorable attitudes precede active learning or arise because of it.

Self-reported engagement may comprise seeking explanations, responding to feedback, and practicing independently, all consistent with behavioral and cognitive engagement theory (Fredricks et al., 2004; Bond et al., 2020). However, the survey measured students' accounts of such involvement, not task logs, classroom observation, or gains in Arabic proficiency.

Ethical awareness had the highest average (M = 4.05), but this was a rating of awareness rather than a direct measure of disclosure, plagiarism avoidance, or verification. Strong alpha coefficients do not by themselves establish that awareness and reported responsible use are separate constructs; further item-level and multi-method validation would strengthen the claim (Nguyen et al., 2023).

Relationships among Perception, Engagement, Ethical Awareness, and Responsible AI Use

Pairwise correlations were positive (Table 4). The largest was between perceptions and engagement ($r = .795$, 95% Fisher-z CI [.642, .887], $p < .001$), a large association in this sample. Ethical awareness correlated with engagement ($r = .454$, $p = .003$) and with responsible AI use ($r = .501$, $p = .001$). These are unadjusted associations, not estimates of causal effects.

Table 4. Pearson Correlations among the Main Constructs

Construct	1	2	3	4
1. Students' perceptions	—			
2. Learner engagement	.795***	—		
3. Ethical awareness	.489**	.454**	—	
4. Responsible AI use	.558***	.575***	.501**	—

Data Source: Online questionnaire, 2026. N = 40. **p < .01. ***p < .001.

The high perception–engagement correlation is practically meaningful but also calls for caution: shared item phrasing, common-method response tendencies, or construct overlap could inflate associations. Although the concepts have distinct theoretical definitions, the present sample is too small for strong claims of discriminant validity. Future studies should examine the item wording, factor structure, and cross-method evidence before interpreting these scales as fully distinct.

Positive perceptions were associated with reported responsible use ($r = .558$, $p < .001$), but this does not show that optimism causes ethical behavior. Students' declared confidence, engagement, and ethical intentions may move together partly because they were rated on the same questionnaire. This possibility offers an alternative explanation to technology acceptance alone.

The association between ethical awareness and engagement ($r = .454$) shows that the variables covary without adjustment. Whether awareness accounts for unique variation in engagement after perceptions are included is a separate question answered by the regression model below. The distinction prevents interpreting a significant correlation as a significant independent regression effect.

Predicting Learner Engagement in AI-Supported Arabic Learning

The two-predictor model was statistically significant, $F(2, 37) = 32.52$, $p < .001$, with $R^2 = .637$ (adjusted $R^2 = .618$). In this dataset, the two predictors jointly accounted for 63.7% of the observed variance in engagement; this in-sample proportion is not a predictive accuracy estimate for new learners. With only 40 purposively recruited respondents, the explained variance describes the fitted sample and should not be generalized as the proportion of engagement explained in the broader population.

Perceptions showed an adjusted association with engagement ($B = 0.748$, $SE = 0.113$, $\beta = .753$, $t = 6.632$, $p < .001$, 95% CI [0.519, 0.976]). Ethical awareness did not have a detectable independent association ($B = 0.071$, $SE = 0.094$, $\beta = .086$, $t = 0.761$, $p = .452$, 95% CI [−0.118, 0.261]). A nonsignificant coefficient does not prove the absence of every possible relationship; it indicates that a unique association was not established in this model and sample.

Table 5. Multiple Regression Predicting Learner Engagement

Predictor	B	SE	β	t	p
Constant	0.665	0.410	—	1.623	.113
Students' perceptions	0.748	0.113	.753	6.632	< .001
Ethical awareness	0.071	0.094	.086	0.761	.452

Data Source: Reanalysis of the 40-case questionnaire dataset (2026). Dependent variable: engagement; $R^2 = .637$, adjusted $R^2 = .618$, $F(2, 37) = 32.52$, $p < .001$. Coefficient 95% CIs: perceptions [0.519, 0.976]; ethical awareness [−0.118, 0.261].

Model checks from the archived responses were as follows: Shapiro–Wilk test on residuals $p = .387$; Breusch–Pagan heteroskedasticity test $p = .169$; and VIF = 1.31 for each predictor. Four cases exceeded the conventional Cook's distance screening value of 4/N, so inference warrants caution. With HC3 robust standard errors, perceptions remained associated with engagement ($B = 0.748$, 95% CI [0.366, 1.129], $p < .001$), whereas ethical awareness did not (95% CI [−0.208, 0.351], $p = .609$). Leave-one-out analysis retained a significant perception coefficient, but these diagnostics cannot resolve sampling bias or shared-method measurement.

Because reported responsible AI use was the fourth construct, a supplementary, explicitly exploratory regression also treated it as an outcome with perceptions and ethical awareness as predictors. That model had $R^2 = .380$. Perceptions showed $B = 0.572$ ($p = .009$; HC3 $p = .033$), whereas ethical awareness showed $B = 0.347$ (conventional $p = .050$; HC3 $p = .204$). The latter estimate was not robust to the choice of standard error; neither this model nor the correlations establish observed ethical behavior. The primary inference of this paper therefore remains the adjusted association between perceptions and engagement. Because the responsible-use measure is self-reported, this secondary model should not be interpreted as a behavioral validation of responsible AI practices.

Responsible AI in Arabic Language Learning: Pedagogical and Ethical Implications

The central contribution is a distinction between perceived usefulness-related evaluations, engagement, and ethical awareness within one preliminary Arabic-learning survey. The strong adjusted perception coefficient and nonsignificant ethical-awareness coefficient are not inconsistent with the positive bivariate correlations: they answer different statistical questions. A plausible explanation is that immediate translation and feedback make AI attractive, while ethical knowledge concerns standards for use rather than the amount of engagement. This mechanism is interpretive and requires experimental or longitudinal testing (Chan & Hu, 2023; Nguyen et al., 2023). Importantly, this dataset does not test whether ethical instruction changes behavior; pedagogical claims about its value are recommendations grounded in educational ethics, not causal findings of this survey.

Arabic teachers could use the observed preference for translation to assign three-stage tasks: learners first produce their own reading or translation, then compare it with an AI output, and finally annotate and justify revisions

using a dictionary or grammar reference. This process is particularly suited to verb-pattern selection, agreement, and ambiguous lexical meaning. Such proposals extend task-design principles for Arabic instruction (Ryding, 2013); they were not tested as an intervention in this study.

AI literacy instruction should separate technical operation from critical verification. For example, students could examine whether a generated sentence reflects Modern Standard Arabic or dialect, whether gender and number agreement is accurate, and whether culturally or religiously marked vocabulary is appropriate. These examples make Arabic-specific implications tangible, but the present questionnaire did not assess students' actual success on them (Moustafa et al., 2026).

Institutions can operationalize academic-integrity guidance by stating which forms of AI help are permitted, requiring a short acknowledgment of substantive AI assistance, and assessing students' explanation of revisions rather than only the final text. Teachers should make expectations explicit for translations, essays, and AI-assisted feedback while retaining human judgment in evaluation (Jeon & Lee, 2023; H. Wang et al., 2024).

Limitations are substantial. The purposive N = 40 sample is concentrated in one study program (38/40) and semester (28/40 in semester eight), with no documented recruitment denominator or institutional coverage. All focal measures are self-reports from one survey, inviting common-method and social-desirability bias. Alpha values above .94 and r = .795 highlight the need to check item redundancy and discriminant validity. No verifiable records support claims of completed formal expert validation or a separate pilot test, and such claims are therefore not made. No Arabic output was scored, and the design cannot establish temporal or causal direction. Influential cases and unstable exploratory coefficients further limit inference.

The most defensible application is thus a testable pedagogical proposition rather than a general claim of effectiveness: Arabic teachers can pair useful AI assistance with explicit verification tasks and transparent-use expectations, then evaluate whether such practices improve independent performance. Replication should recruit diverse institutions and levels, preregister hypotheses, validate separate scales, and triangulate survey responses with learning logs and rated Arabic productions.

Conclusion

This preliminary student survey found widespread AI use for translation and course tasks alongside favorable self-reported perceptions, engagement, and ethical awareness. It answers the descriptive and correlational questions by documenting significant links among the four scales, especially the perception–

engagement association (r = .795). These results describe responses from 40 participants, not the broader population of Arabic learners.

For the main regression question, perceptions independently predicted engagement (B = 0.748, p < .001), but ethical awareness did not (B = 0.071, p = .452), despite its positive unadjusted association with engagement and responsible AI use. Ethical awareness therefore cannot be claimed as an independently significant predictor of engagement here; nor can the nonsignificant result establish that ethical instruction is unnecessary. The study contributes this distinction rather than evidence of a causal mechanism. The limited methodological contribution is the transparent comparison of conventional and heteroskedasticity-robust inference alongside sensitivity checks, not the validation of a new measurement instrument.

In practice, AI-assisted Arabic tasks should require students to check morphology, register, and meaning and disclose substantial machine assistance. These are pedagogical recommendations, not outcomes demonstrated by this survey. Future research should test them with diverse larger samples, independently documented expert review and piloting, validated measures, longitudinal or intervention designs, and scored Arabic learning products. The findings remain hypothesis-generating because of small-sample, common-method, and construct-validity limitations. The study's theoretical implication is to distinguish acceptance-related perceptions from ethical awareness; its practical implication is to evaluate student explanations and linguistic corrections rather than assuming that frequent AI use signifies learning.

Acknowledgment

The authors would like to thank the students who voluntarily participated in this study and provided responses to the questionnaire. The authors also acknowledge the academic support of Institut Agama Islam Al-Zaytun Indonesia during the preparation of this manuscript.

Author Contributions Statement

IS conceptualized the study, developed the research design, analyzed the data, and drafted the manuscript. WI contributed to data collection, literature review, and manuscript revision. HM contributed to international collaboration, academic discussion, and final manuscript review. All authors read and approved the final version of the manuscript.

AI Writing Statement

During the preparation of this manuscript, the authors used ChatGPT to assist with language editing, grammar checking, and idea organization. The

authors carefully reviewed, verified, and edited the content generated with the assistance of AI and take full responsibility for the final content of this manuscript.

References

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI and Ethics, 2(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Alasmari, T. (2025). Artificial Intelligence and M-Learning in Arabic Countries: Innovations, Trends, and Regional Perspectives. International Journal of Interactive Mobile Technologies (iJIM), 19(05), 170–194. https://doi.org/10.3991/ijim.v19i05.52735

Alkaabi, M. H., & Almaamari, A. S. (2025). Generative AI Implementation and Assessment in Arabic Language Teaching. International Journal of Online Pedagogy and Course Design (IJOPCD), 15(1), 1–18. https://doi.org/10.4018/IJOPCD.368037

Allaithy, A., & Zaki, M. (2026). Evaluation of AI-generated reading comprehension materials for Arabic language teaching. Computer Assisted Language Learning, 39(5–6), 1296–1328. https://doi.org/10.1080/09588221.2025.2474037

Al-Shaboul, I. A., Ali, A. A., Kariem, A. A., Zarovna, I., Glushchenko, T., & Khasawneh, M. A. S. (2025). Bridging the gap: The role of artificial intelligence in enhancing Arabic language learning, translation, and speech recognition. Research Journal in Advanced Humanities, 6(2). https://doi.org/10.58256/17f30w24

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Frontiers in Public Health, 6. https://doi.org/10.3389/fpubh.2018.00149

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. International Journal of Educational Technology in Higher Education, 17, Article 2. https://doi.org/10.1186/s41239-019-0176-8

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence, 4, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542–570. https://doi.org/10.1111/ejed.12533

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Education and Information Technologies, 28(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

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., … 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

Khaldun, I. (1958). The Muqaddimah: An introduction to history (F. Rosenthal, Trans.; Vols. 1–3). Princeton University Press.

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

Koo, M., & Yang, S.-W. (2025). Likert-Type Scale. Encyclopedia, 5(1), 18. https://doi.org/10.3390/encyclopedia5010018

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. Computers and Education: Artificial Intelligence, 7, 100266. https://doi.org/10.1016/j.caeai.2024.100266

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. Proceedings of the 2020 CHI Conference on Human

Factors in Computing Systems, CHI '20, 1–16. https://doi.org/10.1145/3313831.3376727

Moustafa, A. H. A., Fayez Al-Hamad, M., Asif Qureshi, M., & Qadir, J. (2026). Generative AI for Learning and Teaching Arabic: Opportunities, Challenges, and Ethical Considerations. IEEE Access, 14, 7379–7395. https://doi.org/10.1109/ACCESS.2026.3651864

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. Education and Information Technologies, 28(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Ponto, J. (2015). Understanding and Evaluating Survey Research. Journal of the Advanced Practitioner in Oncology, 6(2), 168–171. https://doi.org/10.6004/jadpro.2015.6.2.9

Ryding, K. C. (2013). Teaching and Learning Arabic as a Foreign Language: A Guide for Teachers. Georgetown University Press. https://muse.jhu.edu/pub/171/monograph/book/62533

Sabilar, R. M., Ulin, N. M. A., Muhammad, H., Abdur, R. M., Ahmad, N., & Din, A. G. bin M. (2026). From CEFR to AI-Assisted Learning: Contemporary Transformations in Arabic Language Teaching for Non-Native Speakers. ATHLA : Journal of Arabic Teaching, Linguistic and Literature, 7(1), 37–56. https://doi.org/10.22515/athla.v7i1.14295

Sapawi, M. S. M., & Yusoff, N. M. R. N. (2025). Integrating technology into the Arabic language curriculum: A systematic review of trends, strategies and cultural dimensions. Social Sciences & Humanities Open, 12, 101974. https://doi.org/10.1016/j.ssaho.2025.101974

Schiff, D. (2022). Education for AI, not AI for Education: The Role of Education and Ethics in National AI Policy Strategies. International Journal of Artificial Intelligence in Education, 32(3), 527–563. https://doi.org/10.1007/s40593-021-00270-2

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, Article 15. https://doi.org/10.1186/s40561-023-00237-x

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward A Unified View1.

Management Information Systems Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. Computers and Education: Artificial Intelligence, 7, 100326. https://doi.org/10.1016/j.caeai.2024.100326

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 124167. https://doi.org/10.1016/j.eswa.2024.124167