

TESIS

**ANALISIS PENGARUH PANJANG CHUNK DAN OVERLAP
TERHADAP AKURASI JAWABAN SMALL LANGUAGE
MODEL (SLM) PADA DOKUMEN PUBLIK BERBAHASA
INDONESIA**

Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-2 Pada
Program Studi Ilmu Komputer

Oleh:

**RIJALUL FIKRI
NPM. 2024171004**



**UNIVERSITAS INDO GLOBAL MANDIRI
FAKULTAS ILMU KOMPUTER DAN SAINS
PROGRAM STUDI ILMU KOMPUTER
PROGRAM MAGISTER
PALEMBANG
2026**



FAKULTAS ILMU KOMPUTER DAN SAINS
ILMU KOMPUTER PROGRAM MAGISTER

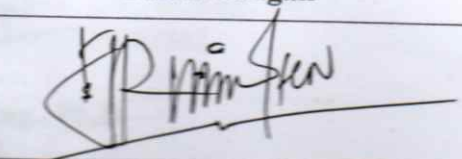
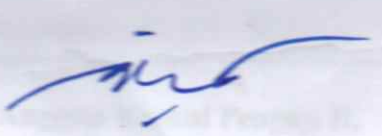
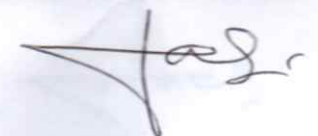
PENGESAHAN TESIS

JUDUL TESIS

**ANALISIS PENGARUH PANJANG CHUNK DAN OVERLAP
TERHADAP AKURASI JAWABAN SMALL LANGUAGE
MODEL (SLM) PADA DOKUMEN PUBLIK BERBAHASA**

Nama Mahasiswa : Rijalul Fikri
NPM : 2024171004
Program Studi : Ilmu Komputer Program Magister
Konsentrasi : Computational Intelligence

Telah Diuji Dalam Sidang Tesis
Pada Tanggal 29 Juli 2026

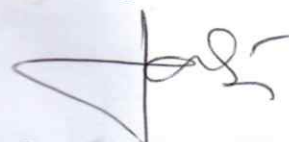
No	Komisi Penguji Sidang Tesis,	Tanda Tangan
01	Rudi Heriansyah, S.T., M.Eng., Ph.D NIDN: 0229047502 Ketua Komisi Penguji / Pembimbing Utama	
02	Dr. Rendra Gustriansyah, S.T., M.Kom NIDN: 0220087301 Anggota Komisi Penguji I	
03	Dr. Gasim, S.Kom., M.Si NIDN: 0217067301 Anggota Komisi Penguji II	

Disetujui oleh;
Plt Dekan,
FAKULTAS ILMU KOM & SAINS



Dr. Herri Setiawan, S.Kom., M.Kom
NIDN. 0213036901

Diketahui oleh;
Ka.Prodi Ilmu Komputer Program
Magister



Dr. Gasim S.Kom., M.Si.,
NIDN. 0217067301



FAKULTAS ILMU KOMPUTER DAN SAINS
ILMU KOMPUTER PROGRAM MAGISTER

LEMBAR PERSETUJUAN TESIS

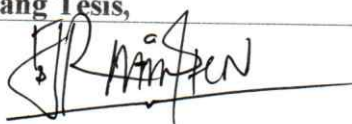
Tesis ini diajukan oleh,

Judul : **ANALISIS PENGARUH PANJANG CHUNK DAN OVERLAP
TERHADAP AKURASI JAWABAN SMALL LANGUAGE
MODEL (SLM) PADA DOKUMEN PUBLIK BERBAHASA**
Nama Mahasiswa : **Rijalul Fikri**
NPM : **2024171004**
Program Studi : **Ilmu Komputer Program Magister**

Telah dipertahankan di hadapan Komisi Penguji Sidang Tesis, dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer pada Program Studi Magister Ilmu Komputer Fakultas Ilmu Komputer dan Sains Universitas Indo Global Mandiri.

Komisi Penguji Sidang Tesis,

Ketua Komisi Penguji /
Pembimbing Utama,


Rudi Heriansyah, S.T., M.Eng., Ph.D
NIDN: 0229047502

Anggota Komisi Penguji I


Dr. Rendra Gustriansyah, S.T., M.Kom
NIDN. 0220087301

Anggota Komisi Penguji II


Dr. Gasim, S.Kom., M.Si
NIDN. 0217067301

Disahkan oleh:

Ka.Prodi Ilmu Komputer Program
Magister,


Dr. Gasim S.Kom., M.Si.,
NIDN. 0217067301

Plt Dekan,
FAKULTAS ILMU KOM & SAINS


Dr. Herri Setiawan, S.Kom., M.Kom
NIDN. 0213036901

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, tesis ini dipersembahkan kepada:

Kedua orang tua dan keluarga tercinta,
atas doa, kasih sayang, dan dukungannya.

Serta semua pihak yang telah memberikan bantuan
dalam proses penyelesaian studi ini.

ABSTRAK

Penelitian ini menganalisis pengaruh panjang *chunk* dan *overlap* terhadap akurasi jawaban *Small Language Model (SLM)* dalam sistem *Retrieval-Augmented Generation (RAG)* pada dokumen publik berbahasa Indonesia. Eksperimen dilakukan dengan tiga model SLM, yaitu *Qwen2.5-3B-Instruct*, *Sailor2-3B-Chat*, dan *Llama-3.2-3B-Instruct-Indonesian*, dengan sembilan konfigurasi *chunking* (128/256/512 token $\times$ 0/10/20% *overlap*) dan 50 pasangan pertanyaan-jawaban. Evaluasi menggunakan metrik generasi leksikal deterministik (*Exact Match*, *Token F1*, *LCS F1*, *BLEU-1*) dan metrik *retrieval* (*Recall@k*), dengan validasi pendamping menggunakan RAGAS (*judge* Kimi *kimi-k2.7-code*, skala 0–10) pada subset 30 jawaban. Signifikansi pengaruh diuji dengan ANOVA faktorial. Hasil menunjukkan bahwa *chunk* 256 token memberikan keseimbangan terbaik antara kelengkapan konteks dan presisi *retrieval* untuk ketiga model. Secara kuantitatif, konfigurasi terbaik menghasilkan *Token F1* 0.591 untuk Qwen (cs256_ol10), 0.532 untuk Sailor (cs256_ol20), dan 0.525 untuk Llama-Indonesian (cs256_ol20). Uji ANOVA mengonfirmasi ukuran *chunk* berpengaruh signifikan terhadap akurasi ($p = 0,032$), sedangkan *overlap* tidak signifikan ($p = 0,264$). Validasi RAGAS pada 30 jawaban menunjukkan *faithfulness* 8.4–9.2, *answer relevancy* 7.7–9.1, *context precision* 4.4–5.8, dan *context recall* 9.0–9.6. Efek *overlap* tidak konsisten; konfigurasi terbaik per model adalah Qwen cs256_ol10, Sailor cs256_ol20, dan Llama-Indonesian cs256_ol20. Analisis latensi menunjukkan *chunk* 512 memiliki latensi tertinggi (rata-rata 201.9 detik per jawaban), sementara konfigurasi terbaik berada pada 96.2–146.6 detik per jawaban. Penelitian ini memberikan panduan praktis dalam membangun sistem RAG berbahasa Indonesia dengan sumber daya komputasi terbatas.

Kata Kunci: *Retrieval-Augmented Generation, Small Language Model, Chunking, Overlap, Bahasa Indonesia, Evaluasi RAG.*

ABSTRACT

This study analyzes the effect of chunk length and overlap on the answer accuracy of Small Language Models (SLMs) in Retrieval-Augmented Generation (RAG) systems for Indonesian public documents. The experiment was conducted using three SLMs: Qwen2.5-3B-Instruct, Sailor2-3B-Chat, and Llama-3.2-3B-Instruct-Indonesian, with nine chunking configurations (128/256/512 tokens $\times$ 0/10/20% overlap) and 50 question-answer pairs. Evaluation used deterministic lexical generation metrics (Exact Match, Token F1, LCS F1, BLEU-1) and a retrieval metric (Recall@k), supplemented by RAGAS validation (Kimi kimi-k2.7-code judge, 0–10 scale) on a subset of 30 answers. The significance of the effects was tested using a factorial ANOVA. The results show that 256-token chunks provided the best balance between context completeness and retrieval precision across all three models. Quantitatively, the best configurations achieved Token F1 scores of 0.591 for Qwen (cs256_ol10), 0.532 for Sailor (cs256_ol20), and 0.525 for Llama-Indonesian (cs256_ol20). ANOVA confirmed that chunk size had a significant effect on accuracy ($p = 0.032$), whereas overlap did not ($p = 0.264$). RAGAS validation on 30 answers showed faithfulness 8.4–9.2, answer relevancy 7.7–9.1, context precision 4.4–5.8, and context recall 9.0–9.6. The effect of overlap was inconsistent; the best configuration per model was Qwen cs256_ol10, Sailor cs256_ol20, and Llama-Indonesian cs256_ol20. Latency analysis indicated that 512-token chunks had the highest latency (201.9 seconds per answer on average), while the best configurations ranged from 96.2 to 146.6 seconds per answer. This study provides practical guidance for building Indonesian-language RAG systems with limited computational resources.

Keywords: Retrieval-Augmented Generation, Small Language Model, Chunking, Overlap, Indonesian Language, RAG Evaluation.

PERNYATAAN BUKAN PLAGIARISME

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Rijalul Fikri

NPM : 2024171004

Dengan ini menyatakan bahwa sesungguhnya tesis saya yang berjudul:

ANALISIS PENGARUH PANJANG CHUNK DAN OVERLAP TERHADAP AKURASI JAWABAN SMALL LANGUAGE MODEL (SLM) PADA DOKUMEN PUBLIK BERBAHASA INDONESIA

Tidak pernah diajukan untuk memperoleh gelar kesarjanaan di perguruan tinggi lain dan benar-benar hasil penelitian saya, sesuai kode etik akademik penulisan ilmiah. Pengutipan hasil penelitian orang lain atau penggunaan teori penelitian, saya selalu disebutkan sumber kutipan yang bersangkutan. Jika saya melakukan plagiarisme dan melanggar pernyataan tersebut, maka saya bersedia dikenakan sanksi akademik yang berlaku di Universitas Indo Global Mandiri Palembang.

Demikianlah pernyataan ini saya buat dan saya tanda tangani secara sadar dan sebenarnya, untuk dapat digunakan sebagaimana mestinya.

Palembang, 12 Agustus 2026

Yang Menyatakan,



Rijalul Fikri
M. 2024171004

**DEKLARASI HAK CIPTA DAN PERSEMBAHAN LICENSI TESIS
TERBATAS KEPADA PROGRAM MAGISTER ILMU KOMPUTER
UNIVERSITAS IGM**

Dengan ini, saya selaku penulis dan pemilik hak cipta kekayaan intelektual berupa tesis yang berjudul:

**ANALISIS PENGARUH PANJANG CHUNK DAN OVERLAP TERHADAP
AKURASI JAWABAN SMALL LANGUAGE MODEL (SLM) PADA
DOKUMEN PUBLIK BERBAHASA INDONESIA**

Dengan sebenarnya memberikan kewenangan terbatas kepada program studi Ilmu Komputer program magister Universitas Indo Global Mandiri Palembang; untuk menyimpan pada perpustakaan, menggandakan dan mensitasi untuk kepentingan penelitian ilmiah.

Demikianlah deklarasi ini dibuat dengan sesungguhnya untuk digunakan sebagaimana mestinya.

Palembang, 12 Agustus 2026

Mendeklarasikan,

 **Rijalul Fikri**
NPM. 2024171004

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga tesis ini dapat diselesaikan. Tesis ini disusun untuk memenuhi salah satu syarat memperoleh gelar Magister (S2) pada Program Studi Ilmu Komputer, Program Magister, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri.

Penulis menyadari bahwa penyelesaian tesis ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Ketua Program Studi Magister Ilmu Komputer beserta jajaran dosen yang telah memberikan ilmu dan wawasan selama masa perkuliahan.
2. Bapak Rudi Heriansyah, S.T., M.Eng., Ph.D., selaku dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan dalam penyusunan tesis ini.
3. Rekan-rekan mahasiswa Program Magister Ilmu Komputer atas kebersamaan, diskusi, dan dukungannya.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu, baik secara langsung maupun tidak langsung, dalam proses penyelesaian tesis ini.

Penulis menyadari bahwa tesis ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk penyempurnaan karya ilmiah ini.

Palembang, 12 Agustus 2026

Yang menyatakan,

Rijalul Fikri

NPM. 2024171004

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
HALAMAN PERSEMBAHAN	iv
ABSTRAK	v
ABSTRACT	vi
PERNYATAAN BUKAN PLAGIARISME	vii
DEKLARASI HAK CIPTA DAN PERSEMBAHAN LICENSI TESIS TERBA- TAS KEPADA PROGRAM MAGISTER ILMU KOMPUTER UNIVERSITAS IGM	viii
KATA PENGANTAR.....	ix
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
BAB I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	2
1.3 Perumusan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Ruang Lingkup	5
1.6 Manfaat Penelitian.....	5
1.6.1 Manfaat Teoritis.....	5
1.6.2 Manfaat Praktis	6
BAB II. KAJIAN PUSTAKA	7
2.1 Landasan Teori	7
2.1.1 <i>Large Language Models (LLM)</i>	7
2.1.2 <i>Small Language Models (SLM)</i>	9
2.1.3 Tokenisasi	10
2.1.4 <i>Retrieval-Augmented Generation (RAG)</i>	11
2.1.5 Representasi Vektor (<i>Embedding</i>) dan Pencarian Kemiripan Semantik	12
2.1.6 Strategi Pemotongan Teks (<i>Chunking Strategies</i>)	13

2.1.7	Dampak Ukuran <i>Chunk</i> dan <i>Overlap</i> terhadap Performa <i>RAG</i>	14
2.1.8	Lanskap NLP Bahasa Indonesia	14
2.1.9	Metrik Evaluasi <i>RAG</i>	15
2.1.10	Analisis Varians (<i>ANOVA</i>)	19
2.1.11	Kesenjangan Penelitian (<i>Research Gap</i>)	20
2.1.12	Kerangka Pemikiran	35
2.1.13	Hipotesis	36
BAB III. METODOLOGI PENELITIAN		37
3.1	Lokasi dan Waktu Penelitian	37
3.2	Populasi dan Sampel	37
3.3	Jenis dan Sumber Data	37
3.3.1	Sumber Data	37
3.3.2	Proses Data	38
3.4	Metode Pengumpulan Data	39
3.5	Definisi Operasional Variabel	40
3.6	Alur Penelitian	41
3.6.1	Pengumpulan Data dan Korpus Pengetahuan	42
3.6.2	Pemilihan Model	42
3.6.3	Strategi <i>Chunking</i> dan <i>Overlap</i>	44
3.6.4	Variabel <i>Overlap</i> (Tumpang Tindih)	45
3.6.5	Desain Sistem <i>Retrieval-Augmented Generation (RAG)</i>	48
3.6.6	Justifikasi Pilihan Rancangan Eksperimen	51
3.6.7	Tantangan Praktis Eksperimen: Tantangan Infrastruktur dan Waktu Proses ...	54
3.6.8	Keterbatasan dan Mitigasi Model: Penggantian SEA-LION-v1-3B	55
3.6.9	Data Pertanyaan untuk Evaluasi	56
3.6.10	Metrik Evaluasi	56
3.6.11	Prosedur Eksperimen	59
BAB IV. HASIL PENELITIAN DAN PEMBAHASAN		60
4.1	Ringkasan Eksperimen	60
4.2	Hasil Evaluasi Lokal	61

4.3	Hasil Evaluasi <i>Retrieval</i>	64
4.4	Hasil Evaluasi RAGAS <i>Subset</i>	67
4.5	Analisis Per Domain	69
4.6	Analisis Latensi	70
4.7	Uji Signifikansi Statistik	72
4.8	Pembahasan	74
4.8.1	Konsolidasi Temuan Utama	74
4.8.2	<i>Trade-off</i> antara <i>Chunk Size</i> dan <i>Retrieval Quality</i>	75
4.8.3	Keterkaitan dengan Fenomena <i>Lost in the Middle</i>	75
4.8.4	Perbedaan Antar Model	76
4.8.5	Implikasi untuk Praktik RAG Berbahasa Indonesia	77
4.9	Benchmark dengan Penelitian Terdahulu	78
4.10	Keterbatasan	80
BAB V. PENUTUP		82
5.1	Kesimpulan	82
5.2	Saran	84
DAFTAR PUSTAKA		85
BAB A. LAMPIRAN		93
1.1	Daftar Pertanyaan dan Jawaban Evaluasi	93
1.1.1	Distribusi Domain	96
1.2	Cuplikan Kode Sistem Eksperimen	97
1.2.1	Pemotongan Dokumen (<i>Chunking</i>)	97
1.2.2	Pembangunan Indeks Vektor FAISS	99
1.2.3	Inferensi RAG dengan Model SLM	101
1.2.4	Evaluasi Lokal dengan Metrik Deterministik	103
1.2.5	Menjalankan Eksperimen	105
1.3	File Data Pendukung	106
1.4	Kartu Bimbingan Tesis	107

DAFTAR TABEL

2.1	Tabel Perbandingan Penelitian Terdahulu	22
3.1	Jumlah <i>Chunk</i> per Konfigurasi Pemotongan	47
3.2	Statistik Panjang <i>Chunk</i> dan Waktu Pemotongan	48
4.1	Skor <i>Token F1</i> untuk Setiap Konfigurasi <i>Chunk</i> dan Model	61
4.2	Hasil Seluruh Metrik Lokal untuk Qwen2.5-3B-Instruct	63
4.3	Hasil Seluruh Metrik Lokal untuk Sailor2-3B-Chat	63
4.4	Hasil Seluruh Metrik Lokal untuk Llama-3.2-3B-Instruct-Indonesian	64
4.5	<i>Recall@k Retrieval</i> untuk Setiap Konfigurasi <i>Chunk</i> (Independen terhadap Model)	65
4.6	<i>Context Precision</i> Lunak (Rata-rata Kemiripan <i>Token Chunk–Ground Truth</i>) per Konfigurasi	65
4.7	Hasil RAGAS <i>Subset</i> pada Konfigurasi Terbaik per Model	68
4.8	Skor <i>Token F1</i> per Domain pada Konfigurasi Terbaik Setiap Model	69
4.9	Latensi vs Akurasi pada Konfigurasi Terbaik per Model	71
4.10	Hasil Uji ANOVA Faktorial terhadap Skor <i>Token F1</i> ($N = 1350$)	73
4.11	Benchmark Hasil Penelitian Ini dengan Penelitian Terdahulu	79
1.1	Daftar 50 Pertanyaan Evaluasi	93

DAFTAR GAMBAR

3.1	Diagram Alur Penelitian	42
3.2	Ilustrasi pemotongan dokumen menjadi <i>chunk</i> dengan <i>overlap</i> . Panjang <i>chunk</i> dan persentase <i>overlap</i> pada gambar bersifat ilustratif; eksperimen menggunakan 128/256/512 token dengan 0/10/20% <i>overlap</i>	45
3.3	Diagram Alur Sistem <i>RAG</i>	51
4.1	Perbandingan <i>Token F1</i> Domain Hukum vs. Berita pada Konfigurasi Terbaik per Model	69
4.2	<i>Trade-off</i> Latensi vs <i>Token F1</i> pada Konfigurasi Terbaik per Model.....	71