JATI JURNAL MAHASISWA TEKNIK INFORMATIKA
INSTITUT TEKNOLOGI NASIONAL MALANG
Kampus 2 : Jl. Raya Karanglo Km.2 Malang
e-ISSN : 2598-828X
Web : https://ejournal.itn.ac.id/index.php/jati Email : jati@scholar.itn.ac.id

Nomor : ITN.103031/III/JATI/2026 Malang, 12 Maret 2026
Lampiran : -
Perihal : Penerimaan Naskah Jurnal JATI

Kepada Yth. :

Bapak / Ibu **Anindia Atikah Putri Tanjung, Hendri Ahmadian, Khairan AR, Fathiah, Nurrizqa**

Dengan hormat, Bersama ini kami sampaikan bahwa naskah Saudara yang berjudul:

PENGEMBANGAN SISTEM PENCARIAN AYAT AL-QUR'AN BERBASIS TOPIK MENGGUNAKAN LARGE LANGUAGE MODEL META AI

Sudah selesai review dan revisi serta sudah dinyatakan diterima, dan akan diterbitkan dalam jurnal JATI (Jurnal Mahasiswa Teknik Informatika) Vol. 10 No. 3, yang dipublikasikan pada edisi Juni 2026, atas perhatian dan kerjasamanya kami ucapkan terimakasih.

Jurnal JATI
Ketua Editor

Joseph Dedy Irawan, ST, MT.
NIP. 197404162005011002

Tembusan :
1. Arsip

PENGEMBANGAN SISTEM PENCARIAN AYAT AL-QUR'AN BERBASIS TOPIK MENGGUNAKAN LARGE LANGUAGE MODEL META AI

Anindia Atikah Putri Tanjung, Hendri Ahmadian, Khairan AR, Fathiah, Nurrizqa
Teknologi Informasi, UIN Ar-Raniry Banda Aceh
Lorong Ibnu Sina No.2, Darussalam, Kopelma Darussalam, Kec. Syiah Kuala, Kota Banda Aceh, Indonesia
anindia.atika@gmail.com

ABSTRAK

Perkembangan teknologi digital telah mengubah cara masyarakat mengakses Al-Qur'an, namun sebagian besar sistem pencarian ayat masih berbasis kata kunci sehingga hasil yang diperoleh cenderung literal dan kurang kontekstual. Permasalahan ini menunjukkan perlunya pendekatan berbasis kecerdasan buatan yang mampu memahami hubungan semantik antar ayat secara tematik. Penelitian ini bertujuan mengembangkan sistem pencarian ayat Al-Qur'an berbasis topik menggunakan *Large Language Model Meta* AI (LLaMA 3.1-8B) yang di *fine-tune* dengan metode *Quantized Low-Rank Adaptation* (QLoRA) dan diintegrasikan dengan *Retrieval-Augmented Generation* (RAG). Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan dataset sebanyak 6.236 ayat beserta tafsirnya. Proses pengembangan meliputi tahap *pre-processing*, *fine-tuning* model, pembangunan basis data vektor menggunakan *embedding*, serta integrasi mekanisme *retrieval* dan generasi teks. Evaluasi *retrieval* menunjukkan nilai *Mean* Recall@k sebesar 0.749 dan *Mean Reciprocal Rank* (MRR) sebesar 1.000, yang mengindikasikan sistem mampu menempatkan ayat relevan pada peringkat pertama secara konsisten. Evaluasi *generative* menggunakan ROUGE menghasilkan nilai ROUGE-1 F1 sebesar 0.1551, ROUGE-2 F1 sebesar 0.0863, dan ROUGE-L F1 sebesar 0.1338. Hasil tersebut menunjukkan bahwa sistem mampu menghasilkan makna ayat yang selaras dengan tafsir referensi secara ringkas dan kontekstual.

Kata kunci : *Sistem Pencarian Ayat Al-Qur'an, Large Language Model, LLaMA 3, QLoRA, Retrieval-Augmented Generation (RAG)*

1. PENDAHULUAN

Al-Qur'an merupakan pedoman hidup dan sumber utama ajaran Islam yang memuat nilai moral, hukum, serta prinsip kehidupan yang bersifat universal dan relevan sepanjang masa. Sebagai petunjuk bagi seluruh umat manusia, Al-Qur'an menjadi rujukan utama dalam pembentukan aqidah, ibadah, dan muamalah[1]. Dalam konteks kehidupan modern, manusia menghadapi tantangan besar dalam memahami dan mengamalkan ajaran Islam secara komprehensif, terutama seiring dengan pesatnya perkembangan teknologi informasi yang memengaruhi pola pikir, perilaku, serta cara memperoleh pengetahuan. Perkembangan era digital telah mengubah cara manusia mengakses dan menyebarkan informasi, termasuk dalam bidang keilmuan Islam, sehingga integrasi teknologi dalam pendidikan Islam menjadi kebutuhan yang tidak terpisahkan[2].

Integrasi teknologi telah membuka peluang besar untuk memperluas akses terhadap sumber-sumber keilmuan Islam serta memperkaya metode penyampaian materi keagamaan. Namun, perubahan ini juga menghadirkan sejumlah tantangan, seperti keterbatasan literasi digital, risiko penyebaran informasi yang tidak valid, serta perlunya menjaga nilai-nilai spiritual dalam lingkungan digital[3]. Dalam praktiknya, transformasi Al-Qur'an di era digital menunjukkan bahwa sebagian besar aplikasi Al-Qur'an masih mengandalkan pencarian berbasis kata kunci *(keyword-based search*), sehingga hasil pencarian cenderung bersifat literal dan kurang kontekstual. Kondisi ini menyebabkan keterbatasan dalam menangkap hubungan makna antar-ayat yang seharusnya dipahami secara tematik.

Sejumlah penelitian menunjukkan bahwa pendekatan leksikal murni memiliki keterbatasan dalam merepresentasikan makna semantik ayat Al-Qur'an, sementara metode berbasis representasi semantik mampu meningkatkan relevansi hasil pencarian[4]. Perkembangan *Natural Language Processing* (NLP) dan pemanfaatan *word embedding* serta model pembelajaran mendalam telah digunakan untuk mengklasifikasikan ayat Al-Qur'an ke dalam tema-tema tertentu, namun masih menghadapi keterbatasan dalam memahami konteks semantik yang kompleks antar-ayat. Seiring dengan itu, kemajuan *Artificial Intelligence*, khususnya *Large Language Models* (LLM), menawarkan kemampuan pemrosesan bahasa alami yang lebih kontekstual, terutama ketika dikombinasikan dengan pendekatan *Retrieval-Augmented Generation* (RAG) yang mengintegrasikan sumber pengetahuan eksternal secara sistematis[5].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sistem pencarian ayat Al-Qur'an berbasis topik dengan memanfaatkan model LLaMA 3 yang di *fine-tune* menggunakan metode *Quantized Low-Rank Adaptation* (QLoRA) dan didukung oleh pendekatan *Retrieval-Augmented Generation* (RAG). Studi ini menilai kemampuan sistem dalam menampilkan hasil pencarian ayat yang relevan secara semantik dan kontekstual sesuai dengan tema yang dimasukkan oleh pengguna. Evaluasi dilakukan untuk melihat sejauh mana integrasi LLM,

QLoRA, dan RAG mampu mengatasi keterbatasan pencarian berbasis kata kunci dalam konteks kajian Al-Qur'an digital.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pemanfaatan *Large Language Model* (LLM) dan *Retrieval-Augmented Generation* (RAG) dalam sistem tanya jawab dan pencarian informasi berbasis teks telah banyak diteliti. Berbagai penelitian menunjukkan bahwa integrasi LLM dengan mekanisme *retrieval* mampu meningkatkan relevansi dan akurasi hasil pencarian. Rachmat *et al.*[6] melakukan *fine-tuning* model Mistral 7B menggunakan metode *Quantized Low-Rank Adaptation* (QLoRA) yang dikombinasikan dengan RAG, dengan evaluasi menggunakan metrik ROUGE. Hasil pengujian menunjukkan bahwa pendekatan RAG menghasilkan skor di atas 0,5 pada sebagian besar pertanyaan uji, sedangkan metode *fine-tuning* mencapai skor ROUGE 1,0 pada seluruh data pengujian. Meskipun efektif, penelitian tersebut masih terbatas pada domain administrasi pendidikan dan belum diterapkan pada teks keagamaan yang memiliki struktur semantik lebih kompleks.

Dalam konteks kajian Islam, penelitian oleh Syah *et al.*[7] mengembangkan sistem tanya jawab hadis berbasis web dengan menerapkan pendekatan RAG menggunakan Langchain yang terintegrasi dengan model GPT-4 dan menyimpan *vector* pada ChromaDB. Evaluasi menggunakan metrik BERTScore menghasilkan F1 rata-rata sebesar 0,7962, sementara evaluasi berbasis kuesioner Likert menunjukkan tingkat kepuasan pengguna sebesar 89,4%. Hasil tersebut menunjukkan bahwa pendekatan RAG mampu meningkatkan relevansi jawaban dan menyertakan sumber rujukan hadis secara eksplisit.

Namun, kajian yang secara khusus mengintegrasikan LLM *open-source* yang di-*fine-tune* menggunakan *Quantized Low-Rank Adaptation* (QLoRA) dengan pendekatan RAG untuk pencarian ayat Al-Qur'an berbasis topik secara semantik dan kontekstual masih terbatas. Oleh karena itu, penelitian ini memfokuskan pada pengembangan sistem pencarian ayat Al-Qur'an berbasis topik dengan memanfaatkan model LLaMA 3 yang di-*fine-tune* menggunakan QLoRA dan didukung oleh mekanisme RAG.

2.2. Sistem Pencarian Ayat Al-Qur'an

Sistem pencarian ayat Al-Qur'an secara umum dikategorikan menjadi pencarian berbasis kata kunci, berbasis semantik, dan *Cross-Language Information Retrieval* (CLIR). Pencarian kata kunci bergantung pada kecocokan literal antara kueri dan teks, sedangkan pendekatan semantik menitikberatkan pada kesesuaian konsep melalui ontologi atau sinonim, dan CLIR mendukung pencarian lintas bahasa melalui proses penerjemahan. Meskipun pendekatan semantik lebih kontekstual, mayoritas sistem pencarian Al-Qur'an digital masih didominasi metode berbasis kata kunci[8].

2.3. Arsitektur *Transformers*

Arsitektur *Transformers* merupakan fondasi *Large Language Model* (LLM) karena mampu memproses data secara paralel melalui mekanisme *attention* yang menangkap hubungan global antar token. Meskipun secara umum terdiri atas encoder dan decoder, banyak model modern seperti GPT dan LLaMA hanya menggunakan decoder. Kemampuannya dalam merepresentasikan struktur dan konteks bahasa menjadikan Transformer sebagai dasar pengembangan model bahasa berskala besar[9].

2.4. *Large Language Model* (LLM)

Large Language Model (LLM) merupakan model bahasa dalam kecerdasan buatan yang dirancang untuk memahami dan menghasilkan teks berbasis bahasa alami. Model ini dilatih menggunakan data berskala besar dengan jumlah parameter yang signifikan, sehingga mampu menghasilkan keluaran yang kontekstual dan gramatikal. Berbasis teknologi *Natural Language Processing* (NLP), LLM dapat menjalankan berbagai tugas seperti menjawab pertanyaan, merangkum, menerjemahkan, serta melakukan penalaran dan generalisasi pengetahuan untuk menghasilkan respons yang relevan dan akurat.[5].

2.5. *Large Language Model Meta AI* (LLaMA 3)

Large Language Model Meta AI (LLaMA 3) merupakan generasi terbaru model bahasa yang dikembangkan oleh Meta dengan menggunakan arsitektur *decoder-only Transformer* yang dioptimalkan untuk meningkatkan pemahaman konteks dan efisiensi pemrosesan. Model ini tersedia dalam beberapa varian parameter, termasuk 8B dan 70B, serta mendukung konteks panjang hingga 128k token, dengan pelatihan berbasis *multilingual high-quality datasets* yang memungkinkan performa kompetitif pada berbagai tugas pemahaman bahasa dan penalaran lintas domain. Dalam penelitian ini, model LLaMA 3.1-8B digunakan sebagai fondasi pengembangan sistem pencarian ayat Al-Qur'an berbasis topik[10].

2.6. *Quantized Low-Rank Adaptation* (QLoRA)

QLoRA merupakan pendekatan *parameter-efficient fine-tuning* (PEFT) yang mengombinasikan kuantisasi bobot model dasar ke dalam presisi 4-bit berbasis *NormalFloat* (NF4) dengan mekanisme *Low-Rank Adaptation* (LoRA) sebagai adaptor berdimensi rendah. Integrasi kedua teknik tersebut memungkinkan proses *fine-tuning* pada model Bahasa berskala besar dengan kebutuhan memori dan komputasi yang jauh lebih rendahm namun tetap mempertahankan kinerja yang sebanding dengan *fine-tuning* presisi 16-bit[11].

2.7. *Retrieval-Augmented Generation* (RAG)

Retrieval-Augmented Generation (RAG) merupakan pendekatan generatif dalam *Information*

Retrieval yang menggabungkan proses *retrieval* dan *text generation* untuk meningkatkan kualitas keluaran model dengan memanfaatkan sumber pengetahuan eksternal. Pendekatan ini memungkinkan model bahasa mengakses dokumen relevan sebagai konteks tambahan, sehingga efektif pada sistem *question answering* dan pencarian informasi yang memerlukan pemahaman kontekstual serta rujukan eksplisit. Dalam sistem pencarian ayat Al-Qur'an, RAG mengintegrasikan kemampuan penalaran LLM dengan mekanisme pengambilan informasi untuk menghasilkan jawaban yang lebih akurat, kontekstual, dan berbasis sumber[12].

2.8. Evaluasi Model

Evaluasi model pada sistem berbasis RAG dilakukan melalui dua komponen utama, yaitu evaluasi *retrieval* dan evaluasi *generative.* Evaluasi *retrieval* mengukur kemampuan sistem dalam menemukan dokumen atau ayat yang relevan terhadap kueri, sedangkan evaluasi *generative* menilai tingkat kesesuaian teks jawaban yang dihasilkan model dengan referensi yang tersedia.

1. Evaluasi *Retrieval*
   a. Recall@k
   Recall@k merupakan metrik yang digunakan untuk mengukur berapa banyak dokumen relevan yang berhasil ditemukan dalam k hasil teratas (*top-k retrieval*). Metrik ini menilai sejauh mana sistem mampu memasukkan dokumen yang benar dalam daftar hasil pencarian teratas.

$$Recall@k = \frac{\text{Dokumen Relevan dalam Top}-k}{\text{Total Dokumen Relevan}} \quad (1)$$

   Nilai *Recall@k* berada pada rentang 0 hingga 1. Semakin mendekati 1, semakin baik kemampuan sistem dalam menemukan dokumen relevan pada posisi teratas hasil pencarian[12].

   b. *Mean Reciprocal Rank* (MRR)
   Mean Reciprocal Rank (MRR) adalah metrik evaluasi kinerja untuk model *Question Answering* (QA) yang digunakan untuk mengukur kualitas peringkat hasil pencarian dengan mempertimbangkan posisi kemunculan pertama jawaban atau dokumen yang relevan. MRR menghitung kebalikan (*reciprocal*) dari peringkat pertama dokumen yang benar untuk setiap kueri, kemudian dirata-ratakan terhadap seluruh kueri yang diuji [13].
   Rumus MRR dinyatakan sebagai berikut:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i} \quad (2)$$

   Di mana |Q| adalah jumlah pertanyaan yang dievaluasi, rank adalah peringkat pertama dokumen relevan untuk kueri ke-i.

2. Evaluasi *Generative*
   Salah satu metrik yang umum digunakan dalam tugas pemrosesan Bahasa alami adalah *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE). Metrik ini membandingkan tingkat kemiripan antara teks yang dihasilkan sistem dengan teks referensi yang dibuat secara manual. Nilai ROUGE berada pada rentang 0 hingga 1, dimana nilai yang lebih tinggi menunjukkan Tingkat kesamaan yang lebih besar antara teks hasil model dan referensi [6].
   a. ROUGE-1
   ROUGE-1 mengukur tingkat kesesuaian *unigram* (kata tunggal) yang tumpang tindih antara teks hasil model dan teks referensi. Metrik ini mengevaluasi seberapa banyak kata yang sama muncul pada kedua teks. Secara umum, presisi, *recall*, dan F1-*score* pada ROUGE-1 dapat dirumuskan sebagai:

$$\text{ROUGE} - 1(Recall) = \frac{unigram\ matches}{unigram\ in\ references} \quad (3)$$

$$\text{ROUGE} - 1(Precision) = \frac{unigram\ matches}{unigram\ in\ outputs} \quad (4)$$

$$\text{ROUGE} - 1(F1) = \frac{Precision\ x\ Recall}{Precision+Recall} \quad (5)$$

   b. ROUGE-2
   ROUGE-2 mengukur tingkat kesesuaian *bigram* (dua kata berurutan) antara teks hasil model dan referensi. Berbeda dengan ROUGE-1 yang hanya mempertimbangkan kata tunggal, ROUGE-2 mempertimbangkan urutan dua kata sehingga lebih sensitif terhadap struktur kalimat. Perhitungan presisi, *recall*, dan F1-*score* pada ROUGE-2 menggunakan rumus yang sama dengan ROUGE-1, namun unit yang dibandingkan adalah bigram.

   c. ROUGE-L
   ROUGE-L didasarkan pada konsep *Longest Common Subsequence* (LCS), yaitu urutan kata terpanjang yang muncul secara berurutan pada teks referensi dan teks hasil model. Metrik ini tidak hanya memperhatikan kemunculan kata, tetapi juga mempertimbangkan urutan relatifnya dalam kalimat. Secara umum, *recall* dan *precision* berbasis LCS dapat dirumuskan sebagai:

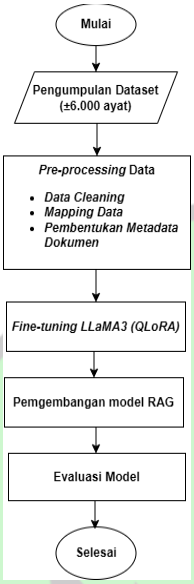
$$\text{ROUGE} - \text{L}(Recall) = \frac{Length\ of\ LCS}{unigram\ in\ references} \quad (6)$$

$$\text{ROUGE} - \text{L}(Precision) = \frac{Length\ of\ LCS}{unigram\ in\ outputs} \quad (7)$$

$$\text{ROUGE} - \text{L}(F1) = \frac{Precision\ x\ Recall}{Precision+Recall} \quad (8)$$

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menguji sistem pencarian ayat Al-Qur'an berbasis topik menggunakan *Large Language Model Meta AI* (LLaMA 3) yang di *fine-tune* dengan metode *Quantized Low-Rank Adaptation* (QLoRA) serta diintegrasikan dengan mekanisme *Retrieval-Augmented Generation* (RAG).



Gambar 1. Diagram Alir Penelitian

Alur tahapan penelitian ditunjukkan pada Gambar 1, yang menggambarkan proses penelitian secara berurutan dan terstruktur. Proses dimulai dari tahap pengumpulan dataset dengan jumlah sekitar ±6.000 ayat yang digunakan sebagai data utama. Selanjutnya, dilakukan tahap *pre-processing* yang mencakup *data cleaning* untuk memastikan kualitas teks, *mapping dataset* untuk menyesuaikan struktur data, serta pembentukan metadata dokumen guna memperkaya informasi kontekstual setiap dokumen.

Tahap berikutnya meliputi *fine-tuning* menggunakan LLaMA 3 dengan metode QLoRA untuk mengadaptasi model terhadap domain Al-Qur'an secara efisien. Setelah itu dilakukan pengembangan model RAG dengan membangun basis data vektor dari *embedding* dokumen ayat guna mendukung proses pencarian semantik berdasarkan kueri pengguna. Tahap akhir adalah evaluasi model, yang dilakukan untuk menilai kinerja sistem dalam menghasilkan jawaban yang relevan dan akurat berdasarkan metrik evaluasi berbasis kesamaan semantik.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan dataset Al-Qur'an dan Tafsir Bahasa Indonesia dari repositori *HuggingFace* "ronnieaban/alquran", yang bersumber dari data resmi Al-Qur'an dan Tafsir Kementerian Agama Republik Indonesia (Quran Kemenag). Dataset ini mencakup teks Arab, transliterasi, terjemahan, serta berbagai bentuk tafsir yang mendukung analisis semantik dan pengembangan sistem pencarian ayat berbasis topik menggunakan *Large Language Model* (LLM).

4.2. *Pre-Processing*

Tahap *pre-processing* dilakukan untuk memastikan kualitas dan konsistensi data sebelum digunakan pada proses *fine-tuning* dan pengembangan model RAG. Tahapan ini terdiri atas tiga proses utama.

a. *Data Cleaning*
Proses *data cleaning b*ertujuan untuk menormalisasi teks dengan menghapus karakter baris baru (*newline*) serta merapikan spasi berlebih menggunakan ekspresi regular.

b. *Mapping* Dataset
Pada tahap ini, setiap ayat dipetakan menjadi satu dokumen dengan menggabungkan bagian terjemahan dan tafsir wajiz dalam satu representasi teks. Proses ini menghasilkan 6.236 dokumen yang digunakan sebagai dasar pembentukan embedding dan penyimpanan dalam *vector database*.

Tabel 1. Hasil *Mapping* Dataset

Komponen	Isi Dokumen
Format dokumen	Terjemahan + Tafsir Wajiz
Jumlah dokumen	6.236
Contoh dokumen	Terjemahan: Dengan nama Allah Yang Maha Pengasih lagi Maha Penyayang. Tafsir: Aku memulai bacaan Al-Qur'an dengan menyebut nama Allah ...

Tabel 1 dapat dilihat bahwa setiap dokumen telah disusun dalam satu representasi teks yang memuat terjemahan dan tafsir sebagai konteks pendukung pencarian berbasis topik.

c. Pembentukan Metadata Dokumen
Selain membentuk dokumen teks utama, setiap ayat juga dilengkapi matadata untuk mendukung proses *retrieval* dan penyajian informasi ayat secara terstruktur. Metadata disimpan dalam bentuk pasangan *key value* yang memuat atribut identitas dan deskripsi ayat.

Tabel 2. Metadata Dokumen Ayat Al-Quran

Atribut	Isi Dokumen
Surah	Al-Fātiḥah
Ayat	1
Topic	*None*
Arabic	بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيْمِ
Latin	Bismillāhir-raḥmānir-raḥīm(i).
Translation	Dengan nama Allah Yang Maha Pengasih lagi Maha Penyayang.
Tafsir Wajiz	Aku memulai bacaan Al-Qur'an dengan menyebut nama Allah ...

Tabel 2 dapat dilihat bahwa setiap dokumen memiliki atribut identitas yang lengkap, sehingga mendukung proses pencarian ayat berdasarkan surah, nomor ayat, maupun topik tertentu dalam sistem RAG.

4.3. *Fine-tuning* LLaMA3

Pada tahap ini dilakukan proses *fine-tuning* model LLaMA-3.1-8B menggunakan pendekatan QLoRA. Model dimuat dalam mode kuantisasi 4-bit untuk meningkatkan efisiensi penggunaan memori selama pelatihan. Proses ini bertujuan mengadaptasi model agar mampu memahami pola penjelasan tafsir ayat Al-Qur'an dalam format *instruction-response.*

Konfigurasi adapatasi dilakukan dengan menambahkan modul LoRA pada beberapa komponen utama arsitektur *Transformer.* Pendekatan ini memungkinkan pelatihan dilakukan secara efisien dengan hanya memperbarui Sebagian kecil parameter model tanpa melakukan *full fine-tuning.*

Tabel 3. Konfigurasi Model dan QLoRA

Parameter	Nilai
Model dasar	LLaMA-3.1-8B
Quantization	4-bit
Rank (r)	16
LoRA *alpha*	16
LoRA d*ropout*	0,05
Target *modules*	q_proj, k_proj, v_proj, o_proj, gate_proj, up_proj, down_proj
Trainable parameters	41.943.040 (±0,52%)

Tabel 3 dapat dilihat bahwa hanya sebagian kecil parameter yang dilatih, sehingga proses adaptasi menjadi lebih efisien dari sisi komputasi. Proses pelatihan dilakukan menggunakan 200 data dalam format *instruction-response* selama dua epoch dengan total 100 pelatihan. Model menunjukkan penurunan nilai *training loss* secara bertahap hingga akhir pelatihan.

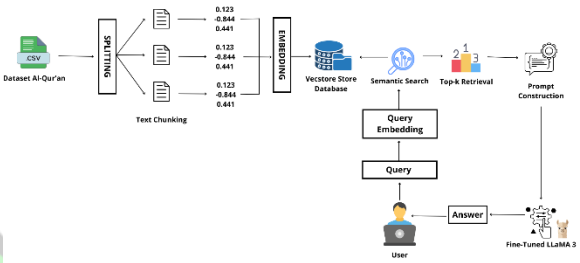
Tabel 4. Konfigurasi dan Hasil Pelatihan

Parameter	Nilai
Jumlah data latih	200
Epoch	2
Total step	100
Batch size	1
Gradient accumulation	4
Learning rate	2×10^{-4}
Optimizer	AdamW 8-bit
Final training loss	0,8805

Tabel 4 dapat disimpulkan bahwa model berhasil menyesuaikan parameter adaptif terhadap pola data tafsir yang digunakan dalam proses pelatihan.

4.4. Pengembangan Model RAG

Pengembangan model *Retrieval-Augmented Generation* (RAG) dilakukan dengan mengintegrasikan sistem pencarian semantik berbasis *vector store* dengan model LLaMA3 hasil *fine-tuning.* Pendekatan ini memungkinkan sistem menghasilkan jawaban yang kontekstual dan tetap terikat pada sumber ayat serta tafsir yang relevan.



Gambar 2. Arsitektur sistem RAG

Berdasarkan Gambar 2, sistem RAG diawali dengan pembentukan dokumen dari dataset yang telah melalui tahap *pre-processing.* Dokumen tersebut kemudian dipecah menjadi beberapa *chunk* agar sistem dapat memproses teks dalam bagian yang lebih kecil dan terfokus sebelum direpresentasikan dalam bentuk vector menggunakan model *embedding.* Seluruh representasi vector disimpan dalam basis data *Chroma* guna mendukung proses pencarian semantik berbasis kemiripan.

Pada saat pengguna memasukkan kueri, sistem terlebih dahulu melakukan normalisasi topik berdasarkan metadata *tafsir_theme_group.* Kueri kemudian diperkaya (*query enrichment*) agar selaras dengan konteks tematik yang terdeteksi. Selanjutnya, sistem melakukan *similarity search* pada basis data vector dengan mempertimbangkan filter topik, dilanjutkan dengan deduplikasi berdasarkan pasangan surah dan ayat untuk menghindari redundansi. Dokumen terpilih dikonstruksi menjadi konteks terstruktur yang memuat identitas ayat, terjemahan, dan tafsir, sebelum diberikan kepada model LLaMA3 hasil *fine-tuning* untuk menghasilkan makna tematik secara terkontrol dan berbasis sumber.

Tabel 5. Statistik dokumen chunking

Komponen	Nilai
Jumlah dokumen awal	6.236
Total *chunk*	15.945
Chunk size	400 karakter
Chunk overlap	80 karakter

Tabel 5 dapat dilihat bahwa proses *chunking* meningkatkan granularitas dokumen sehingga pencarian semantik dapat dilakukan secara lebih presisi pada level potongan teks, bukan hanya pada level ayat secara utuh.

Tabel 6. Konfigurasi *embedding* dan *vectorstore*

Komponen	Spesifikasi
Model *embedding*	all-MiniLM-L6-v2
Metode penyimpanan	*Chroma vector database*
Tipe pencarian	*Semantic similarity search*

Persist directory	Lokal

Tabel 6 dapat diketahui bahwa sistem menggunakan model *Sentence Transformers* untuk mengubah teks menjadi representasi vektor numerik yang kemudian disimpan dalam basis data Chroma guna mendukung pencarian berbasis kemiripan semantik.

Tabel 7. Statistik dan contoh topik

Komponen	Nilai
Jumlah topik unik	1.300 topik
Sumber topik	Metadata *tafsir_theme_group*
Contoh topik	Golongan Munafik, Golongan Kafir, Golongan Orang yang Bertakwa, Adab berpakaian, makan dan minum
Fungsi	Normalisasi dan filter *retrieval*

Tabel 7 terlihat bahwa sistem memanfaatkan klasifikasi tematik berbasis tafsir untuk mengarahkan proses *retrieval.* Normalisasi topik memungkinkan kueri pengguna dipetakan ke kategori tematik yang sesuai sehingga pencarian menjadi lebih berfokus dan relevan.

4.5. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja sistem pencarian ayat Al-Qur'an berbasis topik. Evaluasi dibagi menjadi dua aspek utama, yaitu evaluasi *retrieval* dan evaluasi *generative.*

1. Evaluasi *Retrieval*
   Evaluasi *retrieval* menggunakan metrik Recall@k dan *Mean Reciprocal Rank* (MRR). Recall@k mengukur proporsi dokumen relevan yang berhasil ditemukan dalam k hasil teratas, sedangkan MRR mengukur posisi kemunculan pertama dokumen relevan dalam daftar hasil *retrieval*. Pengujian dilakukan terhadap sepuluh kueri tematik yang merepresentasikan variasi topik dalam dataset.

Tabel 8. Hasil evaluasi *retrieval*

Kode	Ground truth	Retrieved	Recall@k	MRR
Q1	3	2	0,667	1.000
Q2	13	4	0,308	1.000
Q3	5	4	0,800	1.000
Q4	2	2	1.000	1.000
Q5	3	3	1.000	1.000
Q6	4	3	0.750	1.000
Q7	3	3	1.000	1.000
Q8	4	3	0.750	1.000
Q9	12	5	0.417	1.000
Q10	5	4	0.800	1.000
Rata-rata			0.749	1.000

Keterangan *query:*

- Q1: Adab berpakaian, makan dan minum
- Q2: Golongan munafik
- Q3: Golongan orang bertakwa
- Q4: Hukum khamar, dan berjudi
- Q5: Adab memasuki rumah orang lain
- Q6: Adab mendengarkan Al-Qur'an dan zikir
- Q7: Akibat kufur dan syirik
- Q8: Durhaka kepada ibu bapak
- Q9: Kisah Fir'aun
- Q10: Iman dan kafir

Tabel 8 dapat dilihat bahwa sistem secara konsisten menempatkan ayat relevan pada peringkat pertama hasil retrieval, yang ditunjukkan oleh nilai MRR sebesar 1.000 pada seluruh kueri. Evaluasi dilakukan menggunakan *Recall@k* (k=5) dengan nilai rata-rata sebesar 0.749, yang menunjukkan bahwa sebagian besar ayat relevan berhasil ditemukan dalam lima hasil teratas, meskipun pada beberapa topik dengan jumlah ground truth yang lebih besar, tidak seluruh ayat berhasil di retrieval.

2. Evaluasi *Generative*
   Evaluasi *generative* dilakukan untuk mengukur kesesuaian teks makna yang dihasilkan model terhadap tafsir referensi menggunakan metrik ROUGE, yang meliputi ROUGE-1, ROUGE-2, dan ROUGE-L. Pengujian dilakukan terhadap 36 pasangan teks hasil generasi dan teks referensi tafsir.

Tabel 9. Hasil evaluasi *generative*

Metrik	*Precision*	*Recall*	F1-*Score*
ROUGE-1	0.3343	0.1115	0.1551
ROUGE-2	0.1420	0.0669	0.0863
ROUGE-L	0.2749	0.0978	0.1338

Berdasarkan Tabel 9, nilai ROUGE-1 F1-score sebesar 0.1551 menunjukkan adanya tingkat kesamaan leksikal antara teks hasil generasi dan referensi tafsir, meskipun dalam skala moderat. Nilai recall yang relatif rendah mengindikasikan bahwa teks hasil generasi bersifat lebih ringkas dibandingkan teks tafsir referensi yang umumnya lebih panjang dan deskriptif. Namun demikian, precision yang lebih tinggi menunjukkan bahwa kata-kata yang dihasilkan model memiliki relevansi terhadap isi tafsir.

Berdasarkan hasil evaluasi, sistem yang diusulkan memperoleh nilai *Mean Recall*@k sebesar 0.749 dan MRR sebesar 1.000 yang menunjukkan bahwa sistem mampu menempatkan dokumen relevan pada peringkat teratas secara konsisten. Jika dibandingkan dengan penelitian Ramadhani *et al.*[14] yang melaporkan MRR sebesar 0,83 pada chatbot berbasis RAG, capaian penelitian ini menunjukkan bahwa mekanisme *retrieval* berbasis topik tafsir mampu menghasilkan presisi peringkat yang stabil. Pada aspek *generative,* sistem memperoleh ROUGE-1

F1 sebesar 0.1551, ROUGE-2 F1 sebesar 0.0863, dan ROUGE-L F1 sebesar 0.1338. Capaian tersebut lebih rendah dibandingkan Haq *et al.*[15] yang melaporkan ROUGE-L sebesar 0,251 pada sistem RAG *fine-tuning* Komodo-7B, namun lebih tinggi dibandingkan Eman *et al.*[16] yang memperoleh ROUGE-L sebesar 0,0680 pada chatbot nutrisi medis berbasis RAG. Variasi skor ini menunjukkan bahwa performa evaluasi berbasis ROUGE sangat dipengaruhi oleh karakteristik domain, panjang referensi, serta strategi konstruksi konteks. Dalam penelitian ini, keluaran model difokuskan pada perumusan makna tematik secara ringkas dalam satu kalimat, sehingga tingkat kesamaan leksikal terhadap teks tafsir yang panjang menjadi lebih terbatas.

5. KESIMPULAN

Berdasarkan hasil penelitian, sistem pencarian ayat Al-Qur'an berbasis topik yang dikembangkan dengan memanfaatkan model LLaMA3 yang di *fine-tuning* menggunakan QLoRA dan diintegrasikan dengan mekanisme *Retrieval-Augmented Generation* (RAG) mampu meningkatkan relevansi pencarian secara semantik dan kontekstual. Evaluasi *retrieval* menunjukkan nilai *Mean Recall@*k sebesar 0.749 dan MRR sebesar 1.000, yang mengindikasikan bahwa sistem secara konsisten menempatkan ayat relevan pada peringkat pertama hasil pencarian. Sementara itu, evaluasi *generative* menggunakan ROUGE menghasilkan nilai ROUGE-1 F1 sebesar 0.1551, ROUGE-2 F1 sebesar 0.0863, dan ROUGE-L F1 sebesar 0.1338, yang menunjukkan bahwa model mampu menghasilkan makna ayat yang selaras dengan tafsir referensi meskipun dalam bentuk yang lebih ringkas dan abstraktif.

DAFTAR PUSTAKA

[1] Heriah Fitria and Alwizar, "Kajian Pustaka tentang Isi dan Fungsi Al-Qur'an sebagai Pedoman Hidup Umat Islam," *Al-Zayn : Jurnal Ilmu Sosial & Hukum*, vol. 3, no. 2, pp. 1163–1172, Jun. 2025, doi: 10.61104/alz.v3i2.1240.

[2] A. Muid, B. Arifin, and A. Karim, "Peluang dan Tantangan Pendidikan Pesantren di Era Digital (Studi Kasus di Pondok Pesantren Al-Islah Bungah Gresik)," *Modeling: Jurnal Program Studi Pgmi*, vol. 11, no. 1, pp. 512–530, 2024.

[3] J. Basire and Nurdin, "Transformasi Kajian Islam di Era Digital: Peran Kecerdasan Buatan dalam Mendorong Pendidikan Islam yang Progresif dan Humanis," *Prosiding Kajian Islam dan Integrasi Ilmu di Era Society 5.0 (KIIIES 5.0)*, vol. 4, no. 1, pp. 36–41, 2025.

[4] L. Trisnawati, N. A. Binti Samsudin, S. K. Bin Ahmad Khalid, E. F. Bin Ahmad Shaubari, S. Sukri, and Z. Indra, "A proposed semantic keywords search engine for Indonesian Qur'an translation based on word embedding," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 35, no. 2, p. 987, Aug. 2024, doi: 10.11591/ijeecs.v35.i2.pp987-995.

[5] M. A. Hasbi, R. Imanda, and M. Fathan Fauzan, "Implementasi Chatbot Berbasis Large Language Model Untuk Pencarian Skripsi Mahasiswa Terintegrasi dengan Whatsapp," *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 5, no. 1, pp. 148–167, Jun. 2025, doi: 10.29240/arcitech.v5i1.13974.

[6] H. Rachmat, H. Riza, and T. F. Abidin, "Fine-Tuning Large Language Model (LLM) to Answer Basic Questions for Prospective New Students at Syiah Kuala University Using the Retrieval-Augmented Generation (RAG) Method," in *2024 Ninth International Conference on Informatics and Computing (ICIC)*, Medan, Indonesia: IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/ICIC64337.2024.10956296.

[7] M. I. Syah, N. S. Harahap, Novriyanto, and S. Sanjaya, "Penerapan Retrieval Augemented Generation Menggunakan Langchain Dalam Pengembangan Sistem Tanya Jawab Hadis Berbasis Web," *ZONAsi: Jurnal Sistem Informasi*, vol. 6, no. 2, pp. 370–379, May 2024, doi: 10.31849/zn.v6i2.19940.

[8] Moh. Abd. A. Hidayat and T. Hidayat, "Implementasi Pencarian Semantik dalam Tafsir Al-Quran dengan Algoritma Cosine Similarity dan Large Language Models," Other thesis, Nusa Putra University, 2024. [Online]. Available: https://repository.nusaputra.ac.id/id/eprint/1329/

[9] M. A. K. Raiaan *et al.*, "A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges," *IEEE Access*, vol. 12, pp. 26839–26874, 2024, doi: 10.1109/ACCESS.2024.3365742.

[10] R. K. Nasirin and R. D. Indahsari, "Model Penilaian Jawaban Esai Berbasis Semantic Understanding Menggunakan LLaMa dan Text Similarity," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 11, no. 2, Aug. 2025, [Online]. Available: https://jurnal.untan.ac.id/index.php/jepin/article/view/92529

[11] H. Rajabzadeh *et al.*, "QDyLoRA: Quantized Dynamic Low-Rank Adaptation for Efficient Large Language Model Tuning," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, Miami, Florida, US: Association for Computational Linguistics, Nov. 2024, pp. 712–718. doi: 10.18653/v1/2024.emnlp-industry.53.

[12] I. G. N. A. Jayarana, I. G. W. Darma, I. W. A. Juliantara, and I. M. A. W. Putra, "Study Literatur Information Retrieval Model: Teknik Dan Aplikasi," *Jurnal Sutasoma (Science Teknologi Sosial Humaniora)*, vol. 3, no. 2, pp. 61–69, Jun. 2025, doi: 10.58878/sutasoma.v3i2.392.

[13] T. I. Ramadhan, A. Supriatman, and T. R. Kurniawan, "Evaluasi dan Implementasi Indobert

Question Answering (QA) pada Domain Spesifik Menggunakan Mean Reciprocal Rank," *Jurnal Algoritma*, vol. 21, no. 1, pp. 180–188, May 2024, doi: 10.33364/algoritma/v.21-1.1542.

[14] T. Ramadhani, N. Q. Nada, and N. D. S, "Penerapan Metode Retrieval-Augmented Generation (RAG) Pada Chatbot E-Commerce Berbasis Gemini Ai," *Jurnal Ilmiah ILKOMINFO – Ilmu Komputer & Informatika*, vol. 8, no. 2, pp. 310–311, Jul. 2025.

[15] A. A. Haq, "Sistem Tanya Jawab Layanan Administrasi Kependudukan dengan Retrieval Augmented Generation Komodo-7B," *Journal of Computer and Information System (J-CIS)*, vol. 8, no. 2, pp. 85–96, Oct. 2025, doi: https://doi.org/10.31605/jcis.v8i2.

[16] E. T. Eman, T. N. Fatyanosa, and A. F. Aji, "Analisis Perbandingan Metode Chunking dalam Chatbot Berbasis Retrieval-Augmented Generation Rekomendasi Terapi Nutrisi Medis Pasien," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 1, Jan. 2026, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15948