PENGEMBANGAN CHATBOT HUKUM INDONESIA BERBASIS RAG DENGAN EVALUASI KINERJA RETRIEVAL DAN KONSISTENSI FAKTUAL

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

AAN TRIYA VINANTA

22.11.4880

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

PENGEMBANGAN CHATBOT HUKUM INDONESIA BERBASIS RAG DENGAN EVALUASI KINERJA RETRIEVAL DAN KONSISTENSI FAKTUAL

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

AAN TRIYA VINANTA

22.11.4880

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

SKRIPSI

PENGEMBANGAN CHATBOT HUKUM INDONESIA BERBASIS RAG DENGAN EVALUASI KINERJA RETRIEVAL DAN KONSISTENSI FAKTUAL

yang disusun dan diajukan oleh

Aan Triya Vinanta

22.11.4880

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2026

Dosen Pembimbing,

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

HALAMAN PENGESAHAN

SKRIPSI

PENGEMBANGAN CHATBOT HUKUM INDONESIA BERBASIS RAG DENGAN EVALUASI KINERJA RETRIEVAL DAN KONSISTENSI FAKTUAL

yang disusun dan diajukan oleh

Aan Triya Vinanta

22.11.4880

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Theopilus Bayu Sasongko, S.Kom., M.Eng **NIK. 190302375**	
I Made Artha Agastya, S.T., M.Eng., Ph.D. **NIK. 190302352**	
Anna Baita, S.Kom., M.Kom. **NIK. 190302290**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aan Triya Vinanta
NIM : 22.11.4880

Menyatakan bahwa Skripsi dengan judul berikut:

PENGEMBANGAN CHATBOT HUKUM INDONESIA BERBASIS RAG DENGAN EVALUASI KINERJA RETRIEVAL DAN KONSISTENSI FAKTUAL

Dosen Pembimbing : Anna Baita, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Januari 2026

Yang Menyatakan,

Aan Triya Vinanta

HALAMAN PERSEMBAHAN

Dengan penuh kerendahan hati dan rasa syukur yang mendalam, tugas akhir ini saya persembahkan untuk

1. Kedua orang tua tercinta, yang senantiasa mendoakan, memberikan dukungan tanpa henti, serta melimpahkan kasih sayang yang tiada batas, sehingga penyusunan tugas akhir ini dapat diselesaikan dengan baik.
2. Seluruh keluarga besar, atas semangat, dorongan, dan kebersamaan yang terus menjadi sumber inspirasi serta motivasi utama.
3. Dosen pembimbing, yang telah memberikan bimbingan, arahan, dan ilmu pengetahuan berharga sepanjang proses penyusunan penelitian ini hingga selesai.
4. Teman-teman Seperjuangan, yang telah menjadi teman diskusi yang solutif, pendengar keluh kesah yang sabar, dan partner di tengah kepenatan. Kebersamaan, canda tawa, dan semangat pantang menyerah yang kita lalui bersama membuat perjalanan akademis yang terjal ini terasa lebih ringan dan bermakna.

Saya berharap penelitian ini dapat memberikan kontribusi yang bermanfaat bagi pengembangan ilmu pengetahuan serta membawa dampak positif bagi Masyarakat luas.

KATA PENGANTAR

Segala dan uji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas limpahan rahmat, petunjuk, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dengan penuh rasa hormat dan kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu, memberikan dukungan, serta bimbingan selama proses penyusunan skripsi ini. Ucapan itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika.
4. Ibu Anna Baita, S.Kom., M.Kom., selaku dosen pembimbing yang telah memberikan bimbingan, arahan kepada penulis.
5. Kedua orang tua, keluarga besar, dan semua pihak yang memberikan semangat, dukungan, serta doa yang tiada habisnya.

Akhir kata, penulis berharap skripsi ini dapat memberikan manfaat bagi semua pihak yang membutuhkan.

Yogyakarta, 22 Januari 2026

Aan Triya Vinanta

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR LAMBANG DAN SINGKATAN

AI	Artificial Intelligence
BM25	Best Match 25
CLI	Command Line Interface
CoT	Chain-of-Thought
HNSW	Hierarchical Navigable Small World
ITE	Informasi dan Transaksi Elektronik
JSON	JavaScript Object Notation
KUHP	Kitab Undang-Undang Hukum Pidana
LLM	Large Language Model
MRR	Mean Reciprocal Rank
NLP	Natural Language Processing
PDP	Pelindungan Data Pribadi
QA	Question Answering
RAG	Retrieval-Augmented Generation
TF-IDF	Term Frequency-Inverse Document Frequency
UU	Undang-Undang
K	Batas peringkat teratas (cut-off rank) dalam evaluasi retrieval
N	Jumlah total sampel atau pertanyaan dalam pengujian
Q,	Himpunan pertanyaan (Query)
$\sum$	Notasi Sigma (Penjumlahan)
$\in$	Elemen dari (anggota himpunan)

Θ	Sudut antara dua vektor (dalam rumus Cosine Similarity)
@	At, digunakan untuk menunjuk batas peringkat
‖A	Magnitude (panjang) dari vektor A

DAFTAR ISTILAH

Artificial Intelligence (AI)	:	Cabang ilmu komputer yang mensimulasikan kecerdasan manusia dalam mesin, mencakup pembelajaran, penalaran, dan pemecahan masalah.
Asynchronous	:	Metode pemrograman di mana proses eksekusi tidak saling memblokir; sistem dapat menangani tugas lain sembari menunggu proses selesai.
Augmentation	:	Proses menambahkan data atau informasi relevan ke dalam prompt sebelum dikirim ke model generator untuk memperkaya konteks.
Best Match 25 (BM25)	:	Algoritma ranking fungsi probabilistic untuk information retrieval yang memperkirakan relevansi dokumen berdasarkan frekuensi kata kunci (keyword matching).
Candidate Pool	:	Sekumpulan dokumen awal yang berhasil diambil oleh sistem retrieval sebelum dilakukan reranking.
Chatbot	:	Program komputer yang dirancang untuk mensimulasikan percakapan dengan pengguna manusia melalui teks atau suara.
Chunking	:	Proses memecah teks dokumen panjang menjadi potongan-potongan kecil (chunks) agar lebih mudah diproses dan disimpan dalam database vektor.
Contextual Enrichment	:	Teknik memperkaya potongan data (chunk) dengan informasi tambahan dari konteks sekitarnya untuk meningkatkan pemahaman model.
Contextual Reasoning	:	Kemampuan model AI untuk memahami nuansa, latar belakang, dan hubungan antar informasi dalam teks untuk menghasilkan jawaban logis.

Cosine Similarity	:	Metrik matematika untuk mengukur kemiripan antara dua vektor (dokumen) dengan menghitung kosinus sudut di antara keduanya.
Cross-Encoder	:	Arsitektur model transformer yang memproses kueri dan dokumen secara bersamaan untuk menghasilkan skor relevansi yang sangat akurat (biasanya untuk reranking).
Cross Score	:	Nilai relevansi yang dihasilkan oleh model Cross-Encoder yang menilai hubungan semantik mendalam antara kueri dan dokumen.
Decoder	:	Komponen dalam arsitektur Transformer yang bertugas menghasilkan output (generasi teks) berdasarkan representasi yang diproses oleh Encoder.
Deep Learning	:	Sub-bidang Machine Learning yang menggunakan jaringan saraf tiruan berlapis (multi-layered) untuk mempelajari pola data yang kompleks.
Dense Vector	:	Representasi vektor numerik padat yang menangkap makna semantik suatu teks, dihasilkan oleh model embedding.
Document Structuring	:	Proses mengorganisasi data teks mentah ke dalam format terstruktur (seperti JSON) agar mesin dapat memproses hierarki informasi.
Domain Coverage	:	Cakupan seberapa luas atau lengkap topik (misal: jumlah Undang-Undang) yang dikuasai oleh sistem knowledge base.
Dot Product	:	Operasi aljabar perkalian antara dua vektor yang digunakan untuk menghitung skor kemiripan dalam pencarian vektor.

Ecommerce	:	Perdagangan elektronik; domain di mana sistem rekomendasi dan pencarian serupa dengan teknologi RAG sering diterapkan.
Embedding	:	Proses konversi data teks menjadi representasi vektor angka di ruang dimensi tinggi.
End-to-End	:	Pendekatan sistem di mana model dilatih atau bekerja untuk menyelesaikan tugas dari input awal hingga output akhir tanpa langkah manual terpisah.
Encoding	:	Proses mengubah data input (teks) menjadi format representasi internal (vektor) yang dapat dipahami oleh mesin.
Faithfulness	:	Metrik evaluasi untuk mengukur seberapa akurat dan konsisten jawaban model terhadap konteks yang diberikan.
Final Score	:	Nilai akhir gabungan dari berbagai skor (Lexical, Semantic, Metadata) yang menentukan peringkat akhir dokumen.
Fine-Tuning	:	Proses melatih kembali model pre-trained dengan dataset spesifik untuk meningkatkan kinerjanya pada tugas tertentu.
Fixed-Size Chunking	:	Metode pemotongan teks berdasarkan jumlah karakter atau token yang tetap dan konsisten.
Framework	:	Kerangka kerja perangkat lunak yang menyediakan struktur dasar untuk membangun aplikasi.
Gemini Flash	:	Varian model bahasa besar (LLM) dari Google yang dioptimalkan untuk latensi rendah dan efisiensi biaya.
Generative AI (Gen AI)	:	Jenis kecerdasan buatan yang mampu menciptakan konten baru (teks, gambar) berdasarkan pola data pelatihan.

Goldset	:	Kumpulan data referensi berkualitas tinggi (pertanyaan & jawaban benar) yang digunakan sebagai standar kebenaran.
Grounding	:	Proses menambatkan jawaban AI pada fakta atau sumber data nyata untuk mencegah informasi yang mengada-ada.
Hallucination	:	Fenomena ketika model menghasilkan informasi yang terdengar meyakinkan tetapi faktanya salah atau tidak ada dalam sumber data.
Hybrid Filtering	:	Metode penyaringan hasil pencarian yang menggabungkan kriteria skor vektor dan metadata.
Hybrid Retrieval	:	Strategi pencarian yang menggabungkan pencarian kata kunci (Sparse) dan pencarian makna (Dense) untuk akurasi lebih tinggi.
Information Retrieval (IR)	:	Ilmu yang mempelajari teknik pencarian dokumen atau informasi yang relevan dari koleksi data yang besar.
Keyword Matching	:	Teknik pencarian dokumen berdasarkan kecocokan kata per kata secara harfiah.
Knowledge Base	:	Basis pengetahuan terpusat berisi data terstruktur/tidak terstruktur yang digunakan sistem RAG sebagai sumber informasi.
Knowledge Graph	:	Representasi pengetahuan dalam bentuk grafik yang menghubungkan entitas dengan relasinya.
Knowledge-Grounded QA	:	Sistem tanya jawab yang jawabannya didasarkan secara ketat pada dokumen pengetahuan eksternal.
Large Language Model (LLM)	:	Model pembelajaran mesin yang dilatih pada dataset teks skala masif sehingga mampu memahami dan menghasilkan bahasa manusia.

Lexical Score	:	Nilai relevansi yang dihitung berdasarkan kecocokan kata kunci, seperti skor BM25.
Machine Translation	:	Penggunaan AI untuk menerjemahkan teks dari satu bahasa ke bahasa lain secara otomatis.
Mean Reciprocal Rank (MRR)	:	Metrik evaluasi retrieval yang menghitung rata-rata kebalikan dari peringkat dokumen relevan pertama yang ditemukan.
Merge Deduplicate	:	Proses menggabungkan hasil dari beberapa metode pencarian dan menghapus dokumen duplikat.
Metadata	:	Data yang menjelaskan data lain
Metadata Score	:	Nilai tambahan yang diberikan pada dokumen berdasarkan kecocokan
ms-marco-MiniLM-L-6-v2	:	Model Sentence Transformer populer yang sering digunakan untuk tugas Semantic Search dan Reranking.
Natural Language Processing (NLP)	:	Cabang AI yang berfokus pada interaksi antara komputer dan bahasa alami manusia.
Neural Network	:	Jaringan saraf tiruan yang meniru cara kerja otak manusia untuk memproses informasi.
Noise	:	Data tidak relevan yang dapat mengganggu kualitas hasil retrieval atau generation.
Non-linear	:	Hubungan data kompleks di mana output tidak berbanding lurus sederhana dengan input.
Normalization Score	:	Proses menyetarakan skala angka dari berbagai metode agar bisa digabungkan secara adil.
Normalisasi Snorm de Boer	:	Metode normalisasi spesifik untuk menyeimbangkan distribusi angka yang berbeda skala.
Open Domain QA	:	Sistem tanya jawab yang dapat menjawab pertanyaan tentang topik apa pun tanpa batasan domain.
Outdated Knowledge	:	Informasi dalam model AI yang sudah kadaluarsa.

Overlap	:	Bagian teks yang tumpang tindih antar chunk berurutan untuk menjaga konteks kalimat.
Pinecone / Qdrant	:	Layanan Vector Database yang dirancang untuk menyimpan dan mencari embedding vektor.
Pipeline	:	Rangkaian proses berurutan.
Precision Retrieval	:	Ukuran seberapa banyak dokumen yang diambil benar-benar relevan dengan kebutuhan pengguna.
Preprocessing	:	Tahap awal pengolahan data mentah sebelum digunakan model.
Pre-trained Model	:	Model AI yang sudah dilatih sebelumnya pada dataset besar dan siap digunakan.
Raw Data	:	Data mentah yang belum diproses atau dimodifikasi.
Recursive Chunking	:	Teknik pemotongan teks secara hierarkis agar potongan teks tetap koheren.
Reranking	:	Proses mengurutkan ulang hasil pencarian awal menggunakan model presisi tinggi.
Retrieval-Augmented Generation (RAG)	:	Teknik menggabungkan model generatif dengan sistem pencarian dokumen eksternal.
Retriever	:	Komponen sistem yang bertugas mencari dan mengambil dokumen relevan dari Knowledge Base.
Robust	:	Kemampuan sistem untuk tetap bekerja dengan baik meskipun menghadapi input yang tidak ideal atau error.
Rule-Based Scoring	:	Metode penilaian berdasarkan aturan manual yang ditetapkan.
Self-Attention	:	Mekanisme Transformer yang memungkinkan model menimbang pentingnya kata berbeda dalam satu kalimat.
Sentiment Analysis	:	Teknik NLP untuk mengidentifikasi emosi dalam teks.

Size	:	Ukuran panjang maksimum karakter atau token per potongan teks.
Sparse Vector	:	Representasi vektor yang sebagian besar nilainya nol.
Text Embedding	:	Representasi vektor dari teks yang menangkap makna semantiknya.
Text Extraction	:	Proses mengambil teks murni dari format dokumen (PDF/HTML).
Text Normalization	:	Proses standarisasi teks (lowercase, hapus tanda baca).
TF-IDF	:	Statistik numerik yang mencerminkan pentingnya kata dalam dokumen relatif terhadap kumpulan dokumen.
Token	:	Unit dasar teks (kata/karakter) yang diproses model bahasa.
Transformer	:	Arsitektur Deep Learning modern berbasis attention mechanism, dasar dari LLM.
Untraceable Reasoning	:	Kelemahan model AI di mana alur logika keputusannya sulit dilacak.
Vector Database	:	Database khusus untuk menyimpan dan query data vektor embedding.
Watermark	:	Pola untuk menandai asal-usulnya.
Weighted Sum Model	:	Metode penggabungan skor (Sparse + Dense) dengan bobot tertentu.
Weighting	:	Pemberian bobot prioritas pada komponen tertentu dalam perhitungan skor.
Whitespace	:	Karakter spasi, tab, atau baris baru dalam teks.

INTISARI

Rendahnya pengetahuan hukum masyarakat Indonesia sering kali menyebabkan kesulitan dalam mencari rujukan pasal yang tepat, yang berujung pada keputusan keliru dan ketidakmerataan akses keadilan, sementara penggunaan *Large Language Model* (LLM) murni berisiko tinggi menghasilkan halusinasi tanpa sumber rujukan eksplisit. Penelitian ini bertujuan mengatasi permasalahan tersebut dengan mengembangkan *chatbot* hukum menggunakan metode *Retrieval-Augmented Generation* (RAG) yang memproses delapan Undang-Undang utama melalui teknik *recursive chunking* dan menyimpannya dalam *Qdrant Vector Database*. Sistem menggabungkan pencarian semantik dan leksikal yang disempurnakan dengan mekanisme *reranking* menggunakan model *cross-encoder* ms-marco-MiniLM-L-6-v2, serta menghasilkan jawaban melalui model Gemini-2.5 Flash. Hasil evaluasi terhadap 42 kueri *gold set* dari Hukumonline menunjukkan nilai *Recall@k* sebesar 0,619 dan *Mean Reciprocal Rank* (MRR) 0,351 yang mengindikasikan efektivitas penempatan dokumen relevan pada peringkat awal, serta skor *faithfulness* 0,478 yang mencerminkan tantangan penalaran pada kueri alami. Kontribusi penelitian berupa rancangan *pipeline* RAG yang *robust* ini dapat dimanfaatkan oleh masyarakat umum dan praktisi sebagai alat bantu verifikasi hukum yang akurat untuk meminimalisir kesalahan interpretasi.

Kata kunci: RAG, hukum Indonesia, *reranking, faithfulness, LLM.*

ABSTRACT

The lack of legal knowledge among Indonesian society often leads to difficulties in finding precise article references, resulting in wrong decisions and unequal access to justice, while the use of pure Large Language Models (LLMs) carries a high risk of producing factual hallucinations without explicit source references. This study aims to address these issues by developing a legal chatbot using the Hybrid Retrieval-Augmented Generation (RAG) method, which processes eight major Laws through recursive chunking techniques and stores them in the Qdrant Vector Database. The system combines semantic and lexical search refined by a reranking mechanism using the ms-marco-MiniLM-L-6-v2 cross-encoder model, and generates answers via the Gemini-1.5 Flash model. Evaluation results on 42 gold set queries from Hukumonline.com showed a Recall@k value of 0.619 and a Mean Reciprocal Rank (MRR) of 0.351, indicating the effectiveness of placing relevant documents in the top ranks, along with a faithfulness score of 0.478 reflecting reasoning challenges in natural queries. The contribution of this research is a robust RAG pipeline design that can be utilized by the general public and practitioners as an accurate legal verification tool to minimize interpretation errors.

Keyword**:** *RAG, Legal Chatbot, reranking, faithfulness, LLM.*