

**IMPLEMENTASI TEKNIK RETRIEVAL AUGMENTED
GENERATION MENGGUNAKAN LARGE LANGUAGE
MODEL BERBASIS VECTOR DATABASE UNTUK
CHATBOT PAJAK**



SKRIPSI

**AXEL FERDINAND
00000055952**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**IMPLEMENTASI TEKNIK RETRIEVAL AUGMENTED
GENERATION MENGGUNAKAN LARGE LANGUAGE
MODEL BERBASIS VECTOR DATABASE UNTUK
CHATBOT PAJAK**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**AXEL FERDINAND
00000055952**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Axel Ferdinand
Nomor Induk Mahasiswa : 00000055952
Program Studi : Informatika

Skripsi dengan judul:

Implementasi Teknik Retrieval Augmented Generation Menggunakan Large Language Model Berbasis Vector Database Untuk Chatbot Pajak

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, Jumat, 06 Juni 2025



(Axel Ferdinand)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI TEKNIK RETRIEVAL AUGMENTED GENERATION MENGUNAKAN LARGE LANGUAGE MODEL BERBASIS VECTOR DATABASE UNTUK CHATBOT PAJAK

oleh

Nama : Axel Ferdinand
NIM : 00000055952
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

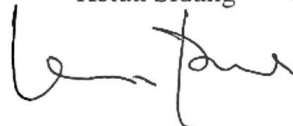
Telah diujikan pada hari Jumat, 11 Juli 2025

Pukul 10.00 s/s 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Dr. Ivransa Zuhdi Pane, B.Eng.,
M.Eng.)

NIDN: 08812520016

Penguji



(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)

NIDN: 0818038501

Pembimbing



(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

Ketua Program Studi Informatika,



(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Axel Ferdinand
NIM : 00000055952
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Implementasi Teknik Retrieval
Augmented Generation Menggunakan
Large Language Model Berbasis
Vector Database Untuk Chatbot Pajak

Menyatakan dengan sesungguhnya bahwa saya bersedia (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, Jumat, 06 Juni 2025

Yang menyatakan


Axel Ferdinand

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: IMPLEMENTASI TEKNIK RETRIEVAL AUGMENTED GENERATION MENGGUNAKAN LARGE LANGUAGE MODEL BERBASIS VECTOR DATABASE UNTUK CHATBOT PAJAK dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan laporan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan laporan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Airin Suryani Maitri yang mendukung saya untuk menyelesaikan tugas akhir ini.

Semoga skripsi dapat bermanfaat bagi para pembaca.

Tangerang, Jumat, 06 Juni 2025



Axel Ferdinand

IMPLEMENTASI TEKNIK RETRIEVAL AUGMENTED GENERATION MENGUNAKAN LARGE LANGUAGE MODEL BERBASIS VECTOR DATABASE UNTUK CHATBOT PAJAK

Axel Ferdinand

ABSTRAK

Banyak masyarakat yang kesulitan untuk memahami regulasi pajak. Dari banyaknya Undang - Undang dan Peraturan Pemerintah sampai dengan kurangnya ahli tenaga kerja seperti konsultan pajak yang dapat membantu masyarakat memahami tentang wajib pajak pribadi. Oleh karena itu, kita perlu memanfaatkan AI seperti *Generative AI* yang dapat membuat *chatbot* untuk berinteraksi dengan masyarakat dalam menjawab pertanyaan pajak. Namun, *chatbot* sendiri tidak cukup karena *chatbot* masih dapat memberikan jawaban yang halusinasi dan tidak faktual. Oleh karena itu, dibutuhkan teknik Retrieval Augmented Generation yang dapat menggabungkan *chatbot* dengan pencarian dokumen di *vector database*. Hasil penelitian menunjukkan bahwa teknik RAG hanya dapat menjawab pertanyaan umum tentang wajib pajak pribadi dengan baik. Pengujian dilakukan dengan wawancara konsultan pajak dan hasil yang didapatkan adalah *chatbot* dapat menjawab pertanyaan tentang wajib pajak pribadi dengan baik tetapi *chatbot* tidak dapat sepenuhnya menggantikan konsultan pajak. Pengujian yang lain dilakukan dengan evaluasi G-Eval yaitu LLM menilai *input* dan *output chatbot* dan didapatkan *score* sebesar 71%. *Score* yang didapatkan menunjukkan *chatbot* masih dapat dikembangkan lagi.

Kata kunci: *Chatbot, Large Language Model, Pajak, Retrieval Augmented Generation, Vector Database*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

**IMPLEMENTATION OF AUGMENTED GENERATION RETRIEVAL
TECHNIQUE USING LARGE LANGUAGE MODEL BASED ON VECTOR
DATABASE FOR TAX CHATBOT**

Axel Ferdinand

ABSTRACT

Many people find it difficult to understand tax regulations. From many laws and government regulations to the lack of labor experts such as tax consultants who can help the public understand about personal taxpayers. Therefore, we need to utilize AI such as Generative AI that can create a chatbot to interact with the public in answering tax questions. However, chatbots alone are not enough because chatbots can still provide answers that are hallucinatory and not factual. Therefore, a Retrieval Augmented Generation technique is needed that can combine chatbot with document search in vector database. The research results show that the RAG technique can only answer general questions about personal taxpayers well. Tests were conducted with tax consultant interviews and the results obtained were that the chatbot could answer questions about personal taxpayers well but the chatbot could not fully replace tax consultants. Another test was carried out with the evaluation of G-Eval, which is LLM assessing the input and output of the chatbot and obtained a score of 71%. The score obtained shows that the chatbot can still be developed further.

Keywords: Chatbot, Large Language Model, Retrieval Augmented Generation, Tax, Vector Database

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	5
2.1 Artificial Intelligence (AI)	5
2.2 Natural Language Processing (NLP)	5
2.3 Generative AI (GenAI)	6
2.4 Transformer	6
2.5 Embedding	10
2.6 Large Language Model (LLM)	12
2.7 Cosine Similarity	12
2.8 Vector Database	14
2.9 Retrieval Augmented Generation (RAG)	16
2.10 Metode Evaluasi G-Eval	18
BAB 3 METODOLOGI PENELITIAN	20
3.1 Alur Penelitian	20
3.2 Pengumpulan Data	21
3.3 Perancangan	22
3.3.1 Rancangan Vector Database	22
3.3.2 Rancangan RAG	23
3.3.3 Rancangan Chatbot	24
3.3.4 Rancangan Aplikasi	25
3.4 Evaluasi	26
3.5 Dokumentasi	27
BAB 4 HASIL DAN DISKUSI	28
4.1 Spesifikasi System	28
4.2 Implementasi	29
4.2.1 Implementasi Vector Database	29
4.2.2 Implementasi Agent	32
4.2.3 Implementasi Aplikasi	37

4.3	Pengujian Wawancara	41
4.4	Pengujian Evaluasi G-Eval	44
4.5	Pengujian Perbandingan Jawaban	46
4.6	Pengujian Similarity	48
BAB 5	SIMPULAN DAN SARAN	57
5.1	Simpulan	57
5.2	Saran	57
DAFTAR PUSTAKA		59



DAFTAR TABEL

Tabel 2.1	Hasil perhitungan <i>softmax</i> sebelum dikali value	10
Tabel 2.2	Contoh <i>One Hot Encoding</i>	11
Tabel 2.3	Perbandingan <i>metrics</i> untuk evaluasi <i>chatbot</i> [1]	19
Tabel 4.1	Daftar pertanyaan dan jawaban oleh chatbot yang keliru . .	42
Tabel 4.2	Tabel perbandingan hasil LLM dengan <i>chatbot</i>	47
Tabel 4.3	Daftar pengujian similarity	48
Tabel A1	List pertanyaan dan jawaban oleh chatbot pajak	71
Tabel A2	Transkrip wawancara dengan konsultan pajak	87



DAFTAR GAMBAR

Gambar 2.1	Mekanisme <i>transformer</i> [2]	6
Gambar 2.2	Plot dua dimensi pada <i>text embedding</i>	11
Gambar 2.3	<i>Vector</i> plot hasil <i>cosine simialrity</i>	14
Gambar 2.4	Ilustrasi <i>Hierarchical Navigable Small World graph</i>	15
Gambar 2.5	Flow RAG	16
Gambar 2.6	Pipeline LLM dan RAG [3]	17
Gambar 3.1	Flowchart Alur Penelitian	21
Gambar 3.2	Flowchart rancangan <i>vector database</i>	23
Gambar 3.3	Flowchart rancangan RAG	24
Gambar 3.4	Flowchart rancangan chatbot	25
Gambar 3.5	Flowchart rancangan aplikasi	26
Gambar 4.1	Contoh data yang disimpan pada <i>vector database</i> Qdrant	32
Gambar 4.2	Hasil <i>embedding</i> teks halaman 190 - 195	32
Gambar 4.3	Tampilan <i>tracing agent</i>	38
Gambar 4.4	Tampilan aplikasi sebelum <i>chat</i>	40
Gambar 4.5	Tampilan aplikasi setelah <i>chat</i>	41
Gambar 4.6	Hasil perhitungan <i>cosine</i> pada <i>input</i> dan <i>content</i>	54
Gambar 4.7	Visualisasi <i>Vector Search</i> dengan PCA 2D	55



DAFTAR KODE

Kode 4.1	Source code inisiasi Qdrant dan <i>embedding</i>	29
Kode 4.2	Source code <i>create collection</i> dan <i>list</i> dokumen	30
Kode 4.3	Source code memuat dokumen dan penyimpanan di <i>vector</i> <i>database</i>	30
Kode 4.4	Source code membuat <i>prompt</i> untuk <i>agent</i>	33
Kode 4.5	Source code membuat LLM	34
Kode 4.6	Source code membuat <i>tools</i> untuk <i>agent</i>	35
Kode 4.7	Source code membuat <i>schema</i> untuk <i>tool</i>	36
Kode 4.8	Source code membuat <i>agent</i>	36
Kode 4.9	Source code implementasi <i>tracing</i> menggunakan langchain	37
Kode 4.10	Source code membuat aplikasi	38
Kode 4.11	Source code membuat fungsi dalam <i>agent</i>	39
Kode 4.12	Source code membuat evaluasi dataset	44
Kode 4.13	Source code mengevaluasi dataset menggunakan GEval	45
Kode 4.14	Source code melihat <i>score</i> dari hasil pencarian pada <i>vector</i> <i>database</i>	48
Kode 4.15	Code menghitung nilai cosine similarity	54



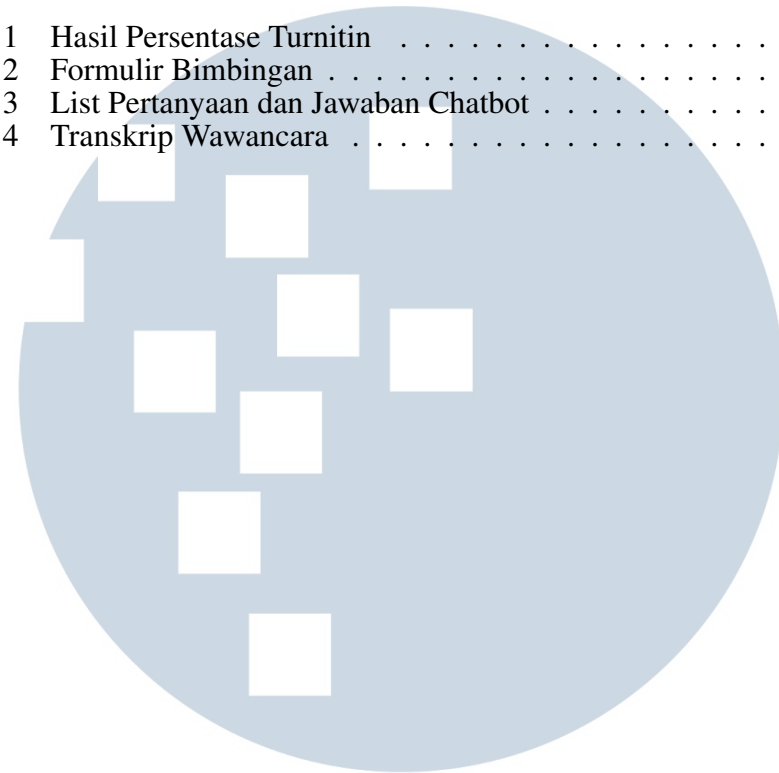
DAFTAR RUMUS

Rumus 2.1	<i>Attention score</i>	7
Rumus 2.8	Inisiasi variabel pada perhitungan <i>attention</i>	8
Rumus 2.19	Contoh perhitungan <i>attention</i>	9
Rumus 2.20	<i>Cosine similarity</i>	12
Rumus 2.27	Contoh perhitungan <i>cosine similarity</i>	13



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	62
Lampiran 2	Formulir Bimbingan	69
Lampiran 3	List Pertanyaan dan Jawaban Chatbot	71
Lampiran 4	Transkrip Wawancara	87



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA