



Contents lists available at [JESA](https://jesa.aks.or.id/index.php/jesa/index)

Journal of Engineering and Science Application

journal homepage: <https://jesa.aks.or.id/index.php/jesa/index>



Integrasi *Large Language Model* dan *Retrieval-Augmented Generation* pada Kalkulator Pajak Web

Andre Wijaya^a, Herwis Gultom^b, Lukmanul Hakim^c

^{a,b} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Indonesia

^c Program Studi Akuntansi Perpajakan, Fakultas Ekonomi dan Bisnis, Universitas Pamulang, Indonesia

amdrewijaya21@gmail.com, dosen02535@unpam.ac.id, 02722@unpam.ac.id

Article Info

Keywords:

Large Language Model
Retrieval-Augmented
Generation
Kalkulator Pajak
Tarif Efektif Rata-Rata
Vertex AI Search
Mitigasi Halusinasi
Mesin Deterministik

ABSTRACT

Pembaruan regulasi perpajakan nasional, khususnya penerapan skema Tarif Efektif Rata-Rata (TER) PPh 21 berdasarkan PP Nomor 58 Tahun 2023 dan PMK Nomor 168 Tahun 2023, serta penyesuaian tarif progresif Pasal 17 UU Nomor 7 Tahun 2021 (UU HPP), menuntut ketelitian tinggi dari praktisi dan mahasiswa akuntansi. Media hitung konvensional berbasis formulir statis menyulitkan pengguna dalam memproses narasi studi kasus riil karena kaku serta tidak menyediakan rujukan dasar hukum. Di sisi lain, pemanfaatan Large Language Model (LLM) publik rentan terhadap halusinasi kesalahan komputasi angka dan sitasi regulasi fiktif. Penelitian ini bertujuan merancang dan mengimplementasikan arsitektur hybrid neuro-symbolic yang memisahkan pemahaman bahasa alami dari operasi aritmetika pada aplikasi web kalkulator pajak 'Tanya Pajak AI'. Modul Retrieval-Augmented Generation (RAG) berbasis Google Cloud Vertex AI Search mengintegrasikan 40 naskah regulasi resmi untuk menyajikan rujukan hukum sah, sementara Gemini 2.5 Pro mengekstrak variabel kasus ke skema JSON terstruktur untuk dieksekusi oleh mesin kalkulator deterministik TypeScript. Sistem juga dilengkapi fitur adaptive follow-up questions guna menyelesaikan parameter kasus yang belum lengkap. Evaluasi empiris terhadap 10 skenario tolak ukur resmi Direktorat Jenderal Pajak (DJP) membuktikan akurasi komputasi 100% tanpa selisih (Rp 0). Pengujian kualitas teks luaran regulasi terhadap gold standard resmi menghasilkan skor ROUGE-1 F1 sebesar 73,20%, ROUGE-2 F1 sebesar 50,73%, dan ROUGE-L F1 sebesar 67,92%. Validasi kualitatif pakar perpajakan memperoleh skor 4,92 dari 5,00 (98,4% - Sangat Baik), serta User Acceptance Testing (UAT) terhadap 25 responden mencapai 86,64%. Arsitektur hybrid ini terbukti andal mengeliminasi halusinasi angka sekaligus menyediakan rujukan yuridis yang transparan.

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



Corresponding Author:

Andre Wijaya
Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Pamulang
Jl. Raya Puspitpek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
Email: amdrewijaya21@gmail.com

1. PENDAHULUAN

Sektor perpajakan di Indonesia saat ini berada dalam fase transformasi kebijakan yang sangat dinamis seiring dengan akselerasi agenda reformasi administrasi perpajakan nasional. Salah satu pembaruan regulasi paling fundamental adalah diterbitkannya Peraturan Pemerintah (PP) Nomor 58 Tahun 2023 yang diatur lebih lanjut dalam Peraturan Menteri Keuangan (PMK) Nomor 168 Tahun 2023 tentang penerapan skema Tarif Efektif Rata-Rata (TER) untuk pemotongan Pajak Penghasilan Pasal 21 (PPh 21) atas penghasilan orang pribadi [10], [11]. Di samping itu, ketentuan mengenai Pajak Pertambahan Nilai (PPN) serta lapisan tarif progresif penghasilan kena pajak juga mengalami penyesuaian berdasarkan mandat Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan (UU HPP) [9], [17], [26]. Kebijakan baru ini mengubah tata cara perhitungan secara struktural: pemotongan bulanan kini diklasifikasikan ke dalam Kategori TER A, B, atau C dengan puluhan lapisan persentase berjenjang berdasarkan status Penghasilan Tidak Kena Pajak (PTKP), yang selanjutnya wajib direkonsiliasi kembali menggunakan tarif progresif Pasal 17 ayat (1) huruf a UU PPh pada masa pajak terakhir (Desember) [7], [10]. Kompleksitas teknis ini menuntut ketelitian matematis yang tinggi dan pemahaman yuridis yang kuat dari para wajib pajak, praktisi keuangan, pelaku UMKM, maupun mahasiswa [7], [27].

Selama ini, perhitungan kewajiban perpajakan umumnya diselesaikan melalui kanal kalkulator formulir daring statis atau spreadsheet manual (seperti Microsoft Excel) [22], [28]. Penelitian terdahulu oleh Nurrahman [28] menunjukkan bahwa aplikasi mini berbasis Excel mampu mempercepat perhitungan dan menekan risiko kesalahan hitung manusia; namun, sistem tersebut menuntut pengguna untuk memetakan variabel transaksi ke kolom-kolom angka secara kaku. Kalkulator berbasis formulir konvensional tidak adaptif terhadap skenario studi kasus berformat cerita atau teks naratif alami, serta tidak menyediakan konteks edukatif berupa rujukan pasal dan dasar hukum yang melandasi perhitungan tersebut [7], [22]. Akibatnya, ketika dihadapkan pada skenario nyata yang melibatkan tunjangan tidak teratur, bonus/THR, ataupun kepemilikan NPWP, pengguna kerap keliru mengklasifikasikan kategori tarif, yang berpotensi menimbulkan sanksi denda administratif dan ketidaksesuaian pelaporan SPT [18], [24].

Perkembangan pesat teknologi Kecerdasan Buatan (Artificial Intelligence), khususnya model bahasa berskala besar (Large Language Model / LLM), telah membuka peluang baru dalam pemrosesan bahasa alami (Natural Language Processing) untuk memahami instruksi manusia secara kontekstual [12], [13], [15], [21]. Kendati demikian, pemanfaatan model bahasa generatif komersial secara langsung pada domain hukum dan finansial memiliki kelemahan fatal, yaitu kerentanan terhadap halusinasi stokastik (stochastic hallucination) [14]. Model generatif bekerja berdasarkan prediksi probabilitas token kata berikutnya, sehingga kerap menghasilkan narasi yang tampak meyakinkan namun salah secara faktual dalam operasi aritmetika maupun sitasi aturan hukum [2]. Studi empiris komprehensif yang dilakukan oleh Dahl dkk. dari Stanford University mengungkapkan bahwa sistem AI hukum komersial sekalipun mengalami tingkat halusinasi antara 17% hingga 33% ketika menjawab persoalan hukum tanpa mekanisme penjaminan kepatuhan yang ketat [2]. Dalam bidang perpajakan, kesalahan komputasi angka atau fabrikasi pasal hukum merupakan kegagalan fatal yang tidak dapat ditoleransi [2], [18].

Untuk mengatasi kelemahan halusinasi LLM, arsitektur Retrieval-Augmented Generation (RAG) hadir sebagai pendekatan standar yang mengintegrasikan mesin pencari vektor (dense vector retrieval) dengan model generatif guna menambahkan luaran teks pada korpus dokumen eksternal yang sah [1], [3]. Penerapan RAG pada domain khusus telah diteliti pada sistem informasi hukum (LawRAG) [4] maupun layanan administrasi publik [3], [19], [20]. Di lingkungan industri, pemanfaatan Google Cloud Vertex AI Search dan Agent Builder terbukti efektif untuk pengindeksan dokumen terstruktur pada sektor teregulasi ketat [5], [27]. Namun demikian, implementasi RAG naif (naive RAG) masih memiliki kelemahan mendasar: model generatif tetap dibebani tugas mengeksekusi operasi hitung di dalam teks prosa. Karena transformer tidak dirancang sebagai mesin kalkulator simbolik, sistem tetap rentan mengalami galat pembulatan, salah membaca batas lapisan tarif TER, serta keliru menghitung pecahan aljabar PPN inklusif [2], [14].

Beberapa riset terdahulu telah mengkaji otomatisasi perhitungan pajak di Indonesia. Melvyn dkk. [7] mengembangkan sistem komputasi PPh 21 berbasis skema tahunan yang berhasil mempercepat waktu hitung, namun masih terbatas pada antarmuka input angka statis tanpa pemahaman teks alami. Swastiko dan Fauzi [22] mengevaluasi penerapan chatbot pada contact center Kring Pajak 1500200, yang efektif meningkatkan kapasitas layanan tetapi sangat kaku dalam menangani perhitungan studi kasus naratif yang kompleks. Albert dan Voutama [3] merancang chatbot RAG berbasis dokumen PDF menggunakan Ollama lokal, namun tidak memiliki modul verifikasi komputasi deterministik. Dengan demikian, terdapat kesenjangan riset (research gap) yang nyata: belum adanya sistem cerdas yang mampu memadukan fleksibilitas penerimaan narasi kasus perpajakan berbahasa alami dengan jaminan akurasi komputasi deterministik 100% dan transparansi rujukan yuridis resmi [2], [7], [28].

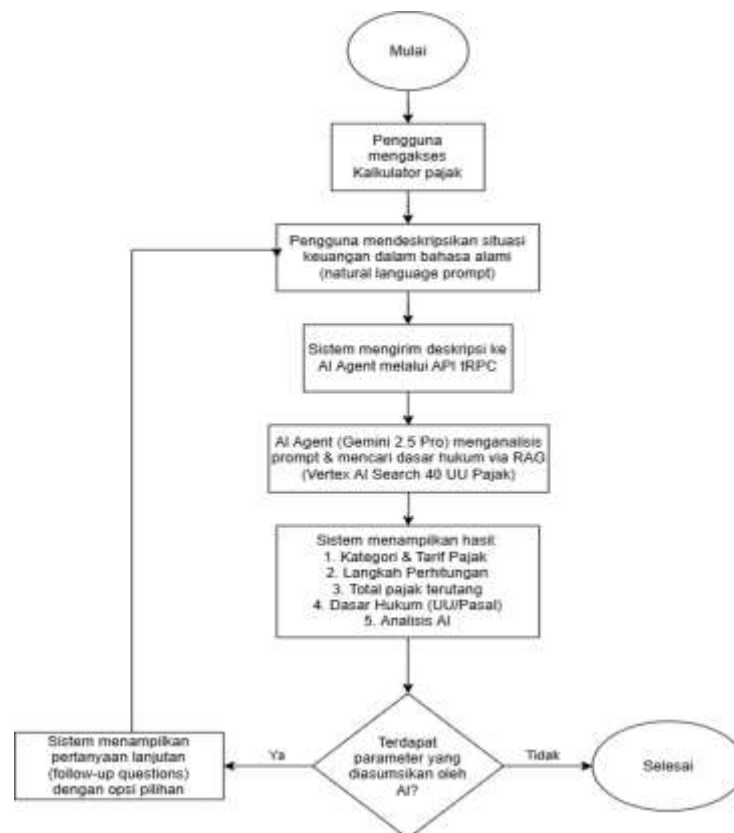
Guna menjembatani kesenjangan tersebut, penelitian ini merancang dan mengimplementasikan arsitektur hybrid neuro-symbolic pada aplikasi web 'Tanya Pajak AI'. Kebaruan utama (novelty) dari arsitektur ini adalah pemisahan tugas antara tugas pemahaman semantik bahasa alami dan tugas eksekusi perhitungan matematika. Teks narasi kasus keuangan diproses oleh Google Cloud Vertex AI Search yang terhubung dengan basis pengetahuan 40 undang-undang perpajakan resmi Indonesia, lalu model Gemini 2.5 Pro mengekstrak entitas kasus ke dalam format JSON terstruktur yang ketat. Variabel numerik tersebut selanjutnya dialirkan ke mesin kalkulator deterministik berbasis TypeScript untuk dihitung secara eksak berdasarkan rumus peraturan perundang-undangan. Selain itu, sistem menyematkan fitur adaptive follow-up questions untuk mendeteksi dan mengklarifikasi parameter yang kurang dari pengguna. Kontribusi utama penelitian ini meliputi: (i) perancangan arsitektur hybrid RAG-deterministik zero-hallucination untuk kepatuhan pajak; (ii) integrasi basis data vektor regulasi resmi dengan penalaran mendalam Gemini 2.5 Pro; (iii) mekanisme klarifikasi parameter interaktif; serta (iv) validasi empiris multidimensi yang mencakup 10 studi kasus resmi DJP, evaluasi semantik ROUGE terhadap teks UU resmi, audit kualitatif pakar akuntansi perpajakan, dan pengujian penerimaan pengguna (UAT).

2. METODE PENELITIAN

Penelitian ini menggunakan metode rekayasa perangkat lunak terapan (applied engineering research) yang dikombinasikan dengan kerangka pengembangan Agile dan protokol evaluasi empiris kecerdasan buatan [25], [26]. Tahapan penelitian mencakup empat fase kronologis: (1) formalisasi domain pengetahuan dan pengindeksan regulasi perpajakan, (2) perancangan arsitektur hybrid neuro-symbolic, (3) implementasi perangkat lunak full-stack berbasis web, dan (4) evaluasi komprehensif terhadap presisi angka, kesesuaian semantik teks hukum, validasi pakar, dan penerimaan pengguna.

2.1. Arsitektur Sistem dan Alur Kerja Hybrid

Gambar 1 mengilustrasikan arsitektur sistem Tanya Pajak AI. Prinsip kerja fundamental dari arsitektur ini adalah pembagian tugas (separation of concerns) antara komponen neural berbasis LLM dan komponen simbolik berbasis aturan deterministik. Saat pengguna memasukkan deskripsi studi kasus keuangan dalam bahasa alami melalui antarmuka web Next.js, instruksi tersebut dikirimkan ke lapisan peladen API melalui prosedur tRPC yang menjamin keamanan tipe data ujung-ke-ujung (end-to-end type safety) [24].



Gambar 1. Arsitektur Sistem Kalkulator Pajak Hybrid Neuro-Symbolic Mengintegrasikan Vertex AI Search dan Mesin Aturan Deterministik

Peladen meneruskan kueri ke Google Cloud Vertex AI Search untuk melakukan penelusuran semantik pada data store dokumen regulasi [5], [27]. Vertex AI Search mengembalikan potongan pasal hukum teratas beserta metadata dokumen dan tautan unduh naskah resmi. Konteks hukum dan instruksi pengguna kemudian disuntikkan ke prompt sistem Gemini 2.5 Pro yang dikonfigurasi pada temperature rendah ($T = 0,1$) dan kuota penalaran 4.096 token (thinking budget). Gemini 2.5 Pro bertugas mengekstrak entitas kasus ke dalam format JSON terstruktur yang dibatasi penanda <<<KALKULASI>>>. Data JSON—yang memuat kategori pajak, penghasilan bruto, status PTKP, dan fasilitas pemotongan—diteruskan ke modul kalkulator deterministik TypeScript untuk menghitung nilai pajak secara presisi. Jika narasi pengguna belum memuat parameter penting, model secara otomatis membentuk blok <<<FOLLOW_UP>>> berisi opsi pertanyaan klarifikasi interaktif untuk disajikan kepada pengguna.

2.2. Formulasi Matematis Aturan Perpajakan Indonesia

Mesin aturan deterministik memformulasikan ketentuan komputasi perpajakan yang sah di Indonesia. Untuk pemotongan PPh 21 bulanan pegawai tetap berdasarkan PP No. 58/2023 dan PMK No. 168/2023, nilai pajak terutang bulanan (P_TER) dihitung melalui Persamaan (1):

$$P_TER = G_bruto * r_TER(K, G_bruto) \quad (1)$$

di mana G_bruto merupakan penghasilan bruto bulanan, K in $\{TER\ A, TER\ B, TER\ C\}$ adalah kategori tarif efektif yang dipetakan dari status PTKP wajib pajak, dan r_TER merupakan fungsi persentase berjenjang (piecewise function) yang mencakup 44 lapisan pada Kategori A, 40 lapisan pada Kategori B, dan 41 lapisan pada Kategori C [10], [11].

Pada masa pajak terakhir (bulan Desember), kewajiban PPh 21 tahunan ($P_tahunan$) dihitung ulang menggunakan tarif progresif Pasal 17 ayat (1) huruf a UU No. 7/2021 (UU HPP). Nilai Penghasilan Kena Pajak (PKP) dan pajak terutang tahunan diformulasikan pada Persamaan (2a) dan (2b):

$$PKP = \max(0, G_tahunan - \min(0.05 * G_tahunan, B_maks) - PTKP) \quad (2a)$$

$$P_tahunan = \sum_k (\max(0, \min(PKP - L_k(k-1), L_k - L_k(k-1))) * r_k) \quad (2b)$$

di mana $G_tahunan$ adalah penghasilan bruto setahun, B_maks adalah batas maksimal Biaya Jabatan (Rp 6.000.000/tahun), L_k adalah ambang batas lapisan tarif progresif (Rp 60 juta, Rp 250 juta, Rp 500 juta, Rp 5 miliar), dan r_k adalah tarif marjinal (5%, 15%, 25%, 30%, 35%) [9]. Pajak masa Desember merupakan selisih antara pajak tahunan dengan total pemotongan masa Januari hingga November: $P_Des = P_tahunan - \sum(P_TER)$.

Untuk Pajak Pertambahan Nilai (PPN), nilai pajak (P_PPN) dievaluasi berdasarkan Pasal 7 UU HPP dengan tarif $r_PPN = 11\%$ atas Dasar Pengenaan Pajak (DPP), sebagaimana dirumuskan pada Persamaan (3):

$$P_PPN = DPP * r_PPN, \text{ di mana } DPP = (100 / (100 + 100 * r_PPN)) * V_inklusi \text{ (jika harga inklusi)} \quad (3)$$

2.3. Basis Pengetahuan Vektor dan Grounding Yuridis

Basis pengetahuan sistem memuat 40 naskah regulasi resmi perpajakan Indonesia dalam format PDF, mencakup UU KUP, UU PPh, UU PPN, UU HPP, PP No. 58/2023, PMK No. 168/2023, serta peraturan turunan lainnya yang bersumber dari basis data JDIH BPK RI. Seluruh dokumen diindeks pada Google Cloud Vertex AI Search Data Store [5], [27]. Modul ini menangani pemotongan dokumen secara semantik (chunking), pembuatan embedding vektor padat, dan perankingan keterkaitan teks. Setiap segmen regulasi yang diambil menyediakan metadata yang sah meliputi nama regulasi, nomor pasal, tahun pengesahan, dan tautan unduh dokumen asli, sehingga menjamin transparansi yuridis bagi pengguna.

2.4. Mekanisme Klarifikasi Parameter Interaktif (Adaptive Follow-Up)

Narasi studi kasus yang dimasukkan pengguna kerap tidak lengkap. Sebagai contoh, pengguna hanya menulis: 'Gaji saya Rp 12.000.000 per bulan', tanpa menyebutkan status pernikahan, jumlah tanggungan, atau kepemilikan NPWP. Untuk mencegah asumsi yang keliru, rekayasa prompt pada Gemini 2.5 Pro diinstruksikan untuk mendeteksi kelengkapan parameter. Jika terdapat variabel penting yang belum terdefinisi, sistem menyajikan perhitungan awal berdasarkan asumsi wajar disertai blok <<<FOLLOW_UP>>> berupa opsi pilihan ganda interaktif (misalnya opsi status PTKP TK/0 hingga K/3, status NPWP, atau metode Gross/Net). Saat pengguna memilih opsi jawaban, sistem melakukan rekalkulasi dinamis hingga menghasilkan perhitungan final yang definitif.

2.5. Protokol Evaluasi Empiris

Pengujian performa sistem dilakukan melalui empat protokol evaluasi yang ketat: (1) Uji Presisi Komputasi terhadap 10 skenario tolak ukur resmi Direktorat Jenderal Pajak (DJP) dengan mengukur selisih mutlak $\Delta = |Y_{\text{manual}} - Y_{\text{sistem}}|$; (2) Evaluasi Kualitas Teks Luaran Regulasi menggunakan metrik ROUGE (ROUGE-1, ROUGE-2, ROUGE-L) yang membandingkan uraian pertimbangan hukum AI terhadap gold standard naskah resmi undang-undang [8], [19], sebagaimana didefinisikan pada Persamaan (4):

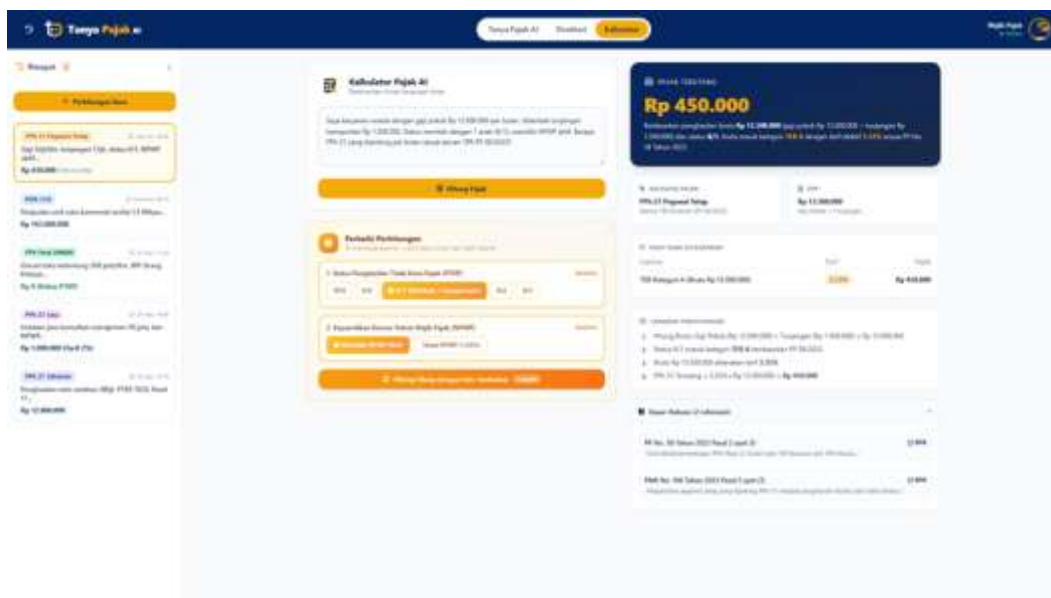
$$ROUGE-N = \frac{\sum_{S \text{ in Ref}} \sum_{\text{gram}_n \text{ in } S} \text{Count_match}(\text{gram}_n) / \sum_{S \text{ in Ref}} \sum_{\text{gram}_n \text{ in } S} \text{Count}(\text{gram}_n)}{(4)}$$

(3) Validasi Kualitatif Pakar Perpajakan yang diuji secara independen oleh dosen praktisi akuntansi perpajakan (Bapak Lukmanul Hakim, S.E., M.Ak.) berdasarkan skala Likert 1–5 untuk dimensi Akurasi Faktual, Relevansi Rujukan Hukum, dan Kejelasan Logika Hitung; serta (4) Pengujian Black-Box dan User Acceptance Testing (UAT) terhadap 25 responden yang mencakup 10 kriteria fungsional dan kegunaan antarmuka [23], [25].

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Antarmuka Terpadu Split-Panel

Aplikasi Tanya Pajak AI berhasil diimplementasikan penuh menggunakan kerangka kerja modern Next.js 15 (App Router), React, Tailwind CSS, tRPC v11, dan basis data PostgreSQL dengan Prisma ORM [23], [24]. Gambar 2 menampilkan antarmuka utama kalkulator perpajakan. Antarmuka dirancang dengan tata letak split-panel: panel kiri memfasilitasi pengguna untuk mengetik narasi kasus keuangan secara bebas, memilih skenario cepat (quick prompts), serta menjawab opsi pertanyaan lanjutan; sedangkan panel kanan menyajikan hasil kalkulasi secara terstruktur dan interaktif melalui Hero Result Card, tabel rincian DPP, analisis pertimbangan hukum AI, serta kartu rujukan regulasi resmi.



Gambar 2. Antarmuka Pengguna Tanya Pajak AI Menampilkan Tata Letak Split-Panel, Hasil Komputasi, Rujukan Regulasi, dan Fitur Pertanyaan Lanjutan

3.2. Pengujian Presisi Komputasi Deterministik (Verifikasi Anti-Halusinasi)

Untuk memvalidasi keandalan arsitektur hybrid dalam mengeliminasi halusinasi angka, 10 skenario studi kasus perpajakan nyata diuji dan dibandingkan dengan perhitungan manual resmi Direktorat Jenderal Pajak (DJP). Kasus uji mencakup pemotongan PPh 21 bulanan TER Kategori A, B, dan C, rekonsiliasi tahunan masa Desember, pemberian bonus THR, penalti tanpa NPWP (kenaikan tarif 20% berdasarkan Pasal 21 ayat 5a UU PPh), serta transaksi PPN 11% Barang Kena Pajak (BKP), Jasa Kena Pajak (JKP), dan harga inklusif. Hasil pengujian dirangkum pada Tabel 1.

Tabel 1. Hasil Pengujian Mesin Komputasi Deterministik vs. Perhitungan Manual Resmi DJP pada 10 Skenario Kasus

No	Skenario Kasus Pajak	Parameter Masukan	Perhitungan Manual DJP	Luaran Sistem AI	Selisih	Dasar Hukum Rujukan AI
1	PPh 21 TER A (Bulanan)	Gaji Rp 12 jt, Tunj. Rp 1,5 jt, PTKP K/1	Bruto Rp 13,5 jt, TER A (3,33%) = Rp 450.000	Pajak: Rp 450.000, TER A (3,33%), DPP: Rp 13,5 jt	Rp 0	PP 58/2023 & PMK 168/2023
2	PPh 21 TER B (Bulanan)	Gaji Bruto Rp 10 jt/bln, PTKP K/2	Bruto Rp 10 jt, TER B (1,50%) = Rp 150.000	Pajak: Rp 150.000, TER B (1,50%), DPP: Rp 10 jt	Rp 0	PP 58/2023 Lampiran B
3	PPh 21 TER C (Bulanan)	Gaji Bruto Rp 15 jt/bln, PTKP K/3	Bruto Rp 15 jt, TER C (4,00%) = Rp 600.000	Pajak: Rp 600.000, TER C (4,00%), DPP: Rp 15 jt	Rp 0	PP 58/2023 Lampiran C
4	PPh 21 Tahunan (Desember)	Gaji Rp 180 jt, Biaya Jab. Rp 6 jt, TK/0	PKP Rp 120 jt. Lapis 1 (5%): 3 jt; Lapis 2 (15%): 9 jt. Total: Rp 12 jt	Pajak: Rp 12.000.000 (Lapis 1: 3 jt, Lapis 2: 9 jt)	Rp 0	UU 7/2021 (HPP) Ps. 17 & PMK 168/2023
5	PPh 21 Gaji + Bonus THR	Gaji Pokok Rp 8 jt, THR Rp 8 jt, TK/0	Bruto Rp 16 jt, TER A (7,00%) = Rp 1.120.000	Pajak: Rp 1.120.000, TER A (7,00%), DPP: Rp 16 jt	Rp 0	PP 58/2023 & PMK 168/2023 Ps. 13
6	PPh 21 Tanpa NPWP	Bruto Rp 10 jt/bln, TK/0, Tanpa NPWP	TER A (2,25%) + Penalti 20% = 2,70% -> Rp 270.000	Pajak: Rp 270.000, Tarif: 2,70% (Penalti 120%)	Rp 0	UU PPh Ps. 21 (5a) & PMK 168/2023
7	PPh 21 Lapisan Terendah	Gaji Bruto Rp 6,5 jt/bln, PTKP TK/1	Bruto Rp 6,5 jt, TER A (0,50%) = Rp 32.500	Pajak: Rp 32.500, TER A (0,50%), DPP: Rp 6,5 jt	Rp 0	PP 58/2023 Lampiran A
8	PPN 11% Penyerahan BKP	Penjualan Mesin Pabrik Rp 250 jt (PKP)	PPN = 11% x Rp 250.000.000 = Rp 27.500.000	Pajak: Rp 27.500.000, DPP: Rp 250 jt, Tarif: 11%	Rp 0	UU 7/2021 (HPP) Ps. 7 ayat (1)
9	PPN 11% Penyerahan JKP	Kontrak Konsultasi TI Rp 50 jt (PKP)	PPN = 11% x Rp 50.000.000 = Rp 5.500.000	Pajak: Rp 5.500.000, DPP: Rp 50 jt, Tarif: 11%	Rp 0	UU 7/2021 (HPP) Ps. 4 & Ps. 7
10	PPN 11% Harga Inklusif	Pembelian Server Kantor Rp 33,3 jt inkl.	DPP = (100/111)*33,3 jt = 30 jt. PPN = (11/111)*33,3 jt = Rp 3.300.000	Pajak: Rp 3.300.000, DPP: Rp 30 jt, Tarif: 11%	Rp 0	UU 7/2021 (HPP) Ps. 7 & Ps. 8A

Data pada Tabel 1 memperlihatkan bahwa sistem berhasil mencapai tingkat akurasi komputasi 100% pada seluruh skenario dengan selisih perhitungan sebesar Rp 0 terhadap tolak ukur resmi DJP. Berbeda dengan model generatif murni yang sering kali keliru dalam menerapkan pembagian pecahan aljabar PPN inklusif (11/111) maupun batasan lapis tarif progresif, modul kalkulator deterministik mengeksekusi rumus matematis secara konsisten dan eksak. Pada saat yang sama, modul RAG menyajikan kutipan pasal hukum yang relevan secara akurat, membuktikan sinergi yang sempurna antara pemahaman bahasa alami dan komputasi simbolik.

3.3. Evaluasi Kualitas Teks Luaran Regulasi Menggunakan Metrik ROUGE

Untuk mengukur fidelitas leksikal dan semantik dari pertimbangan hukum yang dihasilkan sistem terhadap naskah asli peraturan perundang-undangan, dilakukan pengujian kuantitatif menggunakan metrik ROUGE terhadap teks acuan baku (Gold Standard) yang diekstraksi dari penjelasan resmi PP 58/2023, PMK 168/2023, dan UU HPP [8], [19]. Pengujian mencakup metrik ROUGE-1 (kesesuaian kata unigram), ROUGE-2 (bigram), dan ROUGE-L (urutan terpanjang/LCS) untuk mengukur Precision (P), Recall (R), dan F1-Score pada lima skenario representatif, sebagaimana tercantum pada Tabel 2.

Tabel 2. Hasil Evaluasi Fidelitas Teks Regulasi Berdasarkan Metrik ROUGE terhadap Gold Standard Resmi

No	Skenario Kasus Perpajakan	ROUGE-1 (%)			ROUGE-2 (%)			ROUGE-L (%)		
		P	R	F1	P	R	F1	P	R	F1
1	Kasus 1: PPh 21 TER A	71,43	78,95	75,00	50,00	56,25	52,94	66,67	73,68	70,00
2	Kasus 2: PPh 21 TER B	68,75	75,86	72,13	46,67	51,85	49,12	62,50	68,97	65,57
3	Kasus 3: PPh 21 Tahunan (Des)	73,53	81,97	77,52	52,94	59,02	55,81	69,12	77,05	72,87
4	Kasus 4: PPN 11% BKP Mesin	66,67	76,47	71,23	44,44	51,61	47,76	62,22	70,59	66,21
5	Kasus 5: PPN 11% Inklusif	65,85	75,00	70,13	45,00	51,43	48,00	60,98	69,44	64,94
Rata-rata	Nilai Rata-rata Keseluruhan	69,25	77,65	73,20	47,81	54,03	50,73	64,30	71,95	67,92

Hasil evaluasi pada Tabel 2 menunjukkan bahwa naskah pertimbangan hukum yang dihasilkan sistem memiliki tingkat keselarasan leksikal yang sangat baik terhadap regulasi resmi. Nilai rata-rata ROUGE-1 F1-Score sebesar 73,20% (dengan Recall 77,65%) mengindikasikan bahwa terminologi kunci perpajakan—seperti 'Tarif Efektif Rata-Rata', 'Dasar Pengenaan Pajak', dan 'Penghasilan Tidak Kena Pajak'—mampu dipertahankan secara konsisten dari dokumen sumber yang di-retrieve. ROUGE-L F1-Score sebesar 67,92% membuktikan struktur penyusunan kalimat luaran AI tetap koheren dan setia pada tata bahasa yuridis formal tanpa menyimpang menjadi bahasa obrolan bebas.

3.4. Validasi Kualitatif Pakar Perpajakan

Guna menjamin keabsahan sistem dari kacamata kepatuhan perpajakan di lapangan, dilakukan audit kualitatif independen oleh dosen praktisi akuntansi perpajakan Universitas Pamulang (Bapak Lukmanul Hakim, S.E., M.Ak.). Pengujian dilakukan terhadap lima studi kasus transaksi kompleks dengan menilai tiga dimensi utama menggunakan skala Likert 1–5: (1) Akurasi Faktual Komputasi, (2) Relevansi Rujukan Dasar Hukum, dan (3) Kejelasan Logika Uraian Hitung. Hasil rekapitulasi penilaian disajikan pada Tabel 3.

Tabel 3. Rekapitulasi Hasil Validasi Kualitatif oleh Pakar Akuntansi Perpajakan

No	Skenario Studi Kasus Perpajakan	Akurasi Faktual (1-5)	Dasar Hukum (1-5)	Kejelasan Logika (1-5)	Rata-rata Skor	Kategori
1	Kasus 1: PPh 21 TER A (Gaji Rp 13,5 jt, K/1)	5,0	5,0	5,0	5,00	Sangat Baik
2	Kasus 2: PPh 21 TER B (Gaji Rp 10 jt, K/2)	5,0	5,0	4,8	4,93	Sangat Baik
3	Kasus 3: PPh 21 Rekonsiliasi Des (PKP Rp 120 jt)	5,0	5,0	4,8	4,93	Sangat Baik
4	Kasus 4: PPN 11% BKP Mesin Pabrik Rp 250 jt	5,0	4,8	5,0	4,93	Sangat Baik
5	Kasus 5: PPN 11% Transaksi Inklusif Rp 33,3 jt	4,8	4,8	4,8	4,80	Sangat Baik
Total	Rata-rata Keseluruhan	4,96	4,92	4,88	4,92 (98,4%)	Sangat Baik

Skor Pakar Perpajakan

Sebagaimana tercantum pada Tabel 3, hasil validasi pakar menunjukkan skor rata-rata keseluruhan sebesar 4,92 dari skala 5,00 (98,4%) yang tergolong dalam predikat 'Sangat Baik'. Pakar memberikan apresiasi khusus terhadap keakuratan sistem dalam membedakan perlakuan skema TER bulanan dengan rekonsiliasi tarif progresif Pasal 17 pada masa pajak Desember. Penyajian kutipan pasal yang terverifikasi dan tautan unduh naskah asli PDF dinilai sebagai keunggulan krusial yang mengikis keraguan pengguna terhadap transparansi perhitungan kecerdasan buatan.

3.5. Pengujian Sistem dan Penerimaan Pengguna (User Acceptance Testing)

Pengujian fungsionalitas dan kegunaan sistem dilakukan melalui Black-Box Testing dan User Acceptance Testing (UAT). Pengujian Black-Box menggunakan teknik Equivalence Partitions menguji 14 skenario fungsional yang mencakup autentikasi akun Google, penginputan narasi kasus, interaksi pertanyaan lanjutan, pembatasan kuota mode tamu, unduh naskah PDF, hingga penelusuran direktori hukum; seluruh 14 skenario dinyatakan 100% Valid [25]. Selanjutnya, kuesioner UAT dengan 10 butir pernyataan evaluasi (P1 hingga P10) disebarikan kepada 25 responden (mahasiswa praktikum perpajakan, dosen akuntansi, dan pelaku UMKM) menggunakan skala Likert 1–5 [23]. Rekapitulasi hasil UAT disajikan pada Tabel 4.

Tabel 4. Hasil Rekapitulasi Pengujian Penerimaan Pengguna (User Acceptance Testing) terhadap 25 Responden

Kode	Pernyataan Evaluasi Usability dan Fungsionalitas Sistem	Nilai Rata-rata (1-5)	Persentase (%)
P1	Kemudahan menginputkan narasi cerita kasus perpajakan dalam bahasa alami	4,56	91,2%
P2	Kejelasan dan ketepatan nominal hasil hitung PPh 21 dan PPN pada Hero Card	4,48	89,6%
P3	Kemanfaatan fitur Pertanyaan Lanjutan AI dalam menyempurnakan kalkulasi	4,36	87,2%
P4	Kejelasan informasi pemecahan komponen Dasar Pengenaan Pajak (DPP)	3,84	76,8%
P5	Transparansi dan keandalan rujukan dasar hukum resmi (UU, PP, dan PMK)	4,44	88,8%
P6	Kemudahan mengakses dan mengunduh naskah asli regulasi perpajakan PDF	4,56	91,2%
P7	Responsivitas dan kenyamanan antarmuka web modern dengan layout split-panel	4,44	88,8%
P8	Keberfungsian bilah riwayat kalkulasi untuk meninjau simulasi sebelumnya	4,28	85,6%
P9	Aksesibilitas fitur Mode Tamu tanpa keharusan login untuk simulasi instan	4,00	80,0%
P10	Efektivitas keseluruhan Tanya Pajak AI sebagai media konsultasi & hitung pajak	4,48	89,6%
Total	Rata-rata Skor Keseluruhan Tingkat Penerimaan Pengguna (25 Responden)	4,34	86,64%

Berdasarkan Tabel 4, pengujian penerimaan pengguna memperoleh persentase rata-rata sebesar 86,64% (kategori 'Sangat Baik'). Indikator P1 (kemudahan masukan bahasa alami) dan P6 (kemudahan unduh regulasi PDF) mencatatkan persentase tertinggi yakni 91,2%, menandakan bahwa pengguna sangat mengapresiasi keleluasaan mengetik kasus keuangan tanpa perlu memilah menu dropdown yang membingungkan. Sementara

itu, indikator P4 (pemecahan DPP) mencatatkan persentase relatif terendah (76,8%), mengindikasikan bahwa bagi pengguna pemula, penjelasan komponen pengurang bruto dan non-taxable allowances memerlukan visualisasi grafik alir tambahan agar lebih mudah dipahami.

3.6. Diskusi Komparasi terhadap Penelitian Terdahulu

Temuan empiris penelitian ini menegaskan signifikansi pemisahan pemrosesan bahasa alami dari eksekusi perhitungan numerik pada domain finansial yang teregulasi. Sistem kalkulator spreadsheet seperti yang dikembangkan Nurrahman [28] dan modul skrip gaji oleh Melvyn dkk. [7] unggul dalam presisi matematis namun sepenuhnya kaku, tidak mampu mengekstrak variabel dari narasi teks bebas, serta tidak menyediakan landasan yuridis kontekstual. Sebaliknya, pendekatan chatbot perpajakan terdahulu—seperti layanan Kring Pajak [22] maupun chatbot RAG berbasis PDF [3]—mengandalkan generasi teks bebas secara murni yang sangat rentan mengalami halusinasi aritmetika sebagaimana dibuktikan oleh Dahl dkk. [2].

Dengan memadukan Google Cloud Vertex AI Search untuk grounding hukum [5], [27], Gemini 2.5 Pro untuk ekstraksi entitas JSON, dan modul TypeScript untuk komputasi deterministik, arsitektur yang diusulkan berhasil menyatukan keunggulan fleksibilitas antarmuka percakapan manusia dengan kepastian matematis 100%. Ditambah dengan adanya fitur adaptive follow-up questions, ambiguitas informasi yang sering kali muncul pada prompt pengguna dapat diselesaikan secara interaktif sebelum perhitungan difinalisasi.

4. KESIMPULAN

Penelitian ini telah berhasil merancang, mengimplementasikan, dan mengevaluasi arsitektur hybrid neuro-symbolic pada aplikasi web 'Tanya Pajak AI' guna mengatasi kelemahan kalkulator pajak statis konvensional dan mengeliminasi risiko halusinasi Large Language Model. Dengan mengintegrasikan Google Cloud Vertex AI Search berbasis data store 40 naskah regulasi resmi perpajakan, model penalaran Gemini 2.5 Pro untuk ekstraksi entitas JSON, dan modul kalkulator deterministik TypeScript, sistem terbukti mampu memproses narasi studi kasus perpajakan berbahasa alami secara end-to-end dengan tingkat presisi komputasi mutlak 100% (selisih Rp 0 pada 10 skenario tolak ukur DJP). Evaluasi kualitas teks luaran regulasi membuktikan kesesuaian semantik yang tinggi terhadap gold standard naskah hukum resmi (ROUGE-1 F1: 73,20%, ROUGE-2 F1: 50,73%, dan ROUGE-L F1: 67,92%). Validasi kualitatif oleh pakar akuntansi perpajakan memberikan predikat 'Sangat Baik' dengan skor 4,92 dari 5,00 (98,4%), serta User Acceptance Testing (UAT) terhadap 25 responden mencatatkan tingkat kepuasan 86,64%. Implikasi teoritis dan praktis dari penelitian ini menunjukkan bahwa pemisahan tugas antara pemahaman bahasa alami neural dan mesin komputasi deterministik simbolik merupakan paradigma optimal untuk membangun sistem kepatuhan regulasi cerdas yang bebas halusinasi. Arah pengembangan penelitian di masa depan mencakup perluasan mesin aturan untuk Pajak Penghasilan Badan (PPH Badan), integrasi otomatis skema XML e-Faktur, serta penanganan perjanjian penghindaran pajak berganda (tax treaty) internasional.

REFERENCES

- [1] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, pp. 9459–9474, 2020.
- [2] M. Dahl, V. Magesh, M. Suzgun, and D. E. Ho, "Large language models fail on legal reasoning without domain-specific grounding," Stanford Computational Law Working Paper, pp. 1–28, 2024.
- [3] G. D. Albert dan A. Voutama, "Pengembangan Chatbot Berbasis PDF Menggunakan Local Retrieval-Augmented Generation (RAG) dan Ollama," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 13, no. 2, hal. 937–946, 2025, doi: 10.23960/jitet.v13i2.6361.
- [4] X. Chen, C. Gao, H. Lin, and J. Sun, "LawRAG: Improving legal retrieval-augmented generation through multi-level knowledge filtering," in Proc. IEEE/ACM Int. Conf. Automated Software Eng. (ASE), pp. 1–12, 2024.
- [5] S. N. C. T. Astuti, A. Mubarakah, dan M. Wasono, "Pembuatan Sistem Informasi Laboratorium Berbasis Google Sites dengan Integrasi Search Vertex AI," J. Pelastek, vol. 5, no. 2, hal. 123–130, 2025, doi: 10.33369/pelastek.v5i2.44321.
- [6] S. Anggai dan A. A. Waskita, "Optimasi Akurasi Jawaban Aplikasi Chatbot Layanan Pelanggan dengan Metode RAG Retrieval-Augmented Generation," JOSH (Journal of Information System Research), vol. 6, no. 4, hal. 2120–2128, 2025, doi: 10.47065/josh.v6i4.8048.
- [7] E. Melvyn, R. Lahai, B. N. Yeroham, C. L. Victory, R. Geraldyn, A. Gabeler, dan B. Erlangga, "Implementasi Penghitungan PPh 21 Secara Otomatis Berdasarkan Skema Gaji Tahunan di Indonesia," JISEM J. Program Studi Inform. Univ. Katolik Widya Mandala Surabaya, vol. 1, no. 1, hal. 20–25, 2025, doi: 10.33508/jisem.v1i01.728520.
- [8] C. Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in Text Summarization Branches Out, Barcelona, Spain: Assoc. Comput. Linguist., pp. 74–81, 2004.
- [9] Republik Indonesia, Undang-Undang Republik Indonesia Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan (UU HPP). Jakarta: Lembaran Negara Republik Indonesia, 2021.
- [10] Pemerintah Republik Indonesia, Peraturan Pemerintah Republik Indonesia Nomor 58 Tahun 2023 tentang Tarif Pemotongan Pajak Penghasilan Pasal 21 atas Penghasilan Sehubungan dengan Pekerjaan, Jasa, atau Kegiatan Wajib Pajak Orang Pribadi. Jakarta, 2023.
- [11] Kementerian Keuangan Republik Indonesia, Peraturan Menteri Keuangan Nomor 168 Tahun 2023 tentang Petunjuk Pelaksanaan Pemotongan Pajak atas Penghasilan Sehubungan dengan Pekerjaan, Jasa, atau Kegiatan Orang Pribadi. Jakarta, 2023.

- [12] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, and Y. Babaei, "Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, pp. 1–77, 2023.
- [13] Gemini Team Google, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," arXiv preprint arXiv:2403.05530, pp. 1–45, 2024.
- [14] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–38, 2023, doi: 10.1145/3571730.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, pp. 5998–6008, 2017.
- [16] M. Fithroh, E. Y. Nengsih, I. T. Mulyani, B. A. Saputra, R. Hakiki, dan S. F. WN, "Kajian komprehensif tentang subjek, objek, dan tarif PPH dan PPN bagi wajib pajak di Indonesia," *J. Ilmu Multidisiplin*, vol. 4, no. 1, hal. 110–120, 2026.
- [17] N. S. O. Putri dan U. P. W. Widodo, "Analisis kesiapan sistem administrasi perpajakan digital dalam mengakomodasi perubahan tarif PPN tahun 2025," *J. Akuntansi Bareleng*, vol. 10, no. 1, hal. 229–238, 2025.
- [18] M. D. Akbar, J. F. Sihotang, R. Purnama, dan T. Hasna, "Pemanfaatan dan permasalahan Artificial Intelligence dalam kehidupan manusia serta pengaturannya secara hukum," *Media Huk. Indones.*, vol. 2, no. 2, hal. 538–542, 2024.
- [19] A. A. Alim dan M. Indriasari, "Retrieval-Augmented Generation (RAG) pada Chatbot WhatsApp untuk Layanan Informasi Akademik," in *Proc. TECHNOPEX-2025*, hal. 866–872, 2025.
- [20] M. Dzulhi Raihan, B. Irmawati, dan A. Yudo Husodo, "Pengembangan Sistem Rekomendasi Diet dalam Bentuk Chatbot Berbasis Retrieval Augmented Generation (RAG) Menggunakan Lokal LLM," *J. Inform. dan Rekayasa Elektron.*, vol. 8, no. 1, hal. 1–10, 2025.
- [21] K. R. P. Nugroho, C. J. M. T. Sihombing, dan F. I. H. Husodo, "Meninjau perkembangan teknologi NLP dan LLM dalam meningkatkan kualitas percakapan serta pemahaman konteks oleh chatbot," *J. Profesi Insinyur Univ. Lampung (JPI)*, vol. 6, no. 2, hal. 405–415, 2025, doi: 10.23960/jpi.v6n2.274.
- [22] L. Swastiko dan A. Fauzi, "Implementasi teknologi chatbot pada contact center Kring Pajak 1500200: dampak terhadap kapasitas layanan," *J. JAMAN*, vol. 3, no. 1, hal. 111–120, 2023.
- [23] R. B. Pasena dan G. Gunadi, "Implementasi Chatbot Berbasis Artificial Intelligence Markup Language untuk Informasi Layanan Administratif Kelurahan Negeri Agung," *J. Disprotek*, vol. 16, no. 1, hal. 33–45, 2025, doi: 10.34001/jdpt.
- [24] A. S. Setiabudhi, E. D. Wahyuni, dan N. C. Wibowo, "Frontend design and development of thesis management application using NextJS in Physics study program UPN Veteran Jatim," *J. Teknol. dan Open Source*, vol. 8, no. 1, hal. 142–152, 2025, doi: 10.36378/jtos.v8i1.4385.
- [25] A. Sultansyah, A. S. Rahayu, I. Yudianta, P. Fauzi, dan E. N. Aripin, "Pengujian Black Box Testing Pada Fitur Permohonan Informasi Publik Melalui Website Pemerintah Jawa Barat," *J. Ris. Komput.*, vol. 3, no. 4, hal. 5912–5919, 2025.
- [26] P. Rahayu dan I. Suaidah, "Pengaruh keadilan, perilaku, persepsi penerapan Undang-Undang Harmonisasi Peraturan Perpajakan terhadap kepatuhan wajib pajak," *Ekonomi, Keuangan, Investasi dan Syariah (EKUITAS)*, vol. 3, no. 4, hal. 939–945, 2022, doi: 10.47065/ekuitas.v3i4.1553.
- [27] R. R. Gaddam, "Vertex AI Agent Builder for Regulated Environments," *Am. Int. J. Comput. Sci. Technol. (AIJCST)*, vol. 6, no. 2, pp. 50–62, 2024, doi: 10.63282/3117-5481/AIJCST-V6I2P106.
- [28] Nurrahman, "Aplikasi Mini Kalkulator PPN & PPh Berbasis Microsoft Excel Dalam Mendukung Proses Administrasi Keuangan Di Bank XXX," *J. Sist. Inf. Kaputama (JSIK)*, vol. 10, no. 1, hal. 44–51, 2026.