

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ibadah haji merupakan rukun Islam kelima yang wajib dilaksanakan oleh setiap Muslim yang mampu serta memerlukan pemahaman yang baik terhadap tata cara dan hukum-hukum yang mengaturnya [1]. Pada tahun 2024, jumlah jemaah haji dunia mencapai sekitar 1,83 juta orang, dengan Indonesia konsisten menjadi salah satu kontributor jemaah terbesar setiap tahunnya [1], [2]. Dalam penyelenggaraannya, Kelompok Bimbingan Ibadah Haji (KBIH) berperan penting dalam memberikan layanan administrasi, pembinaan, dan bimbingan manasik kepada jemaah [3]. Mengingat sebagian besar jemaah hanya berhaji sekali seumur hidup, kebutuhan akan layanan informasi yang cepat, akurat, dan mudah diakses, baik terkait aspek fiqh maupun teknis pelaksanaan menjadi semakin penting [1].

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*), khususnya *chatbot* berbasis *Natural Language Processing* (NLP), telah banyak dimanfaatkan untuk meningkatkan kualitas pelayanan publik melalui penyediaan informasi secara otomatis dan *real-time*, dengan implementasinya di Indonesia terbukti meningkatkan efektivitas layanan dan nilai publik [4], [5]. Relevansi ini semakin kuat seiring transformasi digital Arab Saudi dalam mendukung jutaan jemaah melalui layanan berbasis AI [2]. Oleh karena itu, *chatbot* berpotensi menjadi solusi penyampaian informasi dan bimbingan haji yang dapat diakses kapan saja dan di mana saja [6].

Studi pendahuluan melalui wawancara dengan Kepala Pembimbing KBIH Miftahul Ulum menunjukkan bahwa meskipun berbagai media informasi (grup WhatsApp, Instagram, *website*, materi manasik) telah digunakan, kebutuhan informasi jemaah masih tinggi. Hal ini disebabkan oleh keterbatasan layanan konsultasi manual yang hanya beroperasi pukul 09.00–17.00, padahal pertanyaan jemaah kerap muncul di luar jam tersebut, serta buku manasik, modul, dan FAQ yang belum mampu menjawab persoalan *fiqh* yang membutuhkan penjelasan kontekstual dan komunikasi dua arah. Narasumber, yang telah berkecimpung dalam pelayanan haji sejak 2003,

menegaskan bahwa kebutuhan layanan informasi 24 jam terus meningkat seiring meningkatnya literasi digital jemaah.

Berbagai penelitian telah mengembangkan *chatbot* pada domain haji dan Islam, mulai dari BERT dan DistilBERT untuk klasifikasi pertanyaan, sistem RAG untuk informasi kesehatan jemaah, hingga AraBERT pada dataset Hajj-FQA untuk menjawab fatwa haji [7], [8], [9], [1]. Namun, penelitian tersebut umumnya masih berfokus pada pencarian atau klasifikasi informasi, belum optimal dalam menangani pertanyaan yang memerlukan penalaran, pemahaman konteks, dan jawaban yang natural serta komprehensif [1], [9]. *Chatbot* berbasis AI juga masih rentan mengalami halusinasi yang mana menghasilkan informasi meyakinkan namun tidak didukung sumber valid yang krusial pada domain keagamaan karena berpotensi menyebabkan kesalahan pelaksanaan ibadah sesuai syariat [1], [10].

Merujuk pada penelitian terdahulu, penelitian ini mengusulkan *chatbot* bimbingan haji berbasis LLM menggunakan IndoT5 yang dioptimalkan dengan LoRA dan pendekatan *Retrieval-Augmented Generation* (RAG) sebagai strategi mitigasi halusinasi. IndoT5 dipilih sebagai model generator karena merupakan adaptasi *transformer* khusus berbahasa Indonesia yang unggul memahami konteks lokal serta mampu menyusun jawaban alami melalui arsitektur *encoder-decoder* [11], [12]. RAG mengintegrasikan mekanisme *retrieval* dengan *generation* untuk memastikan setiap respons didasarkan pada dokumen rujukan yang valid dan faktual, sehingga memitigasi risiko jawaban yang terdengar meyakinkan namun tidak berdasar fakta [11], [13], [14]. Pendekatan ini sangat relevan bagi layanan haji yang menuntut akurasi tinggi dan kepatuhan terhadap sumber otoritatif seperti Al-Qur'an, hadis, serta buku panduan manasik resmi kementerian guna menjaga kepercayaan jemaah [15], [16].

Berdasarkan uraian tersebut, Penelitian ini bertujuan mengembangkan dan mengevaluasi *chatbot* bimbingan haji berbasis IndoT5 dengan teknik LoRA dan pendekatan RAG guna meningkatkan kualitas layanan informasi haji. Sistem ini diharapkan mampu memberikan respons yang akurat, relevan, dan kontekstual, mengurangi risiko halusinasi AI, serta menyediakan bimbingan haji yang dapat diakses berkelanjutan kapan saja dan di mana saja [6].

1.2 Rumusan Masalah

Dari uraian latar belakang di atas, maka dapat dirumuskan masalah sebagai berikut:

1. Bagaimana membangun dan mengukur kinerja model *chatbot* berbasis IndoT5 untuk menjawab pertanyaan layanan haji pada domain tertutup berbahasa Indonesia?
2. Bagaimana merancang dan mengimplementasikan mekanisme *retrieval* bertingkat untuk meningkatkan akurasi jawaban dan memitigasi risiko halusinasi AI pada domain *fiqh* haji yang berisiko tinggi?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan serta mengevaluasi kinerja model *chatbot* layanan haji pada domain tertutup (*closed-domain*) dalam Bahasa Indonesia dengan memanfaatkan arsitektur IndoT5.
2. Merancang dan mengimplementasikan mekanisme *retrieval* bertingkat untuk meningkatkan akurasi jawaban dan memitigasi risiko halusinasi AI pada domain *fiqh* haji.

1.4 Batasan Masalah

Dalam menjaga fokus penelitian ini dan tidak terlalu luas dari pembahasan yang dimaksud, maka penelitian ini membatasi ruang lingkupnya. Oleh karena itu batasan masalah dalam penelitian ini sebagai berikut:

1. Dataset dibatasi pada data asli FAQ KBIH Miftahul Ulum dan buku Tuntunan Manasik Haji dan Umrah Kementerian Agama RI 2024.
2. Model generatif yang digunakan adalah IndoT5-Base (220M parameter) dan menggunakan Teknik *fine-tuning Low-Rank-Adaptation* (LoRA).
3. Chatbot ini dirancang sebagai sistem domain tertutup (*closed-domain*) yang hanya memberikan jawaban dan menerima pertanyaan yang berkaitan dengan haji dan umrah dalam Bahasa Indonesia.
4. *Chatbot* memberikan jawaban berupa teks generatif berbahasa alami, tanpa dukungan *multimodal* (gambar, audio, atau video).

5. Sistem tidak mencakup pembaruan data *real-time* (informasi biaya, kuota, jadwal yang berubah).
6. Evaluasi akurasi konten keagamaan dilakukan melalui validasi model dengan membandingkan *output* sistem terhadap sumber rujukan resmi (buku Tuntunan Manasik Kemenag), dilengkapi validasi manual terbatas oleh pengurus KBIH.
7. Penelitian ini mengadaptasi metodologi CRISP-DM dengan mengecualikan tahap *deployment*, mengingat *output* yang dihasilkan berupa *prototipe* penelitian, bukan sistem yang siap dioperasikan pada lingkungan produksi.

1.5 Manfaat Penelitian

Penulisan penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Pengguna

Dapat mempermudah pengguna dalam mencari informasi haji dan umrah melalui *Chatbot*, sehingga secara signifikan mengurangi waktu tunggu respons manual dan beban administratif staf.

2. Bagi Penulis

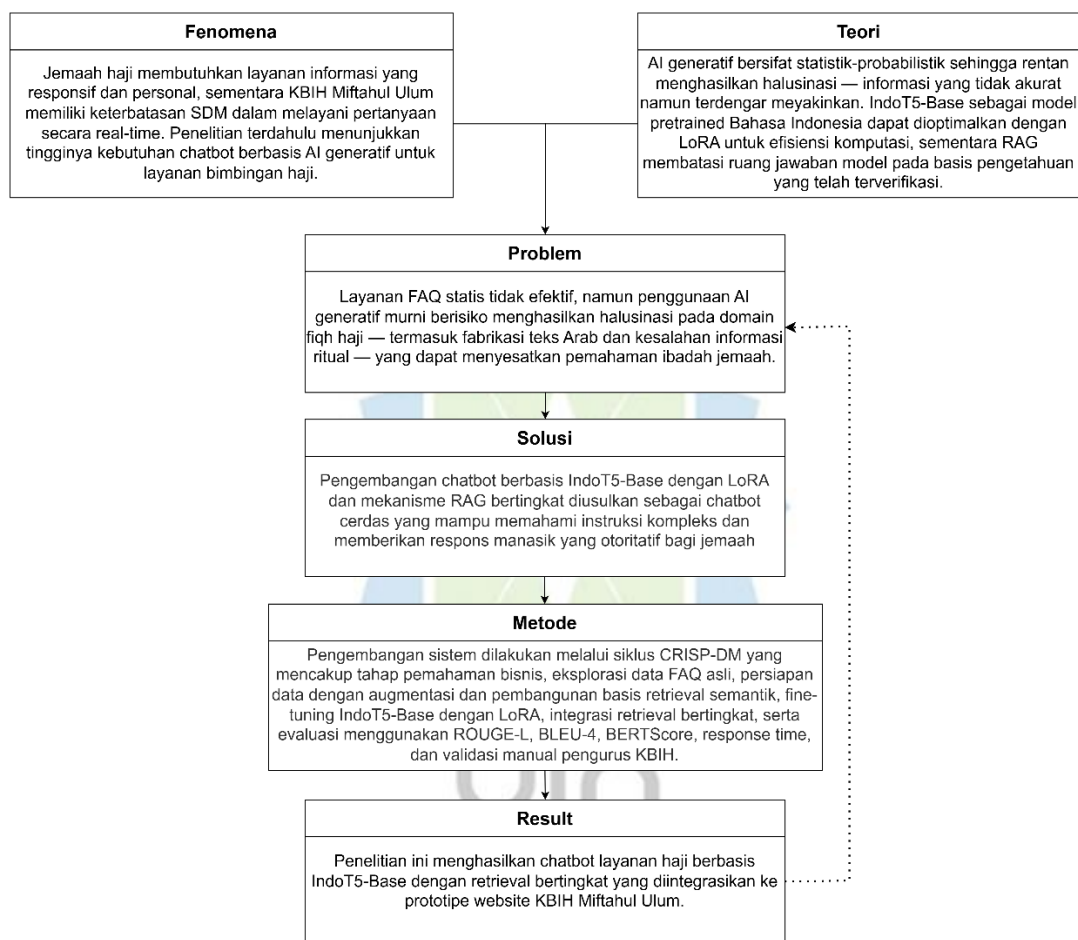
Menjadi sarana penerapan pengetahuan dalam bidang pengembangan *chatbot* berbasis kecerdasan buatan, khususnya menggunakan model IndoT5 dengan teknik LoRA dan pendekatan RAG, serta memperluas pemahaman penulis mengenai implementasi *Natural Language Processing* (NLP) pada layanan informasi haji dan umrah berbasis web.

3. Bagi Akademik

Memperkaya literatur mengenai efektivitas teknik *fine-tuning* model kecil dan RAG untuk domain keagamaan Bahasa Indonesia dalam menangani tugas penalaran logika yang kompleks di domain tertutup (*closed-domain*).

1.6 Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini disusun untuk menjembatani masalah praktis dan landasan teoritis hingga menghasilkan solusi yang tepat, seperti yang ditunjukkan pada Gambar 1. 1.



Gambar 1. 1 Kerangka pemikiran penelitian

Gambar 1. 1 berangkat dari meningkatnya kebutuhan layanan informasi haji yang responsif dan personal, sementara KBIH Miftahul Ulum menghadapi keterbatasan SDM untuk melayani pertanyaan jemaah secara *real-time*. Fenomena ini didukung oleh landasan teoretis bahwa sifat statistik-probabilistik AI generatif membuatnya rentan berhalusinasi yang menghasilkan informasi tidak akurat namun terdengar meyakinkan. Untuk itu, IndoT5-Base dipilih sebagai basis model, dioptimalkan dengan LoRA untuk efisiensi komputasi, sementara RAG digunakan

untuk membatasi ruang jawaban pada basis pengetahuan terverifikasi [17], [18], [19], [20].

Pada praktiknya, dua masalah utama ditemukan: layanan FAQ statis yang tidak mampu memenuhi kebutuhan informasi jemaah yang beragam, dan risiko halusinasi AI yang berbahaya pada domain *fiqh* haji, termasuk fabrikasi teks Arab dan kesalahan informasi ritual yang berpotensi menyesatkan pemahaman ibadah. Penelitian ini menawarkan *chatbot* berbasis IndoT5-Base dengan LoRA dan mekanisme RAG bertingkat: pertanyaan dengan kemiripan tinggi dijawab langsung dari data terverifikasi, kemiripan sedang dijawab dengan bantuan konteks *retrieval*, dan pertanyaan di luar cakupan dijawab dengan pesan *abstain* yang aman.

Implementasi dilakukan melalui metode CRISP-DM, meliputi pengumpulan data FAQ KBIH dan Buku Tuntunan Manasik Haji & Umrah Kemenag RI 2024, persiapan data (augmentasi dan pembangunan basis *retrieval* semantik), *fine-tuning* IndoT5-Base dengan LoRA, integrasi *retrieval* bertingkat, serta evaluasi menggunakan ROUGE-L, BLEU-4, BERTScore, *response time*, dan validasi manual oleh pengurus KBIH [18], [20].

Hasil akhirnya adalah *chatbot* layanan haji berbasis IndoT5-Base dengan *retrieval* bertingkat, terintegrasi ke *prototipe website* KBIH Miftahul Ulum sebagai media informasi digital yang lebih adaptif, interaktif, dan relevan bagi jemaah.

1.7 Sistematika Penulisan

Berikut adalah sistematika penulisan pada penelitian ini yang disusun secara terstruktur:

BAB I PENDAHULUAN

Bab I berisi uraian pendahuluan yang mencakup latar belakang penelitian, rumusan masalah, tujuan dan manfaat penelitian, serta batasan masalah yang digunakan. Selain itu, bab ini juga memaparkan kerangka pemikiran serta sistematika penulisan yang disusun secara terstruktur.

BAB II STUDI PUSTAKA

Bab II menyajikan studi pustaka yang memuat teori-teori dan konsep yang menjadi dasar serta landasan dalam mendukung jalannya penelitian.

BAB III METODOLOGI PENELITIAN

Bab III menjelaskan metodologi yang digunakan dalam penelitian, mulai dari proses perancangan sistem, pemahaman dan pengumpulan data, hingga tahapan pengolahan data yang dilakukan selama penelitian berlangsung.

BAB IV HASIL DAN PEMBAHASAN

Bab IV memaparkan hasil penelitian serta pembahasan mengenai sistem yang telah dirancang dan dikembangkan. Pada bab ini juga dilakukan evaluasi terhadap sistem tersebut.

BAB V PENUTUP

Bab V merupakan bagian akhir yang berisi kesimpulan dari keseluruhan penelitian.

DAFTAR PUSTAKA

Bagian daftar pustaka memuat seluruh sumber rujukan tertulis yang digunakan sebagai acuan dalam penelitian.

LAMPIRAN

Pada bagian lampiran dicantumkan berbagai dokumen pendukung yang digunakan selama proses penyusunan dan perancangan penelitian.