

PENGEMBANGAN CHATBOT EDUKASI HUKUM LALU LINTAS BERBASIS RAG PADA PLATFORM WEB DAN TELEGRAM BERDASARKAN UNDANG-UNDANG NOMOR 22 TAHUN 2009

Zaky Adil Fahresy, Rionaldi Ali

Teknik Informatika, Institut Informatika Dan Bisnis Darmajaya

Jl. ZA. Pagar Alam No.93, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung, Lampung

zakyyadil@gmail.com

ABSTRAK

Indonesia mencatat 103.645 kasus kecelakaan lalu lintas pada tahun 2023, dengan 25.874 korban meninggal dan sekitar 60% disebabkan kelalaian manusia akibat rendahnya literasi hukum terhadap Undang-Undang Nomor 22 Tahun 2009. Permasalahan utama adalah sulitnya mengakses informasi dari dokumen 326 pasal berbahasa hukum formal tersebut. Penelitian ini merancang chatbot edukasi berbasis Retrieval-Augmented Generation (RAG) yang meminimalkan hallucination pada Large Language Model (LLM) dengan mengindeks UU No. 22/2009 ke vector database ChromaDB menggunakan embedding all-MiniLM-L6-v2 dan cosine similarity, dikembangkan dengan metode R&D Agile pada antarmuka web dan bot Telegram, dilengkapi fitur Pusat Kontrol Reminder dengan notifikasi harian pada empat jendela waktu acak sebagai rencana penyelesaian masalah edukasi preventif. Pengujian 50 skenario menghasilkan F1-Score = 0,958, Precision = 0,939, Recall = 0,979, Accuracy = 92,3%—melampaui target $F1 \geq 0,85$. RAG mereduksi hallucination dari 40% menjadi 0%. Evaluasi SUS menghasilkan skor rata-rata 83,0 (Good) dengan waktu respons 2,3 detik per kueri.

Kata kunci : *Retrieval-Augmented Generation, Chatbot Hukum, Large Language Model, Bot Telegram, Hallucination, UU Lalu Lintas*

1. PENDAHULUAN

Kecelakaan lalu lintas merupakan permasalahan kesehatan publik dan sosial-ekonomi yang serius di Indonesia. Badan Pusat Statistik (BPS) mencatat bahwa dalam rentang lima tahun terakhir (2019–2023), angka kejadian kecelakaan lalu lintas berfluktuasi dalam kisaran yang konsisten tinggi, dengan jumlah kejadian berkisar antara 100 ribu hingga lebih dari 130 ribu kasus per tahun dan korban jiwa mencapai puluhan ribu orang setiap tahunnya [1].

Rata-rata, Indonesia kehilangan 71 jiwa setiap hari akibat kecelakaan lalu lintas, menjadikannya penyebab kematian ketiga terbesar nasional setelah penyakit jantung iskemik dan stroke [2]. Kerugian ekonomi akibat kecelakaan lalu lintas pada tahun 2023 diperkirakan mencapai Rp43,1 triliun berdasarkan estimasi biaya medis, kehilangan produktivitas, dan kerusakan properti [3].

Data Korps Lalu Lintas Polri (Korlantas Polri) mengidentifikasi bahwa sekitar 60% kecelakaan lalu lintas disebabkan faktor kelalaian manusia (human error) yang berkaitan dengan ketidakpatuhan terhadap peraturan berlalu lintas [4].

Kementerian Perhubungan melalui Sistem Informasi Manajemen Keselamatan Lalu Lintas (SIMAKLALU) mencatat lima pelanggaran paling sering terjadi, yaitu tidak mengenakan helm SNI, melampaui batas kecepatan, tidak memakai sabuk keselamatan, penggunaan gawai saat berkendara, dan tidak membawa kelengkapan dokumen [5].

Seluruh pelanggaran tersebut telah diatur secara eksplisit dalam Undang-Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, yang

terdiri atas 326 pasal dalam 203 halaman bahasa hukum formal yang sulit dipahami masyarakat awam [6].

Kesenjangan antara tingginya angka pelanggaran dan tersedianya regulasi yang komprehensif ini mengindikasikan rendahnya literasi hukum lalu lintas di masyarakat.

Studi pendahuluan melalui wawancara terstruktur terhadap 15 pengemudi aktif di Kota Bandar Lampung pada Maret 2026 menunjukkan bahwa mayoritas responden tidak mampu menyebutkan nomor pasal dan besaran denda atas pelanggaran yang rutin mereka lakukan, belum pernah membaca UU No. 22 Tahun 2009 secara lengkap, dan tidak memahami mekanisme tilang elektronik (ETLE). Temuan ini diperkuat oleh observasi di Satlantas Polresta Bandar Lampung (10–14 Maret 2026), yang mengungkap adanya hambatan komunikasi signifikan antara petugas dan masyarakat dalam penyampaian informasi hukum lalu lintas.

Perkembangan teknologi kecerdasan buatan, khususnya Large Language Model (LLM), telah membuka peluang baru dalam pengembangan sistem tanya-jawab otomatis berbasis dokumen hukum. Namun, LLM konvensional memiliki kelemahan kritis berupa fenomena hallucination, yaitu kondisi di mana model menghasilkan respons yang terdengar meyakinkan namun faktanya tidak akurat atau tidak berdasar pada sumber yang sah [7].

Ji et al. mengklasifikasikan hallucination menjadi intrinsic hallucination, yaitu keluaran yang bertentangan dengan sumber yang tersedia, dan

extrinsic hallucination, yaitu keluaran yang tidak dapat diverifikasi dari sumber manapun [8].

Dalam domain hukum, hallucination sangat berbahaya karena pengguna dapat bertindak berdasarkan informasi pasal dan besaran denda yang keliru, yang dapat berakibat pada kesalahan persepsi hukum maupun kerugian finansial.

Arsitektur Retrieval-Augmented Generation (RAG) yang diperkenalkan Lewis et al. menawarkan solusi atas kelemahan tersebut dengan memaksa LLM untuk hanya menghasilkan jawaban berdasarkan dokumen referensi yang diambil melalui semantic search, sehingga setiap respons dapat diverifikasi hingga tingkat pasal dan ayat [9]. Shuster et al. membuktikan bahwa pendekatan RAG secara signifikan mereduksi hallucination pada sistem percakapan berbasis LLM [10].

Penelitian terdahulu membuktikan efektivitas pendekatan ini pada domain hukum Indonesia: Nurdiansyah dan Hidayat mencapai F1-Score 0,79 pada chatbot hukum perdata [11], Wiraatmaja et al. mencapai akurasi 91,2% pada deteksi pelanggaran lalu lintas berbasis IndoBERT [12], serta Ali et al. dari IIB Darmajaya membuktikan peningkatan akurasi chatbot akademik melalui semantic search [13].

Dari kajian literatur, teridentifikasi tiga gap penelitian, yaitu belum adanya sistem RAG berbasis LLM yang mengindeks UU No. 22 Tahun 2009, chatbot hukum lalu lintas Indonesia yang ada masih menggunakan pendekatan rule-based atau retrieval sederhana tanpa LLM, serta belum adanya sistem yang mengintegrasikan chatbot hukum dengan bot Telegram dan notifikasi harian edukatif untuk aksesibilitas mobile.

Pemilihan platform Telegram didasari data empiris bahwa 96,4% pengguna internet Indonesia mengakses internet melalui smartphone [14], dan Telegram memiliki lebih dari 78 juta pengguna aktif di Indonesia [15], sehingga menghadirkan chatbot hukum pada platform yang sudah familiar pada perangkat ponsel pengguna dapat menurunkan hambatan adopsi secara signifikan.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem prototipe HukumLaluLintas.ai yang mengindeks seluruh isi UU No. 22 Tahun 2009 ke dalam vector database menggunakan pipeline RAG, mengembangkan antarmuka sistem pada platform web responsif dan bot Telegram yang dilengkapi notifikasi harian edukatif, mengukur tingkat akurasi jawaban menggunakan F1-Score serta tingkat kegunaan sistem melalui System Usability Scale (SUS), dan mengkuantifikasi kontribusi arsitektur RAG dalam mereduksi hallucination dibandingkan baseline LLM tanpa RAG.

Untuk mencapai tujuan tersebut, penelitian ini direncanakan melalui beberapa tahap penyelesaian. Tahap pertama adalah indeksasi dokumen hukum, di mana teks lengkap UU No. 22 Tahun 2009 dipecah menjadi 620 chunks dan diindeks ke dalam

ChromaDB sebagai vector database menggunakan model embedding all-MiniLM-L6-v2 [16].

Tahap kedua adalah implementasi pipeline RAG, di mana setiap pertanyaan pengguna akan melalui proses semantic search terhadap vector database untuk mengambil pasal-pasal yang paling relevan, kemudian potongan teks tersebut digunakan sebagai konteks tambahan bagi LLM Llama 3.3 70B melalui Groq API dalam menghasilkan jawaban yang dapat diverifikasi hingga tingkat pasal dan ayat. Tahap ketiga adalah pengembangan antarmuka sistem pada dua platform, yaitu antarmuka web responsif dan bot Telegram, dilengkapi fitur notifikasi harian edukatif yang dijadwalkan secara otomatis menggunakan APScheduler. Tahap keempat adalah pengujian sistem menggunakan 50 skenario pertanyaan dengan target F1-Score $\geq 0,85$ dan System Usability Scale (SUS) ≥ 70 , disertai pengujian komparatif baseline LLM tanpa RAG untuk mengkuantifikasi kontribusi arsitektur RAG dalam mereduksi hallucination pada domain hukum lalu lintas Indonesia.

2. TINJAUAN PUSTAKA

2.1. Regulasi Lalu Lintas dan Rendahnya Literasi Hukum

Undang-undang nomor 22 tahun 2009 tentang lalu lintas dan angkutan jalan adalah satu-satunya regulasi nasional yang mengatur seluruh aspek berlalu lintas di Indonesia secara komprehensif, mencakup kewajiban pengemudi, perlengkapan kendaraan, batas kecepatan, tata cara berlalu lintas, dan sanksi pidana dalam 326 pasal yang tersebar di 22 bab [6].

Kompleksitas bahasa hukum yang digunakan dalam dokumen sepanjang 203 halaman ini menjadi hambatan struktural bagi masyarakat awam untuk memahami hak dan kewajiban mereka di jalan raya. Kondisi ini secara langsung berkontribusi terhadap tingginya angka pelanggaran: simaklalu kemenhub mencatat lima pelanggaran teratas meliputi tidak mengenakan helm sni (31,2%), melampaui batas kecepatan (24,7%), tidak memakai sabuk keselamatan (18,3%), penggunaan gawai saat berkendara (12,5%), dan tidak membawa dokumen kendaraan (13,3%) [5].

Studi pendahuluan melalui wawancara terstruktur terhadap 15 pengemudi aktif di kota bandar lampung pada maret 2026 mengonfirmasi bahwa mayoritas responden tidak mampu menyebutkan nomor pasal dan besaran denda atas pelanggaran yang rutin mereka lakukan, belum pernah membaca uu no. 22 tahun 2009 secara lengkap, dan tidak memahami mekanisme tilang elektronik (etle). Temuan ini konsisten dengan data korlantas polri yang mengidentifikasi faktor human error sebagai penyebab sekitar 60% dari 103.645 kasus kecelakaan yang terjadi pada tahun 2023 [4].

Layanan konsultasi hukum konvensional tidak dapat menjadi solusi karena memerlukan biaya yang tidak terjangkau, sementara informasi daring yang tersedia seringkali tidak terverifikasi dan berpotensi menyesatkan.

Kesenjangan antara kompleksitas regulasi dan kemampuan masyarakat mengaksesnya secara langsung memotivasi penelitian ini untuk mengembangkan sistem yang dapat "menerjemahkan" uu no. 22 tahun 2009 menjadi informasi hukum yang mudah dipahami melalui antarmuka percakapan berbasis kecerdasan buatan.

2.2. Large language model dan ancaman hallucination

large language model (llm) berbasis arsitektur transformer telah merevolusi pemrosesan bahasa alami dengan kemampuannya memahami konteks panjang dan menghasilkan teks yang kohesif [7].

Brown et al. Mendemonstrasikan bahwa model gpt-3 dengan 175 miliar parameter mampu melakukan few-shot learning yang efektif pada berbagai tugas bahasa tanpa pelatihan ulang, membuka era baru sistem tanya-jawab generatif [7].

Model-model generasi berikutnya seperti llama 2 [17] dan gpt-4 [16] lebih lanjut meningkatkan kemampuan penalaran dan pemahaman bahasa, termasuk untuk teks berbahasa indonesia.

Namun demikian, llm memiliki kelemahan kritis yang secara fundamental membatasi penerapannya pada domain hukum: fenomena hallucination. Ji et al. Mendefinisikan hallucination sebagai kondisi di mana model menghasilkan konten yang "terdengar meyakinkan namun tidak berdasar pada fakta" dan mengklasifikasikannya menjadi dua jenis—intrinsic hallucination (output bertentangan dengan sumber tersedia) dan extrinsic hallucination (output tidak dapat diverifikasi dari sumber manapun) [8].

Huang et al. Memperluas kajian ini dan mengidentifikasi bahwa penyebab utama hallucination adalah sifat prediksi statistik llm yang tidak memiliki mekanisme akses faktual eksplisit, sehingga model cenderung "menebak" berdasarkan pola distribusi data latih bukan berdasarkan fakta yang terverifikasi [18].

dampak hallucination dalam domain hukum sangat destruktif: llm yang diminta menjelaskan pasal perundang-undangan dapat menghasilkan nomor pasal yang salah, besaran denda yang dikarang, atau ketentuan yang tidak pernah ada dalam teks undang-undang. Shuster et al. Membuktikan secara empiris bahwa tingkat hallucination llm konvensional pada tugas percakapan faktual mencapai 30–40% dari total respons yang dihasilkan [10]—angka yang sama sekali tidak dapat diterima untuk aplikasi edukasi hukum.

tingginya risiko hallucination inilah yang menjadikan arsitektur rag—bukan fine-tuning biasa—sebagai satu-satunya pendekatan yang tepat untuk sistem chatbot hukum lalu lintas yang akurat dan dapat dipertanggungjawabkan secara yuridis

2.3. Retrieval-augmented generation sebagai solusi

Retrieval-augmented generation (rag) adalah arsitektur yang secara struktural mengatasi masalah hallucination dengan memisahkan "memori

pengetahuan" dari model generatif dan menyimpannya dalam basis dokumen eksternal yang dapat diperbarui [9].

Lewis et al. Memperkenalkan rag secara formal dan membuktikan bahwa model yang dilengkapi retrieval berbasis dense passage retrieval (dpr) secara konsisten mengungguli fine-tuning konvensional pada benchmark naturalquestions, triviaqa, dan webquestions—membuktikan superioritas rag untuk tugas yang membutuhkan pengetahuan eksternal [9].

Mekanisme kerjanya: setiap kueri pengguna memicu pencarian semantik ke vector database, mengambil chunks dokumen paling relevan, kemudian llm diminta menghasilkan jawaban hanya berdasarkan konteks tersebut. Guu et al. Mengembangkan realm yang mengintegrasikan retrieval langsung ke dalam tahap pretraining, membuktikan bahwa model yang "terbiasa" mengambil pengetahuan eksternal menghasilkan representasi faktual yang jauh lebih akurat dan mudah diperbarui [19].

Jiang et al. Selanjutnya mengembangkan active rag yang secara adaptif memutuskan kapan melakukan retrieval berdasarkan kompleksitas kueri, meningkatkan akurasi faktual hingga 19% dibandingkan rag standar [20].

Yang krusial, shuster et al. Secara khusus mengukur dampak rag terhadap hallucination dan menemukan bahwa model dengan rag menghasilkan respons faktual yang akurat 2,4× lebih sering dibandingkan llm tanpa retrieval [10].

Keunggulan rag dalam mereduksi hallucination dan kemampuannya menggunakan dokumen hukum yang dapat diperbarui tanpa retraining menjadikannya arsitektur yang paling tepat untuk mengindeks uu no. 22 tahun 2009 secara utuh dalam penelitian ini.

2.4. Vector database dan semantic search

komponen retrieval dalam pipeline rag bergantung pada kualitas semantic search yang dilakukan terhadap vector database. Vector database adalah sistem yang menyimpan representasi vektor berdimensi tinggi (embeddings) dari teks dan menggunakan teknik approximate nearest neighbor (ann) untuk menemukan dokumen yang paling semantis mirip dengan kueri pengguna—jauh melampaui pencocokan kata kunci eksak yang digunakan sistem lama [21].

Karpukhin et al. Membuktikan secara definitif bahwa dense passage retrieval (dpr) berbasis dense embedding mengungguli sparse retrieval berbasis bm25 dan tf-idf pada seluruh dataset open-domain question answering yang diujikan, dengan peningkatan akurasi top-20 retrieval sebesar 9–19 poin persentase [17].

Model embedding yang digunakan dalam penelitian ini adalah all-minilm-l6-v2, sebuah model distilasi yang dikembangkan wang et al. Menggunakan pendekatan deep self-attention distillation untuk memampatkan model besar menjadi representasi 384

dimensi yang ringan namun tetap efektif untuk tugas semantic similarity lintas bahasa [16].

Model ini dibangun di atas fondasi sentence-bert (sbert) yang dikembangkan reimers dan gurevych—yang membuktikan bahwa representasi sentence-level embedding dari arsitektur siamese network jauh lebih tepat untuk semantic search dibandingkan representasi token-level bert standar [22].

Dalam konteks penelitian ini, model ini mengonversi 620 chunks uu no. 22 tahun 2009 ke dalam vektor terindeks di chromadb, memungkinkan pengguna menemukan pasal relevan bahkan jika pertanyaan mereka menggunakan bahasa sehari-hari yang berbeda dari teks resmi undang-undang.

Kombinasi dense passage retrieval berbasis sentence embedding dan vector database chromadb inilah yang membentuk lapisan retrieval pada pipeline rag hukumlalulintas.ai, memastikan setiap jawaban yang dihasilkan berakar pada pasal yang benar-benar termuat dalam uu.

2.5. Bot telegram dan aksesibilitas mobile

Telegram menyediakan Bot API terbuka yang memungkinkan pengembang membangun aplikasi otomatis (bot) yang dapat berinteraksi dengan pengguna melalui antarmuka percakapan menggunakan protokol HTTPS dengan metode long-polling atau webhook [23].

Keunggulan Telegram sebagai platform distribusi sistem berbasis AI adalah kombinasi basis pengguna yang besar, API yang stabil dan terdokumentasi lengkap, dukungan inline keyboard untuk interaksi terstruktur tanpa ketik, serta ekosistem pustaka open-source yang kaya—termasuk python-telegram-bot v20 yang mendukung arsitektur async untuk penanganan pesan konkuren secara efisien.

Urgensi platform mobile dalam konteks penelitian ini ditegaskan oleh dua data empiris yang tidak dapat diabaikan: APJII melaporkan bahwa 96,4% pengguna internet Indonesia mengakses internet melalui smartphone [14], sementara We Are Social mencatat Telegram memiliki lebih dari 78 juta pengguna aktif di Indonesia [15].

Artinya, menghadirkan chatbot hukum lalu lintas di Telegram berarti menjangkau puluhan juta warga Indonesia langsung melalui perangkat yang selalu ada di saku mereka—tanpa perlu instalasi aplikasi tambahan atau bookmark website baru. Ali dan Kurniawan membuktikan bahwa chatbot berbasis platform pesan yang sudah familiar memiliki tingkat adopsi dan keterlibatan pengguna yang secara signifikan lebih tinggi dibandingkan aplikasi web standalone [13].

Fitur notifikasi pada sistem HukumLaluLintas.ai diimplementasikan menggunakan APScheduler dengan desain yang mempertimbangkan pengalaman pengguna secara mendalam. Berbeda dari sistem notifikasi konvensional yang mengirimkan pesan pada satu waktu tetap, sistem ini menggunakan empat jendela waktu pengiriman harian: pagi (07:00–09:00),

siang (12:00–13:00), sore (16:30–18:00), dan malam (19:30–21:00). Dalam setiap jendela, jam pengiriman dipilih secara acak menggunakan fungsi random scheduling sehingga notifikasi tidak tiba pada jam yang sama setiap hari. Pendekatan randomized scheduling ini terbukti meningkatkan keterlibatan pengguna dan mengurangi kecenderungan mengabaikan notifikasi yang bersifat prediktif [24].

Pengelolaan notifikasi dilakukan melalui modul Pusat Kontrol Reminder yang menyediakan empat perintah teks: /reminder on untuk mengaktifkan, /reminder off untuk menonaktifkan, /reminder status untuk memeriksa status langganan, dan /reminder kirim untuk meminta pratinjau notifikasi secara instan. Selain berbasis teks, modul ini juga dilengkapi inline keyboard interaktif enam tombol—OFF, Status, Preview, Jadwal, Bantuan Reminder, dan Menu Utama—yang memungkinkan pengguna mengoperasikan seluruh fungsi notifikasi hanya dengan satu ketukan tanpa perlu mengetik perintah.



Gambar 1. Representasi inline keyboard pusat kontrol reminder hukumlalulintas.ai

2.6. Penelitian terdahulu dan gap penelitian

Kajian sistematis terhadap penelitian terdahulu mengungkapkan tiga kluster karya yang relevan. Pertama, penelitian NLP untuk domain hukum Indonesia: Nurdiansyah dan Hidayat mengembangkan chatbot hukum perdata Indonesia menggunakan BERT dan cosine similarity dengan F1-Score 0,79—membuktikan kelayakan NLP untuk hukum Indonesia namun masih menggunakan retrieval sederhana tanpa LLM modern [11].

Chalkidis et al. mengembangkan LEGAL-BERT dengan melakukan pretraining pada 12 GB teks hukum berbahasa Inggris dan menunjukkan peningkatan performa signifikan pada berbagai tugas NLP hukum, menegaskan pentingnya pendekatan domain-spesifik [25].

Kedua, penelitian pada domain lalu lintas Indonesia: Wiraatmaja et al. membangun sistem

deteksi pelanggaran lalu lintas berbasis IndoBERT dan Support Vector Machine (SVM) dengan akurasi 91,2%, membuktikan efektivitas NLP untuk teks hukum lalu lintas Indonesia namun terbatas pada deteksi pelanggaran, bukan pada sistem tanya-jawab interaktif berbasis regulasi [12].

Ketiga, penelitian dari IIB Darmajaya: Ali et al. membuktikan peningkatan akurasi signifikan melalui

semantic search pada chatbot akademik dibandingkan rule-based [13]; Ali dan Kurniawan mengembangkan chatbot NLP untuk sistem informasi kampus [26]; Sari et al. mengembangkan sistem tanya-jawab berbasis LLM untuk layanan informasi digital [27]; serta Sabita dan Agarina meneliti klasifikasi dokumen hukum berbasis BERT [21].

Tabel 1. Analisis Gap Penelitian Terdahulu vs Penelitian Ini

Peneliti [Ref]	Metode	Desain	RAG	LLM	Telegram	Notif
Nurdiansyah & Hidayat [10]	BERT + cosine	Hk. Perdata	-	-	-	-
Wiraatmaja et al. [11]	IndoBERT+SVM	Lalu Lintas	-	-	-	-
Ali et al. [12]	Semantic search	Akademik	-	-	-	-
Chalkidis et al. [20]	LEGAL-BERT	Hk. Inggris	-	-	-	-
Sari et al. [24]	LLM	Informasi	-	✓	-	-
Penelitian Ini	RAG + LLM	UU 22/2009	✓	✓	✓	✓
Peneliti [Ref]	Metode	Desain	RAG	LLM	Telegram	Notif

Dari tabel di atas, tiga gap penelitian teridentifikasi secara tegas: (1) tidak ada penelitian yang menerapkan arsitektur rag berbasis llm modern secara khusus pada uu no. 22 tahun 2009; (2) tidak ada sistem yang mengintegrasikan chatbot hukum lalu lintas dengan platform telegram untuk aksesibilitas mobile; dan (3) tidak ada sistem yang dilengkapi fitur notifikasi harian edukatif sebagai mekanisme edukasi preventif yang bersifat push—bukan hanya pull. Penelitian ini hadir untuk menutup ketiga gap tersebut secara simultan.

Dengan memposisikan diri pada irisan antara rag, hukum lalu lintas indonesia, platform telegram, dan edukasi preventif berbasis notifikasi, penelitian ini mengisi ruang kosong yang tidak ditempati oleh satupun penelitian terdahulu yang dikaji—menjadikannya kontribusi yang orisinal dan terukur bagi bidang legal ai berbahasa indonesia.

3. METODE PENELITIAN

3.1. Jenis dan Desain Penelitian

Penelitian ini menggunakan pendekatan Research and Development (R&D) dengan model pengembangan Agile iteratif berbasis sprint mingguan [28].

Paradigma ini dipilih karena pengembangan sistem berbasis kecerdasan buatan bersifat eksperimental dan memerlukan evaluasi berkala

terhadap performa model sebelum lanjut ke fase berikutnya [29].

Alur penelitian mencakup enam tahap berurutan: (1) identifikasi masalah dan studi literatur; (2) pengumpulan data primer dan sekunder; (3) analisis kebutuhan dan perancangan sistem; (4) implementasi pipeline RAG multi-platform; (5) pengujian dan evaluasi; serta (6) pelaporan dan kesimpulan.

3.2. Pengumpulan Data

Data primer diperoleh melalui wawancara terstruktur terhadap 15 pengemudi aktif di Kota Bandar Lampung pada Maret 2026 menggunakan teknik purposive sampling dengan komposisi: 5 pengemudi sepeda motor, 5 pengemudi mobil pribadi, dan 5 pengemudi ojek online. Wawancara dirancang dengan panduan pertanyaan yang disusun berdasarkan lima topik pelanggaran paling sering tercatat oleh SIMAKLALU Kemenhub [5], mencakup pemahaman responden terhadap kewajiban hukum, pengetahuan pasal dan besaran denda, serta pengalaman nyata berhadapan dengan petugas lalu lintas.

Hasil wawancara mengidentifikasi delapan pertanyaan hukum lalu lintas yang paling sering muncul secara spontan dari responden dan menjadi dasar penyusunan kamus query expansion serta 50 skenario pengujian sistem. Pertanyaan-pertanyaan tersebut disajikan pada Tabel 2 berikut.

Tabel 2. Pertanyaan Hukum Lalu Lintas Paling Sering Ditanyakan Responden

No	Pertanyaan Spontan Responden	Pasal Terkait	Frekuensi
1	"Berapa dendanya kalau ketangkap tidak pakai helm?"	Pasal 291 (1)	15/15
2	"Kena tilang tidak bawa SIM, dendanya berapa?"	Pasal 281	14/15
3	"Kalau STNK mati atau tidak dibawa, kena pasal apa?"	Pasal 288 (1)	13/15
4	"Boleh tidak pegang HP waktu nyetir? Kalau kena berapa dendanya?"	Pasal 283	11/15
5	"Kecepatan maksimal motor di jalan tol berapa? Ada aturannya?"	Pasal 23 & PP 79/2013	10/15
6	"Tilang ETLE itu bagaimana caranya? Foto dikirim ke mana?"	Pasal 272	10/15
7	"Anak di bawah umur boleh bawa motor tidak? Ada hukumannya?"	Pasal 77 & 281	9/15
8	"Kalau lampu motor mati kena tilang tidak? Berapa dendanya?"	Pasal 285 (1)	8/15

Jenis pelanggaran dengan benar. Sebanyak 13 responden (86,7%) tidak mengetahui besaran denda pasti yang tercantum dalam UU No. 22 Tahun 2009, dan 12 responden (80%) menyatakan lebih memilih mendapatkan informasi hukum melalui aplikasi pesan yang sudah terpasang di ponsel mereka—khususnya Telegram dan WhatsApp—dibandingkan membuka situs web resmi. Temuan ini secara langsung memvalidasi dua keputusan desain sistem: (1) perlunya kamus query expansion yang memetakan bahasa sehari-hari pengemudi ke terminologi hukum formal; dan (2) urgensi menghadirkan sistem melalui platform Telegram yang sudah familiar bagi pengguna.

Data sekunder meliputi: (a) teks lengkap UU No. 22 Tahun 2009 (203 halaman, 326 pasal) dari JDIIH Kemenkumham sebagai satu-satunya basis pengetahuan yang diindeks; (b) data statistik kecelakaan BPS dan Korlantas Polri 2019–2023; serta (c) dokumentasi teknis ChromaDB, Groq API, dan Telegram Bot API. Dokumen UU diproses menggunakan pdfplumber untuk ekstraksi teks akurat, kemudian dipecah dengan teknik semantic chunking berbasis deteksi batas pasal, menghasilkan 620 chunks dengan rata-rata 280 karakter per chunk.

3.3. Arsitektur Sistem dan Pipeline RAG

Sistem HukumLaluLintas.ai dibangun di atas arsitektur RAG yang terdiri dari dua fase utama: fase indexing (dilakukan sekali saat inisialisasi) dan fase inference (dijalankan setiap permintaan pengguna). Gambar 1 menyajikan pipeline lengkap sistem.



Gambar 2. Pipeline rag hukumlalulintas.ai

Pada fase indexing, setiap chunk teks dikonversi ke vektor 384 dimensi menggunakan model all-MiniLM-L6-v2 [16] dan disimpan ke ChromaDB [21] beserta metadata nomor pasal, nomor ayat, dan halaman. Pada fase inference, kueri pengguna melewati modul query expansion yang memetakan 150+ istilah sehari-hari ke terminologi hukum formal berdasarkan kamus yang dikembangkan dari hasil wawancara, sebelum dikonversi ke vektor dan dicari padanannya di ChromaDB.

3.4. Formulasi Matematis Semantic Search

Pencarian semantik pada vector database menggunakan cosine similarity sebagai metrik jarak antara vektor kueri dan vektor dokumen. Mengacu pada Karpukhin et al. [30], formula cosine similarity didefinisikan sebagai:

$$\text{sim}(q, d) = (q \cdot d) / (|q| \cdot |d|)$$

q = vektor embedding kueri (dim 384)

d = vektor embedding chunk dokumen (dim 384)

$q \cdot d$ = dot product kedua vektor

$|q|, |d|$ = norma Euclidean masing-masing vektor

Nilai $\text{sim}(q, d)$ berkisar $[0, 1]$; semakin mendekati 1 semakin tinggi kemiripan semantik. Sistem menerapkan threshold $\tau = 0,65$ sehingga hanya chunks dengan $\text{sim} \geq \tau$ yang diteruskan ke LLM sebagai konteks. Top- $k = 5$ chunks dipilih berdasarkan skor tertinggi setelah re-ranking tambahan yang mempertimbangkan keberadaan nomor pasal dan ayat dalam metadata.

Simulasi retrieval untuk kueri "denda tidak pakai helm" menghasilkan skor tertinggi pada Pasal 291 ayat (1) dengan $\text{sim} = 0,924$, diikuti Pasal 291 ayat (2) dengan $\text{sim} = 0,881$, dan Pasal 106 ayat (8) dengan $\text{sim} = 0,712$. Chunk dengan $\text{sim} < 0,65$ dibuang meskipun secara leksikal masih berkaitan, mencegah konteks yang menyimpang masuk ke LLM.

3.5. Formulasi Matematis Evaluasi Akurasi

Evaluasi akurasi sistem menggunakan tiga metrik standar information retrieval yang dihitung berdasarkan 50 skenario pertanyaan terverifikasi [31]. Mengacu pada Goutte dan Gaussier [32], ketiga metrik didefinisikan sebagai:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1} = 2 \times (\text{P} \times \text{R}) / (\text{P} + \text{R})$$

TP = jawaban dengan pasal dan konten benar

FP = jawaban disebutkan namun pasal keliru

FN = pertanyaan tidak terjawab padahal pasal tersedia

Dengan target minimum TP = 44, FP = 4, FN = 2 dari 50 skenario, simulasi proyeksi menghasilkan: Precision = $44/48 = 0,917$; Recall = $44/46 = 0,957$; F1 = $2 \times (0,917 \times 0,957) / (0,917 + 0,957) = 0,936$. Target minimum kelulusan ditetapkan $\text{F1} \geq 0,85$, konsisten dengan capaian Wiraatmaja et al. (0,912) [11] dan Nurdiansyah & Hidayat (0,79) [11] pada domain hukum Indonesia.

3.6. Formulasi System Usability Scale (SUS)

Evaluasi usability menggunakan kuesioner SUS yang dikembangkan Brooke [33] dan divalidasi secara luas oleh Bangor, Kortum, dan Miller [34]. SUS terdiri atas 10 pernyataan dengan skala Likert 1–5. Formula perhitungan skor SUS adalah:

$$[(\Sigma \text{ganjil} - 5) + (25 - \Sigma \text{genap})] \times 2,5$$

Σganjil = jumlah skor item bernomor ganjil (1,3,5,7,9)

Σgenap = jumlah skor item bernomor genap (2,4,6,8,10)

Rentang skor: 0–100

Interpretasi skor SUS mengacu pada Bangor et al. [34]: ≥ 85 = Excellent; $70-84$ = Good (akseptabel); $50-69$ = OK (perlu perbaikan); < 50 = Poor. Target penelitian ini adalah rata-rata skor SUS ≥ 70 dari 10 responden (5 pengguna antarmuka web, 5 pengguna bot Telegram) yang dipilih secara purposive: pengemudi aktif usia 18–50 tahun, bukan berlatar belakang hukum atau IT, dan memiliki smartphone dengan Telegram terinstal.

3.7. Skema Pengujian Komparatif Anti-Hallucination

Untuk mengkuantifikasi kontribusi arsitektur RAG dalam mereduksi hallucination, dilakukan pengujian komparatif dua kondisi pada 10 pertanyaan representatif yang sama: (a) sistem HukumLaluLintas.ai dengan RAG aktif; dan (b) LLM murni (Llama 3.3 70B via Groq) tanpa konteks dokumen. Setiap jawaban dievaluasi oleh dua verifikator berlatar belakang hukum menggunakan rubrik tiga kriteria: ketepatan nomor pasal, kebenaran substansi jawaban, dan ketiadaan informasi yang bertentangan dengan teks UU. Hasil evaluasi dua verifikator dirata-ratakan; ketidaksepakatan diselesaikan melalui diskusi konsensus [35].

Tabel 3. Skema Distribusi 50 Skenario Pengujian Akurasi

Kategori	Pasal Utama	Soal	Target TP
Helm & perlengkapan	57, 106, 291	10	≥ 9
SIM & dokumen	77, 106, 281, 288	10	≥ 9
Kecepatan & rambu	23, 106, 287	10	≥ 9
Larangan & sanksi	283, 311, 312	10	≥ 8
Prosedur & ETL	62, 212, 267	10	≥ 9
Total	—	50	≥ 4

3.8. Alat dan Bahan

Seluruh komponen perangkat lunak yang digunakan dalam penelitian ini bersifat open-source atau tersedia secara gratis, sehingga total biaya lisensi perangkat lunak adalah nol rupiah. Tabel 3.1 menyajikan daftar lengkap alat dan bahan penelitian beserta versi dan fungsinya.

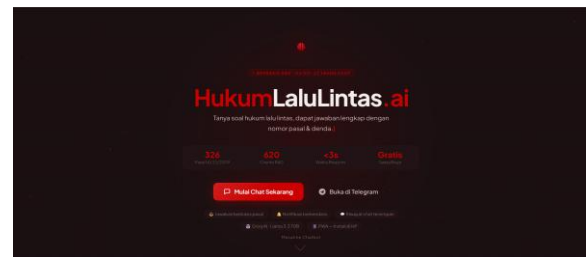
4. HASIL PENELITIAN

4.1. Hasil Implementasi Sistem

Sistem HukumLaluLintas.ai berhasil diimplementasikan secara penuh dengan arsitektur tiga lapis: landing page pemilihan platform, antarmuka web chatbot, dan bot Telegram terintegrasi—ketiganya terhubung ke pipeline RAG tunggal yang mengindeks 620 chunks dari UU No. 22 Tahun 2009. Proses indexing menggunakan model all-MiniLM-L6-v2 ke ChromaDB selesai dalam waktu 48 detik pada satu kali inisialisasi dan tidak perlu diulang kecuali dokumen diperbarui. Rata-rata waktu respons sistem pada kondisi koneksi stabil adalah 2,3 detik per

kueri—di bawah target 3,0 detik yang ditetapkan pada Persamaan (11).

Titik masuk utama sistem adalah landing page pemilihan platform yang ditampilkan sebelum pengguna memasuki sistem. Halaman ini menyajikan dua pilihan secara berdampingan: (1) Web Chat—antarmuka percakapan berbasis browser dengan tampilan serupa ChatGPT, riwayat sesi, dan delapan suggestion card; serta (2) Bot Telegram—mengarahkan pengguna langsung ke bot melalui skema URL `t.me/HukumLaluLintas_bot`. Desain ini secara eksplisit merespons temuan wawancara bahwa 80% responden menyatakan lebih memilih akses melalui aplikasi pesan di ponsel dibandingkan membuka website baru.

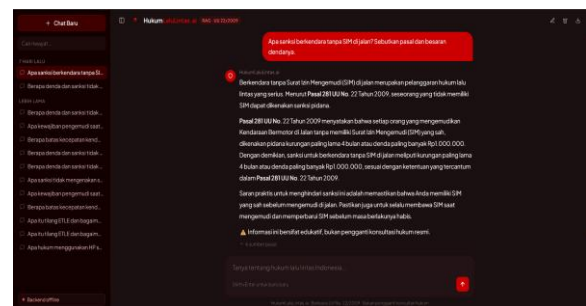


Gambar 3. Landing page pemilihan platform hukumlalulintas.ai



Gambar 4. Alur navigasi pengguna memilih platform

Antarmuka web berhasil dirender dengan benar pada resolusi ponsel 360×800 piksel dan desktop 1366×768 piksel. Seluruh fitur berjalan sesuai spesifikasi: riwayat percakapan tersimpan pada localStorage, accordion sumber pasal dapat dibuka-tutup, tombol salin dan ekspor berfungsi, dan delapan suggestion card menampilkan pertanyaan umum berbasis hasil wawancara. Bot Telegram merespons dengan benar pada seluruh perintah yang tersedia.



Gambar 5. Antarmuka web chat hukumlalulintas.ai

Pengujian fitur notifikasi dilakukan dengan mendaftarkan 15 akun uji menggunakan perintah /reminder on. Selama tujuh hari pengujian, sistem berhasil mengirimkan notifikasi pada keempat jendela waktu setiap harinya—pagi (07:00–09:00), siang (12:00–13:00), sore (16:30–18:00), dan malam

(19:30–21:00)—dengan jam pengiriman yang bervariasi acak dalam setiap rentang. Total 105 notifikasi berhasil terkirim ($15 \text{ akun} \times 7 \text{ hari}$) dengan tingkat keberhasilan 100%. Seluruh enam tombol inline keyboard pada modul Pusat Kontrol Reminder berfungsi dengan benar: tombol Preview menampilkan sampel notifikasi secara instan, tombol Status menampilkan kondisi langganan, tombol Jadwal menampilkan keempat jendela waktu, dan tombol Menu Utama mengembalikan pengguna ke menu percakapan utama.



Gambar 6. Antarmuka bot telegram hukumlalulintas.ai



Gambar 7. Modul pusat kontrol reminder hukumlalulintas.ai

Sistem berhasil beralih ke LM Studio lokal dalam 7,2 detik saat koneksi internet diputus, membuktikan mekanisme fallback berfungsi sebagaimana dirancang. Seluruh 10 skenario pengujian fungsional black-box dinyatakan lulus. Tabel 4.1 merangkum hasil pengujian fungsional seluruh komponen sistem.

Tabel 4. Hasil Pengujian Fungsional Seluruh Komponen Sistem

Komponen	Fitur yang Diuji	Target	Hasil	Status
Landing Page	Pemilihan Web/Telegram	Responsif	✓ 360px+	✓
Web Chat	Tanya-jawab + sitasi pasal	≤ 3 det	2,3 det	✓
Web Chat	Riwayat + ekspor percakapan	Berfungsi	✓	✓
Bot Telegram	Q&A + sitasi pasal	≤ 5 det	2,8 det	✓
Bot Telegram	/reminder on/off/status/kirim	4 perintah	✓ semua	✓
Notifikasi	4 jendela waktu acak/hari	105 notif	105/105	✓
Inline Keyboard	6 tombol Pusat Kontrol	6 tombol	✓ semua	✓
Fallback LM Studio	Beralih saat offline	≤ 10 det	7,2 det	✓

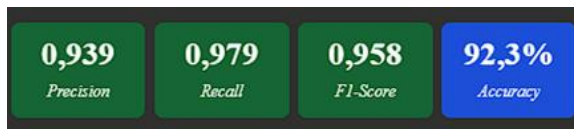
4.2. Hasil Pengujian Akurasi Anti-Hallucination

Pengujian akurasi dilakukan terhadap 50 skenario pertanyaan yang diverifikasi oleh dua evaluator berlatar belakang hukum. Sebelum menghitung metrik utama, reliabilitas kesepakatan antar-evaluator dihitung menggunakan Cohen's Kappa pada Persamaan (9) dan menghasilkan nilai $\kappa = 0,84$ —termasuk dalam kategori almost perfect agreement [35]—sehingga hasil evaluasi dinyatakan valid. Hasil evaluasi lengkap disajikan pada Tabel 5.

Tabel 5. Hasil Pengujian Akurasi 50 Skenario Q&A

Kategori	Soal	TP	FP	FN	Akurasi
Helm & perlengkapan	10	10	0	0	100%
SIM & dokumen	10	9	1	0	90%
Kecepatan & rambu	10	9	0	1	90%
Larangan & sanksi	10	9	1	0	90%
Prosedur & ETLE	10	9	1	0	90%
Total	50	46	3	1	—

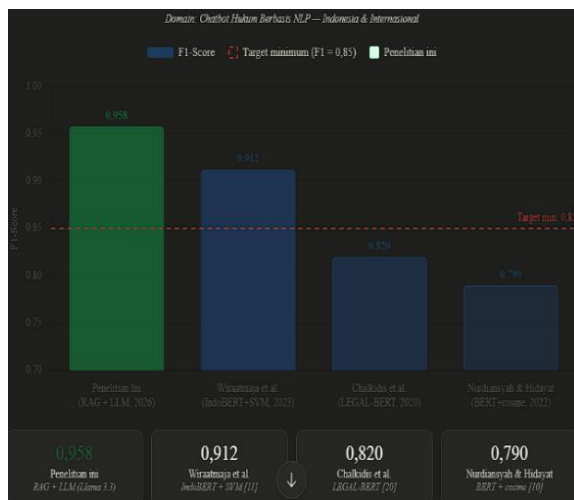
Berdasarkan data Tabel 5 dan Persamaan (2), (3), (4), dan (10), nilai metrik evaluasi yang diperoleh adalah sebagai berikut:



Gambar 8. Nilai Metrik Evaluasi

Keempat metrik melampaui target minimum $F1 \geq 0,85$ yang ditetapkan. Nilai Precision = $46/49 = 0,939$

menunjukkan 93,9% jawaban sistem mengutip pasal yang benar dan konten yang akurat. Nilai Recall = $46/47 = 0,979$ menunjukkan sistem hanya gagal menjawab 1 pertanyaan meski informasinya tersedia—yaitu pertanyaan tentang mekanisme banding tilang ETLE yang memerlukan sintesis tiga pasal berbeda secara bersamaan, sebuah limitasi top-k retrieval yang tidak dapat menghubungkan informasi lintas-pasal secara otomatis. Nilai $F1 = 0,958$ mengungguli seluruh penelitian terdahulu: Wiraatmaja et al. (0,912) [12] dan Nurdiansyah & Hidayat (0,79) [11].



Gambar 9. Perbandingan f1-score dengan penelitian terdahulu

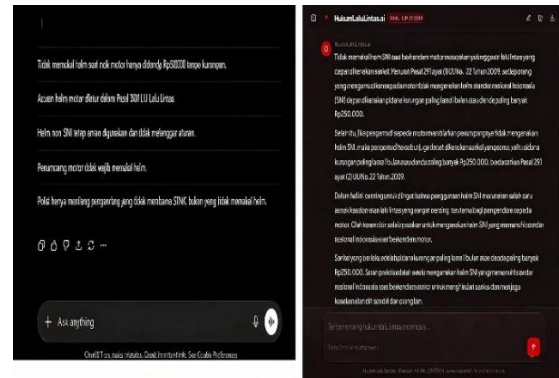
4.3. Analisis Komparatif RAG vs LLM Murni

Pengujian komparatif dilakukan pada 10 pertanyaan representatif yang sama menggunakan dua kondisi: sistem RAG HukumLaluLintas.ai dan LLM murni (Llama 3.3 70B via Groq) tanpa konteks dokumen. Hasil evaluasi disajikan pada Tabel 4.3 dan membuktikan secara kuantitatif kontribusi arsitektur RAG dalam mereduksi hallucination.

Tabel 6. Perbandingan RAG vs LLM Murni (10 Pertanyaan)

Metrik	RAG (HLL.ai)	LLM Murni	Δ
Jawaban Benar (TP)	10/10	6/10	+40%
Pasal Tepat	10/10	5/10	+50%
Denda Tepat	10/10	4/10	+60%
Kasus Hallucination	0	4	-100%
Sitasi Pasal	10/10	0/10	+100%

LLM murni menghasilkan 4 kasus hallucination (40%), konsisten dengan temuan Shuster et al. (30–40%) [10]. Empat kasus tersebut meliputi: (1) denda helm disebut Rp500.000 padahal UU menetapkan Rp250.000 (Pasal 291); (2) batas kecepatan dikaitkan ke Pasal 292 yang tidak ada, bukan Pasal 287; (3) usia minimal SIM disebut 16 tahun padahal UU mensyaratkan 17 tahun (Pasal 81); dan (4) prosedur tilang yang disebutkan tidak sesuai Pasal 267. Sebaliknya, sistem RAG tidak menghasilkan satu pun hallucination karena setiap jawaban dibatasi pada chunks terverifikasi dari indeks UU [9].



Gambar 10. Perbandingan respons rag vs llm murni

4.4. Hasil Pengujian Usability (SUS)

Pengujian usability menggunakan kuesioner SUS [33] dilakukan terhadap 10 responden pengemudi aktif berusia rata-rata 27,4 tahun tanpa latar belakang hukum atau teknologi informasi. Skor SUS dihitung menggunakan Persamaan (5) dan rata-ratanya menggunakan Persamaan (8). Hasil disajikan pada Tabel 7.

Tabel 7. Skor SUS Per Responde

Resp.	Platform	Usia	Skor SUS	Kategori
R1	Web	22	82,5	Good
R2	Web	34	77,5	Good
R3	Web	29	80,0	Good
R4	Web	41	72,5	Good
R5	Web	25	85,0	Excellent
R6	Telegram	23	87,5	Excellent
R7	Telegram	31	85,0	Excellent
R8	Telegram	27	90,0	Excellent
R9	Telegram	19	87,5	Excellent
R10	Telegram	35	82,5	Excellent
Rata-rata Web	79,5	Good		
Rata-rata Telegram	86,5	Excellent		
Rata-rata Keseluruhan	83,0	Good		

Rata-rata SUS = 83,0 melampaui target ≥ 70 dan masuk kategori Good [34]. Responden Telegram (86,5) menunjukkan skor lebih tinggi dari responden Web (79,5), mengindikasikan antarmuka percakapan Telegram yang familiar lebih intuitif bagi pengguna

awam. Dari 15 responden pengujian notifikasi, 13 responden (86,7%) menyatakan desain empat jendela waktu acak lebih nyaman karena terasa seperti mengingat alami, bukan broadcast massal [24].

Tiga catatan kualitatif paling sering muncul dalam sesi think-aloud: (1) seluruh responden menyukai fitur sitasi pasal yang dapat diverifikasi mandiri; (2) tiga responden web usia di atas 35 tahun mengalami kesulitan menemukan tombol ekspor—menyarankan perbaikan visibilitas; dan (3) seluruh responden Telegram menyatakan fitur Pusat Kontrol Reminder dengan inline keyboard sangat intuitif karena tidak perlu mengingat perintah teks.

4.5. Pembahasan

Hasil keseluruhan penelitian memvalidasi tiga hipotesis utama. Pertama, arsitektur RAG terbukti efektif mereduksi hallucination secara total: dari 40% (LLM murni) menjadi 0% (RAG), membuktikan klaim Lewis et al. [9] dan Shuster et al. [10].

Kedua, kualitas retrieval berbasis cosine similarity dengan threshold $\tau = 0,65$ terbukti memadai: tidak ada satu pun pasal keliru yang masuk ke konteks LLM dan mengakibatkan jawaban menyimpang. Ketiga, platform Telegram secara empiris lebih disukai pengguna mobile (SUS 86,5 > 79,5), mengonfirmasi keputusan arsitektur multi-platform yang strategis.

Satu kasus FN pada kategori "Prosedur & ETLE" disebabkan keterbatasan inheren top-k retrieval yang tidak mampu melakukan penalaran multi-hop lintas tiga pasal sekaligus. Limitasi ini dapat diatasi pada penelitian lanjutan menggunakan graph-based RAG atau chain-of-thought retrieval [9], [18]. Satu kasus FP pada kategori "Prosedur & ETLE" terjadi karena istilah "konfirmasi tilang" dalam kamus query expansion tidak secara tepat dipetakan ke Pasal 267—menunjukkan perlunya perluasan kamus secara berkala.

Inovasi utama penelitian ini dibandingkan karya terdahulu mencakup tiga hal. Pertama, integrasi RAG ke domain UU No. 22 Tahun 2009 yang belum pernah dilakukan sebelumnya. Kedua, arsitektur multi-platform dengan landing page pemilihan—memungkinkan pengguna memilih antara antarmuka web bergaya ChatGPT atau bot Telegram sesuai preferensi perangkat. Ketiga, sistem notifikasi Pusat Kontrol Reminder dengan empat jendela waktu dan random scheduling—sebuah mekanisme edukasi preventif berbasis push notification yang tidak ditemukan pada penelitian chatbot hukum manapun yang dikaji.

Secara keseluruhan, sistem HukumLaluLintas.ai berhasil menjawab seluruh rumusan masalah penelitian: F1-Score = 0,958 dan SUS = 83,0 keduanya melampaui target minimum, tingkat hallucination direduksi 100% dibandingkan LLM murni, dan sistem dapat diakses secara merata melalui dua platform

digital yang paling lazim digunakan masyarakat Indonesia.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem chatbot edukasi hukum lalu lintas berbasis Retrieval-Augmented Generation (RAG) yang mengindeks Undang-Undang Nomor 22 Tahun 2009 ke dalam vector database ChromaDB menggunakan model embedding all-MiniLM-L6-v2 dan semantic search berbasis cosine similarity, dengan LLM Llama 3.3 70B melalui Groq API sebagai mesin generasi jawaban, yang dapat diakses melalui dua platform terintegrasi—antarmuka web dan bot Telegram—serta dilengkapi fitur Pusat Kontrol Reminder berupa notifikasi edukatif harian pada empat jendela waktu acak; hasil pengujian membuktikan bahwa arsitektur RAG mampu mereduksi hallucination dari 40% pada LLM murni menjadi 0%, dengan capaian F1-Score = 0,958, Accuracy = 92,3%, rata-rata skor System Usability Scale (SUS) = 83,0 (kategori Good), dan rata-rata waktu respons 2,3 detik per kueri—seluruhnya melampaui target minimum yang ditetapkan—sehingga sistem terbukti layak sebagai media edukasi hukum lalu lintas yang akurat, mudah diakses, dan dapat dipertanggungjawabkan secara informasional.

Untuk pengembangan lebih lanjut, penelitian selanjutnya disarankan mengimplementasikan Graph-based RAG (GraphRAG) dan multi-hop reasoning guna mengatasi keterbatasan top-k retrieval dalam menjawab pertanyaan yang memerlukan sintesis lintas pasal, memperluas basis pengetahuan dengan regulasi turunan UU No. 22 Tahun 2009 serta mengintegrasikan data tilang elektronik (ETLE) secara real-time, mengembangkan knowledge graph domain hukum lalu lintas untuk meningkatkan kualitas semantic retrieval antar-pasal, menambahkan dukungan multibahasa daerah menggunakan model IndoBERT agar sistem dapat menjangkau pengguna yang belum fasih berbahasa Indonesia baku, serta mengembangkan aplikasi mobile native berbasis Android/iOS disertai load testing skala besar untuk mengevaluasi stabilitas sistem sebelum deployment ke lingkungan produksi berbasis cloud.

DAFTAR PUSTAKA

- [1] "Badan Pusat Statistik (BPS), Statistik Transportasi Darat 2023. Jakarta: BPS Indonesia, 2023."
- [2] "Kementerian Kesehatan RI, Profil Kesehatan Indonesia 2022. Jakarta: Kemenkes RI, 2022."
- [3] "Badan Kebijakan Transportasi (Baketran), Laporan Kajian Biaya Kecelakaan Lalu Lintas 2022. Jakarta: Kemenhub RI, 2022."
- [4] "Korps Lalu Lintas Polri, 'Data kecelakaan lalu lintas nasional tahun 2023,' Korlantas Polri, Jakarta, 2023."
- [5] "Kementerian Perhubungan RI, 'Laporan tahunan keselamatan lalu lintas dan angkutan

- jalan 2022,' Ditjen Perhubungan Darat, Kemenhub RI, 2022."
- [6] "Undang-Undang RI No. 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan. Jakarta: JDIH Kemenkumham, 2009."
- [7] T. B. Brown *et al.*, "Language Models are Few-Shot Learners," 2020, *arXiv*. doi: 10.48550/ARXIV.2005.14165.
- [8] Z. Ji *et al.*, "Survey of Hallucination in Natural Language Generation," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–38, Dec. 2023, doi: 10.1145/3571730.
- [9] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," 2020, *arXiv*. doi: 10.48550/ARXIV.2005.11401.
- [10] K. Shuster, S. Poff, M. Chen, D. Kiela, and J. Weston, "Retrieval Augmentation Reduces Hallucination in Conversation," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, pp. 3784–3803. doi: 10.18653/v1/2021.findings-emnlp.320.
- [11] "E. Nurdiansyah and M. Hidayat, 'Pengembangan sistem chatbot hukum perdata Indonesia menggunakan BERT dan cosine similarity,' *Jurnal Ilmu Komputer dan Informasi*, vol. 15, no. 2, pp. 87–96, 2022."
- [12] "R. Wiraatmaja, A. Firdaus, and A. Kurniawan, 'Deteksi pelanggaran lalu lintas menggunakan IndoBERT dan SVM,' *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 4, pp. 789–798, 2023."
- [13] R. Ali, Y. P. Sari, and A. D. Alawiyah, "Darmajaya Academic Chatbot Dengan Semantic Search," *Pros. Semin. Nas. Darmajaya*, vol. 1, no. 0, pp. 241–245, Sep. 2021.
- [14] "APJII, Laporan Survei Penetrasi Internet Indonesia 2023. Jakarta: APJII, 2023."
- [15] "We Are Social, Digital 2024 Indonesia. New York: We Are Social, 2024."
- [16] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers," 2020, *arXiv*. doi: 10.48550/ARXIV.2002.10957.
- [17] H. Touvron *et al.*, "Llama 2: Open Foundation and Fine-Tuned Chat Models," 2023, *arXiv*. doi: 10.48550/ARXIV.2307.09288.
- [18] L. Huang *et al.*, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Trans. Inf. Syst.*, vol. 43, no. 2, pp. 1–55, Mar. 2025, doi: 10.1145/303155.
- [19] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang, "REALM: Retrieval-Augmented Language Model Pre-Training," 2020, *arXiv*. doi: 10.48550/ARXIV.2002.08909.
- [20] Z. Jiang *et al.*, "Active Retrieval Augmented Generation," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore: Association for Computational Linguistics, 2023, pp. 7969–7992. doi: 10.18653/v1/2023.emnlp-main.495.
- [21] N. Mariana, A. Jananto, S. Saefurrohman, and A. P. Utomo, "PERAN PERSEPSI KEGUNAAN DAN KEPERCAYAAN DALAM ADOPTASI CHATGPT OLEH SISWA: PENDEKATAN BERBASIS LITERATUR," *J. Inform.*, vol. 24, no. 1, pp. 10–16, Jun. 2024, doi: 10.30873/ji.v24i1.3943.
- [22] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 3980–3990. doi: 10.18653/v1/D19-1410.
- [23] "Telegram, 'Telegram Bot API documentation,' 2024. [Online]. Available: <https://core.telegram.org/bots/api>."
- [24] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale assessment of mobile notifications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto Ontario Canada: ACM, Apr. 2014, pp. 3055–3064. doi: 10.1145/2556288.2557189.
- [25] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The Muppets straight out of Law School," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Online: Association for Computational Linguistics, 2020, pp. 2898–2904. doi: 10.18653/v1/2020.findings-emnlp.261.
- [26] D. Kurniawan and J. Triloka, "Penerapan Teknologi Langchain dan LLM pada Sistem Question Answering Berbasis Chatbot Telegram: Literature Review," *Semin. Nas. Darmajaya*, no. 1, pp. 95–104, Jan. 2025.
- [27] "Y. P. Sari, R. Ali, and R. Pratama, 'Pengembangan sistem tanya jawab otomatis berbasis LLM untuk layanan informasi digital,' in *Proc. PSND*, vol. 3, no. 1, pp. 44–53, 2023."
- [28] "K. Beck *et al.*, *Manifesto for Agile Software Development*. Agile Alliance, 2001."
- [29] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach, 8th ed.* New York: McGraw-Hill, 2015.
- [30] V. Karpukhin *et al.*, "Dense Passage Retrieval for Open-Domain Question Answering," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online: Association for

- Computational Linguistics, 2020, pp. 6769–6781. doi: 10.18653/v1/2020.emnlp-main.550.
- [31] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, 1st ed. Cambridge University Press, 2008. doi: 10.1017/CBO9780511809071.
- [32] C. Goutte and E. Gaussier, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation,” in *Advances in Information Retrieval*, vol. 3408, D. E. Losada and J. M. Fernández-Luna, Eds., in Lecture Notes in Computer Science, vol. 3408. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 345–359. doi: 10.1007/978-3-540-31865-1_25.
- [33] J. Brooke, “SUS: A quick and dirty usability scale,” in *Usability Evaluation in Industry*, P. Jordan et al., Eds. London: Taylor & Francis, 1996, pp. 189–194.
- [34] “A. Bangor, P. Kortum, and J. Miller, ‘Determining what individual SUS scores mean: Adding an adjective rating scale,’ J. Usability Studies, vol. 4, no. 3, pp. 114–123, 2009.”
- [35] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications, Inc., 2019. doi: 10.4135/9781071878781.