

**IMPLEMENTASI VISION LANGUAGE MODEL,
EMBEDDING, DAN SMALL LANGUAGE MODEL UNTUK
PEMAHAMAN GAMBAR EDUKATIF DI DESA WISATA
TIGARAKSA**



LAPORAN PRO-STEP PENELITIAN

Nicholas Prawira

00000068231

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

2025

**IMPLEMENTASI VISION LANGUAGE MODEL,
EMBEDDING, DAN SMALL LANGUAGE MODEL UNTUK
PEMAHAMAN GAMBAR EDUKATIF DI DESA WISATA
TIGARAKSA**



LAPORAN PRO-STEP PENELITIAN

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Sistem Informasi

Nicholas Prawira

00000068231

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

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Nicholas Prawira

Nomor Induk Mahasiswa : **00000068231**

Program studi : Sistem Informasi

Laporan PRO-STEP Penelitian dengan judul:

**IMPLEMENTASI VISION LANGUAGE MODEL,
EMBEDDING, DAN SMALL LANGUAGE MODEL UNTUK
PEMAHAMAN GAMBAR EDUKATIF DI DESA WISATA
TIGARAKSA**

merupakan hasil karya saya sendiri bukan plagiat dari karya 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 PRO-STEP, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk laporan PRO-STEP yang telah saya tempuh.

Tangerang, 17 Desember 2025



Nicholas Prawira

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Nicholas Prawira

Nomor Induk Mahasiswa : 00000068231

Program Studi : Sistem Informasi

Jenjang : S1

Jenis Karya : Skripsi

Judul Karya Ilmiah : **IMPLEMENTASI VISION
LANGUAGE MODEL, EMBEDDING, DAN SMALL
LANGUAGE MODEL UNTUK PEMAHAMAN GAMBAR
EDUKATIF DI DESA WISATA TIGARAKSA**

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, 17 Desember 2025

Yang menandatangani.



Nicholas Prawira

* Pilih salah satu opsi tanpa harus menghapus opsi yang tidak dipilih

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

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan Pro-Step ini dengan judul: **Implementasi *Vision AI, Embedding, dan Small Language Model* untuk Pemahaman Gambar Edukatif di Desa Wisata Tigaraksa** dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar S.Kom Jurusan Sistem Informasi Pada Teknik & Informatika 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. Bapak Dr. Ir. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.S, selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom, selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Ahmad Faza, S.Kom., M.T.I., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Tuhan Yesus Kristus, yang senantiasa memberikan penyertaan, kekuatan, dan hikmat dalam setiap proses yang saya jalani
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Saya berharap karya ilmiah ini dapat memberi manfaat bagi berbagai pihak—baik sebagai sumber informasi.

Tangerang, 17 Desember 2025



Nicholas Prawira

IMPLEMENTASI VISION LANGUAGE MODEL, EMBEDDING, DAN SMALL LANGUAGE MODEL UNTUK PEMAHAMAN GAMBAR EDUKATIF DI DESA WISATA TIGARAKSA

Nicholas Prawira

ABSTRAK

Penelitian ini berfokus pada pengembangan sistem chatbot multimodal berbasis *Vision AI, Embedding, dan Small Language Model (SLM)* untuk mendukung pemahaman gambar edukatif di Desa Wisata Tigaraksa. Pemilihan topik ini didasarkan pada kebutuhan guru dan siswa terhadap media pembelajaran visual yang tidak hanya menarik tetapi juga mampu dijelaskan melalui teknologi AI. Tantangan utama selama proses penelitian meliputi akurasi metadata visual yang belum stabil, keterbatasan model vision dalam memahami gambar T2I secara mendalam dan risiko halusinasi ketika model bahasa menghasilkan penjelasan tanpa konteks yang kuat.

Untuk mengatasi kendala tersebut, diterapkan pipeline Vision → Embedding → Retrieval-Augmented Generation (RAG) → SLM. BLIP digunakan untuk menghasilkan caption dan deskripsi visual, OCR untuk mengekstraksi teks dalam gambar, serta Sentence Transformer untuk embedding yang disimpan dalam Supabase Vector. Seluruh metadata ini menjadi dasar retrieval sebelum diproses oleh SLM GPT-OSS 20B melalui Groq API sehingga respons yang dihasilkan tetap faktual, aman dan sesuai konteks pembelajaran.

Melalui pendekatan ini, penelitian berhasil menghasilkan prototipe chatbot edukatif yang mampu membantu guru dan siswa memahami gambar dengan lebih baik, sekaligus memberikan kontribusi dalam pengembangan media pembelajaran berbasis AI di lingkungan desa.

Kata kunci: *Vision AI, Embedding, RAG, Small Language Model, Chatbot Multimodal.*

IMPLEMENTATION OF VISION LANGUAGE MODEL, EMBEDDING, AND SMALL LANGUAGE MODEL FOR UNDERSTANDING EDUCATIONAL IMAGES IN TIGARAKSA TOURISM VILLAGE

Nicholas Prawira

ABSTRACT (English)

This project focuses on developing a multimodal chatbot powered by Vision AI, Embedding, and a Small Language Model (SLM) to support the understanding of educational images in Desa Wisata Tigaraksa. This topic was chosen to address the need among teachers and students for visual learning materials that are not only engaging but can also be automatically interpreted and explained through AI. Several challenges emerged during implementation, including inconsistent visual metadata accuracy, the limited ability of vision models to fully understand T2I-generated images, and the potential for hallucination when language models generate responses without adequate contextual grounding.

To overcome these issues, the study implemented a complete Vision → Embedding → Retrieval-Augmented Generation (RAG) → SLM pipeline. BLIP was used to generate captions and visual descriptions, OCR extracted textual elements within the images, and Sentence Transformer produced embeddings stored in Supabase Vector. These metadata components served as retrieval evidence before being processed by the GPT-OSS 20B SLM through the Groq API, ensuring that the final responses remained factual, safe, and aligned with educational context.

This approach resulted in an educational chatbot prototype capable of assisting teachers and students in understanding images more effectively, contributing to the advancement of AI-based learning tools within the village learning environment.

Keywords: *Vision AI, Embedding, RAG, Small Language Model, Multimodal Chatbot*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	2
HALAMAN PENGESAHAN	3
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	4
KATA PENGANTAR	5
ABSTRAK	6
<i>ABSTRACT (English)</i>	7
DAFTAR ISI	8
DAFTAR TABEL	10
DAFTAR GAMBAR	11
DAFTAR LAMPIRAN	12
BAB I	
PENDAHULUAN	13
1.1. Latar Belakang	13
1.2. Rumusan Masalah	15
1.3. Tujuan Penelitian	16
1.4. Urgensi Penelitian	16
1.5. Luaran Penelitian	17
1.6. Manfaat Penelitian	17
BAB II	
TINJAUAN PUSTAKA	18
2.1 Penelitian Terdahulu	18
2.2 Tinjauan Teori	24
2.2.1 Vision Language Model (VLM)	24
2.2.2 Multimodal Learning	24
2.2.3 Retrieval Augmented Generation (RAG)	25
2.2.4 Optical Character Recognition (OCR)	27
2.2.5 Small Language Model	28
2.3 Algoritma dan Framework	29
2.3.1 Human Centric Design (HCD)	29
2.3.2 Vision–Language Processing Framework (BLIP + OCR)	29
2.3.3 Multimodal Reasoning Framework (Embedding → RAG → SLM)	30
2.3.4 Integrated Multimodal Intelligence Pipeline	31

2.4 Software yang digunakan	32
2.4.1 Python	32
2.4.2 Visual Studio Code	32
2.4.3 Google Collab	33
BAB III	
METODE PENELITIAN	34
3.1 Metode Penelitian	34
3.2 Tahapan Penelitian	36
3.2.1 Discover	36
3.2.2 Define	36
3.2.3 Design	37
3.2.4 Prototype & Test	37
3.2.5 Plan & Implement	38
3.3 Teknik Pengumpulan Data	38
3.3.1 Sumber Dataset	39
3.3.2 Karakteristik Dataset Visual	40
3.4 Teknik Analisis Data	40
3.4.1 Analisis Vision Processing (BLIP, OCR, dan Embedding)	41
3.4.2 Analisis Retrieval Berbasis RAG (Vector Similarity Search)	42
3.4.3 Analisis Jawaban Chatbot (SLM Reasoning Evaluation)	42
BAB IV	
HASIL DAN PEMBAHASAN	44
4.1 Discover	44
4.2 Define	45
4.3 Design	46
4.3.1 Vision Processing Layer	47
4.3.2 Embedding & Vector Store Layer	48
4.3.2.1 Perbandingan Model Embedding	48
4.3.2.2 Perbandingan Vector Store	49
4.3.3 Retrieval Augmented Generation (RAG) Layer	50
4.3.4 SLM Reasoning Layer	52
4.3.5 Metode Arsitektur Sistem	55
4.4 Prototype & Test	57
4.4.1 Implementasi Vision Processing	57
4.4.2 Implementasi Embedding	59
4.4.3 Implementasi Vector Retrieval	60
4.4.4 Implementasi RAG	61
4.4.5 Integrasi SLM	62

4.4.6 Evaluasi	64
4.4.6.1 Evaluasi Vision Processing	64
4.4.6.2 Evaluasi Embedding & Vector	68
4.4.6.3 Evaluasi Retrieval-Augmented Generation (RAG)	69
4.4.6.4 Kesimpulan Evaluasi	70
4.5 Plan & Implement	72
4.5.1 Implementasi Backend	72
4.5.2 Implementasi Frontend	75
4.5.3 Deployment Sistem	79
4.5.4 Evaluasi Prototipe	79
4.6 Diskusi	81
BAB V	
SIMPULAN SARAN	83
5.1 Simpulan	83
5.2 Saran	84
DAFTAR PUSTAKA	86
LAMPIRAN	91



DAFTAR TABEL

Tabel 3.1 Perbandingan Metode Penelitian	35
Tabel 4.1 Tabel Perbandingan Model Vision-Language untuk Captioning	48
Tabel 4.2 Tabel Perbandingan Model Embedding	50
Tabel 4.3 Tabel Perbandingan Vector Store	50
Tabel 4.4 Tabel Perbandingan Teknik Retrieval	52
Tabel 4.5 Tabel Perbandingan Model SLM	54
Tabel 4.6 Tabel Perbandingan Groq API dan Self-Hosted	55
Tabel 4.7 Tabel Evaluasi Prototipe	57
Tabel 4.8 Perbandingan Hasil Penelitian dengan Penelitian Terdahulu	80



DAFTAR GAMBAR

Gambar 2.1 Arsitektur Vision Language Model	25
Gambar 2.1 Ilustrasi Retrieval Augmented Generation	26
Gambar 2.3 Optical Character Recognition (OCR)	28
Gambar 2.4 Manfaat Small Language Models	29
Gambar 2.5 Proses Human Centered Design	30
Gambar 4.1 Arsitektur Sistem Diagram	58
Gambar 4.2 Implementasi Model BLIP-2	60
Gambar 4.3 Implementasi OCR	61
Gambar 4.4 Implementasi Embedding	62
Gambar 4.5 Implementasi Vector Retrieval	62
Gambar 4.6 Integrasi Retrieval-Augmented Generation (RAG)	65
Gambar 4.7 Integrasi Small Language Model (SLM)	65
Gambar 4.8 Mekanisme Pemilihan Role dan Penyusunan Context Prompt pada SLM	66
Gambar 4.9 Contoh Instruksi Prompt	67
Gambar 4.10 Evaluasi Akurasi Caption BLIP-Base	69
Gambar 4.11 Evaluasi Performa OCR	69
Gambar 4.12 Grafik Metadata Completeness pada Vision Processing	70
Gambar 4.13 Distribusi Cosine Similarity Embedding dan Precision@1/3	71
Gambar 4.14 Evaluasi Retrieval Augmented Generation	73
Gambar 4.15 Penyusunan Focused Context Mode	76
Gambar 4.16 Proses Encoding Query User menjadi Embedding	77
Gambar 4.17 Instruction Block untuk Mengontrol Reasoning dan Mengurangi Halusinasi	77
Gambar 4.18 Mekanisme Pemilihan Role (Profesor, Kakak Pintar, Teman Baik, Penjelajah)	78
Gambar 4.19 User Interface	79
Gambar 4.20 UI Pemilihan Role dengan Ikon Visual pada Frontend	80
Gambar 4.21 Fitur Drag-and-Drop Gambar untuk Interaksi Multimodal	81
Gambar 4.23 Desain Panel Chatbot Floating pada Sistem Frontend	82
Gambar 4.24 Deployment Diagram Workflow	83
Gambar 4.25 GitHub sebagai Version Control	84
Gambar 4.26 Deployment Frontend Next.js melalui Vercel	85
Gambar 4.27 Deployment Backend FastAPI melalui Railway	85

DAFTAR LAMPIRAN

Lampiran A Surat Pengantar PRO-STEP 01	97
Lampiran B Kartu PRO-STEP - PRO-STEP 02	98
Lampiran C Daily Task PRO-STEP - PRO-STEP 03	99
Lampiran D Lembar Verifikasi Laporan PRO-STEP - PRO-STEP 04	109
Lampiran E Pengecekan Hasil Turnitin	110
Lampiran F Surat Penerimaan PRO-STEP (LoA)	110
Lampiran G Semua hasil karya tugas yang dilakukan selama PRO-STEP	110
Lampiran H Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)	122

