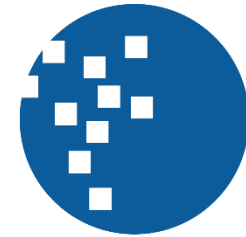


**OPTIMIZING RETRIEVAL-AUGMENTED GENERATION
THROUGH AGENTIC RAG ECOSYSTEM BASED ON
FINE-TUNED BERT CROSS ENCODER AND GPT-4 MODEL**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Faustine Ilone Hadinata

00000054212

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

**OPTIMIZING RETRIEVAL-AUGMENTED GENERATION
THROUGH AGENTIC RAG ECOSYSTEM BASED ON
FINE-TUNED BERT CROSS ENCODER AND GPT-4 MODEL**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom.)

Faustine Ilone Hadinata

00000054212

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

i

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Faustine Ilone Hadinata

Nomor Induk Mahasiswa : 00000054212

Program Studi : Sistem Informasi

Skripsi dengan judul:

“Optimizing Retrieval-Augmented Generation through Agentic RAG Ecosystem
Based on Fine-Tuned BERT Cross Encoder and GPT-4 Model”

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, 27 Mei 2025



(Faustine Ilone Hadinata)

HALAMAN PENGESAHAN

Skripsi dengan judul

*Optimizing Retrieval-Augmented Generation through Agentic RAG Ecosystem
Based on Fine-Tuned BERT Cross Encoder and GPT-4 Model*

Oleh

Nama : Faustine Ilone Hadinata
NIM : 00000054212
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 11 Juni 2025
Pukul 16.00 s.d 17.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut.

Penguji

Pembimbing

Ir. Raymond Wardi Octama, M.CIS
0328046803

Samuel (Ady) Sanjaya, S.T., M.T.
0305049402

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Faustine Ilone Hadinata

NIM : 00000054212

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Optimizing Retrieval-Augmented
Generation through Agentic RAG Ecosystem Based on Fine-Tuned BERT
Cross Encoder and GPT-4 Model

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 3 tahun.

Tangerang, 27 Mei 2025



(Faustine Ilone Hadinata)

* Pilih salah satu

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

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa atas selesainya penulisan laporan skripsi ini dengan judul: *“Optimizing Retrieval-Augmented Generation through Agentic RAG Ecosystem Based on Fine-Tuned BERT Cross Encoder and GPT-4 Model”*, untuk memenuhi salah satu syarat dalam mencapai gelar Sarjana Komputer (S.Kom) Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika di Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, ST, M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Samuel Ady Sanjaya, S.T., M.T., sebagai Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga dan teman-teman saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat bermanfaat bagi para pembaca dan menjadi sumber informasi serta inspirasi.

Tangerang, 27 Mei 2025



(Faustine Ilone Hadinata)

OPTIMALISASI RETRIEVAL-AUGMENTED GENERATION MELALUI EKOSISTEM RAG AGENTIK BERDASARKAN FINE-TUNED BERT CROSS ENCODER DAN MODEL GPT-4

Faustine Ilone Hadinata

ABSTRAK

Pendidikan memainkan peran mendasar dalam pertumbuhan pribadi dan profesional, namun banyak siswa berjuang dengan memilih jurusan yang tepat karena bimbingan yang tidak memadai, yang mengarah pada ketidakpuasan dalam jalur akademis dan karier mereka.

Untuk mengatasi hal ini, kami mengusulkan sistem Agentic Retrieval-Augmented Generation (RAG) yang meningkatkan saran akademis berbasis chatbot dengan mengintegrasikan agen berbasis BERT untuk menyaring dan memvalidasi informasi yang diambil, memastikan respons yang relevan secara kontekstual dan akurat secara faktual. Selain itu, GPT-4 digunakan sebagai komponen Natural Language Generation (NLG) untuk menghasilkan jawaban yang lancar dan terstruktur.

Hasil eksperimen menunjukkan bahwa menggabungkan agen secara signifikan meningkatkan akurasi dan relevansi respons, di mana dari 11 jurusan Skor METEOR menghasilkan 74,33%, kesamaan Jaccard pada 58,77%, dan kesamaan Cosinus pada 94,13%, meningkat masing-masing sebesar 6,54%, 5,57%, dan 3,57%. Skor Relevansi BERT tetap tinggi secara konsisten pada 96,91%. Penerapan menggunakan Django juga diterapkan untuk memungkinkan skenario penggunaan nyata. Meskipun menjanjikan, disarankan agar penelitian berikutnya melibatkan kumpulan data yang lebih besar yang terdiri dari puluhan ribu baris jika memungkinkan untuk mengurangi bias dan memungkinkan agen yang lebih baik.

Kata kunci: Agen, AI, BERT, Chatbot, GPT-4, Retrieval-Augmented Generation

OPTIMIZING RETRIEVAL-AUGMENTED GENERATION THROUGH AGENTIC RAG ECOSYSTEM BASED ON FINE-TUNED BERT CROSS ENCODER AND GPT-4 MODEL

Faustine Ilone Hadinata

ABSTRACT (English)

Education plays a fundamental role in personal and professional growth, yet many students struggle with selecting the right major due to insufficient guidance, leading to dissatisfaction in their academic and career paths.

To address this, we propose an Agentic Retrieval-Augmented Generation (RAG) system that enhances chatbot-based academic advising by integrating a BERT-based agent to filter and validate retrieved information, ensuring contextually relevant and factually accurate responses. Additionally, GPT-4 is employed as the Natural Language Generation (NLG) component to produce fluent, structured answers.

Experimental results show that incorporating the agent significantly enhances response accuracy and relevance, where from 11 majors the METEOR Score resulted at 74.33%, Jaccard similarity at 58.77%, and Cosine similarity at 94.13%, improving by 6.54%, 5.57%, and 3.57%, respectively. The BERT Relevancy score remains consistently high at 96.91%. Deployment using Django is also implemented to allow real-use scenarios. Although promising, it is suggested that the next research involved a larger dataset consisting of tens of thousands of rows if possible to reduce bias and enable a more fine-tuned agent.

Keywords: *Agent AI, BERT, Chatbot, GPT-4, Retrieval Augmented Generation*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xi
INTRODUCTION.....	1
PRELIMINARY WORKS	3
METHODOLOGY.....	4
3.1 Context/Dataset Retrieval.....	4
3.2 Bidirectional Encoder Representations from Transformers (BERT)	4
3.3 Pipeline	6
3.4 User Interface.....	6
EXPERIMENTAL RESULTS.....	6
4.1 Quantitative (METEOR, BERT Relevancy, Cosine Similarity, and Jaccard Similarity)	6
4.2 Qualitative (Human Evaluation).....	10
4.3 Deployment	10
CONCLUSION.....	10
ACKNOWLEDGEMENTS.....	11
REFERENCES.....	11
LAMPIRAN.....	15

DAFTAR TABEL

Table 1. Response Comparison.....	8
Table 2. Evaluation Result Based on Major.....	9



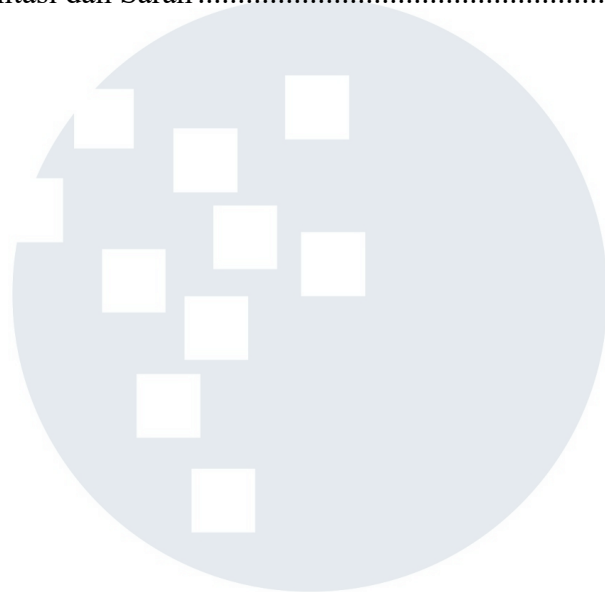
DAFTAR GAMBAR

Fig. 1. Agentic RAG Ecosystem Pipeline.....	4
Fig. 2. Bi-Encoder vs Cross-Encoder Illustration	5



DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	15
Lampiran B Form Konsultasi Bimbingan	18
Lampiran C Letter of Acceptance (LoA)	20
Lampiran D Limitasi dan Saran	21



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA