

**PERANCANGAN CHATBOT DESA PETE TIGARAKSA
MENGUNAKAN *MEDIUM LANGUAGE MODEL* DAN
ARSITEKTUR *RETRIEVAL-AUGMENTATED GENERATION***



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

Sheera Anela

00000075618

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

2025

**PERANCANGAN CHATBOT DESA PETE TIGARAKSA
MENGUNAKAN *MEDIUM LANGUAGE MODEL* DAN
ARSITEKTUR *RETRIEVAL-AUGMENTATED GENERATION***



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Sistem Informasi

Sheera Anela

00000075618

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

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Sheera Anela
Nomor Induk Mahasiswa : 00000075618
Program Studi : Sistem Informasi

menyatakan dengan sesungguhnya bahwa Laporan saya yang berjudul:

PERANCANGAN CHATBOT DESA PETE TIGARAKSA MENGGUNAKAN
MEDIUM LANGUAGE MODEL DAN ARSITEKTUR *RETRIEVAL-
AUGMENTED GENERATION*

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain. Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian daftar pustaka.

Jika di kemudian hari terbukti ditemukan penyimpangan dan penyalahgunaan dalam pelaksanaan kegiatan maupun dalam penulisan laporan ini, saya bersedia menerima konsekuensi untuk dinyatakan **TIDAK LULUS**. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 19 Desember 2025



(Sheera Anela)

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Sheera Anela
NIM : 00000075618
Program Studi : Sistem Informasi
Judul Laporan : Perancangan Chatbot Desa Pete Tigaraksa
Menggunakan *Medium Language Model* dan
Arsitektur *Retrieval-Augmentated Generation*

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan Tugas/Laporan/Project/Tugas Akhir*(pilih salah satu) sebagai berikut (beri tanda centang yang sesuai):

- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
- (3) Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 19 Desember 2025



(Sheera Anela)

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Yang bertanda tangan di bawah ini:

Nama : Sheera Anela
NIM : 00000075618
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : Universitas Multimedia Nusantara
Alamat : Jl. Boulevard, Gading Serpong, Kel. Curug
Sangereng, Kec. Kelapa Dua, Kab. Tangerang, Prop.
Banten, Indonesia, 15810

Email Perusahaan/Organisasi : umn.ac.id

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat di validasi keberadaannya.
2. Jika dikemudian hari, terbukti ditemukan kecurangan/penyimpangan data yang tidak valid di perusahaan/organisasi tempat saya melakukan kegiatan, maka:
 - a. Saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.
 - b. Saya bersedia menerima semua sanksi yang berlaku sebagaimana ditetapkan dalam peraturan yang berlaku di Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

Tangerang, 19 Desember 2025

Mengetahui

Menyatakan



(Ahmad Faza)



(Sheera Anela)

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama : Sheera Anela
NIM : 00000075618
Program Studi : Sistem Informasi
Jenjang : S1
Judul Karya Ilmiah : Perancangan Chatbot Desa Pete Tigaraksa
Menggunakan Medium Language Model dan
Arsitektur Retrieval-Augmentated Generatio

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

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk memublikasikan hasil laporan saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh civitas academica UMN/ publik. Saya menyatakan bahwa laporan yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia memublikasikan hasil laporan ini ke dalam repositori Knowledge Center karena ada data yang bersifat konfidensial.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara.
 - ☐ Embargo publikasi laporan dalam kurun waktu 3 tahun.

Tangerang, 19 Desember 2025



(Sheera Anela)

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan Prostep ini dengan judul: “PERANCANGAN CHATBOT DESA PETE TIGARAKSA MENGGUNAKAN *MEDIUM LANGUAGE MODEL* DAN ARSITEKTUR *RETRIEVAL-AUGMENTED GENERATION*” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer (S.Kom) Program Strata 1 Jurusan Sistem Informasi 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 Prostep ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Prostep ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

Mengucapkan terima kasih kepada:

1. Ir. Andrey Andoko, M.Sc. Ph.D., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Johan Setiawan, S.Kom., MBA., selaku Pembimbing Laporan Prostep yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya laporan Prostep ini.
5. Ahmad Faza, S.Kom., M.T.I., selaku Pembimbing Lapangan yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya proyek Prostep.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan Laporan Prostep.
7. Kepada Desa Pete, sebagai desa wisata yang memberikan saya kesempatan untuk melakukan penelitian ini.
8. Kelompok Prostep Penelitian sebagai tim yang sudah mendukung penyusunan laporan serta pengerjaan proyek ini.

Semoga laporan ini dapat bermanfaat dan dikembangkan lebih lanjut untuk penelitian selanjutnya dan biarlah segala usaha yang dikerahkan disini dapat meningkatkan paparan teknologi terhadap desa lain di Kecamatan Tigaraksa.

Tangerang, 19 Desember 2025



(Sheera Anela)



PERANCANGAN CHATBOT DESA PETE TIGARAKSA MENGUNAKAN MEDIUM LANGUAGE MODEL DAN ARSITEKTUR RETRIEVAL-AUGMENTED GENERATION

(Sheera Anela)

ABSTRAK

Sektor pariwisata merupakan salah satu penggerak utama perekonomian Indonesia, namun pemanfaatan teknologi digital dalam penyediaan layanan informasi pariwisata, khususnya di tingkat desa, masih tergolong terbatas. Kondisi ini menyebabkan akses informasi bagi wisatawan belum optimal dan berdampak pada rendahnya paparan destinasi wisata desa. Penelitian ini bertujuan untuk merancang, mengimplementasikan, dan mengevaluasi sistem informasi pariwisata berbasis *chatbot* yang terintegrasi pada situs web Desa Wisata Pete, Kecamatan Tigaraksa, sebagai bagian dari upaya digitalisasi layanan informasi pariwisata desa.

Penelitian ini menerapkan metodologi CRISP-DM di mana sistem *chatbot* dikembangkan dengan mengombinasikan model bahasa berukuran kecil dan menengah dengan arsitektur *Retrieval-Augmented Generation (RAG)*, sehingga jawaban yang dihasilkan tidak hanya bersifat generatif, tetapi juga berbasis *knowledge base* yang telah tervalidasi. Data diperoleh melalui observasi lapangan, wawancara dengan pihak desa, serta dokumen resmi dan informasi publik, kemudian diproses untuk mendukung mekanisme *retrieval* dan *generation* secara terkontrol.

Hasil evaluasi menunjukkan bahwa sistem *chatbot* berhasil dirancang dengan tingkat keberhasilan *retrieval* sebesar 93,75% dan tingkat *hallucination* sebesar 0% di mana jawaban mampu ditangani dengan *fallback message*. Evaluasi semantik menggunakan *BERTScore* menunjukkan bahwa model bahasa berukuran menengah (0,66) menghasilkan respons yang lebih baik dibandingkan model berukuran kecil (0,65). *Chatbot* juga sudah diintegrasikan ke situs web dan akan dilanjutkan dengan *deployment* ke situs web Kecamatan Tigaraksa ke domain yang tersedia. Hasil penelitian membuktikan bahwa integrasi arsitektur *RAG* dengan model bahasa GPT-2 Medium Indonesian mampu dibangun untuk menangani keterbatasan komputasi dan diterapkan ke dalam situs web untuk pengembangan yang lebih lanjut.

Kata kunci: *chatbot*, digitalisasi pariwisata, desa wisata, gpt-2 medium, *retrieval-augmented generation*

DESIGNING THE PETE VILLAGE CHATBOT IN TIGARAKSA SUBDISTRICT USING A MEDIUM LANGUAGE MODEL AND RETRIEVAL-AUGMENTED GENERATION ARCHITECTURE

(Sheera Anela)

ABSTRACT

The tourism sector is one of the main drivers of Indonesia's economy; however, the utilization of digital technology in providing tourism information services, particularly at the village level, remains limited. This condition results in suboptimal access to information for tourists and contributes to the low exposure of rural tourism destinations. This study aims to design, implement, and evaluate a tourism information system based on a chatbot integrated into the official website of Desa Wisata Pete, Tigaraksa District, as part of efforts to digitalize village-level tourism information services.

This study adopts the CRISP-DM methodology, in which the chatbot system is developed by combining small- and medium-sized language models with the Retrieval-Augmented Generation (RAG) architecture. This approach enables the system to generate responses that are not only generative in nature but also grounded in a validated knowledge base. Data were collected through field observations, interviews with village stakeholders, as well as official documents and publicly available information, which were then processed to support controlled retrieval and generation mechanisms.

The evaluation results demonstrate that the chatbot system achieves strong performance, with a retrieval success rate of 93.75% and a hallucination rate of 0%. Semantic evaluation using BERTScore indicates that the medium-sized language model produces better responses compared to the small-sized model. These findings confirm that integrating the RAG architecture with efficient language models represents an effective and practical solution for developing rural tourism information services under limited computational and infrastructural resources.

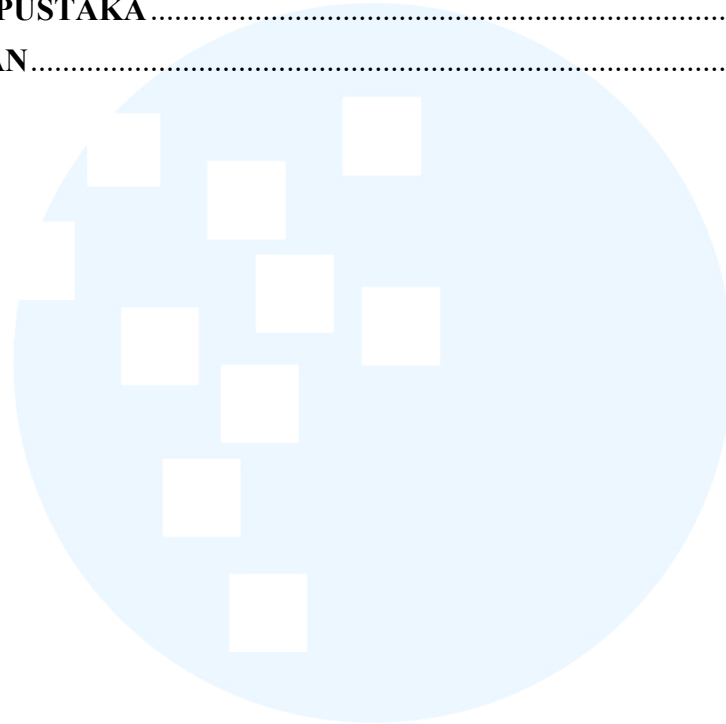
Keywords: chatbot, digital tourism transformation, GPT-2 Medium, retrieval-augmented generation, rural tourism

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Urgensi Penelitian	3
1.5 Luaran Penelitian	4
1.6 Manfaat Penelitian	4
1.6.1 Manfaat Teoritis	4
1.6.2 Manfaat Praktis	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Keperluan <i>Chatbot</i> untuk Pariwisata	6
2.2 Model dan Arsitektur untuk <i>Chatbot</i>	9
2.3 Teori Terkait Penelitian.....	13
2.3.1 <i>Chatbot</i>	13
2.3.2 <i>Retrieval Augmented Generation</i>	13
2.3.3 <i>Language Model</i>	14
2.3.4 Desa Pete.....	15
BAB III METODE PENELITIAN	17
3.1 Metode Penelitian.....	17
3.2 Tahapan Penelitian	19

3.2.1	<i>Business Understanding</i>	23
3.2.2	<i>Data Understanding</i>	23
3.2.3	<i>Data Preparation</i>	24
3.2.4	<i>Modeling</i>	25
3.2.5	<i>Evaluation</i>	26
3.2.6	<i>Deployment</i>	26
3.3	Teknik Pengumpulan Data.....	26
3.4	Teknik Analisis Data.....	27
BAB IV HASIL DAN PEMBAHASAN		28
4.1	<i>Business Understanding</i>	28
4.2	<i>Data Understanding</i>	29
4.3	<i>Data Preparation</i>	30
4.3.1	Data dan Model.....	30
4.3.2	Arsitektur <i>Chatbot</i>	32
4.4	<i>Modeling</i>	34
4.4.1	<i>Chatbot</i>	35
4.4.2	Mengumpulkan Data Pengunjung.....	45
4.4.3	<i>Time Out</i> jika Tidak Dijawab.....	46
4.4.4	Memasukkan data ke <i>Unanswered FAQ</i>	46
4.5	<i>Evaluation</i>	47
4.6	<i>Deployment</i>	49
4.6.1	Fitur <i>Chatbot</i>	55
4.6.2	Fitur Mengumpulkan Data Pengunjung	58
4.6.3	Fitur <i>Time Out</i> Saat Tidak Digunakan	61
4.6.4	Fitur <i>Unanswered FAQ</i>	62
4.7	Pembahasan dan Perbandingan dengan Penelitian Terdahulu ...	64
BAB V SIMPULAN SARAN		67
5.1	Simpulan	67
5.1.1	Tujuan Pertama	67
5.1.2	Tujuan Kedua	67
5.1.3	Tujuan Ketiga.....	68
5.2	Saran.....	68

5.2.1	Stabilisasi dan Pemeliharaan.....	68
5.2.2	Optimalisasi secara Teknis.....	69
5.2.3	Pengembangan Fitur.....	69
DAFTAR PUSTAKA		70
LAMPIRAN.....		74



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu Pentingnya Chatbot.....	6
Tabel 2.2 Penelitian Terdahulu Arsitektur dan Model Chatbot	9
Tabel 3.1 Tabel Perbandingan Metode	17
Tabel 3.2 Lini Masa Pelaksanaan Penelitian.....	19
Tabel 4.1 Tabel Metrik Evaluasi.....	48



DAFTAR GAMBAR

Gambar 2.1 Arsitektur RAG	14
Gambar 3.1 Metode CRISP-DM [32]	17
Gambar 3.2 Flowchart Alur Penelitian sesuai CRISP-DM.....	19
Gambar 3.3 Diskusi dengan Pihak Desa Pete	21
Gambar 3.4 Survei Keliling Lingkungan Desa Pete	21
Gambar 3.5 Tempat Diskusi Penelitian	22
Gambar 4.1 Hasil Transformasi Data Survei menjadi File JSON	31
Gambar 4.2 Alur Arsitektur RAG [39]	32
Gambar 4.3 Kode untuk Fine Tuning	35
Gambar 4.4 Contoh Model yang Dilatih (GPT-2 Medium).....	36
Gambar 4.5 Kode untuk Tokenisasi Data	36
Gambar 4.6 Kode untuk Melatih Model dan Menyimpan Model.....	37
Gambar 4.7 Kode Penarikan Data ke Tabel knowledge_base	38
Gambar 4.8 Tabel knowledge_base	38
Gambar 4.9 Kode Retrieval.....	39
Gambar 4.10 Kode Load Model.....	41
Gambar 4.11 Kode untuk Melakukan Generation	42
Gambar 4.12 Kode untuk Implementasi menggunakan API	44
Gambar 4.13 Data Testing untuk Chatbot.....	48
Gambar 4.14 Flowchart untuk Mengakses Chatbot Desa Pete	51
Gambar 4.15 Halaman Website Kecamatan Tigaraksa Profil Desa Pete.....	52
Gambar 4.16 Button Chatbot Diklik	53
Gambar 4.17 Flowchart Chatbot beserta Fiturnya	54
Gambar 4.18 Halaman Chatbot untuk Percakapan	56
Gambar 4.19 Database pada Tabel knowledge_base	57
Gambar 4.20 Halaman Chatbot saat Mengumpulkan Data User/Pengunjung.....	59
Gambar 4.21 Tabel informasi_pengunjung.....	60
Gambar 4.22 Time Out saat Tidak Digunakan	61
Gambar 4.23 Pertanyaan yang Tidak Dapat Dijawab.....	62
Gambar 4.24 Pertanyaan Dimasukkan ke Tabel unanswered_faq.....	63

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Surat Pengantar / Cover Letter	74
---	----



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA