

Chatbot Informasi dan Kalkulator Pajak Menggunakan Google Vertex AI Berbasis *Retrieval-Augmented Generation* (tanyapajakai.com)

Novandra Anugrah¹, Muhammad Yasser Arafat², Lukmanul Hakim³

^{1,2} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

³ Program Studi Akuntansi Perpajakan, Fakultas Ekonomi dan Bisnis, Universitas Pamulang, Tangerang Selatan, Indonesia

¹novandraanugrah91@gmail.com, ²dosen00680@unpam.ac.id, ³02722@unpam.ac.id

Info Artikel

Riwayat Artikel:

Received June 24, 2026

Revised August 17, 2026

Accepted August 18, 2026

Abstract – Access to tax regulations in Indonesia is often perceived as rigid and difficult for taxpayers to understand. This study implements Tanyapajakai.com, a web-based tax information chatbot and calculator using Retrieval-Augmented Generation (RAG), Vertex AI Search, and Gemini 3.5 Flash over official Indonesian tax regulations. Evaluation combined expert validation on ten scenarios, BERTScore and ROUGE on five reference-answer pairs, and an observational comparison with three general-purpose language-model configurations. Three tax experts produced a mean factual-accuracy score of 4.77 out of 5. Inter-rater reliability reached $ICC(2,k) = 0.47$, indicating moderate agreement under a pronounced ceiling effect. Quantitative evaluation yielded BERTScore F1 of 84.06%, ROUGE-1 of 60.56%, ROUGE-2 of 42.78%, and ROUGE-L of 53.10%. Low-scoring cases were associated primarily with incomplete calculator coverage and outdated monthly income-tax logic. The system supports contextual, source-linked tax information, but its answers remain informational and require verification against official sources or consultation with qualified professionals.

Keywords: Chatbot, Google Vertex AI, Retrieval-Augmented Generation, Tax Calculator, Tax Regulation

Corresponding Author:

Novandra Anugrah

Email:

novandraanugrah91@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak Indonesia – Akses informasi regulasi perpajakan di Indonesia sering dinilai kaku dan sulit dipahami oleh Wajib Pajak. Penelitian ini mengimplementasikan Tanyapajakai.com, aplikasi web chatbot informasi dan kalkulator pajak yang menggunakan Retrieval-Augmented Generation (RAG), Vertex AI Search, dan Gemini 3.5 Flash atas dokumen resmi peraturan perpajakan Indonesia. Evaluasi menggabungkan validasi ahli pada sepuluh skenario, BERTScore dan ROUGE pada lima pasangan jawaban acuan, serta perbandingan observasional dengan tiga konfigurasi model bahasa serbaguna. Tiga ahli perpajakan menghasilkan rata-rata akurasi faktual 4,77 dari skala 5. Reliabilitas antarpemilai mencapai $ICC(2,k) = 0,47$, yang menunjukkan kesepakatan sedang dengan efek batas atas yang kuat. Evaluasi kuantitatif menghasilkan skor F1 BERTScore 84,06%, ROUGE-1 60,56%, ROUGE-2 42,78%, dan ROUGE-L 53,10%. Kasus berskor rendah terutama berkaitan dengan cakupan kalkulator yang belum lengkap dan logika pajak penghasilan bulanan yang belum mutakhir. Sistem mendukung informasi perpajakan yang kontekstual dan dilengkapi tautan sumber, tetapi jawabannya tetap bersifat informatif serta perlu diverifikasi melalui sumber resmi atau konsultasi dengan tenaga profesional.

Kata Kunci: Chatbot, Google Vertex AI, Kalkulator Pajak, Regulasi Perpajakan, Retrieval-Augmented Generation

1. PENDAHULUAN

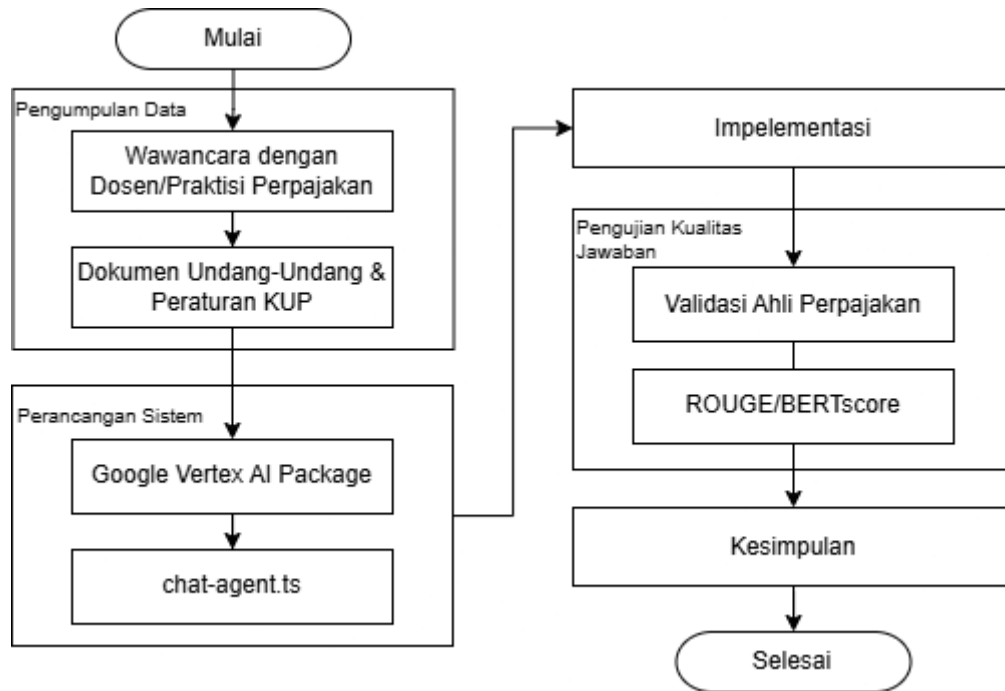
Perkembangan teknologi informasi telah mendorong perubahan dalam penyediaan dan penyebaran informasi pada berbagai sektor, termasuk layanan publik. Salah satu bidang kecerdasan buatan yang berkembang adalah *Natural Language Processing* (NLP), yaitu bidang yang memungkinkan komputer memproses, menganalisis, dan menghasilkan bahasa manusia untuk berbagai aplikasi, seperti pencarian informasi, peringkasan teks, dan sistem penjawab pertanyaan [1]. Perkembangan model bahasa berskala besar atau *Large Language Model* (LLM) selanjutnya memungkinkan sistem menjalankan berbagai tugas pemrosesan bahasa melalui interaksi berbasis teks [2]. Meskipun demikian, penerapan teknologi tersebut pada informasi hukum dan perpajakan memiliki tantangan tersendiri karena dokumen hukum umumnya memiliki struktur bahasa yang kompleks dan memuat terminologi khusus sehingga tidak selalu mudah dipahami oleh masyarakat yang tidak memiliki latar belakang hukum [3]. Dalam konteks perpajakan, pemahaman terhadap regulasi menjadi penting karena pengetahuan perpajakan berkaitan dengan peningkatan kesadaran dan kepatuhan Wajib Pajak [4]. Teknologi *chatbot* dapat dimanfaatkan untuk menyediakan layanan informasi melalui interaksi berbasis percakapan. *Chatbot* memungkinkan pengguna menyampaikan pertanyaan menggunakan bahasa alami dan memperoleh respons tanpa harus menelusuri dokumen secara manual [5]. Penggunaan *chatbot* berbasis LLM berpotensi membuat akses terhadap informasi perpajakan menjadi lebih interaktif, tetapi model generatif masih dapat menghasilkan informasi yang tidak sesuai dengan fakta atau tidak didukung oleh sumber yang dapat diverifikasi. Fenomena tersebut dikenal sebagai halusinasi informasi atau *artificial intelligence hallucination* [6]. Untuk mengurangi risiko tersebut, *Retrieval-Augmented Generation* (RAG) digunakan dengan menggabungkan kemampuan generasi bahasa dari LLM dan proses pengambilan informasi dari sumber pengetahuan eksternal, sehingga sistem dapat mengambil bagian dokumen yang relevan sebagai konteks dalam menghasilkan jawaban [7]. Penelitian ini menerapkan RAG menggunakan Vertex AI Search sebagai layanan pencarian dokumen untuk menghubungkan Gemini dengan kumpulan dokumen milik pengembang sebagai sumber *grounding* [8]. Gemini 3.5 Flash digunakan karena merupakan model Gemini yang mendukung penalaran terhadap permasalahan dan konteks yang kompleks [9].

2. PENELITIAN YANG TERKAIT

Beberapa penelitian terdahulu telah mengembangkan *chatbot* dan sistem penjawab pertanyaan berbasis dokumen. Albert dan Voutama mengembangkan *chatbot* dokumen PDF menggunakan Local *Retrieval-Augmented Generation*, Ollama, ChromaDB, dan model representasi vektor yang dijalankan secara lokal [10]. Saputra et al. mengembangkan aplikasi *question answering* mengenai peraturan perundang-undangan bidang pendidikan dengan memanfaatkan *LangChain*, model bahasa, *BERTScore*, dan *ROUGE Score* [11]. Sementara itu, Faishol dan Ichwani mengembangkan *chatbot* pada situs web konsultasi jasa pajak berbasis Laravel dan ReactJS untuk memberikan jawaban atas pertanyaan umum, membantu penjadwalan konsultasi, serta mendukung layanan pembayaran daring [12]. Penelitian-penelitian tersebut menunjukkan pemanfaatan *chatbot* pada dokumen maupun layanan perpajakan, tetapi belum menggabungkan RAG berbasis Vertex AI Search dan Gemini 3.5 Flash dengan layanan kalkulator pajak dalam satu aplikasi web. Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengimplementasikan Tanyapajakai.com sebagai aplikasi web *chatbot* informasi dan kalkulator pajak berbasis RAG pada ekosistem Google Vertex AI. Basis pengetahuan *chatbot* dibentuk dari dokumen resmi peraturan perpajakan yang digunakan sebagai sumber konteks dalam menjawab pertanyaan pengguna. Kualitas jawaban dievaluasi melalui validasi ahli perpajakan serta pengukuran otomatis menggunakan *BERTScore* dan *ROUGE Score*. *BERTScore* mengukur kesamaan teks berdasarkan representasi kontekstual setiap token [13], sedangkan *ROUGE* mengukur kesamaan berdasarkan tumpang tindih unsur leksikal antara jawaban sistem dan jawaban referensi [14]. Penelitian ini diharapkan memberikan gambaran mengenai penerapan RAG untuk mendukung penyediaan informasi perpajakan yang interaktif, kontekstual, dan berbasis sumber regulasi.

3. METODE PENELITIAN

Tahapan penelitian dalam pengembangan Tanyapajakai.com meliputi pengumpulan data, perancangan sistem penjawab pertanyaan, implementasi aplikasi, evaluasi kualitas jawaban, dan penarikan kesimpulan. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Berdasarkan Gambar 1, penelitian dimulai dengan pengumpulan dokumen peraturan perpajakan dan analisis kebutuhan pengguna. Data yang diperoleh kemudian digunakan untuk merancang arsitektur RAG, mengimplementasikan aplikasi, serta mengevaluasi kualitas jawaban menggunakan pendekatan kualitatif dan kuantitatif.

3.1 Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh kebutuhan sistem dan dokumen yang digunakan sebagai basis pengetahuan. Tahap ini dilakukan melalui wawancara dan pengumpulan dokumen hukum.

a. Wawancara dan Analisis Kebutuhan

Wawancara dilakukan bersama dosen yang juga praktisi dari Program Studi Akuntansi Perpajakan Universitas Pamulang untuk mengidentifikasi kebutuhan informasi perpajakan, menentukan skenario pertanyaan pengguna, serta memahami mekanisme dasar perhitungan pajak yang diterapkan pada fitur kalkulator. Hasil analisis kebutuhan tersebut digunakan untuk menentukan fungsi utama aplikasi, yaitu *chatbot* informasi perpajakan, kalkulator pajak, tampilan sumber referensi, riwayat percakapan, direktori peraturan, serta formulir laporan kesalahan dan saran.

b. Dokumen Hukum Perpajakan

Dokumen hukum diperoleh melalui Basis Data Peraturan JDIH BPK RI yang menyediakan peraturan perundang-undangan dan dokumen hukum dari sumber resmi [15]. Dokumen dikumpulkan dalam format PDF untuk diintegrasikan ke dalam basis pengetahuan aplikasi, meliputi Undang-Undang Ketentuan Umum dan Tata Cara Perpajakan serta peraturan yang berkaitan dengan Pajak Penghasilan, Pajak Pertambahan Nilai, Penghasilan Tidak Kena Pajak, ketentuan perpajakan UMKM, dan Bea Meterai.

3.2 Perancangan Sistem Penjawab Pertanyaan

Sistem penjawab pertanyaan dirancang menggunakan arsitektur RAG yang mengintegrasikan Vertex AI Search dengan Gemini 3.5 Flash. Dokumen perpajakan dimasukkan ke dalam Vertex AI Search *Data Store* agar dapat diproses dan diindeks. Vertex AI Search mendukung proses penguraian dan pembagian dokumen menjadi bagian-bagian teks atau potongan teks untuk membantu proses pengambilan informasi [16].

Secara konseptual, representasi vektor digunakan untuk merepresentasikan informasi teks dalam bentuk vektor dan mendukung pencarian berdasarkan kesamaan semantik [17]. Ketika pengguna mengajukan pertanyaan, sistem mencari bagian dokumen yang relevan dan menggunakannya sebagai konteks bagi Gemini 3.5 Flash dalam menghasilkan jawaban [7], [8]. Alur pemrosesan pertanyaan terdiri atas tahapan berikut:

1. Dokumen peraturan diunggah ke Vertex AI Search *Data Store*.
2. Dokumen diproses, dibagi menjadi bagian teks, dan diindeks.
3. Pengguna mengirimkan pertanyaan melalui aplikasi.
4. Vertex AI Search mengambil bagian dokumen yang relevan.
5. Pertanyaan dan konteks hasil pencarian diteruskan kepada Gemini 3.5 Flash.
6. Gemini 3.5 Flash menyusun jawaban berdasarkan konteks yang diberikan.
7. Jawaban dan metadata sumber dikirimkan kembali ke aplikasi.

Melalui mekanisme tersebut, model tidak hanya mengandalkan informasi yang tersimpan dalam parameter internalnya, tetapi juga menggunakan dokumen peraturan yang telah ditentukan sebagai konteks eksternal.

3.3 Implementasi

Tanyapajakai.com dikembangkan menggunakan Next.js sebagai kerangka kerja aplikasi web berbasis React. Next.js mendukung pengembangan aplikasi web menyeluruh yang menggabungkan komponen antarmuka dan fungsi sisi *server* [18].

Komunikasi antara antarmuka pengguna dan sisi *server* menggunakan tRPC untuk mempertahankan konsistensi tipe data antara sisi klien dan *server* [19]. PostgreSQL digunakan untuk menyimpan data akun pengguna, percakapan, pesan, umpan balik, dan laporan.

Ketika pengguna mengirimkan pertanyaan, sisi *server* meneruskan kueri ke Vertex AI Search untuk mendapatkan konteks dokumen yang relevan. Konteks tersebut kemudian diteruskan kepada Gemini 3.5 Flash untuk menghasilkan jawaban. Respons yang dihasilkan diproses untuk memisahkan teks jawaban dan metadata referensi sebelum disimpan dan ditampilkan kepada pengguna.

Fitur kalkulator pajak diterapkan dalam aplikasi yang sama untuk membantu pengguna melakukan simulasi perhitungan. Proses perhitungan dilakukan menggunakan logika program berdasarkan parameter yang dimasukkan oleh pengguna, sedangkan layanan Gemini digunakan untuk fungsi tanya jawab berbasis dokumen.

3.4 Pengujian dan Evaluasi Kualitas Jawaban

Pengujian dilakukan untuk mengevaluasi kualitas dan kesesuaian jawaban yang dihasilkan *chatbot*. Evaluasi menggunakan pendekatan kualitatif melalui validasi ahli perpajakan dan pendekatan kuantitatif menggunakan *BERTScore* serta *ROUGE Score*.

Validasi ahli digunakan untuk menilai ketepatan substansi perpajakan, sedangkan *BERTScore* dan *ROUGE Score* digunakan untuk mengukur kesamaan jawaban sistem terhadap jawaban referensi dari sisi semantik dan leksikal.

3.5 Validasi Ahli (Kualitatif)

Validasi ahli melibatkan tiga dosen dan praktisi Program Studi Akuntansi Perpajakan Universitas Pamulang. Masing-masing validator menilai sepuluh skenario yang sama secara independen. Skenario mencakup pertanyaan faktual, perhitungan pajak, kasus batas, dan pertanyaan ambigu untuk menguji ketepatan jawaban serta ketangguhan sistem.

Setiap jawaban dinilai berdasarkan akurasi faktual dan kesesuaiannya dengan regulasi menggunakan skala Likert 1 sampai 5[20]. Reliabilitas antarpemilai dihitung dengan koefisien korelasi intrakelas model efek acak dua arah, kesepakatan absolut, dan ukuran rata-rata atau ICC(2,k). Persentase kesepakatan tepat antarpasangan pemilai dan kesepakatan penuh per butir digunakan sebagai ukuran pendamping.

$$\text{Nilai rata - rata} = \text{jumlah seluruh skor} / \text{jumlah pertanyaan}$$

BERTScore dan *ROUGE* (Metrik Kuantitatif)

Evaluasi kuantitatif membandingkan jawaban *chatbot* dengan lima jawaban acuan yang disusun berdasarkan dokumen peraturan resmi. *Dataset* metrik ini dipertahankan terpisah dari sepuluh skenario validasi ahli agar hasil *BERTScore* dan *ROUGE* tetap dapat dibandingkan dengan pengujian awal.

BERTScore mengukur kemiripan antara token pada teks prediksi dan referensi menggunakan representasi kontekstual. Berbeda dengan pengukuran yang hanya mengandalkan kecocokan kata, *BERTScore* dapat mengenali kemiripan representasi kontekstual pada susunan kalimat yang berbeda [13].

$$P_{BERT} = \frac{1}{|\hat{y}|} \sum_{\hat{y}_j \in \hat{y}} \max_{y_i \in y} \hat{x}_j^T x_i$$

$$R_{BERT} = \frac{1}{|y|} \sum_{y_i \in y} \max_{\hat{y}_j \in \hat{y}} \hat{x}_j^T x_i$$

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}}$$

Cakupan mengukur proporsi unsur *n-gram* pada jawaban acuan yang juga ditemukan pada jawaban kandidat. Nilai yang lebih tinggi menunjukkan cakupan informasi acuan yang lebih lengkap, tetapi tidak dengan sendirinya membuktikan kebenaran faktual.

ROUGE Score digunakan untuk mengukur kesamaan leksikal berdasarkan tumpang tindih unsur teks antara jawaban prediksi dan referensi [14]. Jenis pengukuran yang digunakan meliputi: *ROUGE-1* digunakan untuk mengukur kecocokan *unigram* atau kata tunggal, *ROUGE-2* untuk mengukur kecocokan *bigram* atau dua kata yang berurutan, sedangkan *ROUGE-L* digunakan untuk mengukur kesamaan berdasarkan subsekuens bersama terpanjang.

$$R_{ROUGE} = \frac{\sum_{\text{gram}_n \in y} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{\text{gram}_n \in \hat{y}} \text{Count}(\text{gram}_n)}$$

$$P_{ROUGE} = \frac{\sum_{\text{gram}_n \in \hat{y}} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{\text{gram}_n \in y} \text{Count}(\text{gram}_n)}$$

$$F1_{ROUGE} = 2 \frac{P_{ROUGE} \cdot R_{ROUGE}}{P_{ROUGE} + R_{ROUGE}}$$

BERTScore dan *ROUGE* digunakan untuk memberikan sudut pandang evaluasi yang berbeda. *BERTScore* mengukur kedekatan representasi kontekstual, sedangkan *ROUGE* mengukur kesamaan berdasarkan unsur leksikal dan urutan kata.

3.6 Penarikan Kesimpulan

Hasil validasi ahli, *BERTScore*, dan *ROUGE Score* dianalisis secara deskriptif. Penilaian ahli digunakan untuk mengetahui ketepatan substansi perpajakan, sedangkan hasil *BERTScore* dan *ROUGE* digunakan untuk mengukur kemiripan jawaban *chatbot* dengan jawaban referensi. Seluruh hasil tersebut digunakan untuk menilai kualitas jawaban yang dihasilkan melalui penerapan RAG pada Tanyapajakai.com.

4. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil pengumpulan data, implementasi sistem, serta pengujian dan evaluasi kualitas jawaban yang dihasilkan oleh Tanyapajakai.com.

4.1 Pengumpulan Data

Dokumen peraturan perpajakan yang digunakan sebagai basis pengetahuan diperoleh dari Basis Data Peraturan BPK RI <https://peraturan.bpk.go.id/> dalam format PDF. Dokumen tersebut mencakup peraturan yang berkaitan dengan Ketentuan Umum dan Tata Cara Perpajakan, Pajak Penghasilan, Pajak Pertambahan Nilai, Penghasilan Tidak Kena Pajak, dan Bea Meterai. Daftar dokumen yang digunakan dalam sistem disajikan pada Tabel I.

Tabel I. Daftar Dokumen Peraturan Perpajakan.

Jenis Dokumen	Nomor	Tahun	Tentang
Undang-undang (UU)	8	1984	Tentang penangguhan berlakunya PPN
Undang-undang (UU)	12	1994	Tentang pajak bumi dan bangunan
Undang-undang (UU)	8	1997	Tentang Dokumen Perusahaan
Undang-undang (UU)	1	1998	Tentang bea perolehan hak atas tanah dan bangunan
Undang-undang (UU)	16	2000	Tentang ketentuan umum dan tata cara perpajakan
Undang-undang (UU)	17	2000	Tentang pajak penghasilan
Undang-undang (UU)	18	2000	Tentang pajak pertambahan nilai dan PPnBM
Undang-undang (UU)	19	2000	Tentang penagihan pajak dengan surat paksa
Undang-undang (UU)	20	2000	Tentang bea perolehan hak atas tanah dan bangunan
Undang-undang (UU)	24	2000	Tentang Perjanjian Internasional
Undang-undang (UU)	36	2000	Tentang kawasan perdagangan bebas dan pelabuhan bebas
Undang-undang (UU)	14	2002	Tentang pengadilan pajak
Undang-undang (UU)	19	2003	Tentang badan usaha milik negara
Undang-undang (UU)	25	2003	Tentang tindak pidana pencucian uang
Undang-undang (UU)	33	2004	Tentang perimbangan keuangan pusat dan daerah
Undang-undang (UU)	12	2006	Tentang kewarganegaraan Indonesia
Undang-undang (UU)	15	2006	Tentang badan pemeriksa keuangan
Undang-undang (UU)	17	2006	Tentang kepabeaian
Undang-undang (UU)	25	2007	Tentang penanaman modal
Undang-undang (UU)	28	2007	Tentang ketentuan umum dan tata cara perpajakan
Undang-undang (UU)	36	2008	Tentang pajak penghasilan
Undang-undang (UU)	21	2008	Tentang perbankan syariah
Undang-undang (UU)	16	2009	Tentang ketentuan umum dan tata cara perpajakan
Undang-undang (UU)	17	2009	Tentang pemilihan umum legislatif
Undang-undang (UU)	42	2009	Tentang pajak pertambahan nilai dan PPnBM
Undang-undang (UU)	48	2009	Tentang kekuasaan kehakiman
Undang-undang (UU)	51	2009	Tentang peradilan tata usaha negara
Undang-undang (UU)	4	2009	Tentang pertambangan mineral dan batubara
Undang-undang (UU)	5	2011	Tentang akuntan publik
Undang-undang (UU)	2	2012	Tentang pengadaan tanah untuk kepentingan umum
Undang-undang (UU)	11	2016	Tentang pengampunan pajak
Undang-undang (UU)	13	2016	Tentang paten

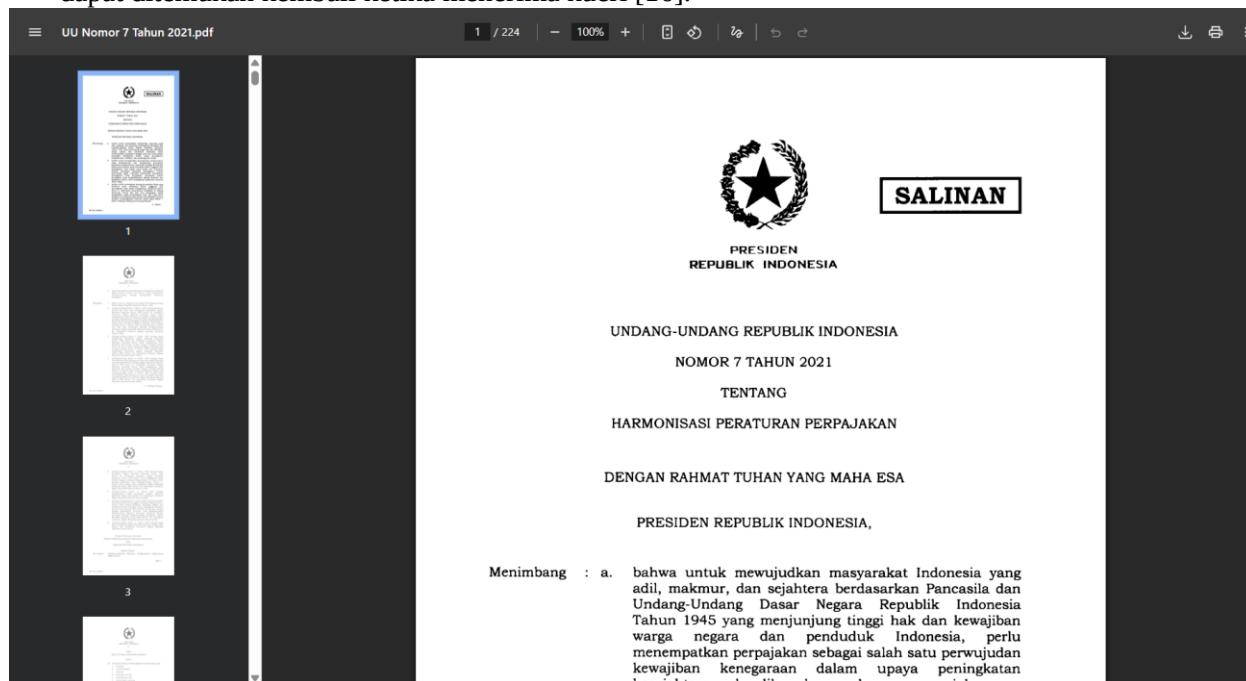
Jenis Dokumen	Nomor	Tahun	Tentang
Undang-undang (UU)	9	2017	Tentang akses informasi keuangan perpajakan
Undang-undang (UU)	9	2018	Tentang penerimaan negara bukan pajak
Undang-undang (UU)	10	2020	Tentang bea meterai
Undang-undang (UU)	7	2021	Tentang harmonisasi peraturan perpajakan
Undang-undang (UU)	1	2022	Tentang hubungan keuangan pusat dan daerah
Undang-undang (UU)	13	2022	Tentang pembentukan peraturan perundang-undangan
Undang-undang (UU)	21	2023	Tentang ibu kota negara
Undang-undang (UU)	6	2023	Tentang cipta kerja
Undang-undang (UU)	1	2025	Tentang badan usaha milik negara

Dokumen pada Tabel I digunakan sebagai sumber konteks eksternal bagi arsitektur RAG. Basis pengetahuan sistem tidak hanya mencakup ketentuan KUP, tetapi juga peraturan yang berkaitan dengan PPh, PPN, PTKP, ketentuan perpajakan UMKM, dan Bea Meterai. Penggunaan dokumen eksternal memungkinkan sistem mengambil informasi dari peraturan yang telah ditentukan sebelum menyusun jawaban. Pendekatan tersebut dimaksudkan untuk meningkatkan keterkaitan jawaban dengan sumber regulasi, bukan untuk menjamin bahwa seluruh jawaban yang dihasilkan model selalu benar.

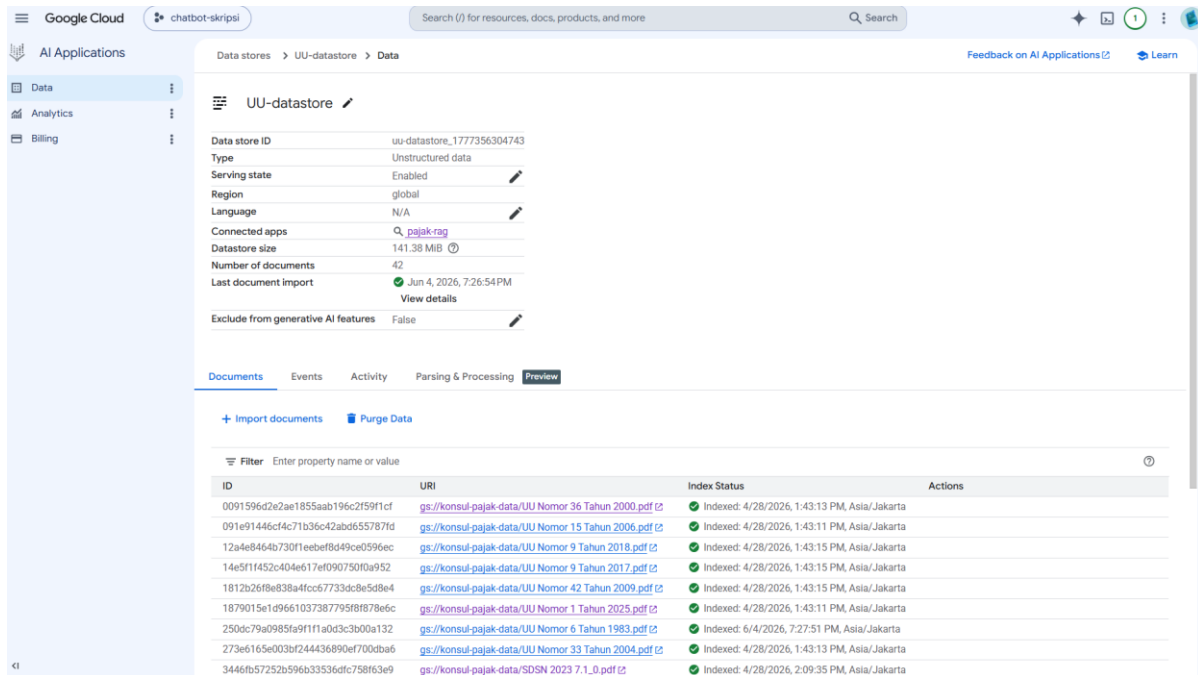
4.2 Analisis dan Perancangan

a. Pemrosesan Data Kontekstual

Dokumen peraturan dalam format PDF diunggah ke Vertex AI Search *Data Store*. Layanan tersebut memproses dokumen melalui penguraian, pembagian teks, dan pengindeksan agar bagian dokumen dapat ditemukan kembali ketika menerima kueri [16].



Gambar 2. Dokumen dalam Format PDF

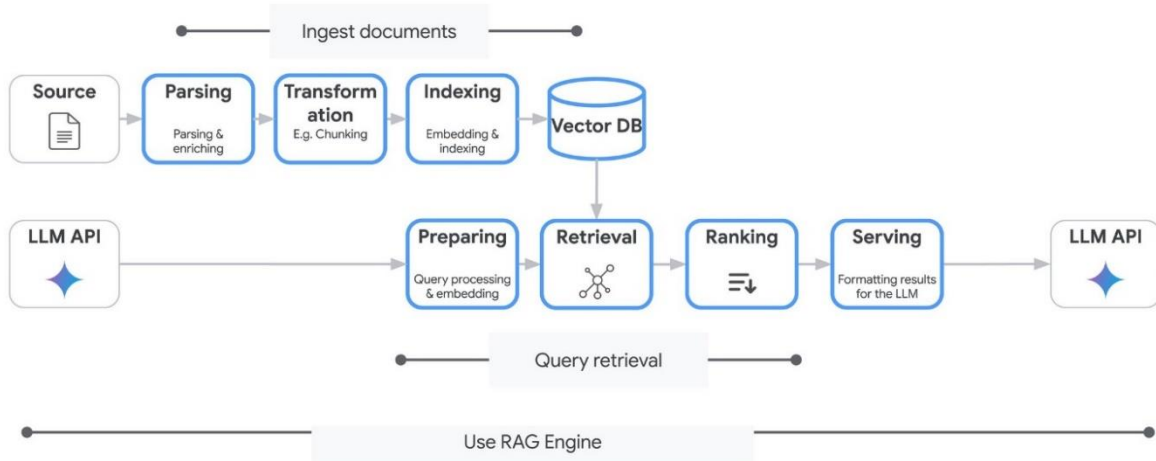


Gambar 3. Google Cloud Data Store dan Proses Pengindeksan

Representasi vektor memungkinkan teks direpresentasikan dalam bentuk vektor sehingga pencarian dapat mempertimbangkan kesamaan makna, tidak hanya kecocokan kata secara persis [17]. Namun, detail algoritma pemeringkatan internal Vertex AI Search tidak ditentukan secara langsung oleh peneliti. Oleh karena itu, penelitian ini tidak menyatakan bahwa seluruh proses pencarian selalu menggunakan satu ukuran tertentu, seperti kemiripan kosinus.

b. Mekanisme Aliran Data RAG

Rancangan fungsi tanya jawab Tanyapajakai.com menggunakan RAG berbasis Vertex AI Search dan Gemini 3.5 Flash. Alur sistem ditunjukkan pada Gambar 4.



Gambar 4. Alur Kerja RAG pada Vertex AI

Arsitektur RAG memuat dokumen peraturan ke *Data Store*, menguraikan dan mengindeksnya, mengambil potongan teks yang relevan untuk setiap pertanyaan, lalu meneruskan pertanyaan beserta konteks kepada Gemini 3.5 Flash. Jawaban dan metadata sumber kemudian ditampilkan pada

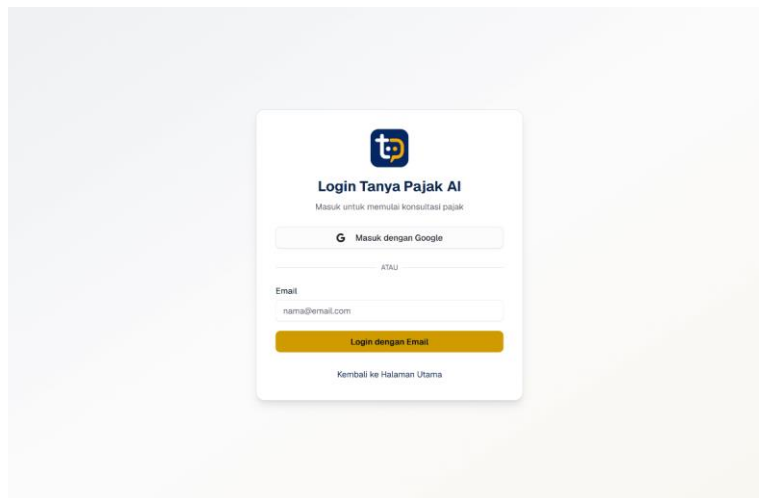
aplikasi. Alur ini menggabungkan pengambilan informasi dan pembangkitan jawaban [7], sedangkan pengaitan konteks melalui Vertex AI Search memungkinkan Gemini menggunakan data milik pengembang [8].

4.3 Implementasi Antarmuka Aplikasi

Tanyapajakai.com diimplementasikan sebagai aplikasi web berbasis Next.js yang menggabungkan layanan *chatbot* informasi dan kalkulator pajak [18]. Aplikasi menyediakan halaman autentikasi, ruang percakapan, sumber referensi, direktori peraturan, kalkulator pajak, serta formulir laporan.

a. Antarmuka Halaman *Login*

Halaman autentikasi digunakan sebagai gerbang akses pengguna. Sistem menyediakan metode autentikasi Google dan kode sandi sekali pakai (OTP) untuk melakukan verifikasi akun.



Gambar 5. Antarmuka Halaman Autentikasi

b. Antarmuka Ruang Percakapan Utama

Antarmuka ruang percakapan menyediakan kolom pertanyaan dan area *dialog* antara pengguna dan sistem. Di bawah kolom masukan ditampilkan penafian bahwa jawaban AI bersifat informatif dan bukan nasihat hukum profesional. Pengguna diarahkan untuk memverifikasi informasi melalui sumber resmi atau konsultan pajak sebelum mengambil keputusan.

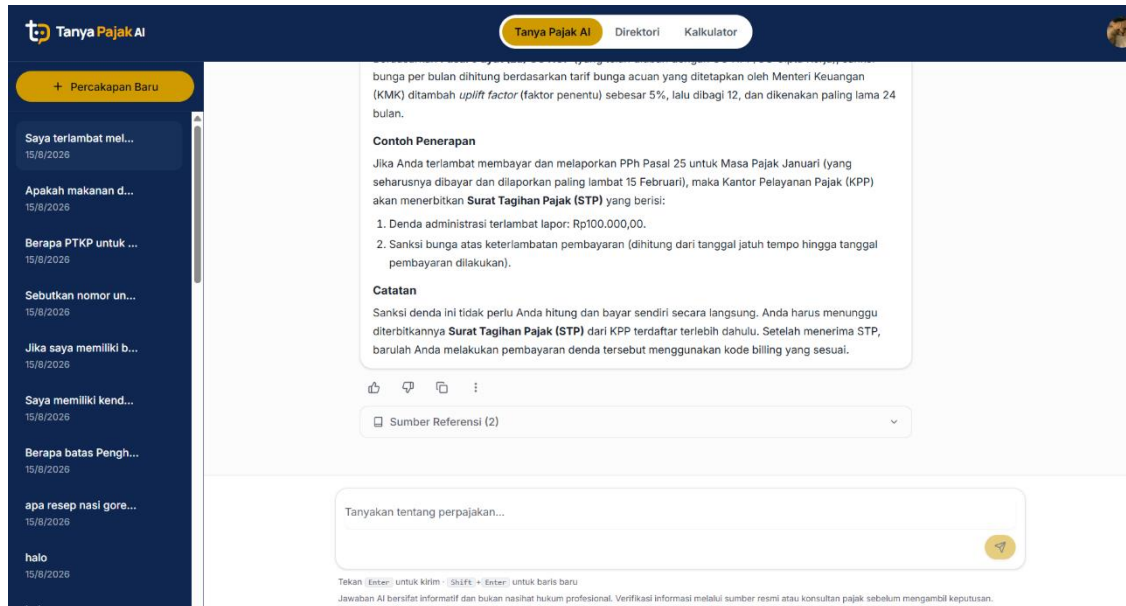


Gambar 6. Antarmuka Ruang Percakapan Utama

c. Antarmuka Menu Referensi

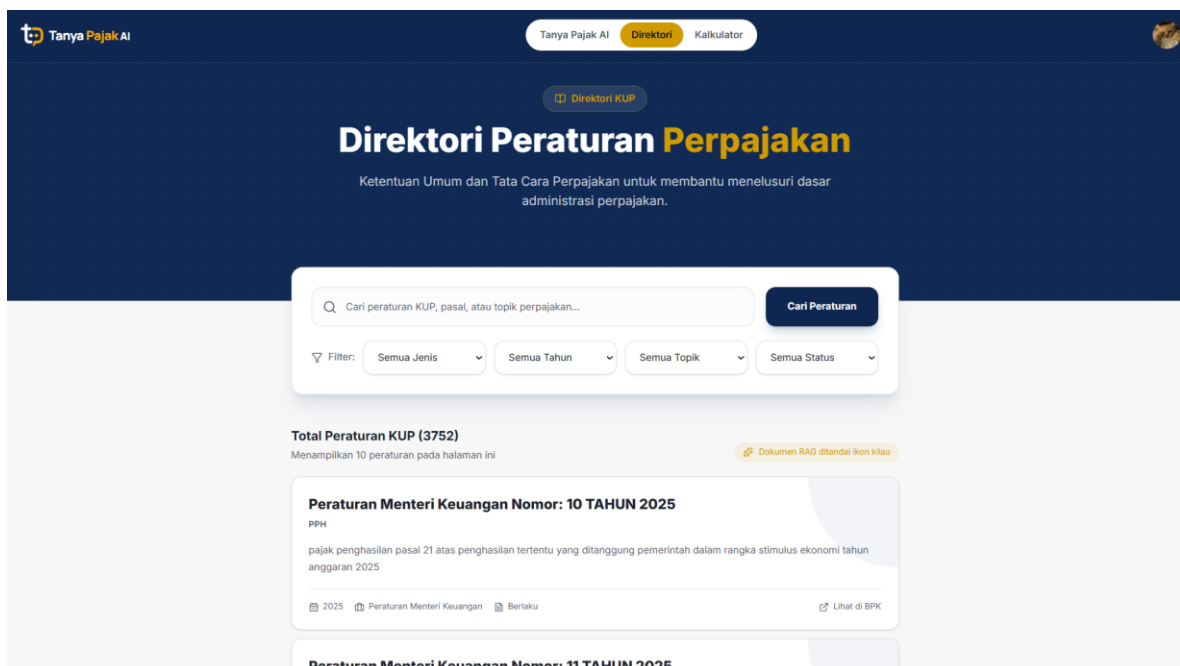
Setiap jawaban dapat disertai tombol sumber referensi. Ketika tombol dipilih, aplikasi menampilkan metadata dokumen yang digunakan sebagai konteks, seperti nomor undang-undang, bagian, atau pasal yang ditemukan. Fitur referensi membantu pengguna meninjau sumber regulasi

yang mendasari jawaban. Meskipun demikian, pengguna tetap perlu memeriksa dokumen resmi apabila jawaban digunakan untuk pengambilan keputusan perpajakan.



Gambar 7. Panel Sampling Referensi Hukum

- d. Antarmuka Halaman Direktori Peraturan
- Halaman direktori menampilkan dokumen peraturan yang telah diintegrasikan ke dalam sistem. Pengguna dapat menelusuri dokumen tanpa harus meninggalkan aplikasi.



Gambar 8. Antarmuka Direktori Peraturan Perpajakan

4.4 Pengujian dan Evaluasi Kualitas Jawaban

Evaluasi kualitas jawaban menggabungkan validasi ahli, metrik semantik dan leksikal, analisis kesalahan, serta pembandingan observasional. Sepuluh skenario pada Tabel II mencakup pertanyaan faktual, perhitungan, kasus batas, dan pertanyaan ambigu.

Tabel II. Daftar Pertanyaan Pengujian

No	Instrumen Pertanyaan Uji Perpajakan
Q1	Berapa batas Penghasilan Tidak Kena Pajak (PTKP) untuk Wajib Pajak orang pribadi berstatus tidak kawin tanpa tanggungan (TK/0)?
Q2	[KASUS AMBIGU] Jika saya memiliki bisnis dengan omzet Rp400.000.000 per tahun bisa tidak dikenai PPh Final UMKM 0,5%?
Q3	[KALKULATOR] Wajib Pajak orang pribadi memiliki omzet usaha Rp600.000.000 dalam satu tahun. Bagaimana perhitungan PPh Final UMKM 0,5%-nya?
Q4	[KASUS AMBIGU] Saya memiliki kendaraan mobil listrik BYD Seal. Berapa pajak yang harus saya bayarkan?
Q5	[KALKULATOR] Saya memiliki bisnis F&B dengan pendapatan bersih Rp50.000.000 per bulan. Berapa pajak yang perlu saya bayarkan?
Q6	Sebutkan nomor undang-undang dan pasal yang mengatur jenis barang yang tidak dikenai PPN, serta tunjukkan sumber regulasi yang dapat diperiksa.
Q7	Berapa PTKP untuk Wajib Pajak berstatus kawin dengan tiga orang tanggungan (K/3)?
Q8	Apakah makanan dan minuman yang disajikan di restoran dikenai PPN? Jelaskan status perpajakannya.
Q9	[KASUS AMBIGU] Saya terlambat melaporkan pajak. Berapa dendanya?
Q10	[KALKULATOR] Saya karyawan dengan gaji Rp10.000.000 per bulan. Berapa PPh Pasal 21 saya?

a. Validasi Ahli (Evaluasi Kualitatif)

Pengujian kualitatif melibatkan tiga dosen dan praktisi Akuntansi Perpajakan sebagai validator materi. Ketiganya menilai sepuluh jawaban yang sama secara independen. Tabel III merangkum rata-rata, catatan, dan reliabilitas antarpemilai.

Tabel III. Hasil Validasi Ahli

Validator	Ringkasan Penilaian	Rata-rata	Catatan Utama
Lukmanul Hakim, S.E., M.Ak.	Sembilan skor 5 dan satu skor 4 pada sepuluh skenario.	4,90	Satu skor 4 tanpa catatan tertulis; penyebab tidak diasumsikan.
Tita Rahmawati, S.E., M.Ak.	Delapan skor 5 dan dua skor 4.	4,80	Kalkulator belum mencakup pajak restoran dan tarif efektif rata-rata (TER).
Siti Asmonah, S.E., M.Ak.	Tujuh skor 5, dua skor 4, dan satu skor 3.	4,60	Penjelasan periodisasi, cakupan pajak restoran, dan logika PPh Pasal 21 perlu diperbaiki.
Gabungan	Total 143 poin dari 30 penilaian.	4,77	Rata-rata tinggi dengan variasi terbesar pada Q10.
Reliabilitas	ICC(2,k) = 0,47; kesepakatan tepat antarpasangan = 70%; kesepakatan penuh per butir = 60%.	Sedang	Banyak skor maksimum menimbulkan efek batas atas; jumlah

			skenario masih terbatas.
-	Skala 1-5; seluruh validator menilai sepuluh skenario yang sama secara independen.	-	Catatan dianalisis secara kualitatif.

Rata-rata validator berturut-turut adalah 4,90; 4,80; dan 4,60, sedangkan rata-rata gabungan mencapai 4,77 dari skala 5. Kesepakatan tepat antarpasangan sebesar 70% dan kesepakatan penuh per butir 60%. Nilai ICC(2,k) sebesar 0,47 menunjukkan reliabilitas sedang; interpretasi perlu hati-hati karena banyak skor berada pada nilai maksimum sehingga terjadi efek batas atas. Analisis jawaban berskor rendah menunjukkan bahwa Q2 memuat penjelasan periodisasi yang keliru meskipun perhitungannya benar, Q5 belum mencakup pajak restoran, dan Q10 belum mengikuti kategori TER serta masih memuat pilihan lama terkait NPWP. Q5 dan Q10 terutama berkaitan dengan *logika* kalkulator, bukan tahap RAG. Tanpa catatan proses pengambilan dokumen, kegagalan pengambilan informasi atau halusinasi tidak dapat dipastikan. Sebagai pembanding observasional, lampiran jawaban aktual membandingkan Tanyapajakai.com dengan ChatGPT tanpa internet, ChatGPT dengan internet, dan DeepSeek V3.2 API. Perbandingan ini bukan ablasi terkontrol karena model, akses sumber, instruksi masukan, dan instrumennya tidak sepenuhnya setara.

b. Analisis Perbandingan Kinerja dengan Metode Lain

Analisis perbandingan dilakukan secara observasional terhadap Tanyapajakai.com dan tiga konfigurasi model bahasa umum berdasarkan jawaban aktual yang tersedia. Penilaian menyoroti ketepatan jawaban, penyertaan sumber regulasi, dan konsistensi perhitungan; karena instrumen serta akses sumber belum sepenuhnya setara, hasilnya digunakan untuk menggambarkan perilaku sistem, bukan membuktikan keunggulan kausal RAG.

Tabel IV. Analisis Pembanding Sistem Secara Observasional

No.	Sistem/Konfigurasi	Ringkasan Temuan Aktual
1	Tanyapajakai.com (RAG)	Jawaban menyertakan sumber regulasi; sebagian besar perhitungan benar, tetapi perhitungan PPh Pasal 21 bulanan perlu dimutakhirkan.
2	ChatGPT tanpa akses internet	Jawaban umumnya benar, tetapi tidak menyertakan tautan sumber dan meminta konteks tambahan pada skenario PPh Pasal 21.
3	ChatGPT dengan akses internet	Memberikan sumber resmi dan ilustrasi TER; hasil dipengaruhi kemampuan penelusuran web.
4	DeepSeek V3.2 API	Kesalahan ditemukan pada dasar perhitungan UMKM, pajak restoran, denda, dan PPh Pasal 21.
5	Batas cakupan	Lampiran memuat 10 skenario, tetapi hanya 7 yang sama atau mendekati instrumen validator terkini; hasil tidak digunakan untuk klaim kausal keunggulan RAG.

Perbandingan menunjukkan manfaat praktis sumber regulasi pada Tanyapajakai.com, tetapi tidak membuktikan bahwa RAG menjadi satu-satunya penyebab perbedaan hasil. Kesetaraan instruksi masukan, versi model, akses internet, dan instrumen perlu dikendalikan pada penelitian lanjutan.

c. Evaluasi Metrik Semantik dan Leksikal (Kuantitatif)

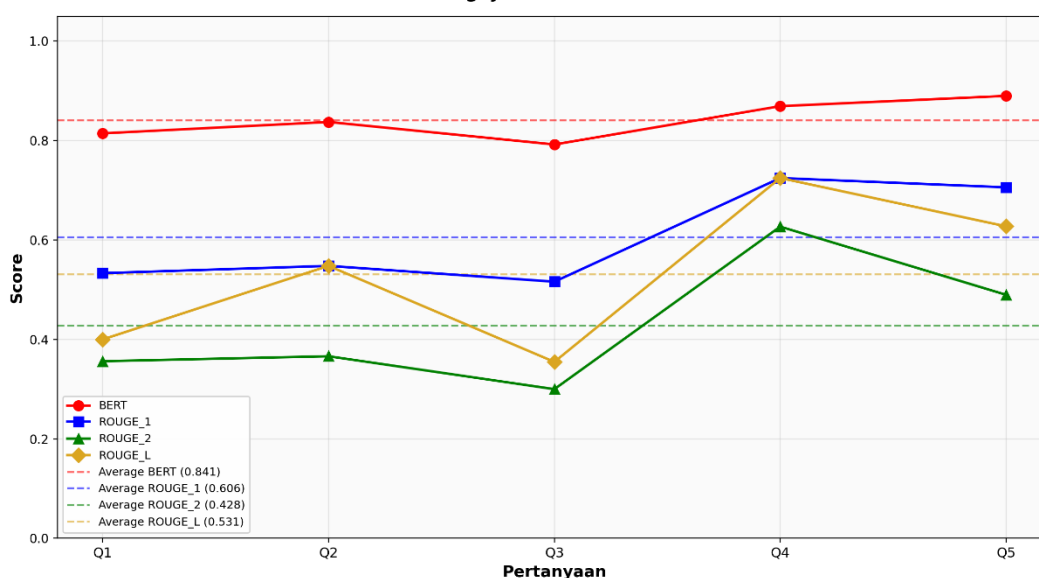
Evaluasi kuantitatif dilakukan pada lima pasangan jawaban sistem dan jawaban acuan menggunakan *BERTScore* dan *ROUGE*. *BERTScore* menilai kemiripan representasi kontekstual, sedangkan *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L* menilai tumpang tindih leksikal; hasilnya disajikan dalam nilai presisi (P), cakupan (R), dan skor F1.

Tabel V. Hasil Evaluasi *BERTScore* dan *ROUGE Score*.

No.	<i>BERTScore</i> (%)			<i>ROUGE-1</i> (%)			<i>ROUGE-2</i> (%)			<i>ROUGE-L</i> (%)		
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
1	77,05	86,38	81,45	38,46	86,96	53,33	25,49	59,09	35,62	28,85	65,22	40,00
2	83,17	84,31	83,74	62,50	48,78	54,79	41,94	32,50	36,62	62,50	48,78	54,79
3	74,94	84,00	79,21	41,03	69,57	51,61	23,68	40,91	30,00	28,21	47,83	35,48
4	83,17	91,02	86,92	59,52	92,59	72,46	51,22	80,77	62,69	59,52	92,59	72,46
5	87,83	90,17	88,98	64,29	78,26	70,59	44,44	54,55	48,98	57,14	69,57	62,75
Rata	81,23	87,18	84,06	53,16	75,23	60,56	37,35	53,56	42,78	47,24	64,80	53,10

Rata-rata skor F1 *BERTScore* mencapai 84,06%. Nilai ini menunjukkan kemiripan representasi kontekstual antara jawaban *chatbot* dan jawaban acuan, bukan bukti kebenaran faktual [13].

Rata-rata skor F1 *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L* berturut-turut adalah 60,56%, 42,78%, dan 53,10%. Nilai *ROUGE-2* yang lebih rendah menunjukkan perbedaan susunan kalimat dan parafrasa terhadap jawaban acuan [14].

Hasil F1 Score Pengujian *BERTScore* dan *ROUGE Score*Gambar 9. Grafik Skor F1 *BERTScore* dan *ROUGE*.

Validasi ahli, metrik otomatis, dan pembandingan observasional memberikan informasi yang berbeda. Penilaian ahli menguji substansi, *BERTScore* dan *ROUGE* mengukur kemiripan pada lima pasangan jawaban acuan, sedangkan pembandingan memperlihatkan perilaku beberapa konfigurasi model. Hasil yang tinggi tidak meniadakan keterbatasan jumlah skenario, efek batas atas, perbedaan instrumen, dan belum tersedianya *log* pengambilan dokumen.

5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan Tanyapajakai.com sebagai aplikasi web *chatbot* informasi dan kalkulator pajak berbasis RAG, Vertex AI Search, dan Gemini 3.5 Flash. Tiga validator yang menilai sepuluh skenario menghasilkan rata-rata 4,77 dari skala 5 dengan ICC(2,k) sebesar 0,47. Evaluasi lima pasangan jawaban acuan menghasilkan skor F1 *BERTScore* 84,06%, *ROUGE-1* 60,56%, *ROUGE-2* 42,78%, dan *ROUGE-L* 53,10%. Analisis kesalahan menunjukkan perlunya pembaruan cakupan kalkulator pajak restoran dan *logika* PPh Pasal 21 berbasis TER. Perbandingan observasional mendukung manfaat penyajian sumber regulasi, tetapi belum dapat membuktikan keunggulan kausal RAG karena konfigurasi pembandingan tidak dikendalikan. Pengujian lanjutan perlu memperluas skenario, menambah validator, menyamakan instrumen pembandingan, dan mencatat proses pengambilan dokumen.

DAFTAR PUSTAKA

- [1] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.
- [2] T. B. Brown *et al.*, "Language Models are Few-Shot Learners," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 1877–1901, 2020.
- [3] Y. Han, A. Ceross, and J. H. M. Bergmann, "The Use of Readability Metrics in Legal Text: A Systematic Literature Review," 2024. [Online]. Available: <http://arxiv.org/abs/2411.09497>
- [4] R. M. Oktaviani, H. Kurnia, Sunarto, and Udin, "The effects of taxpayer knowledge and taxation socialization on taxpayer compliance: the role of taxpayer awareness in developing Indonesian economy," *Accounting*, vol. 6, no. 2, pp. 89–96, 2020, doi: 10.5267/j.ac.2019.12.004.
- [5] E. Adamopoulou and L. Moussiades, "Machine Learning with Applications Chatbots : History , technology , and applications," *Machine Learning with Applications*, vol. 2, p. 100006, 2020, doi: 10.1016/j.mlwa.2020.100006.
- [6] Z. Ji *et al.*, "Survey of Hallucination in Natural Language Generation," *ACM Comput. Surv.*, vol. 55, no. 12, 2023, doi: 10.1145/3571730.
- [7] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, 2020.
- [8] Google, "Grounding with Agent Search." Accessed: Jun. 19, 2026. [Online]. Available: <https://docs.cloud.google.com/gemini-enterprise-agent-platform/models/grounding/grounding-with-vertex-ai-search>
- [9] "Gemini 3.5: frontier intelligence with action." Accessed: Jun. 23, 2026. [Online]. Available: <https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-5/#gemini-3-5-flash>
- [10] G. D. Albert and A. Voutama, "PENGEMBANGAN CHATBOT BERBASIS PDF MENGGUNAKAN LOCAL RETRIEVAL-AUGMENTED GENERATION (RAG) DAN OLLAMA," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6361.
- [11] I. D. Saputra, N. S. Harahap, S. Agustian, M. Fikry, and L. Oktavia, "Aplikasi Web Question Answering Menggunakan Langchain OpenAI Tentang Peraturan Perundang-undangan Bidang Pendidikan," *Journal of Computer System and Informatics (JoSYC)*, vol. 6, no. 1, pp. 293–304, Nov. 2024, doi: 10.47065/josyc.v6i1.6182.
- [12] Adib Faishol and Arief Ichwani, "Implementasi Fitur Chatbot Pada Website Konsultasi Jasa Pajak Online berbasis Laravel dan React JS (Studi Kasus: AMZ Tax Consultant)," *IKRA-ITH Informatika : Jurnal Komputer dan Informatika*, vol. 10, no. 1, pp. 86–95, 2026, doi: 10.37817/ikraith-informatika.v10i1.5706.
- [13] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTSCORE: EVALUATING TEXT GENERATION WITH BERT," in *8th International Conference on Learning Representations, ICLR 2020*, 2020, pp. 1–43.
- [14] M. Barbella and G. Tortora, "Rouge Metric Evaluation for Text Summarization Techniques," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4120317.
- [15] BPK RI, "Database Peraturan | JDIH BPK." Accessed: Jun. 19, 2026. [Online]. Available: <https://peraturan.bpk.go.id/>
- [16] Google, "Parse and chunk documents," 2025. Accessed: Jun. 19, 2026. [Online]. Available: <https://docs.cloud.google.com/generative-ai-app-builder/docs/parse-chunk-documents>
- [17] Google, "Embeddings APIs overview," 2026. Accessed: Jun. 19, 2026. [Online]. Available: <https://docs.cloud.google.com/gemini-enterprise-agent-platform/models/embeddings>
- [18] Vercel, "Next.js Docs | Next.js." Accessed: Jun. 19, 2026. [Online]. Available: <https://nextjs.org/docs>
- [19] Vercel, "tRPC | tRPC." Accessed: Jun. 19, 2026. [Online]. Available: <https://trpc.io/docs>
- [20] A. T. Jebb, V. Ng, and L. Tay, "A Review of Key Likert Scale Development Advances: 1995–2019," *Front. Psychol.*, vol. 12, pp. 1–14, 2021, doi: 10.3389/fpsyg.2021.637547.