Ahlussunnah: Journal of Islamic Education
https://ejournal.stitahlussunnah.ac.id/index.php/JIE

Artificial Intelligence Chatbots for Enhancing Higher Order Thinking Skills in Islamic Education

Silvie Afifatuz Zulfah[1], Nur Fitriyanti[1]
[1]Institut Agama Islam YPBWI Surabaya, Indonesia
silvie.afifahzulfah@gmail.com *

Article Information:
Received June 10, 2026
Revised July 18, 2026
Accepted August 10, 2026

Keywords: *Artificial intelligence, educational chatbot, higher order thinking skills, Islamic education*

Abstract

The development of Artificial Intelligence (AI) has created opportunities for developing more adaptive, interactive, and student-centered learning. However, empirical evidence on the use of AI-powered educational chatbots to support Higher Order Thinking Skills (HOTS) in Islamic Education at the junior high school level remains limited. This study aimed to analyze the differences in students' HOTS between learning supported by an AI-powered educational chatbot and conventional learning. The study employed a quantitative approach using a quasi-experimental method and a pretest-posttest control group design. The sample consisted of 64 eighth-grade students selected using cluster random sampling, with 32 students in the experimental group and 32 students in the control group. Data were collected through HOTS tests, observation sheets, student response questionnaires, and documentation. Data analysis included descriptive statistics, normality and homogeneity tests, an Independent Samples t-test, Normalized Gain (N-Gain), and effect size analysis using IBM SPSS Statistics. The results showed that the experimental group achieved a significantly higher HOTS posttest score ($M = 85.47$) than the control group ($M = 74.16$; $p < .001$). The increase in HOTS was also higher in the experimental group, as indicated by an N-Gain score of 0.64 compared with 0.35 in the control group, with a Cohen's *d* effect size of 1.96. These findings indicate that the use of an AI-powered educational chatbot was associated with improved HOTS, particularly in analysis, evaluation, and creation. The use of chatbots to support interaction, provide feedback, and facilitate the exploration of contextual problems has potential as an innovative learning medium for Islamic Education.

INTRODUCTION

The development of digital technology has transformed various aspects of human life, including education systems. The Fourth Industrial Revolution, characterized by the integration of cyber technologies, the Internet of Things (IoT), big data, and Artificial Intelligence (AI), and subsequently reinforced by the concept of Society 5.0, has encouraged a paradigm shift in education from teacher-centered to student-centered learning. Within this paradigm, technology is no longer viewed

How to cite: Zulfah, S. A., Fitriyanti, N. (2026). Artificial Intelligence Chatbots for Enhancing Higher Order Thinking Skills in Islamic Education. *Ahlussunnah: Journal of Islamic Education*, *5*(2), 1300-1314.

E-ISSN: 2827-9573
Published by: The Institute for Research and Community Service

merely as a medium for delivering information but as an integral component of the learning process that can support learning personalization, enhance interaction, and develop 21st-century competencies, including critical thinking, creativity, communication, and collaboration (Suaidi et al., 2025). Therefore, educational institutions need to develop learning approaches that integrate technological advances with pedagogical strategies to make learning more adaptive, effective, and relevant to the needs of a digital society (Makridakis, 2017).

This transformation has positioned Artificial Intelligence as an increasingly influential technological innovation in education. AI is a branch of computer science that enables systems to perform functions resembling human cognitive abilities through machine learning, data analysis, pattern recognition, and natural language processing, or Natural Language Processing (NLP). In recent years, the development of AI has expanded its application from administrative functions to pedagogical functions that directly support the learning process. Various AI applications, including intelligent tutoring systems, adaptive learning systems, automated assessment, learning analytics, and educational chatbots, can provide more personalized learning experiences by presenting materials according to students' needs, providing real-time feedback, and creating flexible learning environments that are not restricted by space and time (Huda, 2024; Sari Hernawati et al., 2024; Shofiyyah et al., 2024).

This development indicates that AI is not merely a technological innovation but also part of a pedagogical transformation. The integration of AI enables students to engage in more active learning experiences through information exploration, problem solving, data-based discussion, and reflection on learning outcomes. At the same time, teachers can receive support in designing more adaptive learning through the analysis of student characteristics and the provision of faster feedback. Thus, AI has the potential to support the quality of the learning process as well as the development of higher-order thinking competencies.

The importance of developing higher-order thinking skills has increased alongside the growing demand for 21st-century competencies. The ability to merely recall information is no longer sufficient to address the complexity of problems that are rapidly evolving. Students need to be able to analyze information, evaluate alternative solutions, and generate creative solutions based on available evidence. These competencies are referred to as Higher Order Thinking Skills (HOTS), which encompass analysis (C4), evaluation (C5), and creation (C6) in the revised Bloom's Taxonomy developed by Anderson and Krathwohl. HOTS is an important competency in education because it is associated with problem-solving, decision-making, and lifelong learning.

Nevertheless, various international reports indicate that students' higher-order thinking abilities in developing countries, including Indonesia, still require attention. International assessment results indicate that some students continue to experience difficulties in solving problems that require analysis, evaluation, and complex reasoning. This condition suggests that learning still faces challenges in shifting its orientation from factual knowledge acquisition toward the development of critical and creative thinking skills. Therefore, learning innovations that can facilitate higher-order thinking processes have become increasingly important.

The development of HOTS has particular relevance to Islamic Education. As a subject that aims to develop students' faith, morality, and character, Islamic Education requires not only mastery of Islamic concepts but also the ability to understand, interpret, and apply Islamic values in various contexts of life. In the 21st century, Islamic Education should not focus solely on memorizing verses, hadith, or jurisprudential concepts but should also encourage students to think critically about social, ethical, and moral issues emerging in society. Accordingly, students are

expected to make decisions based on rational reasoning as well as moderate and contextual Islamic values (Ahmed, 2020; Saada & Magadlah, 2021).

Theoretically, the development of HOTS in Islamic Education is strongly grounded in constructivist theory. This theory views knowledge not as something directly transferred from teachers to students but as something actively constructed through learning experiences, social interaction, reflection, and processes of meaning negotiation. From a constructivist perspective, learning becomes more meaningful when students are given opportunities to ask questions, test ideas, connect new knowledge with previous experiences, and construct understanding through inquiry. An interactive, collaborative, and feedback-rich learning environment is important for developing higher-order thinking because it encourages students to analyze and evaluate the information they encounter.

These characteristics are compatible with Artificial Intelligence technologies, particularly educational chatbots based on Natural Language Processing. Chatbots enable two-way interaction through natural language-based conversations. Unlike conventional learning media, which tend to operate in a one-way manner, chatbots can respond to students' questions, guide discussions, pose follow-up questions, and provide feedback that encourages reflection. Therefore, AI-based chatbots can function not only as digital sources of information but also as learning support tools that have the potential to facilitate higher-order thinking through dialogic interaction.

Despite the rapid development of educational technology, Islamic Education at the junior high school level still faces various challenges. Classroom learning practices are often dominated by teacher-centered learning, with teachers serving as the primary source of information while students mainly receive material passively. Lecture-based methods are also frequently used to deliver religious content, whereas learning activities that encourage discussion, argumentation, reflection, and problem solving are not always implemented optimally. This condition may cause students to become more accustomed to memorizing concepts, religious evidence, and religious rulings than to developing the ability to analyze, evaluate, and apply Islamic values in complex life situations (Adiyono et al., 2022; Agustini & W, 2021; Alhashmi & Moussa-Inaty, 2021).

This problem is related to the challenges of developing HOTS in Islamic Education. The substance of Islamic teachings encompasses various issues that require higher-order thinking, such as analyzing social phenomena based on Islamic legal principles, evaluating different perspectives emerging in society, and formulating solutions to life problems while remaining grounded in the values of the Qur'an and Hadith. Therefore, Islamic Education needs to be directed not only toward conceptual understanding of religious teachings but also toward students' ability to use such knowledge as a basis for making rational, critical, and responsible decisions. The development of HOTS is important to enable Islamic Education to foster intellectual competencies as well as character that are relevant to the challenges of 21st-century life (Essabane et al., 2022; Kholidah, 2022; Ter Avest et al., 2021).

In this context, the use of Artificial Intelligence-based educational chatbots offers one potential learning alternative. Educational chatbots are AI-based applications that utilize Natural Language Processing technology to establish interactive communication between systems and users through natural language. Unlike static learning media, chatbots can respond to students' questions, guide concept exploration, provide adaptive feedback, and facilitate discussions that encourage reflective thinking. Such interactions can provide students with opportunities to learn through asking questions, clarifying information, comparing alternative answers, and developing arguments based on available evidence (Godazgar, 2025; Shofiyyah et al., 2024).

Pedagogically, the characteristics of educational chatbots are compatible with

the principles of constructivist learning. Dialogic interaction enables students to construct knowledge through exploration and reflection rather than merely receiving information passively. In addition, the ability of AI to provide rapid feedback can support scaffolding, helping students develop their understanding gradually according to their learning needs. Thus, the use of chatbots can not only expand access to information but also potentially support the cognitive processes underlying HOTS development, particularly analysis, evaluation, and creation.

Previous studies have shown that the integration of digital technology can support the quality of Islamic Education. The use of digital learning media has been reported to improve learning motivation, student engagement, the effectiveness of content delivery, and learning outcomes across various learning competencies (Komalasari & Yakubu, 2023; Ma'arif et al., 2022; Putkonen & Poulter, 2023). Other studies have also indicated that learning digitalization provides opportunities for teachers to develop more innovative and collaborative learning strategies that align with the characteristics of digital-generation learners (Huda, 2024; Suaidi et al., 2025). These findings indicate that the use of technology is relevant to efforts to improve the quality of Islamic Education.

Meanwhile, research on HOTS development in Islamic Education has generally focused on the implementation of active learning models, such as Problem Based Learning, Project Based Learning, Inquiry Learning, and Discovery Learning. These models have been reported to support students' abilities to analyze problems, evaluate alternative solutions, and generate solutions based on religious values (Kholidah, 2022; Tuna, 2022; Wakhidah & Erman, 2022). These findings indicate that HOTS development is closely associated with learning strategies that provide students with opportunities to think critically, engage in discussion, and solve problems actively.

At the same time, research on Artificial Intelligence in education has developed significantly in recent years. Much of this research has examined the use of AI for learning personalization, intelligent tutoring services, automated assessment, and the development of educational chatbots as learning support tools. Research findings indicate that chatbots can support learning interaction, provide faster feedback, and facilitate independent learning through natural language-based communication (Godazgar, 2025; Sari Hernawati et al., 2024; Shofiyyah et al., 2024). However, much of this research remains oriented toward technological development, system design, user acceptance, or conceptual discussions of the potential of AI in education.

An analysis of previous studies indicates that a research gap remains to be addressed. First, studies on the integration of AI into Islamic Education have largely discussed concepts, policies, and implementation opportunities without empirical examination of changes in students' cognitive abilities. Second, research on educational chatbots has generally emphasized technical system development and user experience rather than the measurement of higher-order thinking abilities through a quasi-experimental design. Third, research specifically examining the use of Artificial Intelligence-based chatbots in relation to the HOTS of junior high school students in Islamic Education remains limited. Consequently, empirical evidence concerning differences in students' abilities in analysis, evaluation, and creation between those using AI-based chatbots and those receiving conventional instruction remains limited and requires further investigation.

Based on these gaps, this study offers novelty in three main aspects. First, it integrates constructivist theory, the revised Bloom's Taxonomy, and Artificial Intelligence technology into a conceptual framework for explaining HOTS development in Islamic Education. Second, it provides empirical evidence regarding the use of Artificial Intelligence-based educational chatbots through a quantitative approach with a quasi-experimental pretest-posttest control group design, enabling a

comparison of changes in HOTS between the experimental and control groups. Third, it extends research on the application of AI in Islamic Education at the junior high school level, a context that remains relatively limited in the empirical literature.

Theoretically, this study is expected to contribute to the literature on the integration of Artificial Intelligence in Islamic education by providing empirical evidence concerning the use of educational chatbots to support higher-order thinking development. The findings are also expected to strengthen the relevance of constructivist theory to AI-based learning and extend the application of the revised Bloom's Taxonomy in digital learning environments. Practically, the findings are expected to provide a reference for teachers, schools, and learning developers in designing Islamic Education that is more innovative, adaptive, interactive, and oriented toward developing students' critical thinking abilities. In addition, the findings may serve as a consideration for developing strategies for integrating Artificial Intelligence into religious education at various educational levels.

Based on the discussion above, this study aims to analyze differences in Higher Order Thinking Skills between students who receive Islamic Education using an Artificial Intelligence-based educational chatbot and students who receive instruction using conventional methods. Specifically, the study examines changes in students' abilities in analysis (C4), evaluation (C5), and creation (C6) through a quasi-experimental pretest-posttest control group design. The findings are expected to provide empirical evidence regarding the use of Artificial Intelligence as a learning innovation that supports the development of 21st-century competencies and contributes to the transformation of Islamic Education toward learning that is more adaptive, interactive, and technology-based.

METHODS

Research Design

This study employed a quantitative approach with a quasi-experimental design because the researchers could not randomly assign individual students to classes that had already been established by the school (Armstrong & Kepler, 2018; Engkizar et al., 2024; Flannelly et al., 2018; Han, 2018; Holec & Marynowski, 2020). The study used a pretest-posttest control group design involving two groups: an experimental group that received an intervention using an Artificial Intelligence-based educational chatbot and a control group that received instruction using the methods routinely applied by the teacher. This design was employed to measure changes in students' Higher Order Thinking Skills (HOTS) before and after the intervention and to compare the outcomes between the two groups (Mengistu, 2024).

Population and Sample

The study was conducted at a junior high school in Surabaya during the second semester of the 2025/2026 academic year. The study population consisted of all eighth-grade students, totaling 224 students across seven classes. A sample of 64 students was selected using cluster random sampling, with the class serving as the sampling unit because the learning groups had been established prior to the study. Based on the random selection of classes, Class VIII-A was assigned as the experimental group, consisting of 32 students, while Class VIII-B was assigned as the control group, also consisting of 32 students. Cluster random sampling was used to provide each class with an equal opportunity to be selected as the research sample while minimizing disruption to the ongoing learning process at the school (Khoiroh, 2026).

Variables and Intervention

The independent variable in this study was the use of an Artificial Intelligence-based educational chatbot, while the dependent variable was students' HOTS in Islamic Education. HOTS was measured based on the cognitive dimensions of the

revised Bloom's Taxonomy, which include analysis (C4), evaluation (C5), and creation (C6).

During the learning process, the experimental group used the Artificial Intelligence-based chatbot to support discussions, content exploration, feedback provision, and case-based problem solving. Meanwhile, the control group received instruction using the methods routinely applied by the teacher without the use of an Artificial Intelligence-based chatbot. The intervention was implemented to compare changes in HOTS between the two groups.

Instruments and Data Collection

Data were collected using several research instruments. The primary instrument was a HOTS test administered before (pretest) and after (posttest) the intervention. The test was developed based on indicators of analysis, evaluation, and creation in Islamic Education and consisted of reasoned multiple-choice questions and essay questions. In addition to the test, a learning implementation observation sheet was used to ensure that the intervention was implemented in accordance with the planned learning scenario. A five-point Likert-scale student response questionnaire was also administered to obtain information about students' perceptions of using the Artificial Intelligence-based chatbot (Engkizar, Jaafar, Hamzah, Syafril, Ediz, et al., 2026; Hoque et al., 2023; Rokeman, 2024).

Before being used in the study, all instruments underwent content validity assessment by experts in Islamic Education and educational technology. Empirical validity was subsequently tested using the Pearson Product-Moment correlation, while instrument reliability was calculated using Cronbach's Alpha. The instruments were considered suitable for use if they met the validity criteria and had a reliability coefficient greater than 0.70, indicating that they could produce consistent and reliable data (Hair et al., 2022).

Data Analysis

Data were analyzed in stages using IBM SPSS Statistics version 26. Descriptive statistical analysis was initially conducted to describe the characteristics of the data using the mean, standard deviation, minimum, and maximum values. The analysis was then followed by assumption tests, including the Shapiro-Wilk test for normality and Levene's Test for homogeneity of variance (Chapman, 2018; Mustafy & Rahman, 2024; Shapiro et al., 2023). After the assumptions were met, hypothesis testing was conducted using an Independent Samples t-test to determine differences in mean HOTS scores between the experimental and control groups at a 5% significance level. In addition to hypothesis testing, improvements in HOTS were analyzed using Normalized Gain (N-Gain) to determine the level of improvement in scores from the pretest to the posttest in each group.

The magnitude of the intervention effect was analyzed using Cohen's d effect size to provide information about the magnitude of the difference between the two groups. This analysis complemented the statistical significance testing by providing not only information about whether a difference existed but also the magnitude of the difference in HOTS outcomes between the experimental and control groups (Chaokromthong, 2021). All findings were subsequently interpreted based on changes in scores, statistical significance, and the magnitude of the intervention effect in the context of Islamic Education.

RESULT AND DISCUSSION

This study involved 64 eighth-grade students at a junior high school, consisting of 32 students in the experimental group and 32 students in the control group. The experimental group participated in Islamic Education learning using an educational chatbot powered by Artificial Intelligence (AI), whereas the control group received instruction using conventional methods. Before the treatment was administered, both

groups completed a pretest to assess their initial Higher Order Thinking Skills (HOTS). After six instructional sessions, both groups completed a posttest using the same instrument to measure changes in higher-order thinking skills.

Table 1. Distribution of Research Participants

Group	Male	Female	Total
Experimental	17	15	32
Control	16	16	32
Total	33	31	64

Based on Table 1, the composition of participants in the two groups was relatively balanced. The experimental group consisted of 17 male students and 15 female students, whereas the control group consisted of 16 male students and 16 female students. Overall, the participants comprised 33 male students (51.56%) and 31 female students (48.44%). This composition indicates that the distribution of participants by sex was relatively comparable between the two groups.

Balanced participant characteristics are important in experimental research because they can help reduce the potential for selection bias. Although baseline equivalence is not determined solely by sex composition, the relatively balanced distribution provides an indication that the two groups had similar demographic characteristics. Baseline equivalence in academic ability was subsequently examined by comparing the pretest scores of the two groups.

The students' HOTS before and after the treatment were subsequently analyzed using descriptive statistics, including the minimum, maximum, mean, and standard deviation. The results are presented in Table 2.

Table 2. Descriptive Statistics of HOTS Scores

Group	N	Test	Min	Max	Mean	Std. Deviation
Experimental	32	Pretest	48	72	60.38	6.54
Experimental	32	Posttest	74	96	85.47	5.81
Control	32	Pretest	47	71	59.91	6.28
Control	32	Posttest	66	84	74.16	5.67

The results in Table 2 show that the mean pretest score of the experimental group was 60.38, whereas that of the control group was 59.91. The mean difference of 0.47 points indicates that the two groups had relatively similar baseline abilities descriptively. This condition provides a reasonable basis for comparing changes in HOTS following the treatment, although baseline equivalence should ideally also be confirmed through an appropriate statistical test.

After the treatment, both groups experienced an increase in their posttest scores. The mean score of the experimental group increased to 85.47, whereas the control group reached 74.16. Thus, based on the difference between the pretest and posttest scores, the experimental group experienced an increase of 25.09 points, while the control group experienced an increase of 14.25 points. This difference indicates a greater change in scores among students who participated in learning using the educational chatbot powered by AI.

In addition to the increase in mean scores, the standard deviation of the experimental group decreased from 6.54 on the pretest to 5.81 on the posttest. This decrease indicates that score variability within the experimental group became smaller after the treatment. However, a change in standard deviation does not directly demonstrate that improvement occurred uniformly among all participants. This interpretation should therefore be considered together with the distribution of individual scores and other statistical analyses.

Descriptively, these findings indicate that learning using an AI-powered educational chatbot was associated with a greater increase in HOTS than conventional learning. During the learning process, students in the experimental group had opportunities to interact with the chatbot through questions, concept exploration, discussion, and case-based problem solving related to Islamic Education

content. These activities provided opportunities for students to develop analytical, evaluative, and creative thinking processes.

These findings are consistent with the concept of Artificial Intelligence in Education (AIED), which positions AI-based systems as technologies that can support adaptive learning by providing rapid responses and feedback. In this study, the chatbot provided students with opportunities to ask questions, clarify concepts, and explore learning materials more independently. Thus, interaction with the chatbot could expand the learning experience beyond one-way communication.

From a constructivist perspective, the use of a chatbot can be viewed as a form of digital scaffolding that provides stimuli, prompting questions, and feedback during the learning process. Students did not merely receive information but also engaged in constructing and developing their understanding through interaction with the learning environment. Accordingly, the increase in posttest scores can be understood in relation to a more interactive learning process that provided students with opportunities to actively construct knowledge.

These findings are also consistent with previous studies indicating that the use of AI in learning can improve the quality of learning interactions and student engagement (Huda, 2024; Shofiyyah et al., 2024). However, previous studies have largely focused on learning motivation, user perceptions, or the potential implementation of AI in education. The present study extends this body of research by examining changes in HOTS following the implementation of an AI-powered educational chatbot in Islamic Education.

Before hypothesis testing, the research data were first examined to ensure that the assumptions of parametric statistical analysis were satisfied. The prerequisite tests included the Shapiro-Wilk Test for normality and Levene's Test for homogeneity. The results are presented in Tables 3 and 4.

Table 3. Results of the Shapiro-Wilk Normality Test

Group	Data	Statistic	Sig.	Interpretation
Experimental	Pretest	0.973	0.382	Normal
Experimental	Posttest	0.981	0.641	Normal
Control	Pretest	0.976	0.467	Normal
Control	Posttest	0.969	0.298	Normal

The normality test results show that all significance values were greater than 0.05. Therefore, the pretest and posttest data for both groups did not show significant deviations from a normal distribution. The assumption of normality was satisfied, allowing parametric analysis to be used in the subsequent analysis.

Table 4. Results of the Homogeneity Test

Data	Levene Statistic	Sig.	Interpretation
Pretest	0.276	0.601	Homogeneous
Posttest	0.413	0.523	Homogeneous

The homogeneity test results show significance values of 0.601 for the pretest and 0.523 for the posttest. Both values were greater than 0.05, indicating that the variances of the two groups were homogeneous. With the assumptions of normality and homogeneity satisfied, parametric analysis could be continued to examine differences in HOTS between the experimental and control groups.

Next, the difference in HOTS between the two groups was analyzed using an Independent Samples t-test. This test aimed to determine whether there was a significant difference in posttest performance between students who participated in learning using an AI-powered educational chatbot and those who received conventional instruction.

Table 5. Results of Independent Samples t-test

Variable	Mean Experimental	Mean Control	t	Sig. (2-tailed)	Decision
HOTS Posttest	85.47	74.16	7.82	0.000	Ho Rejected

Based on Table 5, the mean *posttest* score of the experimental group was 85.47, whereas that of the control group was 74.16. The mean difference of 11.31 points indicates that the HOTS performance of the experimental group was higher than that of the control group. The Independent Samples t-test yielded a significance value of 0.000, which was lower than the significance level of 0.05. Therefore, H_0 was rejected and H_1 was accepted. These results indicate a significant difference in HOTS between students who participated in learning using an AI-powered educational chatbot and those who received conventional instruction.

These findings indicate that the difference in HOTS performance was not only reflected in the change in mean scores but was also supported by inferential statistical testing. In chatbot-based learning, students had opportunities to engage in dialogue, explore concepts, and solve problems. These activities could encourage students to identify relevant information, evaluate alternative answers, and formulate reasoning based on the concepts being studied.

From the perspective of the revised Bloom's Taxonomy, chatbot-based learning activities can support three dimensions of HOTS: analysis, evaluation, and creation. At the analysis stage (C4), students can be guided to identify relationships among concepts and analyze the causes and consequences of problems in the context of Islamic Education. At the evaluation stage (C5), students can compare alternative answers, provide reasons for their choices, and evaluate issues based on Islamic principles. At the creation stage (C6), students can develop solutions to case-based problems, allowing learning to progress beyond conceptual mastery toward the ability to generate ideas relevant to real-life situations.

These findings can also be explained through constructivist theory, which views knowledge as being actively constructed through interaction with the learning environment. An educational chatbot can function as a medium that provides prompting questions, conceptual clarification, and immediate feedback. This process gives students opportunities to reflect on their answers, correct inaccurate understandings, and construct knowledge based on their learning experiences while interacting with the system.

These findings reinforce the studies by Huda (2024) and Shofiyyah et al. (2024), which reported that Artificial Intelligence can improve learning quality through more adaptive and interactive learning experiences. However, previous studies have largely emphasized conceptual aspects, user acceptance, or the potential application of AI in education. The present study provides empirical evidence regarding differences in HOTS performance following the implementation of an AI-powered chatbot in Islamic Education.

In addition to testing statistical significance, this study analyzed the level of HOTS improvement using Normalized Gain (N-Gain). This analysis was used to describe the level of improvement in students' scores based on the difference between their abilities before and after the learning process.

Table 6. N-Gain Analysis Results

Group	Mean N-Gain	Category
Experimental	0.64	Moderate
Control	0.35	Moderate

The analysis in Table 6 shows that the experimental group obtained a mean N-Gain score of 0.64, whereas the control group obtained a score of 0.35. Based on the categories used in the study, both groups were in the moderate category. Nevertheless, the N-Gain score of the experimental group was higher than that of the control group, with a difference of 0.29 points. These results indicate a greater increase in HOTS among students who participated in learning using the AI-powered educational chatbot.

The N-Gain value of 0.64 for the experimental group indicates that the increase in HOTS occurred at a moderate level over the six instructional sessions.

This finding suggests that the development of HOTS is a process that requires continuous practice and learning experiences. The ability to analyze, evaluate, and generate solutions depends not only on mastery of the subject matter but also on the frequency with which students encounter problems requiring higher-order thinking.

These results are also consistent with the characteristics of AI-based learning, which provide students with opportunities to learn more flexibly. The chatbot allows students to review materials, ask questions, and receive immediate responses. These characteristics can expand learning opportunities and provide students with greater opportunities to develop higher-order thinking skills continuously.

The magnitude of the effect of using the AI-powered educational chatbot was subsequently analyzed using Effect Size (Cohen's d). This analysis aimed to determine the practical magnitude of the difference between the experimental and control groups, rather than relying solely on statistical significance.

Table 7. Effect Size Analysis

Variable	Cohen's d	Interpretation
HOTS	1.96	Very Large Effect

Based on Table 7, the Cohen's d value of 1.96 indicates a very large effect size. This value substantially exceeds the commonly used threshold of 0.80 for a large effect. Therefore, the difference in HOTS between the experimental and control groups was not only statistically significant but also represented a substantial practical difference.

This finding indicates that N-Gain and effect size provide different but complementary information. N-Gain describes the level of score improvement from the initial to the final condition, whereas effect size describes the magnitude of the difference in performance between the experimental and control groups. Therefore, a moderate N-Gain is not inconsistent with a very large effect size.

From the perspective of Artificial Intelligence in Education (AIED), the large effect size can be associated with the chatbot's ability to provide a more interactive and responsive learning experience. The system can provide rapid responses, offer alternative explanations, and support concept exploration through ongoing dialogue during the learning process. These conditions provide students with opportunities to become more actively involved in learning.

These findings indicate that the integration of Artificial Intelligence into Islamic Education can be positioned not only as a technological innovation but also as part of a pedagogical strategy. The added value of the chatbot lies in its ability to support learning interactions, provide feedback, and give students opportunities to actively construct their understanding in accordance with constructivist principles.

Overall, the findings show that the use of an educational chatbot powered by Artificial Intelligence was associated with an increase in students' Higher Order Thinking Skills (HOTS) in Islamic Education. This evidence is reflected in the increase in the mean posttest score of the experimental group, the significant difference identified through the Independent Samples t-test, the higher N-Gain value in the experimental group, and the very large effect size. Taken together, these findings indicate that the implementation of an AI-powered chatbot produced significantly different learning outcomes compared with conventional instruction within the context of this study.

These findings can be explained through constructivist theory, which views knowledge as being constructed through active interaction between learners and their learning environment. In this study, the chatbot served as a medium that supported concept exploration, question formulation, understanding verification, and feedback. This interaction provided students with opportunities to participate more actively in the learning process than in a more teacher-centered instructional pattern.

From the perspective of the revised Bloom's Taxonomy, the use of the chatbot provided opportunities for students to develop analysis (C4), evaluation (C5), and

creation (C6) skills. In the analysis dimension, students could identify relationships among concepts and analyze religious issues in the context of real-life situations. In the evaluation dimension, students could compare alternative solutions, provide logical reasoning, and assess issues based on Islamic values. Meanwhile, in the creation dimension, students could formulate solutions to various cases, allowing learning to progress from conceptual mastery toward the ability to generate relevant ideas.

The findings also support the development of the concept of Artificial Intelligence in Education (AIED), which positions AI as a technology capable of supporting adaptive and personalized learning. The chatbot enabled students to receive responses based on the questions they asked and their learning needs. This flexibility provided opportunities for students to receive feedback, clarify their understanding, and repeat learning processes during instructional activities.

The findings are consistent with the studies by Huda (2024), Shofiyyah et al. (2024), and Komalasari and Yakubu (2023), which reported that the integration of Artificial Intelligence can support learning quality, learning motivation, and student engagement. Nevertheless, much of the previous research has focused on user perceptions, media development, or conceptual discussions of the potential of AI in education. The present study contributes by employing a quasi-experimental design to directly examine differences in HOTS performance between students who used an AI-powered chatbot and those who received conventional instruction.

The difference between this study and previous research is also evident in the context of technology implementation. Some previous studies have been conducted in general subjects or higher education, whereas this study implemented an AI-powered chatbot in Islamic Education at the junior high school level. This context is important because Islamic Education is oriented not only toward conceptual mastery but also toward students' ability to analyze and apply Islamic values to various issues in life. Therefore, the implementation of a chatbot in this context provides an illustration of how AI technology can support more interactive and contextual religious education (Alouzi et al., 2026; Baroud et al., 2025; Engkizar et al., 2026).

The findings also show that the N-Gain value of the experimental group was in the moderate category, whereas the effect size was in the very large category. The difference between these two indicators demonstrates that the level of score improvement and the magnitude of between-group differences represent two distinct aspects. Six instructional sessions resulted in a significant increase in HOTS, but the development of higher-order thinking skills still requires continuous practice and consistent application of appropriate learning strategies.

From a practical perspective, the findings have implications for Islamic Education teachers in utilizing AI-powered chatbots as supplementary learning media. Chatbots can be used to expand opportunities for students to ask questions, explore learning materials, and receive feedback. In this context, teachers continue to play an important role as facilitators who guide students in using technology critically, responsibly, and in accordance with learning objectives. The integration of chatbots may also support differentiated learning by allowing students to explore learning materials according to their individual learning needs.

Theoretically, this study contributes to the field of Islamic educational technology by demonstrating that Artificial Intelligence can be used as a pedagogical instrument to support the development of higher-order thinking skills. The findings extend the application of constructivist theory and the revised Bloom's Taxonomy in AI-based learning contexts and add empirical evidence regarding the implementation of Artificial Intelligence in Education in Islamic Education. These findings can serve as a foundation for developing learning approaches that integrate AI technology with pedagogical strategies oriented toward HOTS.

Nevertheless, this study has several limitations. First, the study was conducted in only one school with a relatively limited sample size, so generalization of the findings to a broader population should be made with caution. Second, the treatment duration of six instructional sessions does not fully capture the long-term effects of using an AI-powered chatbot. Third, the study focused on the cognitive dimension of HOTS; therefore, the effects of the chatbot on affective aspects, religiosity, and collaborative skills were not examined in depth. Future research could involve larger samples, longer implementation periods, and an integration of quantitative and qualitative approaches to obtain a more comprehensive understanding of the effectiveness of Artificial Intelligence in Islamic Education.

CONCLUSION

This study concludes that the use of Artificial Intelligence-powered educational chatbots was associated with improved Higher Order Thinking Skills in Islamic Education, particularly in students' abilities to analyze, evaluate, and create. The findings indicate that chatbot-supported learning can facilitate more interactive and student-centered learning through two-way interaction, immediate feedback, and contextual problem solving, while supporting the development of higher-order thinking. The study contributes empirical evidence to research on Artificial Intelligence in Education by demonstrating the potential application of educational chatbots in Islamic Education. However, the study was limited to one school, a relatively small sample, and a limited implementation period. Future research should involve larger and more diverse samples, longer implementation periods, and different educational levels to examine the consistency of these findings. Further studies could also investigate the use of Artificial Intelligence chatbots in relation to other learning outcomes, including learning motivation, digital literacy, collaborative skills, and character development.

REFERENCES

Adiyono, A., Fadhilatunnisa, A., Rahmat, N. A., & Munawarroh, N. (2022). Skills of Islamic Religious Education teachers in class management. *Al-Hayat: Journal of Islamic Education, 6*(1), 104. https://doi.org/10.35723/ajie.v6i1.229

Agustini, A., & W, U. R. (2021). Strategies of Islamic Religious Education teachers in the development of teaching materials. *Al-Hayat: Journal of Islamic Education, 5*(2), 206. https://doi.org/10.35723/ajie.v5i2.186

Ahmed, A. (2020). Between "Islamic" and "un-Islamic": Navigating religion at an American Islamic high school. *Religious Education, 115*(4), 384–399. https://doi.org/10.1080/00344087.2020.1729682

Alhashmi, M., & Moussa-Inaty, J. (2021). Professional learning for Islamic education teachers in the UAE. *British Journal of Religious Education, 43*(3), 278–287. https://doi.org/10.1080/01416200.2020.1853046

Alouzi, K., Atia, A., Omran, S., Alrumayh, S., Albshkar, H., Alatrish, E., Masuwd, M., & Alfallah, B. (2026). AI and cultural sensitivity: Student perceptions of how AI handles religion, identity, and tradition in texts. *Muaddib: Journal of Islamic Teaching and Learning, 2*(1), 1–21.

Arif, M., & Aziz, M. K. N. A. (2023). Islamic Religious Education learning model in the 21st century: Systematic literature review. *Indonesian Journal of Islamic Education Studies (IJIES), 6*(2), 237–262. https://doi.org/10.33367/ijies.v6i2.4417

Armstrong, C. S., & Kepler, J. D. (2018). Theory, research design assumptions, and causal inferences. *Journal of Accounting and Economics, 66*(2–3), 366–373. https://doi.org/10.1016/j.jacceco.2018.08.012

Baroud, N., Ardila, Y., Akmal, F., & Sabrina, R. (2025). Opportunities and challenges

for Islamic education teachers in using Artificial Intelligence in learning. *Muaddib: Journal of Islamic Teaching and Learning, 1*(2), 1–11. https://muaddib.intischolar.id/index.php/muaddib/article/view/6

Chaokromthong, K., & Sintao, N. (2021). Sample size estimation using Yamane and Cochran and Krejcie and Morgan and green formulas and Cohen statistical power analysis by G*Power and comparisons. *Apheit International Journal of Interdisciplinary Social Sciences and Technology, 10*(2), 76–86. https://so04.tci-thaijo.org/index.php/ATI/article/view/254253

Chapman, S. J. (2018). Review of *Discovering Statistics Using IBM SPSS Statistics*, 4th Edition. *Journal of Political Science Education, 14*(1), 145–147. https://doi.org/10.1080/15512169.2017.1366328

Engkizar, E., Jaafar, A., Sarianto, D., Ayad, N., Rahman, A., Febriani, A., Oktavia, G., Guspita, R., & Rahman, I. (2024). Analysis of Quran education problems in majority Muslim countries. *International Journal of Islamic Studies Higher Education, 3*(1), 65–80. https://doi.org/10.24036/insight.v3i1.209

Engkizar, E., Jaafar, A., Hamzah, M. I., Syafril, S., Ediz, M. H., Efendi, E., Ridha, A., Febriani, A., & Oktavia, G. (2026). Tartil method as an effective pedagogical approach for enhancing students' interest in Quranic learning. *International Journal of Multidisciplinary Research of Higher Education (IJMURHICA), 9*(1). https://doi.org/10.24036/ijmurhica.v9i1.637

Engkizar, E., Jaafar, A., Hamzah, M. I., Syafril, S., Febriani, A., Oktavia, G., & Satrial, A. (2026). Quran teachers' skills as pedagogical foundations: Conceptual and practical insights into nine competencies. *Muaddib: Journal of Islamic Teaching and Learning, 2*(1), 22–38.

Essabane, K., Vermeer, P., & Sterkens, C. (2022). Islamic Religious Education and citizenship education: Their relationship according to practitioners of primary Islamic Religious Education in the Netherlands. *Religions, 13*(9). https://doi.org/10.3390/rel13090826

Flannelly, K. J., Flannelly, L. T., & Jankowski, K. R. B. (2018). Threats to the internal validity of experimental and quasi-experimental research in healthcare. *Journal of Health Care Chaplaincy, 24*(3), 107–130. https://doi.org/10.1080/08854726.2017.1421019

Godazgar, H. (2025). Exploring 'diversity' and 'pluralism': A sociological analysis of religious education textbooks in the 'Islamic Republic' of Iran. *British Journal of Religious Education, 47*(4), 465–484. https://doi.org/10.1080/01416200.2025.2451031

Hair, J., & Alamer, A. (2022). Partial least squares structural equation modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics, 1*(3), 100027. https://doi.org/10.1016/j.rmal.2022.100027

Han, C. (2018). Quantitative research methods in translation and interpreting studies. *The Interpreter and Translator Trainer, 12*(2), 244–247. https://doi.org/10.1080/1750399X.2018.1466262

Holec, V., & Marynowski, R. (2020). Does it matter where you teach? Insights from a quasi-experimental study on student engagement in an active learning classroom. *Teaching & Learning Inquiry, 8*(2), 140–164. https://doi.org/10.20343/teachlearninqu.8.2.10

Hoque, U. S., Akhter, N., Absar, N., Khandaker, M. U., & Al-Mamun, A. (2023). Assessing service quality using SERVQUAL model: An empirical study on some private universities in Bangladesh. *Trends in Higher Education, 2*(1), 255–269. https://doi.org/10.3390/higheredu2010013

Huda, M. (2024). Strengthening religious moderation through the core values of Islamic boarding school education. *Al-Hayat: Journal of Islamic Education, 8*(2).

https://doi.org/10.35723/ajie.v8i2.458
Khoiroh, U., & Anam, S. (2026). Analisis teknik sampling dalam penelitian kuantitatif dan pengaruhnya terhadap kualitas data. *Jurnal Pendidikan Sosial dan Humaniora, 5*(1), 1418–1424. https://publisherqu.com/index.php/pediaqu/article/view/3914
Kholidah, L. N. (2022). Improving students' social responsibility via Islamic Religious Education and social problem-based learning. *Jurnal Pendidikan Agama Islam (Journal of Islamic Education Studies), 10*(2), 163–182. https://doi.org/10.15642/jpai.2022.10.2.163-182
Kibtiyah, A., Idawati, K., & Muaz, Y. A. (2025). Collaboration on local wisdom-based character education between schools and parents of students in Islamic Religious Education units. *Al-Hayat: Journal of Islamic Education, 9*(2), 458–475. https://doi.org/10.35723/ajie.v9i2.175
Kolb, J. (2023). Muslim diversity, religious formation and Islamic religious education: Everyday practical insights into Muslim parents' concepts of religious education in Austria. *British Journal of Religious Education, 45*(2), 172–185. https://doi.org/10.1080/01416200.2021.1911787
Komalasari, M., & Yakubu, A. B. (2023). Implementation of student character formation through Islamic Religious Education. *At-Tadzkir: Islamic Education Journal, 2*(1), 52–64. https://doi.org/10.59373/attadzkir.v2i1.16
Ma'arif, M. A., Rofiq, M. H., & Sirojuddin, A. (2022). Implementing learning strategies for moderate Islamic Religious Education in Islamic higher education. *Jurnal Pendidikan Islam, 8*(1), 75–86. https://doi.org/10.15575/jpi.v8i1.19037
Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, 46–60. https://doi.org/10.1016/j.futures.2017.03.006
Mengistu, M. A., Worku, M. Y., & Melesse, T. (2024). Examining the impact of lesson study on teachers' professional development: A quasi-experimental study. *Bahir Dar Journal of Education, 24*(1), 25–36. https://doi.org/10.4314/bdje.v24i1.3
Mustafy, T., & Rahman, M. T. U. (2024). SPSS. In *Statistics and data analysis for engineers and scientists* (pp. 135–180). https://doi.org/10.1007/978-981-99-4661-7_4
Putkonen, N., & Poulter, S. (2023). Balancing differences through highlighting the common: Religious Education teachers' perceptions of the diversity of Islam in Islamic Religious Education in Finnish state schools. *Religions, 14*(8). https://doi.org/10.3390/rel14081069
Rokeman, N. R. M. (2024). Likert measurement scale in education and social sciences: Explored and explained. *EDUCATUM Journal of Social Sciences, 10*(10), 77–88. https://doi.org/10.37134/ejoss.vol10.1.7.2024
Saada, N., & Magadlah, H. (2021). The meanings and possible implications of critical Islamic religious education. *British Journal of Religious Education, 43*(2), 206–217. https://doi.org/10.1080/01416200.2020.1785844
Sari Hernawati, Hafizh, M., & Muhammad Nur Faizi Arya. (2024). Adjusting the ideal Islamic Religious Education curriculum to the development of AI-based technology. *Progresiva: Jurnal Pemikiran dan Pendidikan Islam, 13*(1), 129–144. https://doi.org/10.22219/progresiva.v13i01.32931
Shapiro, L. N., Gray, M. F., Freitag, C., Taneja, P., Kariya, H., Crane, P. K., O'Hare, A. M., Vig, E. K., & Taylor, J. S. (2023). Expanding the ethnographic toolkit: Using medical documents to include kinless older adults living with dementia in qualitative research. *Journal of Aging Studies, 65*, 101140. https://doi.org/10.1016/j.jaging.2023.101140
Shofiyyah, N. A., Lesmana, O., & Tohari, H. (2024). Metamorphosis of Islamic Religious Education learning method: Classic approach converted by Artificial

Intelligence (AI). *Jurnal Pendidikan: Riset dan Konseptual, 8*(2), 265. https://doi.org/10.28926/riset_konseptual.v8i2.998

Suaidi, S., Faridi, F., & Sunarto, S. (2025). Technology-based digitalization of Islamic Religious Education. *Halaqa: Islamic Education Journal, 9*(2). https://doi.org/10.21070/halaqa.v9i2.1734

Sutrisno, S., & Nasucha, J. A. (2022). Islamic Religious Education project-based learning model to improve student creativity. *At-Tadzkir: Islamic Education Journal, 1*(1), 13–22. https://doi.org/10.59373/attadzkir.v1i1.3

Ter Avest, I., Bakker, C., & Franken, L. (2021). Islamic religious education in the Netherlands. In *Islamic religious education in Europe: A comparative study* (pp. 179–195). Taylor & Francis. https://doi.org/10.4324/9780429331039-11

Tuna, M. H. (2022). The professionalisation of Islamic religious education teachers. *British Journal of Religious Education, 44*(2), 188–199. https://doi.org/10.1080/01416200.2021.1999905

Wakhidah, N., & Erman, E. (2022). Examining environmental education content on Indonesian Islamic religious curriculum and its implementation in life. *Cogent Education, 9*(1). https://doi.org/10.1080/2331186X.2022.2034244