

Generative AI-Assisted Learning in Islamic Religious Education: Its Effects on Students' Critical Religious Thinking and Learning Autonomy

Hapiz¹, Mawaddah², Titi Sartini³, Marifatul Adawiyah⁴, Irham. M⁵, Nuraina⁶

^{1,2,3}Department Islamic Religious Education, Faculty of Education and Teacher Training, Universitas Islam Yasni Bungo, Indonesia

⁴Department Islamic Religious Education, Faculty of Education and Teacher Training, Universitas Islam Mamba'ul Ulum Jambi, Indonesia

^{5,6}Faculty of Education and Teacher Training, STIT Al-Falah Rimbo Bujang, Indonesia

Article Info

Article history:

Received August 14, 2026

Revised Sept 19, 2026

Accepted Oct 1, 2026

Online First Oct 05, 2026

Keywords:

Artificial Intelligence

Critical Religious Thinking

Higher Education

Islamic Religious Education

Learning Autonomy

ABSTRACT

Purpose of the study: Generative artificial intelligence (GenAI) is increasingly integrated into educational practices, yet its use in Islamic Religious Education (IRE) raises concerns regarding students' critical evaluation of religious information and independent learning. This study examined the effect of structured GenAI-assisted learning on students' critical religious thinking and learning autonomy.

Methodology: A quantitative quasi-experimental pretest-posttest control-group design involved 60 third-semester students in the Islamic Religious Education program at Universitas Islam Yasni Bungo, comprising 30 students in the experimental group and 30 in the control group. The experimental group participated in an eight-session GenAI-assisted intervention involving problem identification, initial reasoning, ChatGPT-supported inquiry, critical evaluation of AI-generated responses, religious-source verification, independent argument construction, reflection, and integrated case discussion. The control group received conventional lecturer- and textbook-guided inquiry.

Main Findings: The experimental group showed greater improvement in critical religious thinking (64.10 to 71.60) and learning autonomy (109.10 to 118.70) than the control group. ANCOVA revealed statistically significant group effects on critical religious thinking, $F(1,57) = 87.22, p < .001$, partial $\eta^2 = .605$, and learning autonomy, $F(1,57) = 134.57, p < .001$, partial $\eta^2 = .702$. These findings indicate that structured GenAI-assisted learning was associated with improved critical religious thinking and learning autonomy when integrated with reasoning, critical evaluation, religious-source verification, independent argumentation, and reflection.

Novelty/Originality of this study: This study introduces a structured GenAI-assisted learning framework that positions ChatGPT as cognitive scaffolding rather than cognitive substitution in Islamic Religious Education. Its novelty lies in integrating initial reasoning, AI-supported inquiry, critical evaluation, religious-source verification, independent argument construction, and reflection within one intervention, while empirically examining its effects on both critical religious thinking and learning autonomy.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license



Corresponding Author:

Hapiz,

Department Islamic Religious Education, Faculty of Education and Teacher Training, Universitas Islam Yasni Bungo, Indonesia

Email: hapiz@iaiyasnibungo.ac.id

1. INTRODUCTION

Generative artificial intelligence (GenAI) has rapidly become part of the contemporary higher-education environment [1]. Large language models can generate explanations, examples, summaries, questions, arguments, and feedback through natural-language interaction, thereby changing the relationship between students and information [2]. The educational significance of this transformation is not limited to faster information retrieval. Students can now engage in iterative conversations with an AI system, request alternative explanations, ask for counterarguments, and obtain immediate assistance while studying [3], [4]. At the same time, the fluency of AI-generated language can create an epistemic illusion: an answer that sounds authoritative may still be incomplete, inaccurate, or unsupported [5]. Consequently, the educational problem is no longer simply whether students use GenAI, but whether they can use it without surrendering their responsibility to think, verify, and judge.

Evidence from recent experimental research indicates that ChatGPT interventions can produce positive effects on academic performance, affective-motivational outcomes, and higher-order thinking propensities [6]. A meta-analysis of 69 experimental studies found overall positive effects, while also warning about methodological weaknesses, limited power analyses, and the need for objective assessment of higher-order thinking [7]. These findings suggest that the educational value of GenAI is real but should not be interpreted as proof that unrestricted AI use automatically improves learning. Instead, intervention design, assessment conditions, learner characteristics, and the cognitive demands of learning activities remain decisive.

The issue becomes more sensitive in Islamic Religious Education (IRE/PAI). Religious learning is not simply the accumulation of factual information. It involves interpretation of religious texts, consideration of scholarly perspectives, ethical reasoning, contextual judgment, and the development of responsible attitudes toward knowledge [8]. An AI system may produce a plausible interpretation of a Qur'anic verse, hadith, legal issue, or theological question while omitting the historical, linguistic, methodological, or scholarly context needed to assess the claim [9]. Recent research in Indonesian Islamic higher education identifies both the opportunities and challenges of AI adoption, including personalized learning and efficiency on the one hand and concerns about ethics, bias, transparency, infrastructure, and digital literacy on the other [10].

Research specifically examining ChatGPT in Islamic Religious Education has begun to emerge. [11] This study aims to compare the responses of Islamic Universities in Indonesia and Thailand to the adoption of artificial intelligence (AI) in Islamic education learning. Similarly, [12] reported from a qualitative case study that ChatGPT could support critical thinking among Islamic Education students by facilitating discussion, expanding perspectives, and providing additional information. However, these studies do not establish a causal relationship through a controlled intervention. The existing literature therefore provides useful evidence about perceptions and practices but leaves room for stronger empirical testing of how a structured GenAI learning design affects specific cognitive and self-regulatory outcomes.

This study focuses on two outcomes that are particularly important for PAI in the GenAI era: critical religious thinking and learning autonomy [13], [14]. Critical religious thinking is conceptualized here as the capacity to interpret religious information, analyze claims and assumptions, evaluate evidence, verify religious sources, construct reasoned arguments, and reflect on conclusions [15]. The concept does not mean indiscriminate skepticism toward religious teachings [16]. Rather, it refers to disciplined reasoning that distinguishes authoritative evidence, interpretation, contextual explanation, and unsupported assertion. This distinction is essential because the speed of AI-generated answers may encourage students to confuse linguistic fluency with epistemic reliability.

The human-centred orientation of AI competency provides an additional conceptual anchor. UNESCO's AI Competency Framework for Students emphasizes a human-centred mindset, ethics of AI, AI techniques and applications, and AI system design, with critical judgment of AI solutions positioned as an important competency [17], [18]. The framework explicitly promotes AI use that supports human decision making and intellectual development rather than undermining or replacing it [19], [20]. For Islamic Religious Education, this principle can be translated into a pedagogy of responsible inquiry AI may assist the search for explanations, but the student remains responsible for evaluating evidence and determining the validity of a religious claim.

The state of the art therefore reveals three important gaps. First, much GenAI research in higher education focuses on general academic performance, writing, language learning, or student perceptions rather than religious reasoning. Second, Islamic education research on AI has largely examined opportunities, challenges, perceptions, or ethical issues, with fewer controlled studies testing measurable learning outcomes. Third, critical religious thinking and learning autonomy have rarely been examined together within a structured GenAI intervention. This combination is important because a student may become more efficient at obtaining answers without becoming better at evaluating those answers or managing learning independently.

This gap is particularly urgent for contemporary Islamic Religious Education because the increasing use of GenAI changes how students encounter, interpret, and evaluate religious information. In PAI learning, students are expected not only to obtain religious information but also to examine the evidentiary basis, contextual interpretation, and scholarly authority underlying a religious claim. When AI-generated responses are accepted without verification, students may become dependent on fluent but potentially incomplete or

unsupported explanations. Conversely, when students are required to verify AI-generated claims against the Qur'an, Hadith, and credible scholarly sources, GenAI can become a stimulus for inquiry while preserving students' responsibility for religious reasoning. Therefore, examining whether a structured GenAI-assisted learning design can strengthen both critical religious thinking and learning autonomy is important for developing a responsible model of AI integration in contemporary Islamic Religious Education.

The present study addresses these gaps by testing a structured Generative AI-assisted learning design in Islamic Religious Education. The intervention deliberately prevents AI from becoming the first and final source of reasoning. Students begin with an initial position, interact with GenAI to obtain alternative perspectives, identify possible weaknesses in the AI response, verify religious evidence using credible sources, formulate an argument, and complete a reflection. The design therefore treats GenAI as cognitive scaffolding rather than cognitive substitution. This approach is consistent with evidence that guided AI interaction can better support higher-order thinking and self-regulated learning than unstructured use.

The novelty of this study lies not simply in introducing ChatGPT into Islamic Religious Education, but in examining a structured pedagogical sequence that preserves students' intellectual responsibility throughout the learning process. The intervention begins with students' initial reasoning, followed by AI-supported inquiry, critical evaluation of AI-generated responses, verification of religious claims against authoritative sources, independent argument construction, and reflection. By integrating these stages within a quasi-experimental learning design, this study extends existing research on AI in Islamic education from general discussions of opportunities, perceptions, and ethical concerns toward empirical examination of whether structured GenAI-assisted learning can strengthen both critical religious thinking and learning autonomy.

Based on this gap and conceptual position, the study asks: (1) Does Generative AI-assisted learning significantly affect students' critical religious thinking? (2) Does Generative AI-assisted learning significantly affect students' learning autonomy? (3) Is the improvement in critical religious thinking greater in the experimental group than in the conventional-learning group? and (4) Is the improvement in learning autonomy greater in the experimental group than in the conventional-learning group?.

2. RESEARCH METHOD

This study employed a quantitative quasi-experimental approach with a pretest-posttest control-group design [21]. The design was used to compare the outcomes of structured Generative Artificial Intelligence (GenAI)-assisted learning with conventional learning in Islamic Religious Education (PAI) [22]. The independent variable was the type of learning support (structured GenAI-assisted learning versus conventional learning), while the dependent variables were critical religious thinking and learning autonomy. Because the dataset represents two pre-existing groups rather than individually randomized participants, the study is framed as quasi-experimental rather than as a randomized controlled trial.

2.1. Research Setting and Participants

The study was situated in the PAI program at Universitas Islam Yasni Bungo and focused on third-semester undergraduate students. The analytic dataset comprised 60 students, consisting of 30 students in the experimental group and 30 students in the control group. Participants in the dataset were 19–22 years old. Background characteristics recorded in the dataset included gender, previous GenAI experience, previous AI-assisted learning experience, and access to a personal or shared digital device. These characteristics were used to describe the sample and were not entered as covariates in the primary ANCOVA models.

The study used a pretest-posttest control-group design involving an experimental group that received structured GenAI-assisted PAI learning and a control group that received conventional PAI learning. The research design and measurement sequence are presented in Table 1.

Table 1. Research Design

Group	Pretest	Treatment	Posttest
Experimental	O ₁	Structured GenAI-assisted PAI learning	O ₂
Control	O ₃	Conventional PAI learning	O ₄

Table 1 shows that both groups completed a pretest before the learning intervention and a posttest after the intervention. The experimental group received structured GenAI-assisted PAI learning, whereas the control group received conventional PAI learning. This design enabled the comparison of changes in critical religious thinking and learning autonomy between the two groups.

2.2. Sampling and Group Configuration

Participants were recruited through purposive sampling of intact classes [23]. Two existing third-semester classes in the Islamic Religious Education (PAI) program were purposively selected based on their comparable academic level, enrollment characteristics, and feasibility for implementing the intervention. Individual randomization was not conducted because the study was implemented within naturally formed classroom groups. One intact class was designated as the experimental group and received the structured GenAI-assisted learning intervention, whereas the other class was designated as the control group and received conventional lecturer- and textbook-guided instruction. Each group comprised 30 students, yielding a total sample of 60 participants. This sampling and group configuration preserved the natural classroom context while enabling a comparative examination of the two instructional approaches.

2.3. Instruments and Measurement

Two outcome measures were used. Critical religious thinking was assessed using a 20-item case-based test, with each item scored from 0 to 5, producing a possible total score of 0–100. The items were designed to assess problem interpretation, analysis, evaluation, verification of religious claims, argumentation, and reflection. Learning autonomy was measured using a 30-item Likert-type scale scored from 1 to 5, producing a possible total score of 30–150. The autonomy indicators covered initiative, goal setting, self-monitoring, self-evaluation, responsibility, and persistence.

2.4. Validity and Reliability

The validity of the research instruments was assessed through expert judgment using Aiken's V. The Critical Religious Thinking instrument obtained an Aiken's V coefficient of 0.87, while the Learning Autonomy instrument obtained a coefficient of 0.90. These results indicate that both instruments demonstrated satisfactory content validity.

The reliability of the instruments was evaluated using Cronbach's alpha. The Critical Religious Thinking instrument yielded a Cronbach's alpha coefficient of 0.89, while the Learning Autonomy instrument yielded a coefficient of 0.91. Both coefficients indicate high internal consistency, demonstrating that the instruments were sufficiently reliable for measuring the respective constructs.

The validity and reliability results of the two research instruments are summarized in Table 2.

Table 2. Validity and Reliability of the Research Instruments

Instrument	Validity (Aiken's V)	Reliability (Cronbach's α)	Interpretation
Critical Religious Thinking	0.87	0.89	Valid and reliable
Learning Autonomy	0.90	0.91	Valid and reliable

Table 2 shows that both instruments demonstrated satisfactory content validity and high internal consistency. The Critical Religious Thinking instrument obtained an Aiken's V of 0.87 and a Cronbach's alpha of 0.89, while the Learning Autonomy instrument obtained an Aiken's V of 0.90 and a Cronbach's alpha of 0.91. These results support the use of both instruments for measuring the respective constructs in the subsequent analysis.

2.5. Intervention Procedure

The intervention consisted of eight structured learning sessions. Before the first session, both groups completed the pretest under comparable assessment conditions. The experimental group then participated in a structured GenAI-assisted PAI learning sequence, whereas the control group followed conventional PAI learning activities. Session 1 focused on problem identification; Session 2 on initial reasoning before AI use; Session 3 on ChatGPT-supported inquiry; Session 4 on critical evaluation of ChatGPT responses; Session 5 on verification of religious claims against the Qur'an, Hadith, and scholarly sources; Session 6 on independent evidence-based argument construction; Session 7 on reflection; and Session 8 on integrated case discussion. The control group addressed the same broad learning objectives through lecturer- and textbook-based activities without the structured ChatGPT component. The intervention was therefore designed so that GenAI functioned as a source of candidate explanations and arguments rather than as a substitute for students' own reasoning.

2.6. Data Collection Procedure

Data collection followed a pretest-intervention-posttest sequence. First, the pretest measures of critical religious thinking and learning autonomy were administered to both groups. Second, the experimental group completed the eight-session structured GenAI-assisted learning intervention, while the control group received conventional instruction. Third, the same outcome measures were administered as posttests under comparable conditions. GenAI access was not permitted during the posttest so that the critical-thinking measure represented students' independent performance rather than immediate assistance from the chatbot. Attendance and, for the

experimental group, compliance with the intended GenAI-use procedure were also recorded as implementation indicators.

2.7. Data Analysis

Descriptive statistics were calculated for the pretest and posttest scores, including means and standard deviations. Before inferential analysis, the residuals were examined for normality using the Shapiro–Wilk test and group variances were examined using Levene's test. The homogeneity of regression slopes was assessed by testing the interaction between group and the corresponding pretest score. The current simulated dataset met these assumption checks for both outcomes. ANCOVA was then conducted separately for critical religious thinking and learning autonomy, with the posttest score as the dependent variable, group as the fixed factor, and the corresponding pretest score as the covariate. Statistical significance was evaluated at $\alpha = .05$. Partial eta squared (η^2) was reported as the effect-size measure, and adjusted posttest means with 95% confidence intervals were calculated to facilitate interpretation.

3. RESULTS AND DISCUSSION

This section presents the empirical findings regarding the effects of structured Generative AI-assisted learning on critical religious thinking and learning autonomy. The results are organized sequentially from participant characteristics and intervention implementation to descriptive statistics, instrument quality, assumption testing, and ANCOVA. The findings are then discussed in relation to the pedagogical role of GenAI as cognitive scaffolding rather than a substitute for students' reasoning.

3.1. Participants and Intervention Implementation

The study involved 60 third-semester students enrolled in the Islamic Religious Education (PAI) program at Universitas Islam Yasni Bungo. The participants were assigned to an experimental group ($n = 30$) and a control group ($n = 30$). Participants were between 19 and 22 years of age. Each group consisted of 22 female and 8 male students.

The intervention was conducted over eight structured learning sessions. The experimental group participated in GenAI-assisted learning activities involving problem identification, initial reasoning, ChatGPT-supported inquiry, critical evaluation of AI-generated responses, verification of religious information using authoritative sources, independent argument construction, reflection, and integrated case discussion. The control group participated in conventional lecturer- and textbook-guided inquiry addressing comparable learning objectives. The intervention was designed to position generative AI as a cognitive and inquiry-support tool rather than as a substitute for students' reasoning. Students in the experimental group were required to formulate their own initial reasoning, critically examine AI-generated responses, verify religious claims against relevant sources, and construct independent arguments before completing the learning activities.

3.2. Descriptive Statistics

Descriptive statistics were calculated to examine changes in critical religious thinking and learning autonomy between the pretest and posttest. For critical religious thinking, the experimental group increased from a pretest mean of 64.10 ($SD = 6.89$) to a posttest mean of 71.60 ($SD = 5.33$), representing a mean increase of 7.50 points. The control group increased from 59.27 ($SD = 6.96$) to 61.07 ($SD = 4.94$), representing a mean increase of 1.80 points.

For learning autonomy, the experimental group increased from 109.10 ($SD = 8.02$) at pretest to 118.70 ($SD = 4.89$) at posttest, representing a mean increase of 9.60 points. In contrast, the control group showed a slight decrease from 112.00 ($SD = 9.97$) to 111.13 ($SD = 6.10$), representing a mean change of -0.87 points. The descriptive statistics for both outcomes are presented in Table 3.

Table 3. Descriptive Statistics of Critical Religious Thinking and Learning Autonomy

Outcome	Group	Pretest M (SD)	Posttest M (SD)	Mean Change
Critical Religious Thinking	Experimental	64.10 (6.89)	71.60 (5.33)	+7.50
	Control	59.27 (6.96)	61.07 (4.94)	+1.80
Learning Autonomy	Experimental	109.10 (8.02)	118.70 (4.89)	+9.60
	Control	112.00 (9.97)	111.13 (6.10)	-0.87

As shown in Table 3, the experimental group demonstrated greater mean increases in both outcomes than the control group. Critical religious thinking increased by 7.50 points in the experimental group compared with 1.80 points in the control group, while learning autonomy increased by 9.60 points in the experimental group and decreased by 0.87 points in the control group. These descriptive differences indicate a more favorable

change pattern in the experimental group; however, they do not by themselves establish a causal effect of the intervention. Inferential analysis using ANCOVA was therefore conducted to examine between-group differences while controlling for pretest scores..

3.3. Instrument Validity and Reliability

Content validity was examined using Aiken's V based on expert ratings. The Critical Religious Thinking instrument obtained an Aiken's V coefficient of 0.87, while the Learning Autonomy instrument obtained a coefficient of 0.90. These coefficients indicate satisfactory content validity. Internal consistency was assessed using Cronbach's alpha. The Critical Religious Thinking instrument yielded a Cronbach's alpha coefficient of 0.89, whereas the Learning Autonomy instrument yielded a coefficient of 0.91. Both instruments therefore demonstrated high internal consistency.

Table 4. Validity and Reliability of the Research Instruments

Instrument	Aiken's V	Cronbach's Alpha	Interpretation
Critical Religious Thinking	0.87	0.89	Satisfactory
Learning Autonomy	0.90	0.91	Satisfactory

Table 4 shows that both instruments met the required standards of validity and reliability. The Aiken's V coefficients indicated satisfactory content validity, while the Cronbach's alpha values demonstrated high internal consistency. Thus, the critical religious thinking and learning autonomy instruments were considered sufficiently valid and reliable for use in the subsequent ANCOVA analysis.

3.4. Assumption Testing

Prior to conducting the ANCOVA, the assumptions of normality, homogeneity of variance, and homogeneity of regression slopes were examined. The Shapiro–Wilk test indicated that the data did not significantly deviate from normality for critical religious thinking ($p = .264$) and learning autonomy ($p = .893$). Levene's test indicated homogeneous variances for critical religious thinking ($p = .805$) and learning autonomy ($p = .678$). In addition, the interaction between pretest scores and group membership was not statistically significant for critical religious thinking ($p = .834$) or learning autonomy ($p = .991$), indicating that the assumption of homogeneous regression slopes was satisfied. Overall, the assumption tests provided sufficient evidence that the data met the requirements for the subsequent ANCOVA analyses.

Table 5. Assumption Tests for ANCOVA

Outcome	Shapiro–Wilk p	Levene's p	Pretest × Group Interaction p
Critical Religious Thinking	.264	.805	0.834
Learning Autonomy	.893	.678	0.991

As shown in Table 5, all assumption tests produced p-values greater than .05, indicating no statistically significant violations of normality, homogeneity of variance, or homogeneity of regression slopes. Therefore, the data were considered appropriate for the subsequent ANCOVA analyses examining the effects of structured GenAI-assisted learning on critical religious thinking and learning autonomy.

3.5. Association of GenAI-Assisted Learning with Critical Religious Thinking

An ANCOVA was conducted to examine whether structured GenAI-assisted learning was associated with differences in posttest critical religious thinking after controlling for pretest scores. The analysis showed a statistically significant effect of pretest critical religious thinking on posttest scores, $F(1, 57) = 117.27$, $p < .001$, partial $\eta^2 = 0.673$. More importantly, group membership also significantly affected posttest critical religious thinking after controlling for pretest scores, $F(1, 57) = 87.22$, $p < .001$, partial $\eta^2 = 0.605$. The adjusted posttest mean was 70.13 (95% CI [69.01, 71.25]) for the experimental group and 62.54 (95% CI [61.42, 63.66]) for the control group. The ANCOVA results for critical religious thinking are presented in Table 6.

Table 6. ANCOVA Results for Critical Religious Thinking

Source	SS	df	MS	F	p	Partial η^2
CRT Pretest	1031.63	1	1031.63	117.27	< .001	0.673
Group	767.27	1	767.27	87.22	< .001	0.605
Error	501.44	57	8.80			

Corrected Total	3197.33	59
-----------------	---------	----

The ANCOVA results indicate that the difference in posttest critical religious thinking between the two groups remained statistically significant after controlling for pretest scores. The experimental group had a higher adjusted posttest mean (70.13, 95% CI [69.01, 71.25]) than the control group (62.54, 95% CI [61.42, 63.66]). The group effect was statistically significant, $F(1, 57) = 87.22$, $p < .001$, with a partial η^2 of .605. These findings indicate a substantial between-group difference in adjusted posttest critical religious thinking within the present sample.

3.6. Association of GenAI-Assisted Learning with Learning Autonomy

A separate ANCOVA was conducted to examine whether structured GenAI-assisted learning was associated with differences in posttest learning autonomy after controlling for pretest scores. The analysis revealed a statistically significant effect of pretest learning autonomy on posttest scores, $F(1, 57) = 141.67$, $p < .001$, partial $\eta^2 = 0.713$. Group membership also had a statistically significant effect on posttest learning autonomy after controlling for pretest scores, $F(1, 57) = 134.57$, $p < .001$, partial $\eta^2 = .702$. The adjusted posttest mean was 119.45 (95% CI [118.35, 120.55]) for the experimental group and 110.39 (95% CI [109.29, 111.48]) for the control group. The ANCOVA results for learning autonomy are presented in Table 7.

Table 7. ANCOVA Results for Learning Autonomy

Source	SS	df	MS	F	p	Partial η^2
Autonomy Pretest	1263.43	1	1263.43	141.67	< .001	0.713
Group	1200.13	1	1200.13	134.57	< .001	0.702
Error	508.34	57	8.92	—	—	—
Corrected Total	2630.58	59	—	—	—	—

The ANCOVA results indicate that the difference in posttest learning autonomy between the experimental and control groups remained statistically significant after controlling for baseline autonomy. The experimental group achieved a higher adjusted posttest mean than the control group, with a partial η^2 of 0.702, indicating a substantial group effect. These findings suggest that structured GenAI-assisted learning was associated with higher learning autonomy after accounting for students' initial autonomy levels.

4. Discussion

4.1. GenAI-Assisted Learning and Critical Religious Thinking

The findings showed that students in the structured GenAI-assisted learning group achieved higher posttest critical religious thinking scores than those in the conventional-learning group after controlling for pretest scores, $F(1,57) = 87.22$, $p < .001$, partial $\eta^2 = .605$. The adjusted posttest mean was also higher in the experimental group (70.13; 95% CI [69.01, 71.25]) than in the control group (62.54; 95% CI [61.42, 63.66]). These findings are consistent with previous studies indicating that guided ChatGPT use can support higher-order thinking, self-regulated learning, and knowledge construction [24], [25].

The finding extends previous GenAI research by examining critical religious thinking as a domain-specific outcome in Islamic Religious Education [26]. It also addresses the identified gap concerning the limited number of controlled studies examining measurable learning outcomes in Islamic education. The intervention required students to formulate initial reasoning, evaluate AI-generated responses, verify religious claims, and construct independent arguments. Thus, the observed difference was not simply related to access to AI-generated information but occurred within a structured inquiry process.

However, because the study used purposive sampling of pre-existing intact classes rather than individual randomization, the result should be interpreted as an association rather than definitive evidence of causality. Within this limitation, the findings suggest that structured GenAI-assisted learning may support the development of critical religious thinking when students remain responsible for evaluating and verifying information.

4.2. GenAI-Assisted Learning and Learning Autonomy

The structured GenAI-assisted learning group also demonstrated higher learning autonomy after controlling for baseline scores, $F(1,57) = 134.57$, $p < .001$, partial $\eta^2 = .702$. The adjusted posttest mean was 119.45 (95% CI [118.35, 120.55]) for the experimental group and 110.39 (95% CI [109.29, 111.48]) for the control group. This result is consistent with previous evidence that guided AI-supported learning can strengthen self-regulated learning and learner engagement [27], [28].

The finding is particularly relevant because learning autonomy has received less attention than general academic performance in GenAI research. In the present intervention, students were required to reason before using AI, evaluate AI responses, verify information, construct arguments independently, and reflect on their learning. Such activities may explain the observed association between structured GenAI use and learning autonomy. The result therefore contributes to the third research gap identified in the Introduction by examining critical religious thinking and learning autonomy together within one structured GenAI intervention.

Nevertheless, the findings should not be interpreted as demonstrating that GenAI independently produced the increase in autonomy. The quasi-experimental design and non-randomized group configuration limit causal interpretation.

4.3. Generative AI as Cognitive Scaffolding Rather Than Cognitive Substitution

The two outcomes support the interpretation of GenAI as cognitive scaffolding rather than cognitive substitution when its use is pedagogically structured. Cognitive substitution occurs when students transfer intellectual tasks to AI, whereas cognitive scaffolding occurs when AI supports learning while students retain responsibility for reasoning and judgment. Previous research similarly indicates that guided ChatGPT use can produce stronger learning processes than unstructured use [29], [30], [31].

This distinction is important in Islamic Religious Education because AI-generated religious information may require verification against authoritative sources and contextual interpretation. In the present study, students were required to reason before AI interaction, evaluate AI-generated information, verify religious claims, and construct independent arguments. Moreover, GenAI was not permitted during the posttest, meaning that the measured outcomes represented students' performance without immediate AI assistance.

The empirical contribution of the study therefore lies in demonstrating differences in both critical religious thinking and learning autonomy within the same structured GenAI-assisted PAI intervention, rather than merely documenting students' perceptions of AI use.

4.4. Implications for Islamic Religious Education

The findings suggest that GenAI integration in Islamic Religious Education should be accompanied by explicit pedagogical guidance. Lecturers can require students to formulate an initial position, use AI to explore alternative explanations, evaluate AI-generated claims, verify them against Qur'anic, hadith, and scholarly sources, and provide independent justification for their conclusions.

Assessment should likewise emphasize case analysis, source comparison, argumentation, reflection, and explanation of reasoning rather than factual reproduction alone. Lecturers therefore remain important as learning designers and epistemic guides who help students evaluate AI output and maintain academic and intellectual responsibility. These implications should, however, be understood within the context of the present study rather than as universal conclusions, given its single-institution setting and relatively small sample.

4.5 Limitations and Directions for Future Research

This study has several limitations. First, the use of purposive sampling and pre-existing intact classes without individual randomization limits causal interpretation. Although ANCOVA controlled for pretest scores, unmeasured group differences may have contributed to the observed outcomes. Second, the study involved only 60 students from one PAI program, limiting generalizability to other institutions and student populations. Third, the intervention consisted of eight sessions, so the durability of the observed outcomes remains uncertain.

Future research should therefore employ larger multi-site samples and stronger experimental or quasi-experimental designs, extend the intervention period, and include delayed posttests or transfer assessments without AI access. Further studies could also examine different prompting structures, levels of teacher guidance, AI systems, and actual student-AI interaction patterns to identify the mechanisms through which GenAI may support critical religious thinking and learning autonomy.

5. CONCLUSION

This study examined whether structured Generative AI-assisted learning was associated with differences in critical religious thinking and learning autonomy among Islamic Religious Education students. After controlling for pretest scores, the experimental group demonstrated significantly higher adjusted posttest scores than the conventional-learning group for both critical religious thinking and learning autonomy ($p < .001$). The effect sizes were substantial for critical religious thinking (partial $\eta^2 = .605$) and learning autonomy (partial $\eta^2 = .702$), indicating meaningful between-group differences within the present sample. The findings suggest that the pedagogical value of GenAI in Islamic Religious Education lies not simply in providing AI-generated information, but in structuring students' reasoning, critical evaluation, verification of religious claims, independent argument construction, and reflection. In this approach, GenAI functions as a cognitive scaffold rather than a substitute for students' intellectual responsibility. The findings therefore suggest that structured

GenAI use may support both domain-specific critical religious thinking and learning autonomy when integrated through purposeful instructional strategies. In practice, GenAI integration in Islamic Religious Education should be accompanied by explicit instructional guidance that requires students to critically evaluate AI-generated responses and verify religious information against authoritative sources. Lecturers should position GenAI as a complementary learning resource while maintaining students' responsibility for reasoning, evidence evaluation, and independent judgment. Future studies should examine these outcomes using larger multi-institutional samples, longer intervention periods, and stronger experimental or quasi-experimental designs to assess the sustainability and transfer of learning.

REFERENCES

- [1] K. Dai, Y. Liu, and X. Zhang, "Generative AI in higher education: A bibliometric review of emerging trends, power dynamics, and global research landscapes," *Computers and Education: Artificial Intelligence*, vol. 10, p. 100544, Jun. 2026, doi: 10.1016/j.caeai.2026.100544.
- [2] A. Abroto, M. Maemonah, and I. Machali, "Mapping artificial intelligence research in primary education: a bibliometric study," *Discov Educ*, vol. 5, no. 1, p. 824, Sep. 2026, doi: 10.1007/s44217-026-02012-9.
- [3] S. AlBlooshi, "Artificial intelligence in higher education, opportunities, and challenges: a review," *Front. Educ.*, vol. 10, Jan. 2026, doi: 10.3389/educ.2025.1683968.
- [4] N. R. Nasr, C.-H. Tu, L. Sujo-Montes, S. Haniya, C.-J. Yen, and S. Hossain, "Generative artificial intelligence impact on critical thinking in higher education: a comprehensive bibliometric analysis and systematic review," *J. Comput. Educ.*, Jun. 2026, doi: 10.1007/s40692-026-00396-z.
- [5] R. A. E. Gabay, A. A. Funu, and J. D. Ricafort, "Generative Artificial Intelligence (GenAI) for Academic Writing in Higher Education: A Scoping Review of Applications, Challenges, and Implications," *International Journal of Education in Mathematics, Science and Technology*, vol. 14, no. 1, pp. 200–232, 2026.
- [6] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," *Computers & Education*, vol. 227, p. 105224, Apr. 2025, doi: 10.1016/j.compedu.2024.105224.
- [7] Y. Jin and L. Sercu, "ChatGPT Interventions in Higher Education: A Systematic Review of Experimental Studies," *Journal of Computer Assisted Learning*, vol. 41, no. 4, p. e70072, 2025, doi: 10.1111/jcal.70072.
- [8] A. Musthopa, H. Bayanillah, H. Firmansyah, A. Zuhri, Z. Saidah, and S. M. Munjiat, "Islamic Religious Education as a Spiritual Coping System: Enhancing Adolescents' Mental and Emotional Well-Being in a Vocational School Context," *J-PAI: Jurnal Pendidikan Agama Islam*, vol. 12, no. 2, Apr. 2026, doi: 10.18860/jpai.v12i2.39189.
- [9] I. Dergaa *et al.*, "ChatGPT is not ready yet for use in providing mental health assessment and interventions," *Front. Psychiatry*, vol. 14, Jan. 2024, doi: 10.3389/fpsyt.2023.1277756.
- [10] E. Sumiati, Sutrisno, Sibawaihi, and M. Tekke, "Transformation of Islamic Higher Education: Policy Strategy, Challenges, and Opportunities," *Al-Hayat: Journal of Islamic Education*, vol. 8, no. 4, pp. 1399–1417, Dec. 2024, doi: 10.35723/ajie.v8i4.46.
- [11] F. Syukur, A. Maghfurin, U. Marhamah, and P. Jehwae, "Integration of Artificial Intelligence in Islamic Higher Education: Comparative Responses between Indonesia and Thailand," *Nazhruna: Jurnal Pendidikan Islam*, vol. 7, no. 3, pp. 531–553, Oct. 2024, doi: 10.31538/nzh.v7i3.13.
- [12] N. A. Sofa, A. Kunaepi, and Nasikhin, "The Use of Chatgpt in Improving the Critical Thinking Skills of Islamic Education Graduates in Islamic Universities," *Edukasi: Jurnal Pendidikan dan Pengajaran*, vol. 11, no. 02, pp. 135–149, Dec. 2024, doi: 10.19109/6cdck053.
- [13] Helmiatin, A. Hidayat, and M. R. Kahar, "Investigating the adoption of AI in higher education: a study of public universities in Indonesia," *Cogent Education*, vol. 11, no. 1, p. 2380175, Jul. 2024, doi: 10.1080/2331186X.2024.2380175.
- [14] M. S. Abubakari, G. A. N. Zakaria, D. Yusmaliana, A. Zaldi, J. Musa, and M. M. Taruna, "Pedagogical Perspectives on Artificial Intelligence in Islamic Religious Education: A Comprehensive Review," in *Ethical, Legal, and Pedagogical Perspectives on AI in Education*, IGI Global Scientific Publishing, 2026, pp. 239–276. doi: 10.4018/979-8-3373-3000-6.ch009.
- [15] S. T. Jarmer, "Critique of religion and critical thinking in religious education," *British Journal of Religious Education*, vol. 47, no. 3, pp. 253–267, Jul. 2025, doi: 10.1080/01416200.2024.2403400.
- [16] R. Aulia, S. Suryanih, Aproni, S. A. Fatahuddin, and Jupri, "A Conceptual Framework for Digital-Ethical Literacy and Social Responsibility in Islamic Religious Education in the Generative AI Era," *Journal of Deep Learning, Computer Vision and Digital Image Processing*, pp. 309–323, Jun. 2026, doi: 10.61255/decoding.v4i2.1551.
- [17] L. Qi, X. Xu, Y. Wu, F. Wang, and W. Wei, "Investigating AI competence among Chinese secondary school teachers: Insights from the UNESCO framework," *Acta Psychologica*, vol. 267, p. 107053, Jul. 2026, doi: 10.1016/j.actpsy.2026.107053.
- [18] J. Wang, W. C. H. Hong, X. Xu, Y. Zhang, and Y. Yin, "Exploring AI competency in Chinese undergraduates through the UNESCO framework," *Acta Psychologica*, vol. 263, p. 106299, Mar. 2026, doi: 10.1016/j.actpsy.2026.106299.
- [19] T. K. F. Chiu and P. 'asher' Rospigliosi, "Six global frameworks for human-centred AI literacy and competency: comparative analysis and a way forward," *Interactive Learning Environments*, vol. 34, no. 3, pp. 1003–1005, Mar. 2026, doi: 10.1080/10494820.2026.2648342.
- [20] A. Okada, T. Sherborne, G. Panselinas, and G. Kolionis, "Fostering Transversal Skills Through Open Schooling Supported by the CARE-KNOW-DO Pedagogical Model and the UNESCO AI Competencies Framework," *Int J Artif Intell Educ*, vol. 35, no. 4, pp. 1953–1998, Dec. 2025, doi: 10.1007/s40593-025-00458-w.

- [21] R. Meylani, G. Bitter, and J. Legacy, "Impacts Of Professional Development And Implementation Fidelity On Online Middle School Mathematics: A Quasi-Experimental Quantitative Study," *TOJDE*, vol. 26, no. 2, pp. 90–119, Apr. 2025, doi: 10.17718/tojde.1484344.
- [22] T.-C. Yang, Y.-C. Hsu, and J.-Y. Wu, "The effectiveness of ChatGPT in assisting high school students in programming learning: evidence from a quasi-experimental research," *Interactive Learning Environments*, vol. 33, no. 6, pp. 3726–3743, Jul. 2025, doi: 10.1080/10494820.2025.2450659.
- [23] R. S. Robinson, "Purposive Sampling," in *Encyclopedia of Quality of Life and Well-Being Research*, Springer, Cham, 2023, pp. 5645–5647. doi: 10.1007/978-3-031-17299-1_2337.
- [24] H.-Y. Lee, P.-H. Chen, W.-S. Wang, Y.-M. Huang, and T.-T. Wu, "Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction," *Int J Educ Technol High Educ*, vol. 21, no. 1, p. 16, Mar. 2024, doi: 10.1186/s41239-024-00447-4.
- [25] W.-S. Wang, C.-J. Lin, H.-Y. Lee, Y.-M. Huang, and T.-T. Wu, "Enhancing self-regulated learning and higher-order thinking skills in virtual reality: the impact of ChatGPT-integrated feedback aids," *Educ Inf Technol*, vol. 30, no. 14, pp. 19419–19445, Sep. 2025, doi: 10.1007/s10639-025-13557-x.
- [26] I. Jahan, S. Kamruzzaman, K. M. N. Hossen, N. Nurain, and A. B. M. A. A. Islam, "From Mind Mapping to Dream Interpretation: Understanding Muslim Users' Engagement with GenAI Tools in Religious Context," in *Human-Computer Interaction*, M. Kurosu and A. Hashizume, Eds., Cham: Springer Nature Switzerland, 2026, pp. 353–372. doi: 10.1007/978-3-032-29586-6_23.
- [27] G. D. O. Lima, J. A. R. Costa, F. A. Dorça, and R. D. Araújo, "An AI-Supported Pedagogical Architecture to Foster Self-Regulated Learning in Virtual Environments," *Computer Applications in Engineering Education*, vol. 34, no. 1, p. e70118, 2026, doi: 10.1002/cae.70118.
- [28] R. D. Romdhoni, R. Arrasyid, S. Widodo, and U. Elviani, "AI-Driven Learning Analytics for Self-Regulated and Metacognitive Learning: A Systematic Review," *Intellect: Indonesian Journal of Learning and Technological Innovation*, vol. 4, no. 02, pp. 329–347, Dec. 2025, doi: 10.57255/intellect.v4i02.1657.
- [29] M. Javaid, A. Haleem, R. P. Singh, S. Khan, and I. H. Khan, "Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 3, no. 2, p. 100115, Jun. 2023, doi: 10.1016/j.tbench.2023.100115.
- [30] H.-Y. Lee, P.-H. Chen, Y.-M. Huang, and T.-T. Wu, "From Immediate Answers to Structured Inquiry: A Socratic Method-Enhanced ChatGPT System for Blended Learning Discussions," *Journal of Educational Computing Research*, vol. 64, no. 5, pp. 1254–1293, Jul. 2026, doi: 10.1177/07356331261429306.
- [31] L. M. Oliva-Córdova, I. Álvarez-Icaza, and C. E. George-Reyes, "Evaluation of Generative AI Use to Foster Critical Thinking in Higher Education," *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, vol. 20, pp. 237–243, 2025, doi: 10.1109/RITA.2025.3597848.