Contemporary Readings in Law and Social Justice
ISSN: 1948-9137, e-ISSN: 2162-2752
Vol 16 (1s), 2024
pp. 162 - 172

Utilizing Artificial Intelligence (AI) in Generating Exam Questions for Islamic Studies Courses

Dr. Umm Kulthum Hakum Dawood bin Yahya[1], Dr. Muhammad Madiani[2]

[1]Department of Jurisprudence and its Principles, Faculty of Sharia, King Khalid University, obenyahia@kku.edu.sa

[2]College of Information Technology, United Arab Emirates University

Abstract

Discussions about artificial intelligence (AI) began in 1956; however, its implementation as a tool to simplify human life and add practical and professional value to their work was delayed by many decades. This delay occurred until AI became equipped with various forms of machine learning capable of identifying data patterns, enabling predictive capabilities. AI has proven its ability to transform work methods across various fields. New technologies, such as machine learning and deep learning, contribute to analyzing massive amounts of data quickly and accurately. This facilitates the extraction of patterns and trends that may not be apparent to human intelligence. AI can also generate educational content, including questions, answers, and tests, providing teachers with additional resources and interactive learning methods. Automating highly complex tasks, in addition to routine ones, has become a necessity and a requirement using AI. As this technology continues to evolve, its applications are likely to expand in the future, creating new opportunities and driving societal progress. To assess AI's role in generating questions for religious studies curricula, the researchers used the ChatGPT tool to generate diverse questions after providing it with the curriculum material. A question bank (one hundred questions) was created, combining questions from the course instructor and ChatGPT. Students were asked to answer the questions, indicating whether the question was posed by the professor or ChatGPT. The researchers then analyzed and commented on the results. The research yielded several key findings: First: AI can assist faculty members in automating tasks, such as summarizing material, organizing it, and creating questions and assignments related to the curriculum. However, complete reliance on AI is not yet feasible, as a degree of error was observed in true/false questions and multiple-choice/multiple-answer questions. Second: AI-generated questions were distinguished from the instructor's questions by their diversity and coverage of all curriculum elements. Thus, AI saves instructors time and effort in ensuring that questions encompass the entire syllabus. Thirdly: Artificial intelligence has succeeded in generating questions according to Bloom's taxonomy by using Bloom's action verbs that include: (knowledge, comprehension, application, analysis, synthesis, evaluation), which ensures a variety of levels of thinking in the required responses.

Keywords: Artificial Intelligence, Religious Studies Curricula, Exam Questions, Bloom's Taxonomy.

Received: 25 Jun 2024 **Revised**: 30 Aug 2024 **Accepted**: 15 Sep 2024 **Published**: 25 September 2024

1. Introduction

With the rapid advancement of modern technology, artificial intelligence (AI) has become a valuable tool for researchers and legal scholars in formulating and analyzing questions for legal courses. This technological progress provides an opportunity for educators and academics to alleviate teaching burdens, allowing them to focus more on scientific research and knowledge development.2]

AI is also recognized as an academic discipline dedicated to developing systems capable of exhibiting intelligent behavior. As a result, it plays a crucial role in modern technological advancements, marking a significant transformation in contemporary technology.

This study aims to assess the capability of AI programs to simulate human interaction in disciplines with religious and cultural specificity, such as Islamic Sharia sciences. It seeks to evaluate the effectiveness of AI in generating test questions for legal courses, ensuring accuracy, clarity, comprehensiveness, and a diverse range of difficulty levels. Additionally, the study examines whether these AI-generated questions align with Bloom's taxonomy and fulfill the intended learning outcomes outlined in course descriptions.

1.1 Study Questions

- To what extent can artificial intelligence fulfill the role of a professor in creating questions for Sharia courses?
- What is the accuracy level of AI-generated questions, and what is the margin of error?
- Are students able to distinguish between questions designed by their professor and those generated by artificial intelligence?
- Can artificial intelligence formulate questions in a manner that aligns with learning outcomes and Bloom's taxonomy?

1.2 Study Objectives

- Explore the Investigate the potential of AI in generating questions for Sharia courses.
- Determine the extent to which AI can replicate a professor's role in question formulation.
- Measure the error rate in AI-generated questions.
- Evaluate the accuracy and reliability of AI-generated questions in Sharia courses.
- Examine students' ability to differentiate between questions created by their professor and those generated by AI.
- The primary objective is to determine whether artificial intelligence can produce questions that align with learning outcomes and Bloom's taxonomy.

1.3 Study Limits

To explore the role of AI in generating questions for Islamic courses and assess students' ability to distinguish between professor-created and AI-generated questions, a study was conducted involving 80 female students. An electronic test was designed for the course on fundamental and logical introductions, consisting of 50 questions categorized as follows:

- True or false questions
- Multiple-answer questions
- Multiple-choice questions
- Essay questions

2. Previous Studies

Among the previous studies that we were able to stand on and benefit from:

– Turing Test

Several previous studies have contributed valuable insights to this study. One of the most influential is the Turing Test, proposed by British computer scientist Alan Turing in 1950. This test serves as a foundational reference for philosophical and ethical discussions on AI and consciousness. It remains a key framework for evaluating the development of intelligent AI systems (French, 2000).

The Turing Test operates by comparing the performance of three participants: a human (observer), a machine (AI program), and another human (witness). The observer communicates with both the machine and the human through a text-based interface without knowing which is which. After a period of interaction, the observer must determine which participant is the human. If the observer is unable to distinguish between them, the AI is considered to have "passed" the test, demonstrating human-like intelligence in conversation.

However, it is important to recognize that the Turing Test has limitations. It primarily assesses an AI system's ability to mimic human conversation but does not evaluate other critical aspects of intelligence, such as learning capabilities or problem-solving skills. Consequently, while the test remains historically significant, modern AI research has expanded beyond its scope to develop more comprehensive evaluation methods.

– ChatGPT versus human in generating medical graduate exam multiple choice questions—A multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom)

In this study, 50 doctoral exam questions were generated using ChatGPT, based on Harrison's "Principles of Internal Medicine" and Bailey & Love's "Short Practice of Surgery"(Cheung et al., 2023).

To evaluate the quality of these AI-generated questions, they were reviewed by five judges from various academic backgrounds, representing different international universities.

The study concluded with the following results:

- The AI-generated questions were produced in 20 minutes and 25 seconds, whereas human-created questions required approximately 211 minutes and 33 seconds to develop.
- The results indicated no significant difference in question quality between those generated by ChatGPT and those created by human experts.

– Evaluating Large Language Models in Medical Education: A Comparison of ChatGPT and Human-Created Exam Questions

This study examined the use of ChatGPT 3.5 in reformulating multiple-choice exam questions created by professors. A total of 25 multiple-choice questions were designed by educators, and then revised by ChatGPT for clarity and consistency. The questions were categorized based on their level of difficulty before being presented to students (Laupichler et al., 2024).

Students answered the questions in a randomized manner, and the evaluation was based on three key factors: question difficulty, discrimination power, and question source identification.

The main findings:

– Students correctly identified the origin of the questions 57% of the time.
– Question difficulty: No statistically significant difference was found between human-created and AI-generated questions.
– Discrimination power: Questions designed by human experts exhibited significantly higher discrimination power compared to those generated by ChatGPT.

3. Study methodology

The research adopted a descriptive-analytical approach to examine key concepts and terms, providing clarification and explanation. Additionally, an experimental approach was employed by generating a random question bank, which included both AI-generated questions and those created by the course professor. Furthermore, a comparative approach was utilized to analyze and compare the questions, as well as assess student performance based on their responses.

4. Study Structure

The research structure consists of an introduction, two main chapters, and a conclusion summarizing key findings and recommendations, along with a reference list of sources used in the study.

4.1 Section One: Study Terminology

First: Definition of Artificial Intelligence (AI)

The term "artificial intelligence" (AI) refers to intelligence attributed to human-made creations, distinguishing it from the innate intelligence bestowed upon humans by Allah Almighty. It is sometimes referred to as "machine intelligence" or "non-biological intelligence"(Briber, 2024).

– Linguistic Definition of Intelligence

In the Arabic language, the term "intelligence" originates from the trilateral root (Dhal, Kaf, and a weak letter), which conveys the meanings of sharpness, penetration, and depth in understanding (Al-Qazwini Al-Razi, 1979). The core meaning of the word relates to completion and perfection in cognition (Al-Herawi, n.d).

According to Tahdheeb al-Lughah, intelligence in comprehension is characterized by quick and complete understanding (Al-Harawi, 2001).

In Tahdheeb al-Akhlaq, intelligence is described as the speed and ease with which the mind processes results and their ease for the soul (Maskawayh, n.d).

In al-Kulliyyat: "Intelligence: the intensity of the soul's strength prepared to acquire opinions according to the language." (Al-Kafwi, n.d), and it is the newness of the heart, the speed of insight and nobility. The plural is dhikaa, and dhika yadhiku and dhiki, and its origin is ignition and flame, and a dhiki man; meaning: noble, and intelligence is the ability to understand, or think.

– Terminological Definition of Intelligence

From a technical perspective, intelligence is the innate cognitive ability of the soul that enables the acquisition of conceptual and belief-based knowledge. This cognitive power is commonly referred to as the mind (Al-Barakti, 2003).

– Definition of Artificial Intelligence

Artificial intelligence (AI) refers to the intelligence embedded in machines or systems created and programmed by humans. It enables machines to simulate human intelligence by performing tasks that typically require cognitive abilities, such as reasoning, learning, and problem-solving, mirroring human intellectual processes and achieving what a human being achieves with the intelligence that Almighty has given him (Briber, 2024).

– The Origins and Characteristics of AI

Historical Background: The concept of artificial intelligence (AI) can be traced back to the Dartmouth Conference in 1956, where the term "artificial intelligence" was first introduced by John McCarthy. His vision was to create intelligent machines, particularly computer programs, capable of simulating human cognitive abilities. This conference marked a significant milestone and served as a foundation for the development of AI as a field.

During the 1960s, AI research progressed with the development of programs such as ELIZA, the first conversational program, which laid the groundwork for natural language processing. This era also witnessed advancements in machine learning, shaping the future of AI.AI has been defined in various ways, including:

- Alan Turing described AI as "the ability to achieve human-level performance in cognitive functions"(Watanabe, 2021).
- Another definition describes AI as "the capability of a machine or system to perform activities that require intelligence, such as reasoning and self-awareness"(Abdel Hamid, 2008).
- AI has also been defined as "the science dedicated to building machines and developing computer systems that exhibit intelligence and perform tasks previously exclusive to humans"(Abdel Nour, 2005).

– Important Characteristics of AI (Bonnet, 1993)

AI systems are distinguished by several important characteristics, including:

- Learning Capability – AI systems can learn from past experiences and historical data, improving their performance over time.
- Inference and Decision-Making – These systems can analyze available information and derive conclusions or make decisions based on patterns and logical reasoning.
- Adaptability to New Situations – Unlike traditional machines that strictly follow predefined instructions, AI-powered systems can respond to unforeseen scenarios by utilizing technologies like Machine Learning (ML) and Deep Learning (DL).
- This adaptability enables AI to function effectively in dynamic environments where conditions are constantly changing. Examples include:

- Self-driving cars, which make real-time decisions based on traffic conditions and obstacles.
- Service robots, which adjust their responses to unexpected customer interactions.

Second: Sharia Courses

Sharia courses refer to the educational subjects included in an academic program, whether at the bachelor's or postgraduate level, which students are required to study throughout their program. These courses are typically structured around a thematic framework and have a designated number of teaching hours. Some courses are mandatory, meaning students must complete them to fulfill graduation requirements, while others are elective, allowing students some flexibility in their academic choices.

Third: Course Description

A course description serves as an academic identification document for a specific course. It includes essential details such as the program to which the course belongs, the academic level at which it is taught, and the total number of credit hours. Additionally, it outlines grading distribution, expected learning outcomes, and the knowledge and skills students are expected to acquire upon completing the course. The description also specifies the primary and supplementary learning resources necessary for mastering the course content.

4.2 Section Two: Study Stages

The study was conducted in two main stages:

First Stage – Developing the Question Bank

A diverse question bank was created, consisting of both AI-generated questions from ChatGPT and questions designed by the course professor. The questions were arranged randomly, and after a thorough review, unsuitable questions were eliminated. The selected questions were then finalized as test questions. Once the students completed the test, their responses were analyzed.

Second Stage – Training the AI Model

In this phase, the AI tool was provided with the course material for the Logical and Fundamental Introductions course, along with the course description, which included learning outcomes and examples of previous exam questions created by the professor. ChatGPT was then tasked with generating new questions that closely simulated the style and structure of the professor's questions.

First Stage:

This phase consisted of two main steps:

Step One:

The artificial intelligence tool, ChatGPT, was provided with the course material and tasked with generating 100 questions, including true or false, multiple-choice, multiple-answer, and short-answer formats. During the question-generation process, several observations were made:

1- Accuracy of Questions: ChatGPT produced 100 questions covering fundamental and logical aspects of the course. Upon review, it was found that only three questions contained errors, resulting in a 3% error rate—relatively low and potentially even lower than typical human error rates. Examples of errors included:

- **True or False:** "The principles of jurisprudence are linguistically defined as prevention." (ChatGPT marked this as true).
- **Correct Answer:** The principles of jurisprudence are defined as a compound term.
  - *Jurisprudence (linguistically)*: Understanding.
  - *Origin (linguistically)*: A foundation upon which other concepts are built.
- **Multiple-Choice Question:** *Which of the following terms is defined as "the description surrounding its described that distinguishes it from others"?*

A) Jurisprudence
B) Principles of jurisprudence
C) Analogy
D) Consensus

Correct Answer: Limit, which was not included in the answer choices.

2- Repetition of Questions: ChatGPT sometimes generated duplicate questions, even if they were phrased differently.

3- Use of Unfamiliar Terminology: Some of the terms used by ChatGPT were not part of the course material. For example:

- *Which of the following is considered a type of verbal meaning?*

A) Meaning of presentation
B) Meaning of implication (correct answer)
C) Meaning of indication
D) Meaning of image

The terms meaning of presentation, meaning of indication, and meaning of image were not found in the course notes, making it easier for students to identify the correct answer by elimination.

4- Leading Questions: Some questions provided clues to the correct answer. For instance:

- *Which of the following is considered a type of induction?*

A) Measurement
B) Proof
C) Complete Induction
D) Definition

5- Overly Simplistic Questions: Some questions were too basic for students specializing in the subject. For example:

- *True or False: The subject of the principles of jurisprudence is the human body.* (False).

6- Exclusion of Certain Questions Not Included in the Assignment: ChatGPT generated some questions irrelevant to the assigned material. For example:

- **Question:** *Explain, based on the text, how variations in the methodologies of jurisprudence scholars influence their understanding of the five general principles and their applications.*
- **Answer:** Differences in methodological approaches lead to variations in identifying categories such as genus, class, type, general characteristics, and specificity. These distinctions impact how concepts are defined and classified, ultimately leading to differences in the interpretation of jurisprudential evidence.

Step Two:

In this step, the paid version of ChatGPT was utilized to regenerate the questions while adhering to specific guidelines:

- Ensuring the accuracy of answers before finalizing them.
- Avoid repetition across different question formats.
- Strictly using course-specific terminology without deviation.

The newly generated questions demonstrated improved accuracy compared to the previous set. A random sample of these questions was selected, supplemented with additional questions provided by the course instructor. The finalized set was then uploaded as an electronic assignment for students on Blackboard. After grading the test, the following percentage results were obtained:

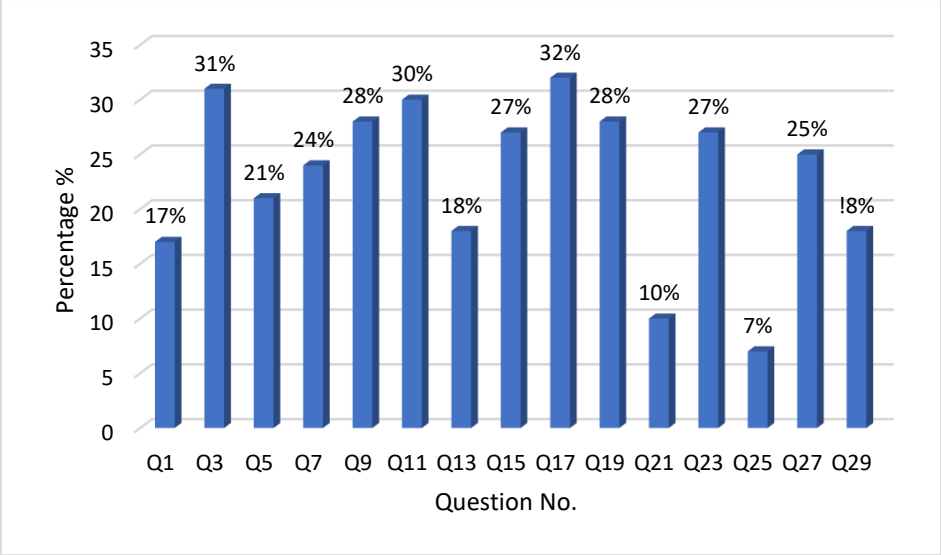

Figure 1: Percentage of correct and incorrect answers in true or false questions after implementing AI-generated and instructor-modified questions on Blackboard.

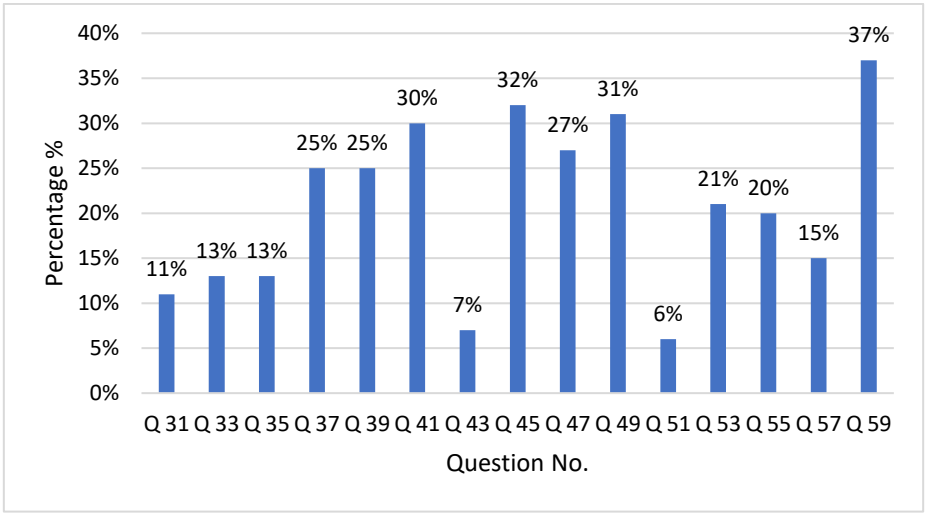

Figure 2: Percentage of correct and incorrect answers in multiple-choice questions after implementing AI-generated and instructor-modified questions on Blackboard.

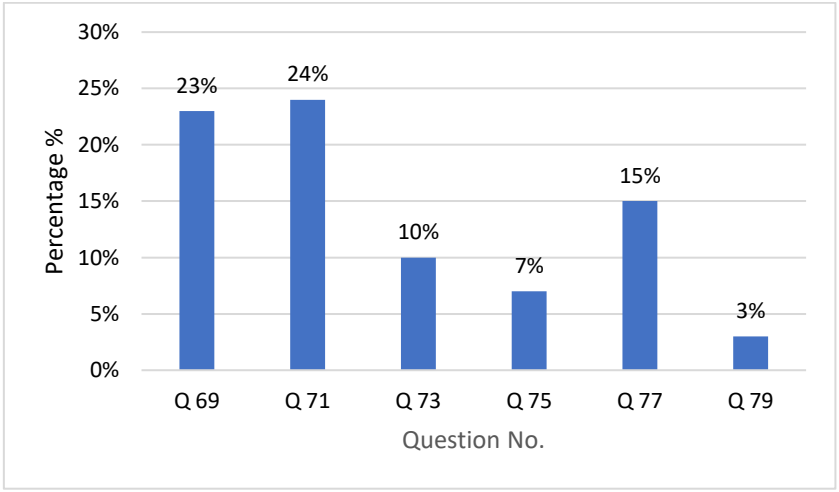

Figure 3: Percentage of correct and incorrect answers in multiple-answer questions after implementing AI-generated and instructor-modified questions on Blackboard.

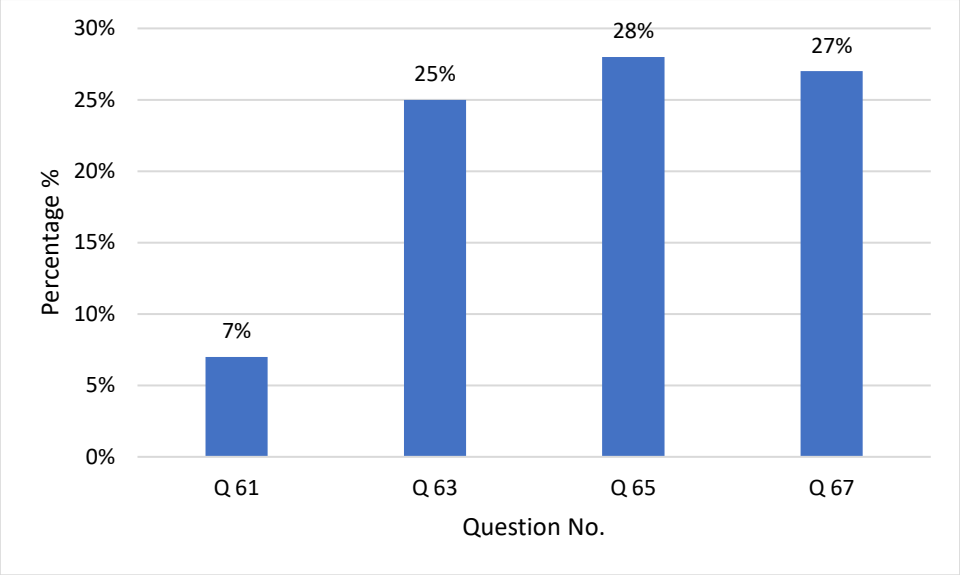

Figure 4: Percentage of correct and incorrect answers in short-answer questions after implementing AI-generated and instructor-modified questions on Blackboard.

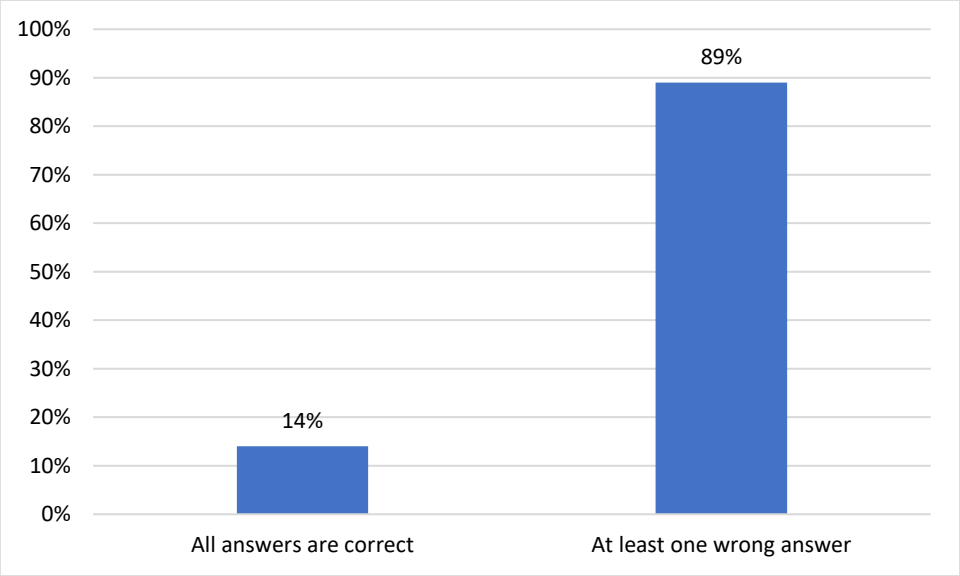

Figure 5: Percentage of students who successfully differentiated between ChatGPT-generated questions and those provided by the course instructor.

Analysis of Results from the Figures (1-5)

- **Coverage and Diversity:** ChatGPT-generated true or false questions were broader and more varied compared to those created by the course instructor.
- **Adherence to Course Material:** ChatGPT demonstrated consistency in aligning questions with the provided course content. However, the professor's expertise and diverse use of terminology resulted in more precise and in-depth questions that often connected to other subjects studied by the students.
- **Unique Aspects of Instructor-Generated Questions:** Unlike ChatGPT, the professor's questions included elements of classroom discussions and additional insights provided during lectures, which were not present in AI-generated questions.
- **Standardized AI Question Formats:** ChatGPT followed a structured approach when generating multiple-choice questions, frequently using formats such as:

- *Which of the following terms is…?*
- *What is…?*
- *Which of the following types is…?*

– **AI Versus Human-Generated Questions:** AI demonstrated an advantage in generating comprehensive question banks that covered all course topics and learning objectives without excessive repetition.

– **Student Differentiation Ability:** Students were unable to clearly distinguish between questions generated by ChatGPT and those created by the course professor.

– **Performance Comparison:** The percentage of correct answers was higher for ChatGPT-generated questions compared to those from the professor, indicating that AI-generated questions were relatively easier.

– **Efficiency for Educators:** AI can significantly reduce the time and effort required by faculty members by generating a diverse and extensive question bank. The instructor's role would then be to verify accuracy and refine the wording to better assess students' analytical and discussion skills.

Second Stage: AI-Assisted Question Generation

In this phase, an artificial intelligence tool was provided with the course material for *Logical and Fundamental Introductions*, including the course description, learning outcomes, and previous exam questions prepared by the instructor. ChatGPT was then tasked with generating new questions that closely resembled those formulated by the course professor.

To enhance the accuracy and efficiency of question generation, the following steps were implemented:

1. Input the course topics and learning outcomes (CLOs).
2. Input the course learning objectives (PLOs).
3. Input the academic program outcomes as approved by the Education and Training Evaluation Commission (KLOs).
4. Input the general knowledge units (GKU), specialized knowledge units (SKU), and the corresponding outcomes (SLOs).
5. Define the criteria for question formulation, which included:

- **Curriculum Coverage:** Ensuring that the questions comprehensively address all learning outcomes.
- **Clarity and Accuracy:** Crafting precise, unambiguous questions.
- **Simplicity:** Using clear and straightforward language.
- **Varied Difficulty Levels:** Distributing questions evenly across easy, medium, and difficult levels.
- **Bloom's Taxonomy Integration:** Use Bloom's Taxonomy verbs (knowledge, comprehension, application, analysis, synthesis, and evaluation) in the questions to ensure diversity in thinking levels.
- **Answer Quality:** Providing detailed model answers, linking them to learning outcomes, specifying question types (memorization, application, analysis, critical thinking), and indicating difficulty levels.

Each unit of the course was sequentially fed into ChatGPT, prompting the tool to generate evenly distributed questions across topics while maintaining the pre-defined difficulty levels (easy, medium, difficult) and Bloom's Taxonomy criteria.

For each question, ChatGPT was required to:

- Provide the correct answer with an explanation.
- Link the question to relevant learning outcomes (SLO-KLO-PLO-CLO).
- Indicate the difficulty level.
- Align it with Bloom's Taxonomy.

Once the AI completed generating questions for the first unit, the same process was repeated for the subsequent units until all course content was covered. Upon evaluation, ChatGPT demonstrated an impressive ability to adapt and refine its outputs, producing highly accurate and well-structured questions. The generated questions effectively covered all course material, adhered to Bloom's Taxonomy, and maintained an appropriate balance of difficulty. Furthermore, the answers included specific references to page and paragraph numbers from the course document uploaded to the application, ensuring precision and alignment with the curriculum.

5. Conclusions and Recommendations

5.1 Conclusion

AI has significantly reshaped various aspects of daily life, influencing fields such as healthcare, finance, gaming, and transportation. Its integration is no longer a luxury but a necessity, driven by the fast-paced demands of modern life, where educators constantly strive to meet their responsibilities within limited timeframes.

While some argue that using AI in education—particularly for generating questions—may negatively impact learning outcomes by enabling students to generate both questions and answers effortlessly, AI reliance cannot be entirely restricted. Instead, students should be guided on how to leverage AI effectively to enhance their creative thinking and develop essential skills that will prepare them for future competitiveness in the job market.

Findings

The study yielded several findings:

- Exclusive reliance on AI for question generation in Islamic courses without human oversight may lead to unfavorable outcomes. Given the unique nature of Islamic studies and their deep connection to human reasoning and interpretation, human involvement remains crucial.
- AI-generated questions in Sharia sciences require ongoing refinement and direct human supervision. Continuous training and correction of AI-generated content can enhance accuracy. AI can also be utilized to analyze previous exam questions and academic discussions, allowing for the development of best practices to optimize its application.
- AI, in its current state, struggles to generate questions that require critical thinking, analysis, and inference in Islamic studies. Such tasks necessitate human expertise, making them beyond the full capabilities of AI at this stage.

5.2 Recommendations

- Specialists in Sharia studies must enhance their understanding of modern artificial intelligence technologies and effectively integrate them into curriculum development. Utilizing AI optimally can contribute to achieving learning objectives and improving educational outcomes in Sharia studies.

Acknowledgment

"The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Group Project under grant number. (RGP.2/ 309/ 45)

References

1. Abdel Hamid, Ahmed Mukhtar. (2008). *Dictionary of Contemporary Arabic*. 1st ed., Alam Al-Kutub, 1429 AH / 2008 AD, p. 818.
2. Abhishake Reddy Onteddu. (2021). AI and Deep Learning-Based Intelligent Drug Recommendation System for Patient Health Monitoring in IoT-Enabled Healthcare. Journal of Informatics Education and Research, 1(3).
3. Abdel Nour, Adel. (2005). *Introduction to Artificial Intelligence*. 1st ed., King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, 2005, p. 7.
4. Al-Barakti, Muhammad Ameem Al-Ihsan Al-Mujaddidi. *Jurisprudential Definitions*. 1st ed., Dar Al-Kutub Al-Ilmiyyah (Reprint of the old edition in Pakistan, 1407 AH / 1986 AD), 1424 AH / 2003 AD.
5. Al-Harawi, Muhammad ibn Ahmad ibn al-Azhari (Abu Mansur). Tahdhib al-Lugha. Edited by Muhammad Awad Mara'b, 1st ed., Dar Ihya al-Turath al-Arabi, Beirut, 2001, p. (10/184).
6. Al-Herawi, Muhammad bin Ahmad bin Al-Azhari (Abu Mansour). *Al-Zahir fi Gharib Alfadh al-Shafi'i (Al-Zahir in the Strange Words of Al-Shafi'i)*. Edited by Masoud Abdul Hamid Al-Saadani, Dar Al-Tala'i', p. 263.
7. Al-Kafwi, Ayoub bin Musa Al-Hussaini Al-Quraimi (Abu Al-Baqa Al-Hanafi) (d. 1094 AH). *Al-Kulliyat: A Dictionary of Linguistic Terms and Differences*. Edited by Adnan Darwish and Muhammad Al-Masry, Al-Risala Foundation, Beirut, p. 456.

8. Al-Qazwini Al-Razi, Ahmad bin Faris bin Zakariya (Abu Al-Hussein). *Mu'jam Maqayis al-Lugha (Dictionary of Language Measures)*. Edited by Abdul Salam Muhammad Har, Dar Al-Fikr, 1399 AH / 1979 AD, p. (2/ 357)
9. Bonnet, Alain. (1993). Artificial Intelligence: Its Present and Future. Translated by Ali Sabry Farghali, National Council for Culture, Arts, and Literature, World of Knowledge Series, Kuwait, 1993, p. 19-21.
10. Briber, Abdel Nour. (2024). *A Purposeful Vision of the Dangers and Prospects of "Artificial Intelligence" and Its Applications*. International Conference: Artificial Intelligence and Its Applications in Islamic Sciences, November 29-30, 2023. Publications of the Laboratory of Jurisprudential and Judicial Studies, University of El Oued, Algeria, 1st ed., 1445 AH / 2024 AD, pp. 57-74.
11. Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., ... & Co, M. T. H. (2023). ChatGPT versus human in generating medical graduate exam multiple choice questions—A multinational prospective study (Hong Kong SAR, Singapore, Ireland, and the United Kingdom). *PloS one*, *18*(8), e0290691.
12. French, R. M. (2000). The Turing Test: the first 50 years. *Trends in cognitive sciences*, *4*(3), 115-122.
13. Laupichler, M. C., Rother, J. F., Kadow, I. C. G., Ahmadi, S., & Raupach, T. (2024). Large language models in medical education: comparing ChatGPT-to human-generated exam questions. *Academic Medicine*, *99*(5), 508-512.
14. Maskawayh, Abu Ali Ahmad ibn Muhammad ibn Yaqub. Tahdhib al-Akhlaq wa Tathir al-A'raq. Edited by Ibn al-Khatib, Maktabat al-Thaqafa al-Diniyah, 1st ed, p. 7.
15. Watanabe, Y. (2021). I, Inventor: Patent Inventorship for Artificial Intelligence Systems. *Idaho L. Rev.*, *57*, 473.