

Penerapan Model AI Gemini pada Pengembangan Chatbot Tanya Fiqih Haji dan Umrah

Ahmad Farisi^{1*}, Chriestina², Dafid³, Adit Jansa⁴

^{1*,2,3,4}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Palembang, Indonesia

Email: ^{1*}ahmadfarisi@mdp.ac.id, ²chriestina@mdp.ac.id, ³dafid@mdp.ac.id,

⁴aditjansa_2226240064@mhs.mdp.ac.id

Abstract

This study aims to develop a hybrid mobile application prototype called “Tanya Fiqih Haji dan Umrah” as a supportive consultation medium for Islamic jurisprudence (fiqh) related to Hajj and Umrah. The application utilizes the Gemini 2.5 Flash model with a Retrieval-Augmented Generation (RAG) approach that is restricted to two official references issued by the Indonesian Ministry of Religious Affairs, namely the Hajj Manasik Infographic Book 2025 and the Hajj Supplications and Dhikr Guide 2025. The system is implemented using a modern web-to-mobile hybrid architecture based on React, TypeScript, Tailwind CSS, and Capacitor, enabling deployment as a lightweight Android application with local conversation history storage. System evaluation was conducted through expert validation on ten test questions representing the core themes of Hajj and Umrah fiqh. The results show that 70% of AI responses were categorized as fully aligned with the references, 10% as partially accurate, and 20% as not aligned. These findings indicate that the RAG mechanism helps maintain consistency between AI-generated answers and authoritative literature, potentially reducing hallucination in generative models, although expert clarification remains necessary in certain cases. This study contributes to the application of RAG-based large language models in the domain of Hajj fiqh and offers a technological support tool that may benefit pilgrims, particularly independent pilgrims.

Keywords: Chatbot, Expert Validation, Gemini 2.5 Flash, Hajj and Umrah Fiqh, Retrieval-Augmented Generation.

Abstrak

Penelitian ini bertujuan mengembangkan prototipe aplikasi mobile hybrid Tanya Fiqih Haji dan Umrah sebagai media bantu konsultasi fiqh berbasis kecerdasan buatan. Aplikasi ini memanfaatkan model Gemini 2.5 Flash dengan pendekatan Retrieval-Augmented Generation (RAG) yang dikunci pada dua rujukan resmi Kementerian Agama RI, yaitu Buku Infografis Manasik Haji 2025 dan Buku Panduan Manasik Haji (Doa & Zikir) 2025. Sistem dibangun menggunakan arsitektur modern web-to-mobile hybrid berbasis React, TypeScript, Tailwind CSS, dan Capacitor, sehingga dapat dijalankan sebagai aplikasi Android yang ringan serta menyimpan riwayat percakapan secara lokal. Evaluasi sistem dilakukan melalui validasi pakar terhadap sepuluh pertanyaan uji yang merepresentasikan materi inti fiqh haji dan umrah. Hasil validasi menunjukkan bahwa 70% jawaban berada pada kategori sesuai rujukan, 10% kurang tepat, dan 20% tidak sesuai rujukan. Temuan ini mengindikasikan bahwa mekanisme RAG membantu menjaga keselarasan jawaban AI dengan literatur otoritatif dan berpotensi mengurangi hallucination pada model generatif, meskipun pada beberapa kasus tetap diperlukan klarifikasi pakar. Penelitian ini berkontribusi pada pengembangan penerapan LLM berbasis RAG dalam domain fiqh haji serta menghadirkan solusi teknologi yang dapat mendukung kebutuhan bimbingan jamaah, khususnya jamaah mandiri.

Kata Kunci: Chatbot, Fiqih Haji dan Umrah, Gemini 2.5 Flash, Retrieval-Augmented Generation, Validasi Pakar.

1. PENDAHULUAN

Transformasi digital dalam sektor pelayanan ibadah haji dan umrah terus berkembang seiring dengan meningkatnya pemanfaatan teknologi mobile, aplikasi berbasis web, dan kecerdasan buatan. Berbagai inovasi sistem informasi telah dikembangkan untuk mendukung proses manasik, administrasi, serta layanan bimbingan jemaah, baik pada tingkat pemerintah maupun lembaga penyelenggara bimbingan ibadah. Meskipun demikian, sejumlah penelitian terkini menunjukkan bahwa permasalahan pada aspek pemahaman fiqih haji masih menjadi tantangan utama di lapangan.

Hasil penelitian (Salim & Farisi, 2024) menunjukkan bahwa jemaah masih menghadapi berbagai kendala selama pelaksanaan ibadah di Tanah Suci, termasuk keterbatasan akses media konsultasi dan minimnya sarana pembelajaran praktis terkait fiqih manasik. Meskipun keberadaan pembimbing KBIHU membantu sebagian permasalahan, kasus kebingungan dalam pengambilan keputusan fiqih tetap ditemukan, terutama pada jemaah yang melaksanakan ibadah secara mandiri dan tidak selalu memiliki akses langsung kepada pembimbing.

Sejumlah penelitian telah mengembangkan aplikasi manasik haji berbasis mobile maupun web sebagai media pembelajaran digital. (Helviansyah, Harahap, Irsyad, & Negara, 2025) membangun sistem tanya jawab fiqih kontemporer berbasis chatbot website menggunakan Gemini, sedangkan (Hardiyanti & Syarifah Sany, 2025) mengembangkan chatbot konsultasi haji dan umrah yang berfokus pada pemberian informasi umum dan administratif. Meskipun demikian, sebagian besar penelitian tersebut masih menekankan aspek penyajian informasi atau dialog generik, belum secara khusus mengevaluasi kesesuaian jawaban chatbot terhadap literatur fiqih yang menjadi rujukan resmi Kementerian Agama serta belum membahas secara eksplisit masalah hallucination pada model bahasa besar.

Di sisi lain, penggunaan LLM generik seperti GPT sering menunjukkan ketidakkonsistenan dalam menjawab pertanyaan fiqih, khususnya ketika model menarik rujukan dari sumber global yang tidak selalu sejalan dengan praktik manasik di Indonesia. Beberapa studi komparatif menunjukkan bahwa Gemini menawarkan kombinasi yang seimbang antara kualitas analisis dan kecepatan respons dibandingkan model lain, sehingga cocok untuk skenario aplikasi real-time dan berbiaya rendah (Sun dkk., 2025). Keunggulan lain Gemini adalah dukungan long context hingga jutaan token yang memungkinkan integrasi dokumen panjang secara utuh ke dalam konteks pemodelan, dengan tingkat akurasi pengambilan konteks yang sangat tinggi (>99%) pada skala jutaan token (Google, 2025a). Hal ini menjadikannya relevan untuk skenario *Long-Context Retrieval-Augmented Generation* (RAG) yang membutuhkan pembacaan keseluruhan isi buku manasik, bukan sekadar potongan teks. Seperti misalnya Buku Infografis Manasik Haji (Khoiri dkk., 2025) dan Buku Panduan Manasik Haji dan Umrah (Ditjen Penyelenggaraan Haji dan Umrah Kementerian Agama RI, 2025) yang dikeluarkan secara resmi oleh Kemenag RI pada Tahun 2025.

Berdasarkan hal tersebut, masih terdapat research gap pada pengembangan chatbot fiqih haji yang (1) secara eksplisit mengendalikan sumber rujukan melalui mekanisme RAG, (2) memanfaatkan kemampuan long context Gemini untuk menginjeksikan isi buku manasik secara utuh, dan (3) mengevaluasi kesesuaian jawaban AI terhadap buku manasik resmi melalui validasi pakar.

Dengan demikian, penelitian ini memberikan dua kontribusi utama. Pertama, dari sisi teknis, penelitian ini mengusulkan arsitektur chatbot fiqih berbasis Long-Context RAG yang memanfaatkan model Gemini 2.5 Flash dengan strategi full-context injection terhadap dua buku manasik resmi Kementerian Agama RI Tahun 2025. Pendekatan ini

menggabungkan keunggulan *long context* Gemini dengan arsitektur aplikasi web-to-mobile hybrid yang ringan dan offline-tolerant. Kedua, dari sisi evaluasi, penelitian ini menyajikan analisis validasi pakar dan analisis kegagalan (*error analysis*) terhadap jawaban AI pada topik fiqh haji tertentu, sehingga memberikan pemahaman yang lebih mendalam mengenai perilaku hallucination dan batas efektivitas pendekatan Long-Context RAG pada domain fiqh haji.

2. METODOLOGI PENELITIAN

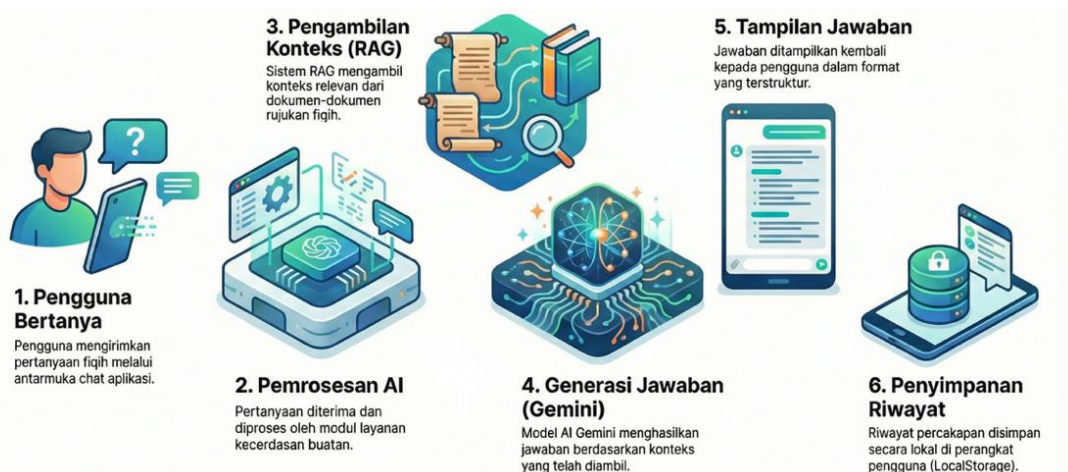
2.1 Desain Penelitian

Penelitian ini termasuk dalam kategori penelitian pengembangan (*research and development*) yang berfokus pada perancangan, implementasi, dan evaluasi prototipe sistem chatbot fiqh haji dan umrah berbasis model AI Gemini dengan pendekatan *Retrieval-Augmented Generation* (RAG). Pendekatan penelitian ini merujuk pada paradigma *Design Science Research* (DSR), yaitu pendekatan ilmiah yang berorientasi pada pengembangan artefak teknologi sebagai solusi terhadap permasalahan nyata, sekaligus memungkinkan evaluasi sistematis terhadap artefak yang dibangun (Hevner, March, Park, & Ram, 2004; Wieringa, 2014).

Dalam konteks penelitian ini, artefak yang dikembangkan berupa: (1) Aplikasi *mobile hybrid* “Tanya Fiqh Haji dan Umrah”, dan (2) Modul RAG yang membatasi jawaban AI hanya pada dua sumber rujukan resmi Kementerian Agama RI tahun 2025. Pemilihan pendekatan DSR dipertimbangkan karena cocok digunakan pada penelitian yang bertujuan menghasilkan solusi teknologi kontekstual serta menilai kegunaan dan kualitasnya melalui proses evaluasi terstruktur (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007).

2.2 Arsitektur Umum Sistem

Aplikasi dikembangkan menggunakan pendekatan *Modern Web-to-Mobile Hybrid*, di mana antarmuka dibangun dengan React dan TypeScript, kemudian dibungkus menjadi aplikasi Android menggunakan Capacitor. Pendekatan *hybrid* dipilih karena memungkinkan aplikasi tetap ringan, mudah dipelihara, dan dapat berjalan baik pada perangkat mobile dengan kebutuhan sumber daya yang terbatas (Moroney, 2017).



Gambar 1. Alur Kerja Sistem
Sumber: (NotebookLM, 2025)

Gambar 1 menjelaskan alur kerja sistem secara umum terdiri dari: (1) Pengguna mengirimkan pertanyaan fiqh melalui antarmuka chat, (2) Pertanyaan diproses pada modul layanan AI, (3) Sistem menerapkan mekanisme RAG untuk mengambil konteks

dari dokumen rujukan, (4) Model Gemini menghasilkan jawaban berdasarkan konteks tersebut, (5) Jawaban ditampilkan kembali dalam format yang terstruktur pada aplikasi, dan (6) Riwayat percakapan disimpan secara lokal melalui *LocalStorage*.

Arsitektur ini dirancang berdasarkan prinsip *lightweight*, *offline-tolerant*, dan *context-aware*, sehingga tetap dapat digunakan dalam kondisi lapangan dengan keterbatasan jaringan internet.

2.3 Spesifikasi RAG

Pada penelitian ini digunakan pendekatan Long-Context RAG (full-context injection). Berbeda dengan RAG standar berbasis vector database yang memerlukan proses chunking, pembuatan embedding, dan pencarian vektor, pendekatan ini memanfaatkan kemampuan *long context* model Gemini untuk membaca seluruh isi dokumen rujukan sekaligus. Gemini 1.5 dan generasi berikutnya mendukung context window hingga jutaan token, sehingga mampu memproses dokumen panjang dengan tingkat akurasi pengambilan konteks yang sangat tinggi (Google, 2025b).

Langkah-langkah implementasi RAG dalam penelitian ini adalah (1) Retrieval. Isi Buku Infografis Manasik Haji 2025 dan Buku Panduan Manasik Haji dan Umrah (Doa & Zikir) 2025 diekstraksi dan dibersihkan, kemudian disimpan sebagai konstanta `HAJJ_CONTEXT` pada berkas `constants.ts` di sisi frontend, (2) Augmentation (full-context injection). Variabel `HAJJ_CONTEXT` dimasukkan secara langsung ke dalam `systemInstruction` saat pemanggilan Gemini API. Dengan cara ini, seluruh isi buku manasik tersedia sebagai konteks global setiap kali pengguna mengajukan pertanyaan. Tidak digunakan proses embedding maupun vector store karena ukuran teks masih jauh di bawah kapasitas context window Gemini, dan (3) Generation. Gemini 2.5 Flash menyusun jawaban berdasarkan kombinasi antara pertanyaan pengguna dan konten `HAJJ_CONTEXT`. Prompt instruction dirancang agar model (a) hanya menjawab berdasarkan isi buku rujukan, (b) menyebutkan apabila pertanyaan di luar cakupan literatur, dan (c) menghindari spekulasi yang berpotensi menimbulkan hallucination.

Dengan demikian, spesifikasi RAG pada penelitian ini secara eksplisit tidak menggunakan model embedding maupun vector database seperti Pinecone atau FAISS, melainkan mengandalkan strategi full-context grounding yang lebih sederhana dan efisien untuk skala dokumen yang relatif kecil.

2.4 Prosedur Pengembangan & Integrasi Gemini Capacitor

Prosedur pengembangan disusun melalui beberapa tahapan analisis kebutuhan, perancangan arsitektur sistem, implementasi aplikasi, integrasi RAG, hingga pembuatan paket aplikasi *mobile* (APK). Prosedur ini terdapat pada Gambar 2.



Gambar 2. Prosedur Pengembangan Aplikasi
Sumber: (NotebookLM, 2025)

Analisis kebutuhan dilakukan melalui kajian literatur dan masukan pakar dari Forum Komunikasi Kelompok Bimbingan Ibadah Haji dan Umrah (FK-KBIHU) Sumatera Selatan mengenai kebutuhan bimbingan fiqih jemaah haji. Perancangan arsitektur sistem yang dilakukan meliputi desain modul *frontend*, modul layanan AI, serta skema RAG. Implementasi aplikasi dilakukan menggunakan React, TypeScript, Tailwind CSS, Gemini API, dan Capacitor. Integrasi RAG dan pengujian konteks sumber dilakukan untuk memastikan model hanya menjawab berdasarkan dokumen rujukan. Dan Pembuatan paket aplikasi *mobile* (APK) dilakukan untuk uji coba penggunaan pada perangkat Android.

Pendekatan pengembangan dilakukan secara iteratif mengikuti prinsip *agile prototyping*, di mana pengujian internal dilakukan pada setiap siklus pengembangan untuk menjaga stabilitas fungsi sistem (Pressman & Maxim, 2020).

Integrasi antara Gemini API dan aplikasi *mobile* dilakukan sepenuhnya di lapisan *frontend*. Komponen React yang berjalan di dalam WebView Capacitor melakukan pemanggilan HTTP ke endpoint Gemini menggunakan *fetch*/Axios dengan menyertakan kunci API yang disimpan secara aman di *environment variable* aplikasi. Respons dari Gemini kemudian diolah menjadi objek Message dan disimpan ke dalam state React serta LocalStorage. Capacitor berperan sebagai jembatan yang mengemas aplikasi web tersebut menjadi APK Android tanpa logika tambahan di sisi native, sehingga arsitektur tetap sederhana namun mampu memanfaatkan fitur perangkat *mobile* seperti penyimpanan lokal dan pengelolaan jaringan.

2.5 Metode Evaluasi Akurasi Jawaban AI

Evaluasi akurasi jawaban dilakukan melalui validasi pakar fiqih / pembimbing KBIHU, dengan metode *expert judgement*. Setiap jawaban AI dinilai berdasarkan tiga kategori, yaitu (1) Sesuai dengan isi buku rujukan, (2) Kurang tepat / perlu klarifikasi, atau (3) Tidak sesuai dengan rujukan.

Pendekatan validasi pakar umum digunakan pada penelitian berbasis konten keagamaan untuk memastikan keselarasan jawaban dengan sumber otoritatif (Cresswell & Cresswell, 2023). Instrumen penilaian akurasi jawaban AI disusun dalam bentuk lembar evaluasi isi, butir penilaian kualitas jawaban, dan catatan koreksi pakar. Data validasi pakar fiqih dianalisis menggunakan persentase tingkat kesesuaian jawaban terhadap sumber rujukan.

Secara formal, tingkat akurasi jawaban AI dihitung berdasarkan proporsi jawaban yang dinilai “sesuai rujukan” terhadap seluruh item pertanyaan uji. Notasi perhitungannya terdapat pada persamaan (1).

$$Akurasi = \frac{N_{sesuai}}{N_{total}} \times 100\% \quad (1)$$

dengan N sesuai adalah jumlah jawaban yang dinilai sesuai dengan buku rujukan dan N total adalah jumlah seluruh pertanyaan uji. Selain itu, distribusi kategori kurang tepat dan tidak sesuai juga dianalisis secara deskriptif untuk mengidentifikasi pola kegagalan sistem.

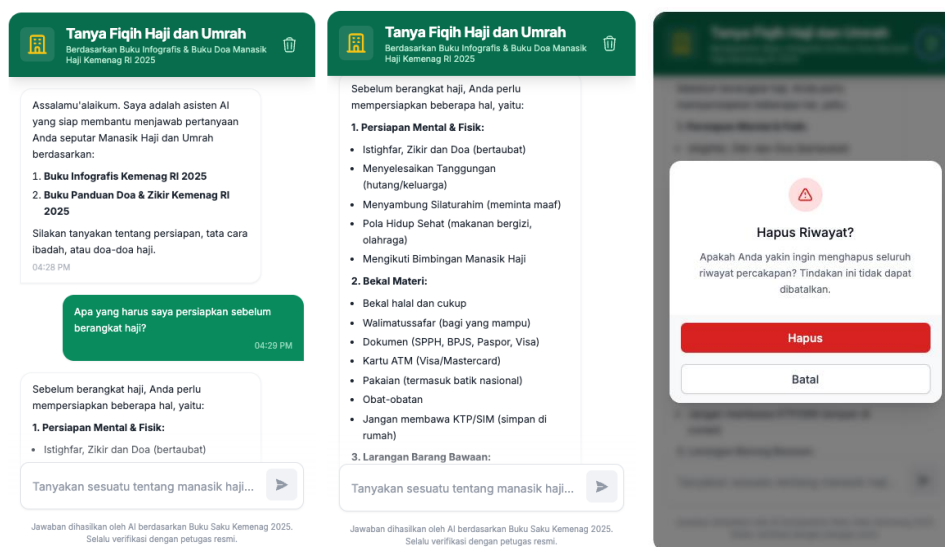
3. HASIL DAN PEMBAHASAN

3.1. Hasil Implementasi Sistem

Hasil pengembangan menunjukkan bahwa prototipe aplikasi *mobile hybrid* “Tanya Fiqih Haji dan Umrah” berhasil diimplementasikan sesuai rancangan arsitektur yang telah dijelaskan pada bagian metodologi. Aplikasi dibangun menggunakan React dan TypeScript, sementara proses pembungkusan ke format APK dilakukan melalui Capacitor sehingga aplikasi dapat dijalankan pada perangkat Android.

Antarmuka aplikasi dirancang dalam bentuk tampilan percakapan (*chat interface*) yang sederhana dan mudah dibaca. Pengguna dapat mengirimkan pertanyaan terkait fiqih haji dan umrah dalam bahasa Indonesia, kemudian sistem memproses pesan tersebut melalui modul layanan chatbot dan modul *Retrieval-Augmented Generation (RAG)*. Jawaban yang dihasilkan oleh model Gemini 2.5 Flash ditampilkan dalam format teks terstruktur menggunakan *React Markdown*, sehingga poin-poin hukum fiqih, langkah ibadah, maupun doa tetap terbaca dengan jelas.

Riwayat percakapan disimpan secara lokal melalui Web Storage API (*LocalStorage*), sehingga pengguna masih dapat mengakses percakapan sebelumnya meskipun aplikasi ditutup dan dibuka kembali. Pendekatan ini sejalan dengan prinsip desain aplikasi ringan (*lightweight system*), karena tidak memerlukan server basis data eksternal dan tetap memungkinkan pengalaman penggunaan yang cepat pada kondisi jaringan terbatas.



Gambar 3. Aplikasi Tanya Fiqih Haji dan Umrah

Secara umum, hasil implementasi menunjukkan bahwa: (1) Aplikasi dapat berjalan stabil sebagai aplikasi Android, (2) Waktu respons chatbot relatif cepat karena pemanfaatan model Gemini 2.5 Flash, dan (3) Sistem mampu menampilkan jawaban fiqih dalam format yang mudah dipahami oleh pengguna.

Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa pendekatan aplikasi mobile berbasis AI dapat meningkatkan aksesibilitas materi keagamaan dan bimbingan ibadah secara interaktif, sepanjang sistem dirancang dengan struktur antarmuka yang sederhana dan mudah digunakan (Hardiyanti & Syarifah Sany, 2025; Helviansyah dkk., 2025).

3.2. Hasil Evaluasi Akurasi Jawaban AI (Validasi Pakar)

Evaluasi akurasi jawaban dilakukan melalui validasi pakar yang berasal dari pembimbing KBIHU / praktisi fiqih haji. Sejumlah pertanyaan uji disusun berdasarkan skenario ibadah yang umum ditemui jemaah, misalnya terkait rukun, wajib, larangan ihram, pelaksanaan thawaf, sa'i, tahallul, serta doa-doa pada tahapan manasik.

Setiap jawaban AI dibandingkan dengan *isi* Buku Infografis Manasik Haji Kemenag RI 2025, dan Buku Doa & Zikir Manasik Haji Kemenag RI 2025. Pakar kemudian memberikan penilaian dalam tiga kategori: (1) Sesuai sumber rujukan, (2) Kurang tepat / memerlukan klarifikasi, dan (3) Tidak sesuai dengan sumber rujukan.

Tabel 1. Hasil Evaluasi Akurasi Jawaban AI (Validasi Pakar)

No	Topik Pertanyaan	Contoh Pertanyaan Uji	Hasil Jawaban AI	Penilaian Pakar	Catatan Validasi
1	Niat Ihram	Kapan niat ihram untuk umrah dibaca?	Dibaca di miqat	Sesuai rujukan	Sudah sesuai teks buku
2	Larangan Ihram	Apakah boleh memakai wewangian dalam keadaan ihram?	Tidak boleh, karena termasuk larangan ihram	Sesuai rujukan	Redaksi jelas
3	Thawaf	Apakah thawaf dimulai dari Hajar Aswad?	Ya, Thawaf dimulai & diakhiri di Hajar Aswad	Sesuai rujukan	Lengkap
4	Thawaf	Apakah boleh berhenti thawaf karena padat?	Jawaban tidak ditemukan dalam rujukan.	Tidak sesuai	Rujukan menjelaskan boleh, tapi tetap menjaga urutan putaran
5	Sa'i	Apakah sa'i wajib tujuh kali lintasan?	Ya, tujuh lintasan dari Shafa ke Marwah	Sesuai rujukan	Valid
6	Sa'i	Apakah perempuan ikut berlari kecil di jalur hijau?	Tidak disunnahkan untuk perempuan	Sesuai rujukan	Tepat
7	Tahallul	Berapa minimal rambut yang dipotong saat tahallul?	Cukup sebagian rambut	Kurang tepat	Buku menegaskan minimal tiga helai
8	Umrah	Sebutkan rukun umrah	Ihram, thawaf, sa'i, tahallul	Sesuai rujukan	Lengkap
9	Wukuf	Apa hukum meninggalkan wukuf di Arafah?	Haji tidak sah karena Wukuf merupakan rukun haji	Sesuai rujukan	Tepat
10	Melempar Jumrah	Apakah boleh diwakilkan bagi lansia?	Jawaban tidak ditemukan dalam rujukan.	Tidak sesuai	Rujukan membolehkan

Berdasarkan hasil evaluasi pada 10 pertanyaan uji, diperoleh bahwa 7 jawaban (70%) berada pada kategori *sesuai rujukan*, 1 jawaban (10%) berada pada kategori *kurang tepat*, dan 2 jawaban (20%) berada pada kategori *tidak sesuai rujukan*. Jawaban pada kategori *kurang tepat* umumnya terkait aspek penekanan redaksi fiqih yang memerlukan klarifikasi tambahan dari pembimbing. Sementara itu, ketidaksesuaian jawaban ditemukan pada topik thawaf dan wakalah pelemparan jumrah, di mana AI belum merujuk secara eksplisit pada penjelasan yang terdapat dalam buku manasik. Secara umum, hasil ini menunjukkan bahwa mekanisme RAG mampu menjaga konsistensi jawaban AI terhadap sumber rujukan resmi, meskipun pada beberapa kasus tetap dibutuhkan peran validasi pakar.

Perlu dicatat bahwa jumlah pertanyaan uji pada penelitian ini masih terbatas pada 10 skenario yang dipilih secara purposif untuk merepresentasikan rukun, wajib, larangan ihram, serta kasus praktis yang sering ditanyakan jemaah. Oleh karena itu, hasil yang diperoleh belum dapat digeneralisasikan secara statistik, melainkan lebih bersifat studi awal (*pilot study*) untuk memotret perilaku sistem pada topik-topik kunci. Oleh karena itu, temuan pada penelitian ini diposisikan sebagai bukti empiris awal (*early empirical evidence*), bukan generalisasi populasi. Penelitian lanjutan direncanakan untuk

memperluas himpunan pertanyaan uji menjadi 30–50 skenario atau lebih sehingga memungkinkan analisis kuantitatif yang lebih kuat dan pengujian yang lebih menyeluruh terhadap variasi kasus fiqih haji.

3.3. Analisis Kegagalan dan Hallucination

Analisis lebih mendalam terhadap dua jawaban yang dikategorikan tidak sesuai rujukan menunjukkan pola kegagalan yang menarik. Pada topik thawaf, model menjawab bahwa “jawaban tidak ditemukan dalam rujukan”, padahal Buku Infografis Manasik Haji menjelaskan bahwa jamaah boleh berhenti sejenak ketika kondisi sangat padat selama urutan putaran tetap terjaga dan jumlah putaran tidak kurang dari tujuh. Pada kasus ini, model gagal “menemukan” penjelasan yang sebenarnya telah ada dalam konteks HAJJ_CONTEXT. Hal ini mengindikasikan bahwa kegagalan bukan disebabkan oleh ketiadaan data, tetapi oleh mekanisme attention model terhadap bagian teks tertentu serta perancangan prompt yang belum cukup menonjolkan aturan thawaf dalam situasi padat.

Kasus kedua terjadi pada topik wakalah pelemparan jumrah bagi jamaah lansia. Model menjawab bahwa hal tersebut tidak ditemukan dalam rujukan, sementara buku manasik secara eksplisit membolehkan pelemparan jumrah diwakilkan bagi jamaah yang uzur, sakit, atau lemah, dengan syarat wakilnya telah melempar untuk dirinya sendiri. Jawaban model dalam kedua kasus ini dapat dikategorikan sebagai bentuk hallucination, yaitu ketika model menyatakan ketiadaan informasi atau memberikan kesimpulan yang tidak selaras dengan teks rujukan yang sebenarnya tersedia.

Dari perspektif teknik, terdapat beberapa kemungkinan penyebab kegagalan. Pertama, meskipun pendekatan Long-Context RAG memberikan akses terhadap keseluruhan isi buku, model tetap harus melakukan pemilihan konteks internal melalui mekanisme self-attention. Jika bagian teks terkait thawaf padat atau wakalah jumrah kurang terekspos oleh struktur paragraf atau kata kunci pada pertanyaan, model dapat salah mengestimasi relevansi konteks. Kedua, prompt instruction pada penelitian ini lebih menekankan pembatasan domain jawaban (“jawab hanya berdasarkan buku”), tetapi belum secara eksplisit menginstruksikan model untuk menghindari klaim “tidak ada di rujukan” kecuali setelah melakukan pemindaian menyeluruh. Hal ini membuka ruang bagi model untuk menghasilkan jawaban yang secara linguistik meyakinkan tetapi tidak sepenuhnya faktual, sebagaimana dipaparkan dalam literatur mengenai hallucination pada LLM (Gemini Team dkk., 2024).

Untuk memitigasi hallucination, penelitian ini merancang prompt yang memuat instruksi eksplisit agar model (1) mengacu pada kutipan teks dari HAJJ_CONTEXT ketika menjawab, dan (2) menyatakan secara terbuka apabila pertanyaan tidak tercakup dalam buku. Strategi ini terbukti efektif pada mayoritas pertanyaan uji, sebagaimana tercermin dari 70% jawaban yang sesuai rujukan. Namun, dua kasus kegagalan menunjukkan bahwa instruksi tersebut belum sepenuhnya cukup: model cenderung “menggeneralisasi” perintah kedua dan terlalu mudah menyimpulkan bahwa informasi tidak tersedia. Ke depan, prompt engineering dapat diperkuat dengan menambahkan instruksi untuk melakukan parafrase langsung terhadap bagian teks yang relevan (misalnya, dengan format kutipan atau ringkasan terarah) sehingga peluang hallucination berupa klaim ketiadaan informasi dapat semakin ditekan.

4. KESIMPULAN

Penelitian ini menghasilkan prototipe aplikasi mobile hybrid “Tanya Fiqih Haji dan Umrah” yang memanfaatkan model Gemini 2.5 Flash dengan pendekatan Long-Context RAG melalui strategi full-context injection terhadap dua sumber rujukan resmi, yaitu

Buku Infografis Manasik Haji 2025 dan Buku Doa & Zikir Manasik Haji 2025. Pendekatan ini memungkinkan model memproses isi dokumen secara utuh tanpa mekanisme vector database, sehingga lebih efisien, ringan, dan sesuai dengan skala korpus yang terbatas pada domain fiqih haji.

Validasi pakar terhadap sepuluh skenario pertanyaan uji menunjukkan bahwa sebagian besar jawaban AI berada pada kategori sesuai rujukan, sementara sebagian kecil kasus masuk kategori kurang tepat dan tidak sesuai. Analisis kegagalan menunjukkan bahwa ketidaktepatan jawaban terutama terjadi pada topik tertentu, seperti thawaf pada kondisi padat dan wakalah pelemparan jumrah, yang mengindikasikan fenomena *false-negative grounding response* pada mekanisme Long-Context RAG, di mana model menyimpulkan bahwa informasi tidak tersedia meskipun sebenarnya terdapat dalam teks rujukan. Temuan ini memberikan pemahaman empiris mengenai dinamika hallucination pada sistem RAG berbasis full-context injection, serta menegaskan pentingnya perancangan prompt yang lebih ketat dan eksplisit untuk meminimalkan bias inferensi model.

Dengan demikian, kontribusi utama penelitian ini terletak pada (1) penerapan dan formulasi arsitektur Long-Context RAG berbasis Gemini 2.5 Flash pada domain fiqih haji, dan (2) penyajian analisis validasi pakar dan *error analysis* yang memberikan perspektif empiris terhadap akurasi jawaban dan pola kegagalan model pada konteks pembatasan sumber rujukan. Adapun keterbatasan penelitian ini terletak pada jumlah skenario uji yang masih terbatas, sehingga penelitian lanjutan akan diarahkan pada perluasan himpunan pertanyaan (≥ 30 –50 item), penguatan instrumen validasi pakar, serta eksplorasi strategi prompt refinement untuk meningkatkan stabilitas *grounded response* pada berbagai variasi kasus fiqih haji.

REFERENCES

- Cresswell, J. W., & Cresswell, J. D. (2023). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches Sixth Edition* (6 ed.). Los Angeles: SAGE Publications, Inc.
- Ditjen Penyelenggaraan Haji dan Umrah Kementerian Agama RI. (2025). *Panduan Manasik Haji dan Umrah (Doa & Zikir)*. Jakarta: Dirjen PHU Kementerian Agama Republik Indonesia.
- Gemini Team, Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., ... Vinyals, O. (2024). *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context*. Diambil dari <http://arxiv.org/abs/2403.05530>
- Google. (2025a). Long context | Gemini API | google AI for developers. Diambil 29 Desember 2025, dari Gemini API Documentation website: <https://ai.google.dev/gemini-api/docs/long-context>
- Google. (2025b, Desember 18). Long context | Generative AI on Vertex AI | Google Cloud Documentation. Diambil 29 Desember 2025, dari Google Cloud Documentation website: <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/long-context>
- Hardiyanti, R., & Syarifah Sany, D. (2025). Pengembangan Chatbot Konsultasi Haji dan Umrah Berbasis RAG dan Prompt Engineering: Studi Kasus Fillah Guide Informasi Artikel Abstrak. *Prosiding Seminar Nasional Teknik Universitas Suryakancana*, 329–337. Cianjur.
- Helviansyah, T., Harahap, N. S., Irsyad, M., & Negara, B. S. (2025). SISTEM TANYA JAWAB BERBASIS CHATBOT WEBSITE MENGGUNAKAN GEMINI AI PADA DATA FIQIH KONTEMPORER. *Journal of Information System Management (JOISM)*, 7(1), 38–47. <https://doi.org/10.24076/joism.2025v7i1.2082>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75–106. <https://doi.org/10.2307/25148625>

- Khoiri, I., Ulinnuha, M., Hafizh, E., Irawan, H., Anggoro, H., Jimun, M., ... Desfianti, W. G. (2025). *Infografis Tuntunan Manasik Haji & Umrah* (Vol. 4509178). Jakarta: Direktorat Bina Haji Dirjen PHU Kementerian Agama Republik Indonesia.
- Moroney, L. (2017). The Definitive Guide to Firebase. Dalam *The Definitive Guide to Firebase: Build Android Apps on Google's Mobile Platform*. Berkeley, CA: Apress. <https://doi.org/10.1007/978-1-4842-2943-9>
- NotebookLM. (2025). Alur Kerja Sistem Chat AI Fiqih Haji.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pressman, R. S., & Maxim, B. R. (2020). *Software Engineering a Practitioner's Approach 9th Edition* (9 ed.). New York: Mc Graw Hill.
- Salim, V., & Farisi, A. (2024). Implementasi Metode Iterative Enhancement dalam Pengembangan Aplikasi Web dan Mobile Terintegrasi bagi Kelompok Bimbingan Ibadah Haji dan Umrah (KBIHU) di Kota Palembang. *JSI: Jurnal Sistem Informasi (E-Journal)*, 16(2), 294–312. <https://doi.org/10.18495/jsi.v16i2.168>
- Sun, R., Hu, X., Shao, Y., Luo, Z., Liu, B., & Cheng, Y. (2025). Using Large Language Models to Analyze Interviews for Driver Psychological Assessment: A Performance Comparison of ChatGPT and Google-Gemini. *Symmetry*, 17(10), 1713. <https://doi.org/10.3390/sym17101713>
- Wieringa, R. J. (2014). Design Science Methodology for Information Systems and Software Engineering. Dalam *Design Science Methodology: For Information Systems and Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-43839-8>