Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

USING AI FOR ISLAMIC JURISPRUDENCE QUESTIONS: COMPARATIVE ANALYSIS OF DOMAIN-SPECIFIC RETRIEVAL-AUGMENTED GENERATION (RAG) VS. GENERIC LLM

Khurram Shahzad, Adnan Ashraf, Muhammad Salama

International Open University, The Gambia

ABSTRACT

The rapid growth of large language models such as ChatGPT, Claude and Llama have unlocked a range of opportunities in the text generation. However, these models are trained on vast unfiltered data from Internet which is useful in many generic applications and can cause fabricated or distorted information on a complex query in a specific domain. In a highly regulated and sensitive domain of Islamic Jurisprudence this deficiency can have serious consequences. To address this issue, this research explored the application of AI on Islamic jurisprudence text using domain-specific AI capabilities which surpasses the capabilities general-purpose Large Language Models (LLMs) like ChatGPT. A POC experiment has been conducted to show how a domain specific AI can outperform generic Large Language Models (LLMs) such as ChatGPT. The results of the experiment demonstrated that a retrieval-augmented generation (RAG) system, fine-tuned on a curated corpus of Islamic jurisprudence, consistently outperformed and provided accurate answers compared to generic ChatGPT. The superior performance is due to data access from verified texts and its ability to retrieve contextual passages for indexing and querying; thereby reducing hallucinations. In addition, future research should also focus on developing and training AI models for various schools of Islamic Jurisprudence which can enable a collaborative "multi-school AI council" that can address the challenges of the contemporary and future jurisprudential issues. Educational *institutions must embed AI literacy into their curricula so that emerging scholars understand both the capabilities and the limitations of these systems and can anticipate the societal impacts of their deployment.*

Keywords: *Artificial Intelligence, AI, Natural Language Processing, NLP, Large Language Model, LLM, Religious Text, Islamic Jurisprudence, Fiqh, FatawaRAG.*

Corresponding author: Khurram Shahzad can be contacted at khurram063@gmail.com

1. INTRODUCTION

Artificial intelligence has demonstrated impressive capabilities that work extremely well in generic problem solving and as well as specialized problem solving. The rapid advancement of large language models (LLMs) has marked a new era in artificial intelligence, including GPT-4, Claude, and Llama which define the current frontier. These models exhibit remarkable capabilities in text generation, translation, and general question-answering.

However, generalized LLMs are trained on vast, often unfiltered, datasets from the internet, they can produce inaccurate or misleading information, a phenomenon known as "hallucination". The advancements in AI have not only improved the efficiency of various tasks but have also raised important questions about the reliability and accuracy of the information generated by these models. The issue of hallucination, where the model generates information that is not based on correct data and presents it in a factual way to the users, is a significant challenge. This can lead to serious consequences in critical applications such as healthcare, legal, and financial sectors. Moreover, the process of updating the knowledge base of these models is not an easy task. It requires extensive retraining, which is both time-consuming and costly.

Another critical issue is the lack of transparency in the reasoning processes of these models. Users often find it challenging to understand how the model arrived at a particular conclusion or answer.

In addition, the main problem occurs when a generic problem-solving AI is used to solve a specialized problem or question; it can sometimes provide correct answers, however, at other times it gives hallucinated responses. The inaccuracy of responses is increased if they require legal analysis. There are two distinct limitations in responses that require legal analysis. 1) attendance to hallucinate and give vague responses using mixed or similar legal rulings or 2) a very surface level answer without going into in-depth analysis and ignoring context, culture and nuisances.

To address these issues, multiple techniques can be used in LLMs to enhance their domain specific knowledge. LLMs can be enhanced and optimized using two different ways:

The first technique that can address this challenge is Retrieval-Augmented Generation (RAG). In this approach, the domain specific knowledge base is indexed and queried in real time. When the user queries, the system retrieves the documents that are related to the context and creates a new prompt that goes into the standard LLM And provides the response based on the documents that are queried in the initial search. And because the knowledge base is Independent and decoupled from the model's internal parameters, it can be updated continuously without retraining the entire model (Izacard & Grave, 2021; Guu et al., 2020; Lewis et al., 2020; Zhang et al., 2021).

The second and slightly more complicated technique is Fine-tuning and existing LLM which is more complicated process than RAG. In this process, the weights of a pre-trained model are updated based on the domain specific knowledge base. While fine-tuning can embed specialized knowledge directly into the model, it can still suffer from the static-knowledge issue and can introduce hallucinations in the responses (Parthasarathy et al., 2024; Liu et al., 2024; Kumar & Shaurya, 2024).

Each of these two approaches have its pros and cons. In this POC we will implement RAG to show that an existing LLM can be specialized in Islamic religious jurisprudence, and that it can answer the questions better than a generic LLM model.

2. REVIEW OF LITERATURE

The rapid advancement of technology is significantly impacting various aspects of human life. The upcoming integration of Artificial Intelligence (AI) and Natural Language Processing (NLP) has the capacity to transform problem-solving methods and solutions. AI, a specialized field within computer science, aims to give machines human-like intelligence ("Artificial Intelligence", 2023). Similarly, NLP, another area of computer science, involves analyzing textual data using specific algorithms. This field uses various computational techniques

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

which can potentially enable computers to understand, interpret, and analyze language like human linguistic comprehension (Liddy, 2001).

AI introduces a novel perspective for tackling issues that require cognitive abilities. AI algorithms empower researchers to cultivate and uncover solutions with completely new perspectives. Notably, an Algorithm constitutes a defined sequence of steps designed to address mathematical quandaries utilizing computational devices ("Algorithm", 2023).

2.1 Natural Language Processing (NLP)

Natural Language Processing, a subfield of AI, encompasses language processing through various stages. It typically operates on three distinct levels to accomplish its objectives: 1) lexical analysis, 2) syntax analysis, and 3) semantic analysis. Numerous technologies are available that can execute one or more of these stages within the domain of natural language processing (Graham, 2019).

Lexical analysis means parsing the sentence into tokens and correcting any token if it is misspelled, while syntax analysis tries to understand the sentence's grammatical structure Graham (2019). For example, if we write "Can Muslim man wear shorts for prayers?", most programs such as Google and Microsoft Word would parse the sentence into tokens, understand its grammatical structure, and find out if any of the words are misspelled.

Google can find that "prayers" is not correctly spelled, and there is a grammatical mistake as well; it will immediately show that the sentence should be written as "Can Muslim men wear shorts for prayers." The same applies to Microsoft Word as well; it can detect spelling and grammatical mistakes, but Google performs an additional step by suggesting sentences with better synonyms stored in its text repository. It might suggest the new sentence as "Can Muslim men wear shorts for Salah?".

Semantic analysis in NLP is even more complicated because it tries to understand the meaning of the sentence and allows the program to recreate the sentence better. Grammarly is a tool that does this analysis and restructures the sentence in a more formal and better way. Grammarly might suggest rewriting the sentence as "Is it allowed for Muslim men to pray in shorts?" Word processing software like Microsoft Word and Google usually do not perform semantic analysis.

Most of the studies performed on Christian religious texts using AI technology found patterns and plotted data for analysis. No specialized AI work has been performed to create intelligence out of the text or exploring AI cognitive abilities on text interpretation. This research aims to understand existing literature using AI and NLP techniques, specifically on religious texts, and derive reasoning from it.

Question-answering (QA) systems have been explored in the technological world for the past few decades. IBM has developed IBM Watson, a system that uses intelligent NLP and AI to answer any question related to the data fed to it. The

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

system was trained on the data and produced models that could help find the answers (Ferrucci et al., 2010; Ferrucci, 2012).

2.2 Supervised and Unsupervised Machine Learning

2.2.1 Supervised Learning

Supervised learning is a method used in machine learning where the computer is given examples that are already "answered". Think of it like a student learning from a teacher who gives both the questions (inputs) and the correct answers (outputs). The aim is for the computer to understand how to connect questions with their answers so that when it sees new, similar questions, it can predict the answers correctly (Delua, 2023).

Let's simplify this with an everyday example: imagine you're learning to identify different types of fruits. Someone shows you several pictures of apples and tells you they are apples. You're also shown pictures of bananas with the label "banana". Over time, you learn to distinguish apples from bananas based on features like color, shape, and size. The next time you see a fruit you've been taught about, you can identify it correctly. Supervised learning works in a similar way but with data. It's like teaching the computer to recognize patterns by giving it lots of labeled examples (Delua, 2023).

Now, why is this useful? For one, it makes training artificial intelligence (AI) more straightforward. Since every piece of data (like a picture of a fruit) comes with a label (like "apple"), the AI doesn't have to guess what it's looking at. This clear guidance helps the AI learn faster and make fewer mistakes when it starts making predictions on its own.

Moreover, supervised learning can be applied to a vast range of tasks, from filtering spam emails by learning what makes an email spam or not, to predicting house prices based on location, size, and other features. Each time, the process is similar: feeds the system lots of examples, and it learns to make predictions.

The power of supervised learning lies in its versatility and efficiency. It is used in healthcare to predict patient outcomes based on their symptoms and medical histories; in finance to assess credit risk, and in everyday apps such as email and social media to filter out undesirable content. As technology advances and we gather more data, supervised learning becomes even more potent, aiding in the resolution of intricate problems by learning from examples, much like humans do.

2.2.2 Unsupervised Learning

Unsupervised learning, a pivotal branch of machine learning, leverages algorithms to sift through and scrutinize data autonomously. This approach diverges from supervised learning by eschewing the need for pre-labeled output data. Instead, it empowers the algorithm to independently discover patterns and deduce insights, making it a self-reliant and insightful method of data analysis. This form of learning is not without its challenges; it demands extensive datasets to derive meaningful and applicable outcomes effectively. As highlighted by Delua (2023), unsupervised learning's essence lies in its ability to navigate through data without predefined guidance, thereby revealing underlying structures and associations that might not be immediately apparent.

Expanding upon its foundational objectives, unsupervised learning endeavors to reveal the intricate and often hidden patterns within data sets without any form of external input or supervision. Its primary goals include clustering, where the algorithm groups similar data points together, and association, a process of identifying relationships between variables within the data. By doing so, it offers a nuanced understanding of data dynamics, facilitating the discovery of natural classifications, novel patterns, and complex correlations that would remain undiscovered through manual analysis or supervised techniques.

Moreover, unsupervised learning plays a crucial role in dimensionality reduction – simplifying datasets by reducing the number of variables under consideration yet preserving the essential characteristics of the data. This aspect is particularly valuable in handling high-dimensional data, enhancing both the efficiency and clarity of data analysis processes.

Another notable application is anomaly detection. Unsupervised learning algorithms can efficiently identify outliers or anomalous data points that deviate significantly from the norm. This capability is invaluable in various domains, such as fraud detection in finance, network security, and monitoring systems' health, where recognizing unusual patterns quickly can prevent potential issues or uncover critical insights.

As unsupervised learning continues to evolve, its potential applications and benefits across diverse fields become increasingly apparent. From enhancing customer segmentation in marketing to advancing genetic research by uncovering hidden genetic patterns, the implications of unsupervised learning are vast and far-reaching. However, the sophistication of unsupervised learning algorithms also necessitates a careful and considered approach to their deployment, ensuring that the insights gleaned are both accurate and actionable (Delua, 2023).

2.2.3 Semi-supervised Learning

Semi-supervised learning is a machine learning approach that lies between supervised and unsupervised learning. In supervised learning, the algorithm learns from a labeled dataset, while unsupervised learning makes sense of data without any labels. Semi-supervised learning leverages a small amount of labeled data along with a larger pool of unlabeled data to train models (Hady & Schwenker, 2013). This method is particularly useful when acquiring a comprehensively labeled dataset is expensive or impractical, a common scenario in many real-world applications.

The core idea behind semi-supervised learning is that, even though the unlabeled data does not have explicit outputs, it can still provide a significant amount of information about the structure of the problem space. This can be especially valuable for tasks like classification, where the algorithm can learn the boundaries between different classes not just from the labeled examples but also by the distribution and relation of the unlabeled data points. Semi-supervised learning techniques often involve strategies like self-training, where the model labels the unlabeled data itself and then retrains on this newly labeled set, and co-training, where multiple models are trained on separate views of the data and then collaborate to label the unlabeled data. These techniques have shown promising results in fields such as image recognition, natural language processing, and bioinformatics, where labeled data can be scarce but unlabeled data is abundant (Hady & Schwenker, 2013).

2.2.4 Large Language Models (LLMs)

A Large Language Model (LLM) is a type of artificial intelligence (AI) that specializes in processing and generating human language. LLMs are designed to understand, interpret, and produce text, allowing them to perform a variety of language-related tasks, such as answering questions, generating content, and translating languages (Bengio, Goodfellow, & Courville, 2016; Radford et al., 2018; Tamkin et al., 2021).

LLMs are considered “large” due to their use of vast amounts of data for training. They are built using advanced machine learning techniques and trained on large corpora of text data—such as books, articles, and websites. This extensive exposure allows them to learn the structures, patterns, and relationships between words, phrases, and concepts in human language. Through training, LLMs become capable of generating human-like text and responding to user inputs in a coherent and contextually appropriate manner (Brown et al., 2020).

The training process of an LLM involves exposing the model to vast amounts of written text. During this phase, the model learns to predict the likelihood of a word or sequence of words occurring based on the context provided by previous words. Over time, the model builds a statistical understanding of language, which enables it to generate text that mirrors human speech or writing (Vaswani et al., 2017).

For instance, when asked a question like, “What is the weather like in New York today?” The LLM does not have direct access to current weather data. Instead, it uses its training to generate a plausible answer based on patterns it has learned from text discussing similar topics (e.g., weather, geography, and common phrases associated with asking about the weather).

LLMs are versatile tools that can perform a range of tasks, including:

- Answering Questions: LLMs can provide responses to factual or explanatory queries based on their training data.
- Text Generation: They can generate essays, stories, articles, and other forms of written content.
- Language Translation: LLMs are capable of translating text from one language to another by understanding and replicating linguistic patterns.
- Conversation Simulation: LLMs can engage in real-time dialogues, offering responses that simulate human conversation.
- Educational Assistance: They can explain complex concepts, provide study help, and assist with homework in various subjects.

LLMs function by analyzing the input text they receive and predicting what comes next based on the patterns they learned during training. For example, given the phrase “Once upon a time,” an LLM might generate the continuation “there was a princess,” as this is a common narrative structure in many texts. The model generates these responses by computing probabilities for each possible word or phrase, selecting the most likely continuation based on context (Vaswani et al., 2017).

It is important to note that LLMs do not possess true comprehension of language in the way humans do. Instead, they rely on mathematical models that predict the next most probable word or sequence, drawing from vast amounts of data.

Large Language Models (LLMs) are advanced AI systems that can generate and comprehend text by analyzing patterns in vast amounts of data. While they offer significant benefits across various fields, their limitations, including the lack of true understanding and the potential for bias, highlight the need for careful use and further research. LLMs represent a powerful tool in the ongoing intersection of technology and human language, capable of transforming many industries and daily activities (Brown et al., 2020).

3. RESEARCH METHODOLOGY - EXPERIMENTAL DESIGN SCIENCE RESEARCH STUDY

"Design Science Research (DSR) is a problem-solving paradigm that seeks to enhance human knowledge via the creation of innovative artifacts. DSR seeks to enhance technology and science knowledge bases via the creation of innovative artifacts that solve problems and improve the environment in which they are instantiated" (Brocke, Hevner, & Maedche, 2020, p. 1). DSR first, utilizes gained knowledge to solve problems, create change or improve existing solutions; and secondly generates new knowledge, insights and theoretical explanations" (Baskerville et al., 2015; Horváth, 2007; Pello, 2018, p. 1).

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

Design Science Research includes six steps or activities including "(1) identification of the problem, defining the research problem and justifying the value of a solution; (2) definition of objectives for a solution; (3) design and development of artifacts (constructs, models, methods, etc.); (4) demonstration by using the artifact to solve the problem; (5) evaluation of the solution, comparing the objectives and the actual observed results from the use of the artifact; and (6) communication of the problem, the artifact, its utility and effectiveness to other researches and practicing professionals (Lapão et al., 2017; Peffers et al., 2007; Pello, 2018; Teixeira et. al., 2017).

3.1 Step 1: Identification of the problem

Using the DSR methodology, an experimental research study will be conducted. First, the research problem is whether AI can be used on Islamic jurisprudence (fiqh) verdicts repository to derive accurate fiqh solutions to posed questions related to Islamic finance and family law.

3.2 Step 2: Identification of the problem and Objectives for Solution

The main objective of this experiment is to prove that the domain specific AI can answer more accurately and intelligently than the generic LLMs. When a specific model is trained based on a domain specific knowledge, it adds value; provides answers accurately; reduces biases and performs better than the generic LLM models.

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

3.3 Step 3: Design and Development of artifacts (constructs, models, methods)

An Islamic finance and family law knowledge base from the Canadian Islamic Fatawa Center repository, was used to train open-source Large Language Models. Accordingly, with the latest AI training methods, we can train these open-source models by feeding Islamic jurisprudence data to create domain specific intelligence, which would answer fiqh questions in this research.

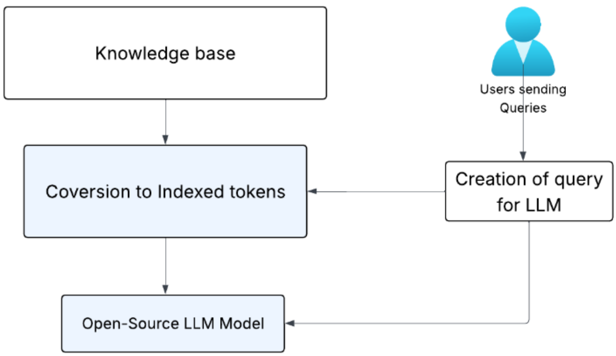


Figure 1. Logical Architecture Overview for Proof-of-Concept experiment

3.4 Step 4: Models Demonstration to Solve the Problem

For this research, domain specific AI solution was built to allow the users to answer Islamic Jurisprudence questions. For example, “what to do if wudu breaks during salah” would be asked, and AI algorithms would use Natural Language Processing to search the textual data and find “experimental” solutions to posed questions. As already stated, general LLMs such as ChatGPT hallucinate or do not answer specific questions accurately. Therefore, a domain specific proof of concept machine is developed for this study that would answer specific fiqh questions because such fiqh knowledge base has been used to train this AI model.

3.5 Step 5: Evaluation of the Solution

AI driven experimental answers are compared to actual answers found in original fiqh books. This allowed the researcher to ensure algorithms created correct AI in answering questions with accuracy and completeness. Data was tested and retested to reach proper solutions to pose questions.

3.6 Step 6: Communication of the Solution with other Professionals

Once the experiments run successfully and solutions are driven from AI created models, the results section was written in detail to share those findings. AI strengths and weaknesses were highlighted and discussed for further improvement in future studies.

4. FatawaRAG: A PROOF-OF-CONCEPT FOR CANADIAN ISLAMIC JURISPRUDENCE

This chapter introduces a proof-of-concept (PoC) system, FatawaRAG, designed to demonstrate that a Retrieval-Augmented Generation (RAG) architecture can overcome the above-mentioned limitations. The primary aim of FatawaRAG is to demonstrate that a domain-specific RAG system can outperform state-of-the-art generic LLMs in terms of accuracy, and transparency. The chosen domain for this PoC is the

Canadian Fatawa Center, a field where precision, verifiability, and access to up-to-date information are not merely desirable, but essential.

The central theme is that by decoupling the knowledge base from the reasoning engine, FatawaRAG can provide more accurate, trustworthy, and auditable responses to legal queries than a general-purpose LLM. While a frontier model draws from a vast but static and opaque internal knowledge store, FatawaRAG will reason over a curated, current, and explicitly cited corpus of Canadian Islamic Jurisprudence queries.

This chapter also details the system's architecture, the experimental design used to test its performance against leading frontier models, and an analysis of the results. It presents a POC application designed for a very narrowly scoped field of Islamic jurisprudence using the data from a Fatawa center.

4.1 The Challenge of Islamic Jurisprudence AI and Limitations of Frontier Models

The Islamic legal domain presents a unique and formidable challenge for artificial intelligence. Legal reasoning is predicated on a structured, evolving body of knowledge comprising statutes, regulations, and case law. The correctness of a legal conclusion is directly tied to the accurate application of these specific source texts.

When applied to legal tasks, general-purpose frontier models exhibit several critical weaknesses:

4.1.1 *Jurisdictional Confusion*

A model trained on a global dataset may struggle to distinguish between the laws of different schools of thought. It might incorrectly apply principles from multiple rulings which can cause a subtle but critical error.

4.1.2 *Hallucination of Legal responses*

The tendency of LLMs to hallucinate is particularly high when the query is technical, and there are multiple options in various schools of thought. A model might generate a flawlessly written Fatawa argument that rests upon a fabricated case citation. This is not just incorrect; it is deeply misleading and undermines the very foundation of Islamic Jurisprudence practice.

4.1.3 *Lack of Transparency and Auditability*

When a frontier model provides a legal summary or conclusion, it is impossible to trace the precise source of its "knowledge". For a Jurist, lawyer, judge, or student, an answer without a verifiable citation is professionally useless.

These limitations make the direct application of general-purpose LLMs in serious legal contexts highly unreliable.

5. SYSTEM ARCHITECTURE: THE FatawaRAG PROOF OF CONCEPT

To address the challenges highlighted above, we designed and implemented FatawaRAG. A RAG system tailored for Islamic Jurisprudence questions asked by people living in Canada. The architecture consists of three core components: a curated legal knowledge base, a hybrid retrieval system, and a generation module guided by a specialized prompting strategy. Below is the Architecture diagram.

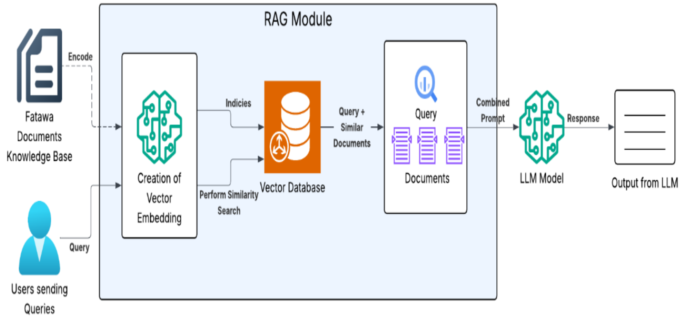


Figure 2. Architecture diagram of the FatawaRAG System

5.1 The Knowledge Base

The foundation of FatawaRAG is a high-quality, curated corpus of legal documents. For this POC, the knowledge base was restricted to a well-defined and manageable subset of Canadian Fatawa center located at https://fatawacenter.com, comprising:

- These documents answer questions related to Canadian context
- Most of the questions answered were based on financial matters
- It is not a complete repository of all the Islamic rulings

These documents were parsed, cleaned, and segmented into logical chunks to prepare them for retrieval. Each chunk retains metadata linking it to its precise source. The source of the information is https://fatawacenter.com/

5.2 Retrieval-Augmented Generation Module

For this POC, we used Llama-3-8B-Instruct, a capable yet efficient model. The prompt structure is critical:

Role: You are a legal research assistant specializing in Islamic Jurisprudence law. Your task is to answer the user's query based *exclusively* on the provided Islamic legal texts. Do not use any external knowledge.

This structured instruction forces the generator to act as a synthesizer of the provided information, rather than as a knowledge source itself.

6. EXPERIMENTAL DESIGN

To evaluate the performance of FatawaRAG against frontier models, we developed a benchmark of three legal questions. These questions were designed by a legal expert to test a range of reasoning skills and reflect realistic queries. These three questions are from these three categories:

a. **Islamic Finance (1 question)**: Question requiring direct interpretation of the Islamic finance (e.g., "Shariah compliant way to purchase a house on mortgage").
b. **Ritual purity (Cleanliness) (1 question):** Question about seeking the ritual purity for observing the prayers (e.g., "What shariah says, if someone gets late and he still needs to do ablution to perform his prayers, but the time of the Salah is very short").

c. **Family Law (1 question):** Complex question requiring the synthesis of information from an Islamic family law (e.g., Is the divorce valid when the husband was in anger).

These three questions were asked to ChatGPT on 3rd July 2025 and FatawaRAG system, and the responses were evaluated for accuracy and completeness.

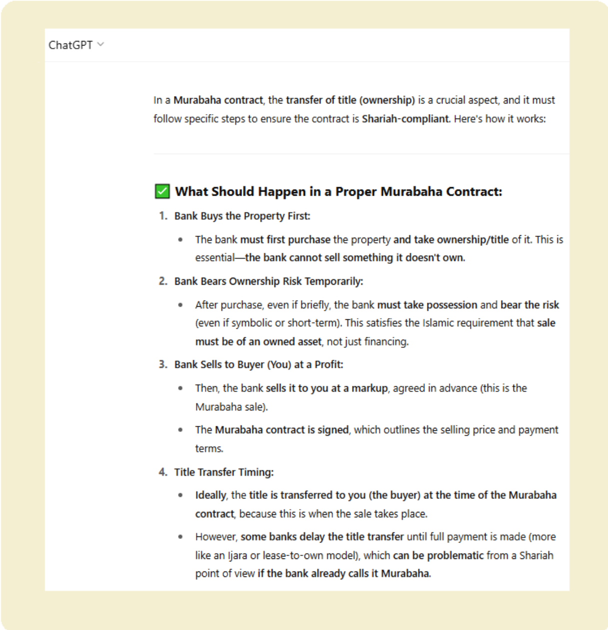
6.1 Question 1

Here is the first prompt that was asked to ChatGPT.

“In a Murabaha contract, does the title have to be transferred to the buyer after the signing of the contract?”

The responses from ChatGPT and FatawaRAG are shown in the images below. Also, in the images, the sections which are incorrect or have concerns are highlighted in red ink.

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

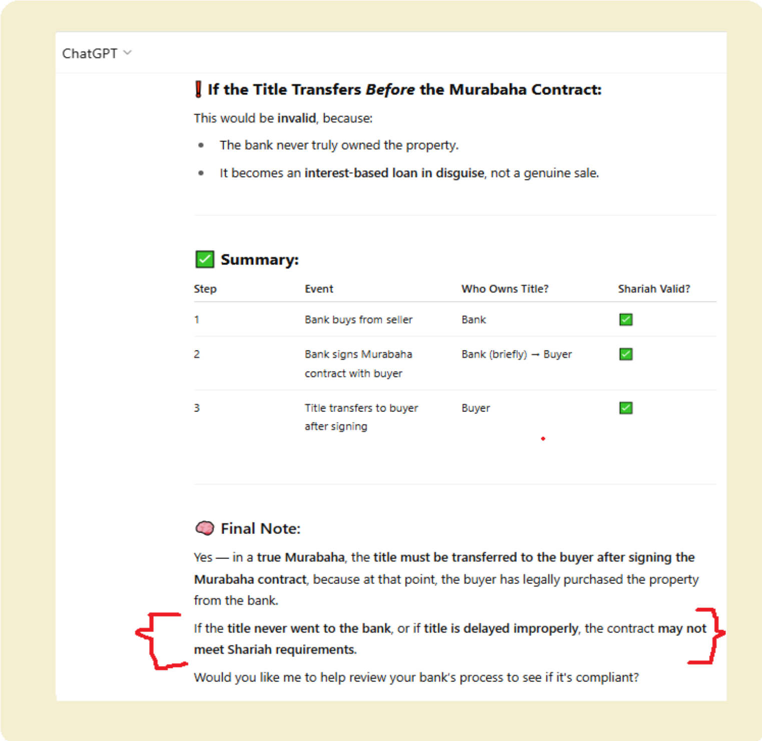

ChatGPT

In a **Murabaha contract**, the **transfer of title (ownership)** is a crucial aspect, and it must follow specific steps to ensure the contract is **Shariah-compliant.** Here's how it works:

✅ **What Should Happen in a Proper Murabaha Contract:**

1. **Bank Buys the Property First:**
   - The bank **must first purchase** the property **and take ownership/title** of it. This is essential—**the bank cannot sell something it doesn't own.**
2. **Bank Bears Ownership Risk Temporarily:**
   - After purchase, even if briefly, the bank **must take possession** and **bear the risk** (even if symbolic or short-term). This satisfies the Islamic requirement that **sale must be of an owned asset**, not just financing.
3. **Bank Sells to Buyer (You) at a Profit:**
   - Then, the bank **sells it to you at a markup**, agreed in advance (this is the Murabaha sale).
   - The **Murabaha contract is signed**, which outlines the selling price and payment terms.
4. **Title Transfer Timing:**
   - **Ideally**, the **title is transferred to you (the buyer) at the time of the Murabaha contract**, because this is when the sale takes place.
   - However, **some banks delay the title transfer** until full payment is made (more like an Ijara or lease-to-own model), which **can be problematic** from a Shariah point of view **if the bank already calls it Murabaha.**

Figure 3. Question 1 - ChatGPT response – 1


ChatGPT

❗ **If the Title Transfers *Before* the Murabaha Contract:**

This would be **invalid**, because:

- The bank never truly owned the property.
- It becomes an **interest-based loan in disguise**, not a genuine sale.

✅ **Summary:**

Step	Event	Who Owns Title?	Shariah Valid?
1	Bank buys from seller	Bank	✅
2	Bank signs Murabaha contract with buyer	Bank (briefly) → Buyer	✅
3	Title transfers to buyer after signing	Buyer	✅

Final Note:

Yes — in a **true Murabaha**, the **title must be transferred to the buyer after signing the Murabaha contract**, because at that point, the buyer has legally purchased the property from the bank.

If the **title never went to the bank**, or if **title is delayed improperly**, the contract **may not meet Shariah requirements.**

Would you like me to help review your bank's process to see if it's compliant?

Figure 4. Question 1 - ChatGPT response section 2

In the image above the marked sections shows that ChatGPT brings this information as a final note which is incorrect according to the Islamic financial jurisprudence. In Islamic financial jurisprudence, the contract would meet Shariah requirements because according to the AAIOFI standards the title-transfer can be delayed. The main thing is the contract if it meets the Shariah compliance and title is the result of the contract. Many jurisdictions the bank don’t want to put the name on the title to avoid double taxations which is completely allowed by AAIOFI standards (Accounting and Auditing Organization for Islamic Financial Institutions (AAOIFI) 2025).

To further clarify the above question, this prompt was used: "what are the aaiofi standards say regarding this". And below are the images that show how ChatGPT responds and gives a mixed information response.

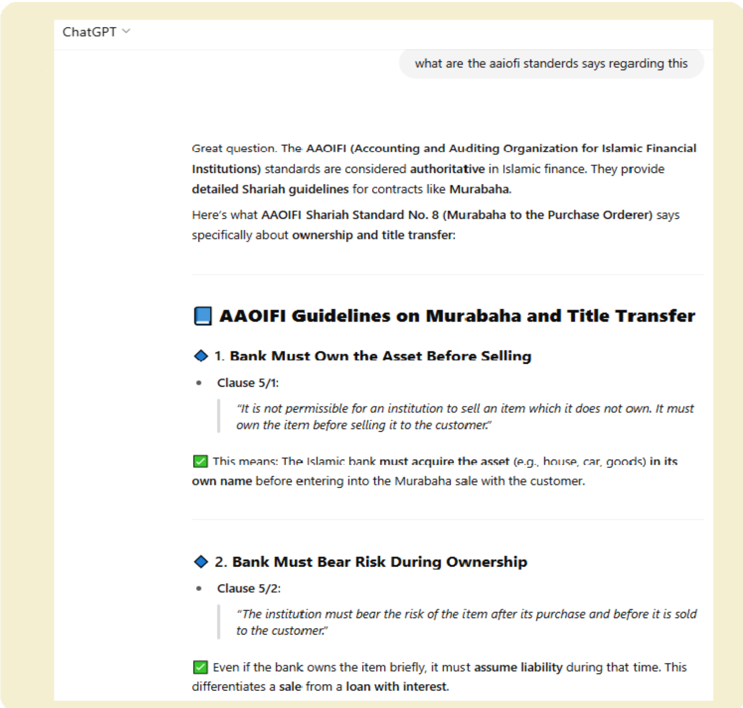
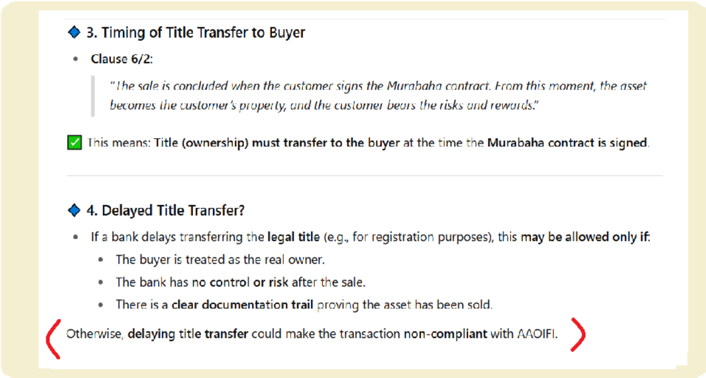


Figure 5. Question 1 - ChatGPT response section 3



🔷 3. Timing of Title Transfer to Buyer

- **Clause 6/2:**

> *"The sale is concluded when the customer signs the Murabaha contract. From this moment, the asset becomes the customer's property, and the customer bears the risks and rewards."*

✅ This means: Title **(ownership) must transfer to the buyer** at the time the **Murabaha contract is signed**.

🔷 4. Delayed Title Transfer?

- If a bank delays transferring the **legal title** (e.g., for registration purposes), this **may be allowed only if:**
  - The buyer is treated as the real owner.
  - The bank has no **control or risk** after the sale.
  - There is a **clear documentation trail** proving the asset has been sold.

Otherwise, **delaying title transfer** could make the transaction **non-compliant** with AAOIFI.

Figure 6. Question 1 - ChatGPT response section 4

The highlighted section in the figure above is incorrect answer according to the AAIOFI standard. The AAIOFI Shariah compliance guidelines clearly allow the bank to delay the transfer of title (Accounting and Auditing Organization for Islamic Financial Institutions (AAOIFI) 2025).

ChatGPT was asked another question, but it provided a very mixed response. Here is the prompt that is asked "Can the bank hold on the title until the last payment of the property"

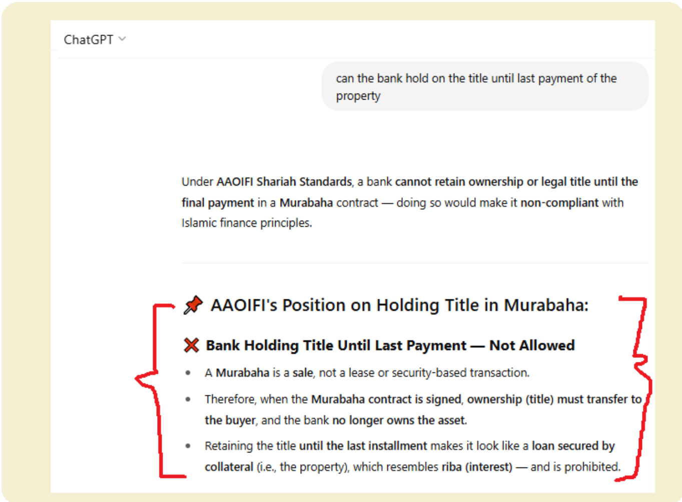


Figure 7. Question 1 - ChatGPT response section 5

As highlighted in the image above, ChatGPT refers to AAOIFI position which is completely incorrect according to their standards in the figure above. Here ChatGPT says that "Retaining the title until the last installment makes it look like a loan secured by collateral (i.e., the property), which resembles *Riba* (interest) — and is prohibited". This statement is against the AAOIFI standards. The correct answer is that yes, the bank *can* hold the title until the last payment of the property because the timing of ownership or title transfer is subject to terms outlined in the contract (Accounting and Auditing Organization for Islamic Financial Institutions (AAOIFI) 2025).

Now the same question is asked to FatawaRAG application and below is the response.

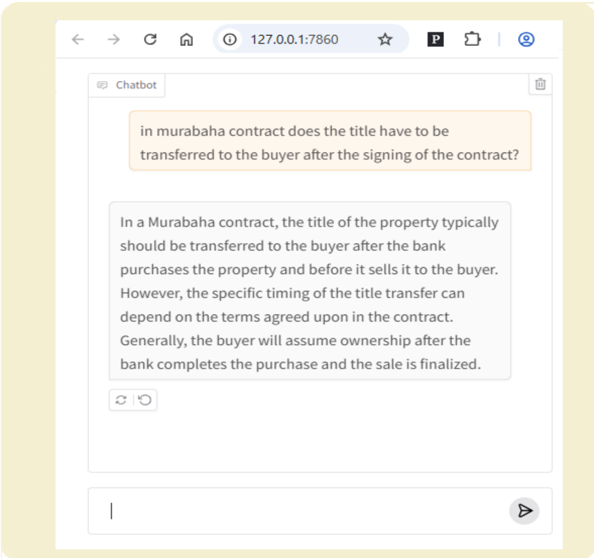


Figure 8. Question 1 - ChatGPT response section 6

The answer given by FatawaRAG is short, concise, clear, and accurate. It does not provide extensive and related details, but it provides short and to the point answer (AAOIFI, 2025).

6.2 Question 2

Here is the second question prompt that was asked to ChatGPT. The prompt is "During travelling, can I make *Tayammun* if I cannot catch the *salah* on time if I get water?". To make the question more technical the travelling concept is added which does not have to do anything with the answer of the main question.

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

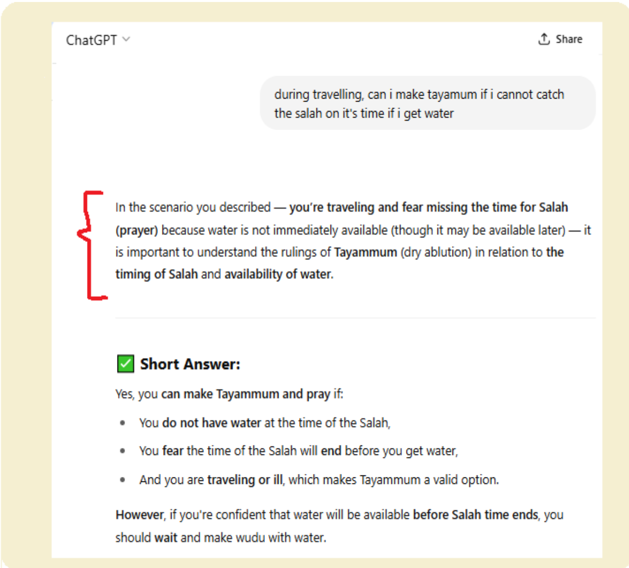


Figure 9. Question 2 - ChatGPT response section 1

The figure above of the response gives the impression that if a person is about to miss a Salah if he goes to perform ablution, he can just do *Tayammun* and complete his salah. This is one of the opinion in the Maliki school of thought by (Al-Hattab, n.d.). Al-Hattab also mentioned in his book *Mawahib al-jalil fi sharh Mukhtasar Khalil* that Imam Zufr, which is a Hanafi scholar, also agrees with this opinion.

Majority of the scholars disagree with the opinion which ChatGPT provided. The Hanbali school holds that Tayammum is not permissible in this circumstance (Ibn Qudamah al-Maqdisi n.d.). The Shafi school also says that it impermissible to do Tayammum in this case (Al-Ansari n.d.). According to the Hanafi

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

tradition, *Tayammum* is likewise not allowed for this case (Ibn Nujaym n.d.). Even, the Maliki school takes the same stance and does not allow tayammum in this situation (Ibn al-Hajib n.d.).

The response from FatawaRAG is clear because it is shown below:

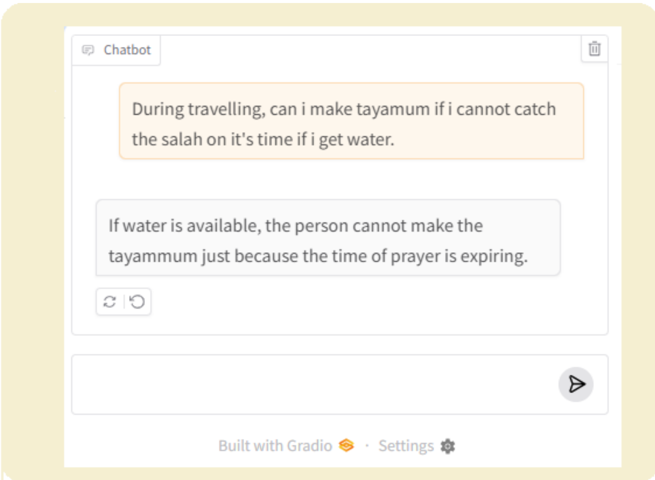


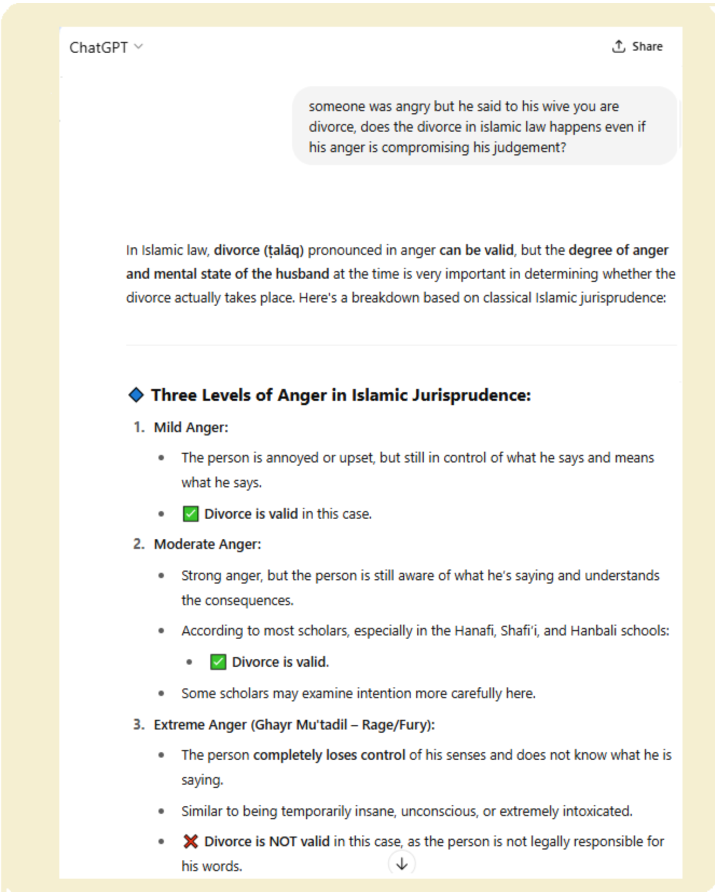
Figure 10. Question 2 - ChatGPT response section 2

The FatawaRAG application responds the opinion which is verified by majority of the scholars from various Islamic Jurisprudence school of thought. It is accepted according to the Hanafi, Shafi and Maliki School of Thoughts (Al-Ansari n.d.; Ibn Nujaym n.d.; Ibn Qudamah al-Maqdisi n.d.). Most Maliki scholars also agreed with this as well (Ibn al-Hajib n.d.).

6.3 Question 3

Here is the third question prompt that was asked to ChatGPT.

"Someone was angry, but he said to his wife you are divorced, does the divorce in Islamic law happen even if his anger is compromising his judgement?"



ChatGPT

Share

someone was angry but he said to his wive you are divorce, does the divorce in islamic law happens even if his anger is compromising his judgement?

In Islamic law, **divorce (ṭalāq)** pronounced in anger **can be valid**, but the **degree of anger and mental state of the husband** at the time is very important in determining whether the divorce actually takes place. Here's a breakdown based on classical Islamic jurisprudence:

🔷 **Three Levels of Anger in Islamic Jurisprudence:**

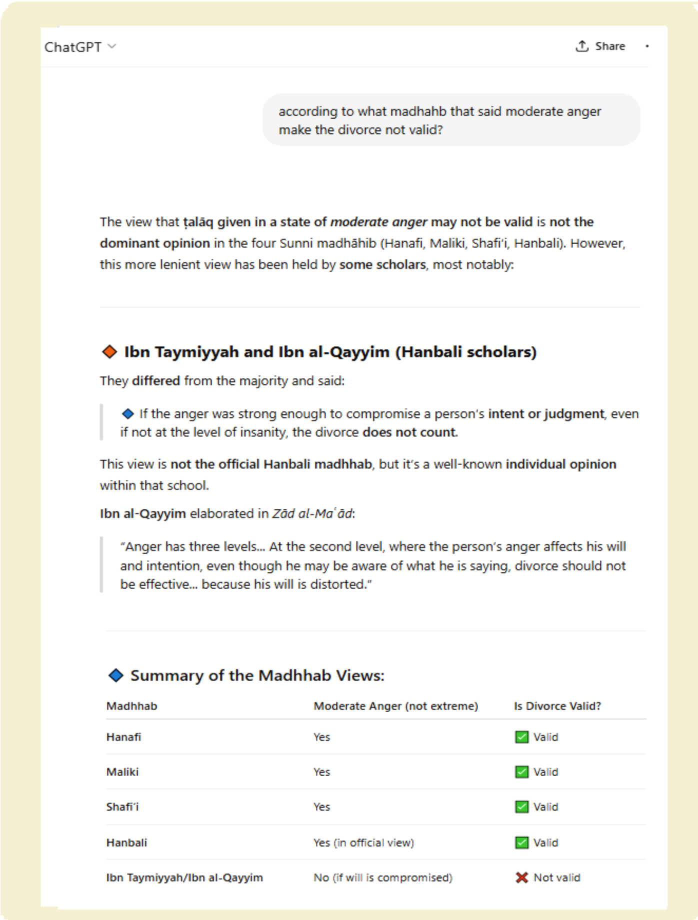
1. **Mild Anger:**
   - The person is annoyed or upset, but still in control of what he says and means what he says.
   - ✅ **Divorce is valid** in this case.
2. **Moderate Anger:**
   - Strong anger, but the person is still aware of what he's saying and understands the consequences.
   - According to most scholars, especially in the Hanafi, Shafi'i, and Hanbali schools:
     - ✅ **Divorce is valid.**
   - Some scholars may examine intention more carefully here.
3. **Extreme Anger (Ghayr Mu'tadil – Rage/Fury):**
   - The person **completely loses control** of his senses and does not know what he is saying.
   - Similar to being temporarily insane, unconscious, or extremely intoxicated.
   - ❌ **Divorce is NOT valid** in this case, as the person is not legally responsible for his words.

Figure 11. Question 3 ChatGPT response section 1

The answer provided by the ChatGPT is that divorce is not valid in extreme anger when the person completely loses control of his senses. This opinion is verified by multiple scholars from various school of Islamic Jurisprudence (Al-Buhuti n.d.; Al-Quduri n.d.; Al-Rahbani n.d.; Ibn Qudamah al-Maqdisi n.d.).

However, when further asked a more technical version of this question, “According to what madhhab that said moderate anger make the divorce not valid?”. Below is the response:

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801



ChatGPT

Share

according to what madhahb that said moderate anger make the divorce not valid?

The view that **ṭalāq given in a state of *moderate anger* may not be valid** is **not the dominant opinion** in the four Sunni madhāhib (Hanafi, Maliki, Shafi'i, Hanbali). However, this more lenient view has been held by **some scholars**, most notably:

Ibn Taymiyyah and Ibn al-Qayyim (Hanbali scholars)

They **differed** from the majority and said:

> If the anger was strong enough to compromise a person's **intent or judgment**, even if not at the level of insanity, the divorce **does not count**.

This view is **not the official Hanbali madhhab**, but it's a well-known **individual opinion** within that school.

Ibn al-Qayyim elaborated in *Zād al-Maʿād*:

> "Anger has three levels… At the second level, where the person's anger affects his will and intention, even though he may be aware of what he is saying, divorce should not be effective… because his will is distorted."

Summary of the Madhhab Views:

Madhhab	Moderate Anger (not extreme)	Is Divorce Valid?
Hanafi	Yes	✅ Valid
Maliki	Yes	✅ Valid
Shafi'i	Yes	✅ Valid
Hanbali	Yes (in official view)	✅ Valid
Ibn Taymiyyah/Ibn al-Qayyim	No (if will is compromised)	❌ Not valid

Figure 12. Question 3 - ChatGPT response section 2

And the response to the second prompt, the summary was correct. The scholars, Ibn Taymiyyah and Ibn al Qayyim differed from Hanbali approach. Al-Mardawi (n.d.) mentioned in his book

Al Insaaf that Ibn Taymiyyah's stance was that extreme anger that compromise the decision it will not work.

If we further ask the question with slightly more technical details with this prompt "but i looked for ibn Taymiyyah and ibn al Qayyim and they just agreed with the Hanbalis, but my mufti said some of late Shafi considered moderate anger makes divorce not valid"

Here is the response from ChatGPT:

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

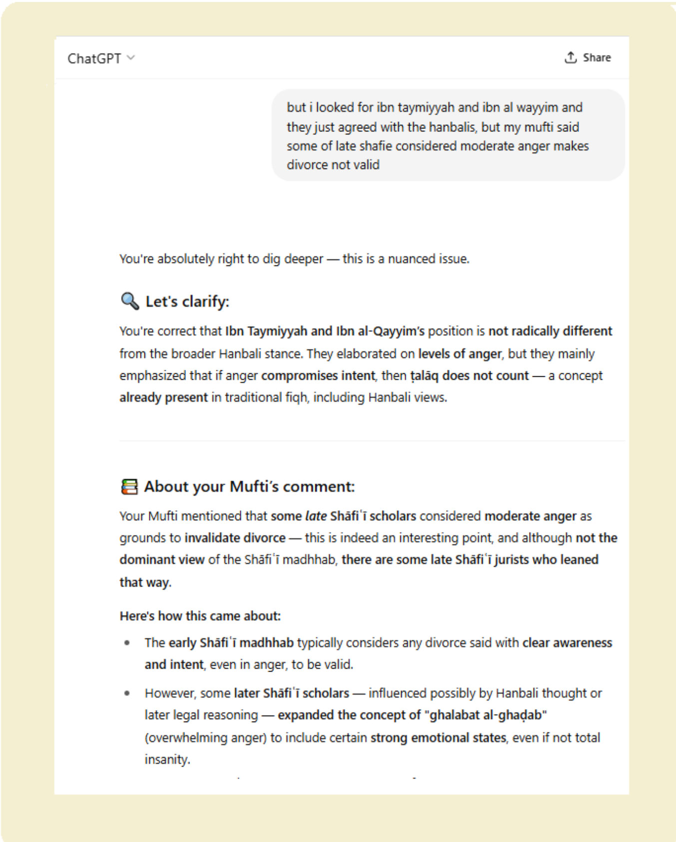
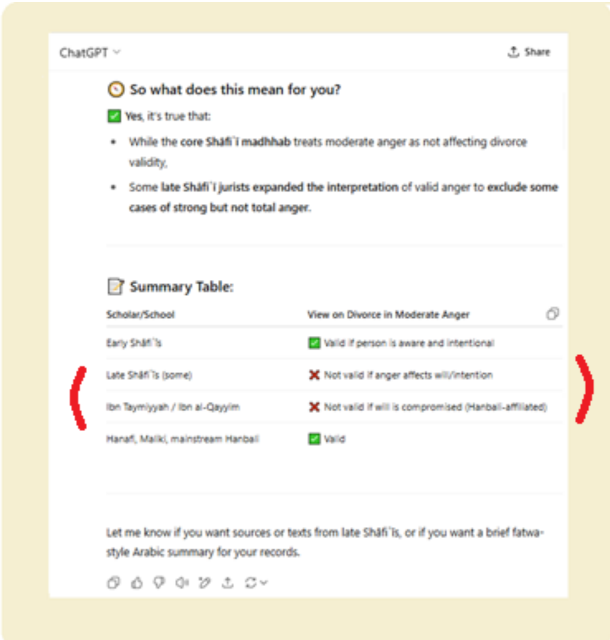


Figure 13. Question 3 - ChatGPT response section 3

Now here in the image above it clarifies that scholars Ibn Taymiyyah and Ibn al Qayyim took the position of the Hanbali school of thought (Al-Mardawi n.d.). The next part of the response shows the summary of the details and in that summary a confused response appears.



ChatGPT

Share

So what does this mean for you?

Yes, it's true that:

- While the **core Shāfiʿī madhhab** treats moderate anger as not affecting divorce validity.
- Some **late Shāfiʿī jurists expanded the interpretation** of valid anger to **exclude some cases of strong but not total anger.**

Summary Table:

Scholar/School	View on Divorce in Moderate Anger
Early Shāfiʿīs	Valid if person is aware and intentional
Late Shāfiʿīs (some)	Not valid if anger affects will/intention
Ibn Taymiyyah / Ibn al-Qayyim	Not valid if will is compromised (Hanbali-affiliated)
Hanafi, Maliki, mainstream Hanbali	Valid

Let me know if you want sources or texts from late Shāfiʿīs, or if you want a brief fatwa-style Arabic summary for your records.

Figure 14. Question 3 - ChatGPT response section 4

None of the late Shafi scholars took that position as mentioned in the image above (Al-Malibari n.d.). Answers like these from LLM models could cause confusion with the user.

Here is the response to the question from FatawaRAG:

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

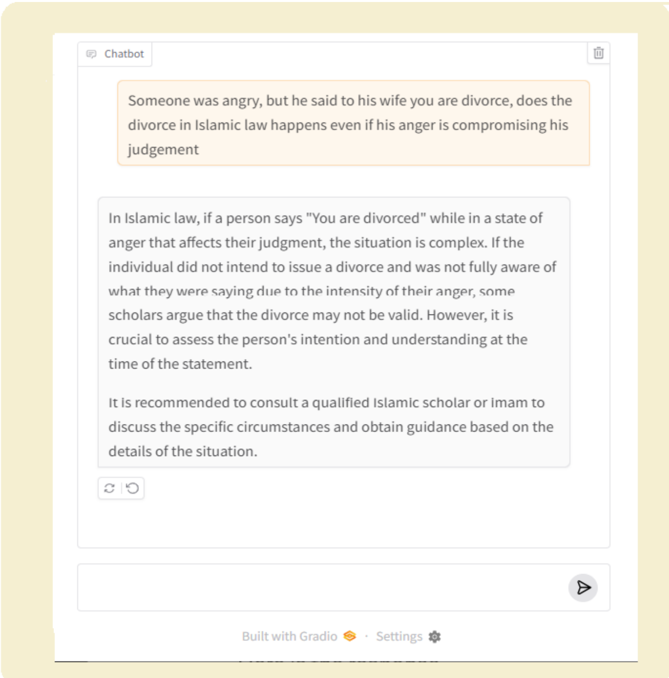


Figure 15. Question 3 - ChatGPT response section 5

In the image above FatawaRAG application show the short and clear answer to the problem. It does not provide extensive and related details which could cause confusion to the users (Al-Mardawi n.d.). Al Azhar also adopted the same opinion in one of their well-known Fatawa number 5513 (Ma'mun 1955).

In all the three questions asked to the ChatGPT yield a detailed and good response, but when a question becomes more technical, it provides answers that are either confusing or mixed with multiple schools of thought in Islamic jurisprudence.

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

7. EVALUATION METRICS

This section provides the details of how these responses were evaluated based on the Likert-type scale grading system for the various criteria:

The evaluators scored each response based on three criteria:

- **Factual Accuracy (Scale 1-5):** Is the answer legally correct according to the provided corpus?
- **Faithfulness (Scale 1-5):** Is the answer exclusively based on the provided source texts (for **FatawaRAG**) or generally accepted legal principles (for frontier models)? A low score indicates hallucination or irrelevant information.
- **Relevance and Clarity (Scale 1-5):** Is the answer well-written and does it directly address the user's question?

8. RESULTS AND ANALYSIS

The above ChatGPT and FatawaRAG responses were evaluated by two respected Canadian scholars who have multiple years of experience giving Islamic Jurisprudence verdicts. They rated each response on a 1 to 5 Likert-type scale grading system, where 1 represented "a very poor response from the system" and 5 represented "a very good response from the system" (Likert, 1932). This qualitative judgement was performed by reputable experts in the field of Islamic Jurisprudence

Table 1. Table shows the evaluation ratings by the expert in the Islamic Jurisprudence

	Question 1 for FatawaRAG	Question 1 for ChatGPT	Question 2 for FatawaRAG	Question 2 for ChatGPT	Question 3 for FatawaRAG	Question 3 for ChatGPT
Factual Accuracy	4	3	5	4	5	3
Faithfulness	5	3	4	4	5	4
Relevance and Clarity	5	3	5	3	5	2
Averages	4.6	3.0	4.6	3.6	5.0	3.0

Source: The author(s) own work.

8.1 Key Observations:

- **Superior Accuracy and Faithfulness:** FatawaRAG achieved near-perfect scores in accuracy and faithfulness. Its answers were consistently grounded in the legal texts provided. The frontier models, while often correct, were prone to subtle inaccuracies and occasionally introduced extraneous concepts.

- **Elimination of Hallucination:** Frontier models were presenting multiple opinions on the question asked which were confused and merged. The prompt might have also led to hallucinations, especially in Question #3.

- **Comparable Clarity:** The frontier models held a slight edge in the fluency and clarity of their prose, which is expected given the scale of their training. However, FatawaRAG's responses were rated as highly clear and relevant, indicating that the quality of the retrieved context was sufficient for the generator to produce high-quality output.

9. QUALITATIVE EVALUATION OF INFERENCE QUALITY USING THE CONSIDERS-THE-HUMAN FRAMEWORK

In addition to the key advantages of the FatawaRAG, there are multiple qualitative and quantitative frameworks that assess the quality of responses from the LLMs. These frameworks evaluate various capabilities of the LLMs by inferring and checking the responses that come out from these LLMs. In this research we used a qualitative evaluation framework to assess the responses from the FatawaRAG system.

ConSiDERS-The-Human framework is a six-pillar scheme proposed by Elangovan et al. (2024) to make human evaluation of generative models on a rigorous scale. ConSiDERS framework evaluates the responses based on six fundamental pillars and assigns ratings to the responses. The six fundamental pillars on which responses are evaluated include:

a. **Consistency:** This metric checks for stable task framing, repeatability and consistency of the response.

b. **Scoring Criteria:** These metric checks clarity, appropriateness, and accuracy of responses.

c. **Differentiating**: This metric checks the ability to highlight capability gaps between systems or prompts.

d. **User Experience:** These metric checks multiple attributes of the users' experience such as perceived usefulness, and speed.

e. **Responsible:** This metric shows whether the proper safety, fairness, bias, or factuality is applied.

f. **Scalability:** This metric shows how easy it is to implement this on a larger scale.

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

9.1 Aggregate Pillar Scores

Table 2. Aggregate Pillar scores for FatawaRAG application vs ChatGPT

Pillar	ChatGPT (μ)	RAG (μ)	Delta (Δ)
Consistency	4.8	4.2	-0.6
Scoring Criteria	4.1	4.7	+0.6
Differentiating	4.5	4.6	+0.1
User Experience	4.8	4.1	-0.7
Responsible	3.8	4.9	+1.1
Scalability	4.9	4.1	-0.8

Source: The author(s) own work.

Applying ConSiDERS-The-Human showed that FatawaRAG application improves inference quality for domain-specific technical tasks by using internal documents, resulting in more accurate and consistent answers with fewer hallucinations. This is very crucial for safety critical applications responses for the Islamic jurisprudence questions.

In “User Experience” the RAG application is clearly behind because increased response latency and less details”. Regarding scalability, the RAG application can provide accurate responses only when an appropriate solution exists within the knowledge base. If the knowledge base lacks the necessary information, the RAG application's performance may be compromised, which represents a significant disadvantage of this approach.

10. DISCUSSION AND IMPLICATIONS

The results of this proof of concept strongly support the thesis that a well-architected, domain-specific RAG system can outperform general-purpose frontier models on specialized tasks. By building FatawaRAG on a curated legal knowledge base, we effectively traded the vast but unreliable knowledge of a model like GPT-4o for the focused, verifiable, and always-current knowledge of our corpus.

The key advantages of the FatawaRAG approach are:

- **Trustworthiness**: The system's outputs are grounded in real, verifiable from the knowledge base which eliminates the risk of factual hallucination.
- **Maintainability**: The knowledge base can be updated instantly by adding new cases and documents which can be added to the index on regular bases to update the abilities of the FatawaRAG application.
- **Transparency**: Every statement in the output can be traced back to its source document, providing a crucial audit trial for legal professionals.

While this PoC was limited to a small subset of Canadian Islamic Jurisprudence questions, the architectural principles are highly generalizable. This approach could be extended to other complex, document-based domains such as medical research, financial compliance, and engineering standards, where accuracy and verifiability are paramount. The future of AI in professional domains does not lie with general Frontier models, but with specialized, RAG-based systems that combine the reasoning power of LLMs with the integrity of curated knowledge, these systems can be highly useful.

The Proof-of-Concept (POC) test showed that a partly incorrect answer about family law could break the rules set by Islam. Getting incorrect responses in Islamic family law, especially divorce issues could lead to serious outcomes. In such cases, a person might change a marital or financial decision based on an inaccurate legal rule, causing personal, familial, religious, or spiritual hardships. Moreover, misinformation can spread quickly. A single hallucinated statement, once posted online, can be shared across social media, reaching thousands of people in a matter of seconds. If it goes uncorrected, it could become accepted as the truth, misleading millions of people in a relatively short period of time.

In addition to AI producing hallucinated responses, it can also be trained on faulty datasets. The experiment conducted showed the importance of fed data in AI models. This experiment showed that data quality determined the accuracy of the model's responses; If the data is altered or contains errors, the model's responses will be wrong. If the data is intentionally corrupted by inserting fake facts to mislead the system, AI can be used to spread harmful misinformation, posing a major challenge in the realm of religion and its dominant practices.

In addition to hallucinated responses or incorrect data, Muslims may start "fatwa shopping". Human beings have a natural tendency to seek answers that align with their personal objectives or desires. Historically, this has been shown where individuals seek answers from religious scholars persistently until they receive a ruling that matches their specific needs. We may see a similar pattern emerging in the rapidly developing world of AI. Muslims, particularly the younger generation, who have AI in their palms would likely engage in such behavior thereby living a life according to their own "religious" terms and desires.

Journal of Integrated Sciences
Volume 6, Issue 3, June 2026
ISSN: 2806-4801

AI can also be misused. One serious problem is that AI could spread wrong or harmful information if it is trained on manipulated or false data. For example, if someone intentionally changes the information or propagates extremist propaganda in a dataset to include fake information or extremist ideas, then AI would produce religiously corrupt arguments, misleading people. Moreover, AI can create persuasive social-media posts, videos, or images that could spread extremist messages very quickly. These problems show that AI can distort the real teachings of Islam or promote a false agenda if fallen into the wrong hands.

11. LIMITATIONS OF GENERALIZATION AND SCALING OF RAG BASED SOLUTIONS

For any Domain specific RAG based solution can solve various problems which is a generic LLM model will not be able to help. But this approach has some advantages and limitations.

a. **Size and Diversity of the Dataset:** With the smaller and concise data set the response will be much more accurate. As soon as the size of the data increases, which may include diversity as well. The responses that will come out of this application might not be as accurate.

b. **Training Language:** A model which is trained on a specific language may not perform properly in the other languages as each language has its own constructions, idioms, legal terminologies which are significantly different between languages.

c. If the knowledge base repository consists of data from multiple schools of jurisprudence; effectiveness may diminish with this approach.

d. The inferences from RAG Application depend on the corpus that is used. If the corpus has missing topics, over representation of a particular scholar or jurisprudence school of thought it could lead to skewed answers.

12. CONCLUSION

The POC FatawaRAG solution has shown that a dedicated, domain-specific knowledge base combined with open-source LLM models can outperform standard generic LLM models. However, the POC currently presents challenges in terms of user experience and scalability; but these issues can be corrected by using more advanced approaches and better hardware. By maintaining a modular, easily updatable, and auditable knowledge base, we can harness the generative strengths of LLMs while addressing their inherent limitations.

REFERENCES

Ab Rahim, S. F., Ab Rahman, M. F., Abdullah Thaidi, H. A., Nik Mohd Azimi, N. N. M. A., & Jailani, M. R. (2025). Artificial Intelligence for Fatwa Issuance: Guidelines And Ethical Considerations. *Journal of Fatwa Management and Research*, *30*(1), 76–100. https://doi.org/10.33102/jfatwa.vol30no1.654

Abdul-Mageed, M., Elmadany, A., & Nagoudi, E. M. B. (2020, January). *ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic*. https://doi.org/10.48550/arXiv.2101.01785

Accounting and Auditing Organization for Islamic Financial Institutions (AAOIFI). (2025). *Shari'ah Standards*. https://aaoifi.com/shariaa-standards/?lang=en

Adeleke, A., Samsudin, N., Mustapha, A., & Mohd Nawi, N. (2017). Comparative Analysis of Text Classification Algorithms for Automated Labelling of Quranic Verses. *International Journal on Advanced Science, Engineering and Information Technology*, *7*, 1419–1427. https://doi.org/10.18517/ijaseit.7.4.2198

Algorithm. (2023). Retrieved from https://www.merriam-webster.com/dictionary/algorithm

Al-Ansari, Z. ibn M. (n.d.). *Asna al-Matalib Sharh Rawd al-Talib*.

Al-Buhuti, M. ibn Y. (n.d.). *Sharh Muntaha al-Iradat*.

Al-Hattab, M. A. A. A. ibn M. (n.d.). *Mawahib al-Jalil fi Sharh Mukhtasar Khalil*.

Al-Malibari, Z. al-D. ibn A. al-Aziz. (n.d.). *Fath al-Mu'in Bi Sharh Qurat Al-'Ain Bi Muhimat Al-Din*.

Al-Mardawi, A. al-Din. (n.d.). *Al-Insaaf*.

Al-Quduri, A. al-H. A. ibn M. (n.d.). *Al-Tajrid, Volume 10*.

Al-Rahbani, M. ibn S. ibn A. U. (n.d.). *Matalib Uli al-Nuha fi Sharh Ghayat al-Muntaha*.

Al-Shatibi, A. I. I. ibn M. (n.d.). *Al-Muwafaqat fi Usul al-Shari'ah* (A. A. Darr\āz, Ed.). Dar al-Hadith.

Alshithisri, A. B. A. (2023). *A Study of the Legal and Fiqhi Impact of AI on Issuing Fatwas*.

Atwell, E., Habash, N., & Shawar, B. A. (2010). *Understanding the Quran: A New Grand Challenge for Computer Science and Artificial Intelligence*. https://www.researchgate.net/publication/215442740

Artificial Intelligence. (2023). Retrieved from https://www.merriam-webster.com/dictionary/artificial%20intelligence

Atwell, E., & Sawalha, M. (2011). *An Artificial Intelligence Approach to Arabic and Islamic Content on the Internet*. https://doi.org/10.13140/2.1.2425.9528

Baskerville, R. L., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production. *MIS Quarterly*, *39*(3), 541–564.

Bengio, Y., Goodfellow, I., & Courville, A. (2016). *Deep Learning*. MIT Press.

Benke, I. (2020). On Implementing Ethical Principles in Design Science Research. *The Human-Computer Interaction Journal*.

Brocke, J. vom, Hevner, A., & Maedche, A. (2020). *Introduction to Design Science Research* (pp. 1–13). https://doi.org/10.1007/978-3-030-46781-4_1

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & others. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., & Amodei, D. (2020, January). *Language Models are Few-Shot Learners*. https://doi.org/10.48550/arXiv.2005.14165

Center, F., & Omer, M. (2025, July). Fatawa Center: - Fatawa Center. In *Fatawa Center*. https://fatawacenter.com/

Delua, J. (2023). *Supervised-vs-unsupervised-learning*. IBM. https://www.ibm.com/topics/supervised-learning

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018, January). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. https://doi.org/10.48550/arXiv.1810.04805

Duha, S., Al-Ahmad, S., & Ahmad, S. (2025). *Using Artificial Intelligence in Issuing Fatwas: A Jurisprudential Study*.

Elangovan, A., Liu, L., Xu, L., Bodapati, S. B., & Roth, D. (2024). *ConSiDERS-The-Human Evaluation Framework: Rethinking Human Evaluation for Generative Large Language Models*. 1137–1160. https://doi.org/10.18653/v1/2024.acl-long.63

ElKomy, M., & Sarhan, A. M. (2022). *Arabic Language Question Answering Over Holy Qur'an Using a Post-Processed Ensemble of BERT-based Models.* http://arxiv.org/abs/2206.01550

Ferrucci, D. A. (2012). Introduction to "This is Watson." *IBM Journal of Research and Development*, *56*, 1:1-1:15. https://doi.org/10.1147/JRD.2012.2184356

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., Schlaefer, N., & Welty, C. (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine*, *31*, 59–79. https://doi.org/10.1609/aimag.v31i3.2303

Graham, B. (2019). Using Natural Language Processing to Search for Textual References. In *Digital Biblical Studies* (Vol. 3, pp. 115–132). Brill Academic Publishers. https://doi.org/10.1163/9789004399297_008

Hady, M. F. A., & Schwenker, F. (2013). Semi-supervised Learning. In M. Bianchini, M. Maggini, & L. C. Jain (Eds.), *Handbook on Neural Information Processing* (Vol. 49, pp. 215–239). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36657-4_7

Hamad, Z. T., Laouar, M., & Issam, B. (2020). *Quran content representation in NLP.* 1–6. https://doi.org/10.1145/3447568.3448552

Hamoud, B., & Atwell, E. (2016). Quran question and answer corpus for data mining with WEKA. *Proceedings of 2016 Conference of Basic Sciences and Engineering Studies, SGCAC 2016*, 211–216. https://doi.org/10.1109/SGCAC.2016.7458032

Ibn al-Hajib, J. al-D. ibn U. (n.d.). *At-Tawdih fi Sharh al-Mukhtasar*.

Ibn Nujaym, Z. al-D. ibn I. (n.d.). *Al-Bahr al-Ra'iq Sharh Kanz al-Daqa'iq*.

Ibn Qudamah al-Maqdisi, A. A. ibn A. (n.d.). *Al-Mughni*.

ISO/IEC 42001:2023 *Information technology—Artificial intelligence—Management system* (Issue ISO/IEC 42001:2023). (2023). ISO/IEC. https://www.iso.org/standard/42001

Izacard, G., & Grave, E. (2021). Leveraging Passage Retrieval for Knowledge-Intensive Language Tasks. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, 1–10.

Jamal, M. H. N., & Zakaria, M. Z. (2023). *Shariah Guidelines on Artificial Intelligence (AI) in Fatwa Issurance*.

Jawawi, D. N. Abang. (2014). *Support Vector Machine based approach for Quranic Words Detection in Online Textual Content*. IEEE.

Kammani, D. A., & Safeena, R. (2014). A review of Quranic computation for e-learning. *International Journal of Web Science*, *2*, 127. https://doi.org/10.1504/IJWS.2014.066415

Knoblauch, H. (2001). Focused Ethnography. *Sozialersinn: Zeitschrift Fur Hermeneutische Sozialforschung*, *6*, 123–141.

Lapão, L., Silva, M. da, & Gregorio, J. (2017). Implementing an online pharmaceutical service using design science research. *BMC Medical Informatics and Decision Making*, *17*. https://doi.org/10.1186/s12911-017-0428-2

Lewis, P., Perez, E., Piktus, A., & al, et. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Proceedings of the 38th International Conference on Machine Learning*, 1000–1010.

Liddy, E. D. (2001). *Natural Language Processing*. https://surface.syr.edu/cgi/viewcontent.cgi?article=1019&context=cnlp

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140), 1–55.

Mahamood, S. F., Fikry, A., Hamzah, M. I., Khalid, M. M., & Bhari, A. (2023). Fiqh Robotic for Artificial Intelligent In Humanoids Used For Therapy, Services And Other Social Activities: An Integration Of Artificial Intelligence (Ai) And Maqasid Shariah. *Journal of Fatwa Management and Research*, *28*(2), 1–13. https://doi.org/10.33102/jfatwa.vol28no2.527

Mamun, H. (1955, August 16). *Fatawa No. 5313*.

Mcdonald, D. (2014). A Text Mining Analysis of Religious Texts. *The Journal of Business Inquiry*, *13*(1), 27–47.

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education* (2nd ed.). Jossey-Bass.

Merriam, S. B. (2002). *Qualitative Research in Practice: Examples for Discussion and Analysis* (1st ed.). Jossey-Bass.

Mohammed, A., & Kora, R. (2019). Deep learning approaches for Arabic sentiment analysis. *Social Network Analysis and Mining*, *9*. https://doi.org/10.1007/s13278-019-0596-4

Nassourou, M. (2021). *Using Machine Learning Algorithms for Categorizing Quranic Chapters by Major Phases of Prophet Mohammad's Messengership*.

Natural Language Processing. (2023). Retrieved from https://www.merriam-webster.com/dictionary/natural%20language

Parthasarathy, V. B., Zafar, A., Khan, A., & Shahid, A. (2024). The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities. *arXiv*. https://arxiv.org/abs/2408.13296

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*. https://doi.org/10.1016/j.infsof.2015.03.007

Shamsudin, A. (2000). *AI Natural Language in Meta-Synthetics of Al-Qur'an*.

Sihotang, M. T., Jaya, I., Hizriadi, A., & Hardi, S. M. (2020). Answering Islamic Questions with a Chatbot using Fuzzy String-Matching Algorithm. *Journal of Physics: Conference Series*, *1566*(1). https://doi.org/10.1088/1742-6596/1566/1/012007

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. https://doi.org/10.48550/arXiv.1706.03762