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,

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

UIGM	**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,	**Plt Dekan,**
Dr. Gasim S.Kom., M.Si.,	**Dr. Herri Setiawan, S.Kom., M.Kom**
NIDN. 0217067301	**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 penggunaaan 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 penyataan ini saya buat dan saya tanda tangani secara sadar dan sebenarnya, untuk dapat digunakan sebagaimana mestinya.

Palembang, 12 Agustus 2026

Yang Menyatakan,

Rijalul Fikri
NPM. 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

DAFTAR TABEL

DAFTAR GAMBAR