IMPLEMENTASI LARGE LANGUAGE MODEL (LLM) UNTUK ANALISIS STATIS DAN KLASIFIKASI MALWARE PADA APLIKASI ANDROID

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer

Disusun oleh:

Lalu Muhammad Syamsul Hadi
22.83.0766

Kepada

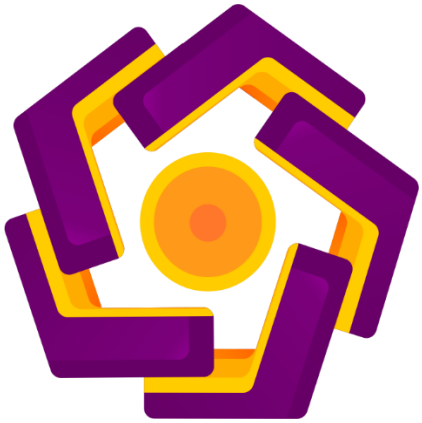
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2025

IMPLEMENTASI LARGE LANGUAGE MODEL (LLM) UNTUK ANALISIS STATIS DAN KLASIFIKASI MALWARE PADA APLIKASI ANDROID

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



Disusun oleh:

Lalu Muhammad Syamsul Hadi
22.83.0766

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

IMPLEMENTASI LARGE LANGUAGE MODEL (LLM) UNTUK ANALISIS STATIS DAN KLASIFIKASI MALWARE PADA APLIKASI ANDROID

yang dipersiapkan dan disusun oleh

Lalu Muhammad Syamsul Hadi

22.83.0766

Telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Desember 2025

Dosen Pembimbing,

Melwin Syafrizal, S.Kom., M.Eng., Ph.D.
NIK. 190302105

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI LARGE LANGUAGE MODEL (LLM) UNTUK ANALISIS STATIS DAN KLASIFIKASI MALWARE PADA APLIKASI ANDROID

yang dipersiapkan dan disusun oleh

Lalu Muhammad Syamsul Hadi

22.83.0766

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Desember 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Jeki Kuswanto, S.Kom., M.kom. **NIK. 190302456**	
Muhammad Koprawi, S.Kom., M.Eng. **NIK. 190302454**	
Melwin Syafrizal, S.Kom., M.Eng., Ph.D. **NIK. 190302105**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Lalu Muhammad Syamsul Hadi
NIM : 22.83.0766

Menyatakan bahwa Skripsi dengan judul berikut:

IMPLEMENTASI LARGE LANGUAGE MODEL (LLM) UNTUK ANALISIS STATIS DAN KLASIFIKASI MALWARE PADA APLIKASI ANDROID

Dosen Pembimbing : Melwin Syafrizal, S.Kom., M.Eng., Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Desember 2025

Yang Menyatakan,

Lalu Muhammad Syamsul Hadi

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan sebagai bentuk rasa terima kasih dan penghargaan yang mendalam kepada semua pihak yang telah memberikan dukungan, bimbingan, dan semangat selama proses penelitian dan penulisan. Perjalanan ini tidak akan terwujud tanpa peran serta orang-orang yang selalu berada di samping saya, memberikan motivasi, dan membantu saya dalam setiap langkah yang diambil. Skripsi ini dipersembahkan untuk:

1. Orang tua tercinta, yang telah memberikan dukungan, kasih sayang, dan semangat tanpa henti sepanjang perjalanan ini.
2. Bapak Melwin Syafrizal, S.Kom.,M.Eng.,Ph.D. selaku dosen pembimbing, yang telah memberikan arahan, bimbingan, dan pengetahuan yang sangat berharga dalam penyelesaian skripsi ini.
3. Teman-teman yang selalu memberikan dukungan moral dan kerja sama yang luar biasa selama penelitian ini.

Semoga karya ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan teknologi.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi yang berjudul "IMPLEMENTASI LARGE LANGUAGE MODEL (LLM) UNTUK ANALISIS STATIS DAN KLASIFIKASI MALWARE PADA APLIKASI ANDROID" ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik Komputer di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa tanpa dukungan dari berbagai pihak, penyelesaian skripsi ini tidak akan dapat terlaksana dengan baik. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Melwin Syafrizal, S.Kom., M.Eng., P.hD. selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan masukan yang sangat berharga selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji, yang telah memberikan kritik, saran, dan evaluasi yang membangun untuk meningkatkan kualitas skripsi ini.
3. Orang tua penulis, yang selalu memberikan dukungan moral, material, dan doa yang tak terhingga sepanjang perjalanan pendidikan ini.

Penulis juga mengucapkan terima kasih kepada teman-teman, rekan sejawat, serta semua pihak yang tidak dapat disebutkan satu per satu, atas dukungan dan bantuan yang telah diberikan.

Semoga skripsi ini dapat memberikan kontribusi positif bagi perkembangan ilmu pengetahuan, khususnya di bidang Cyber Security. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk penyempurnaan karya ini.

Yogyakarta, 22 Desember 2025

Lalu Muhammad Syamsul Hadi

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

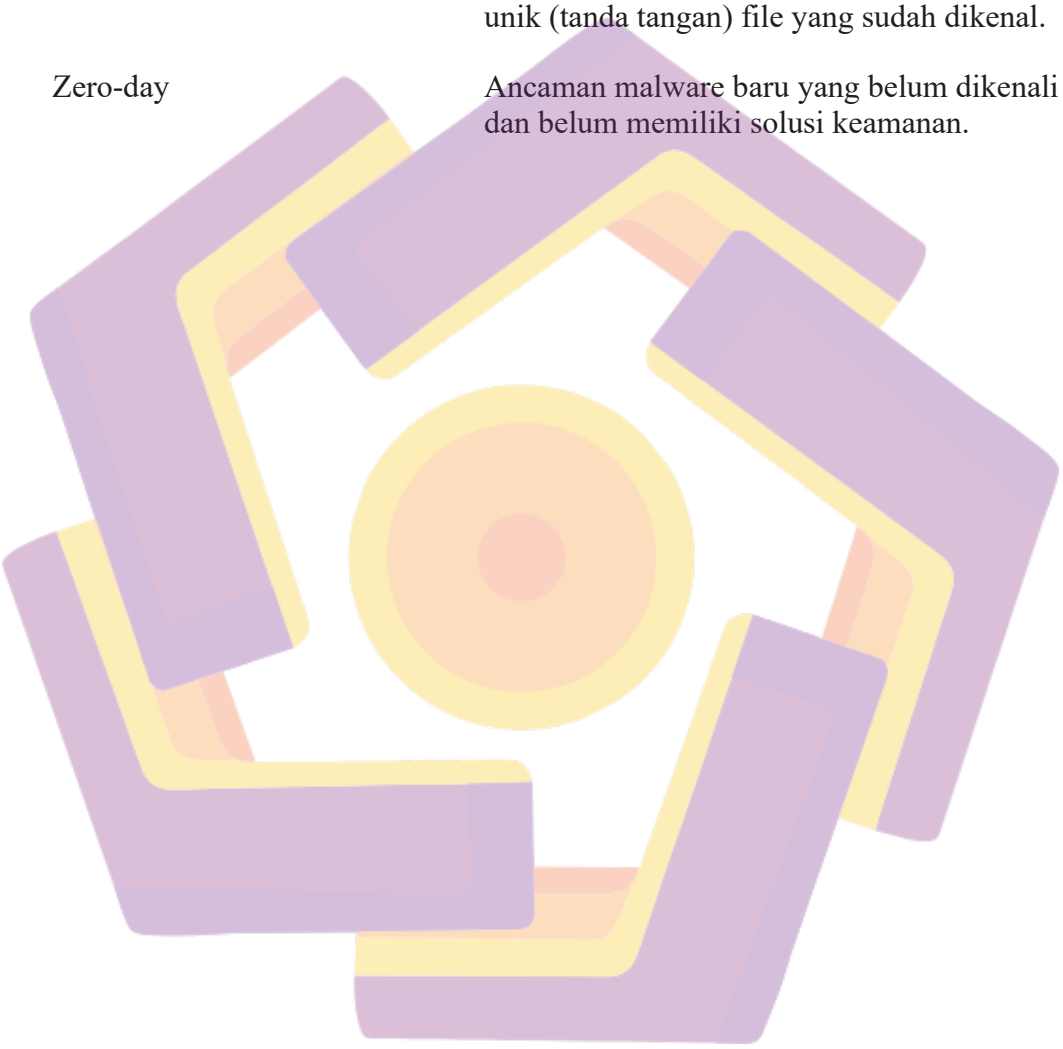
AI	Artificial Intelligence
API	Application Programming Interface
APK	Android Package Kit
AST	Abstract Syntax Tree
C&C / C2	Command and Control
CoT	Chain-of-Thought
DEX	Dalvik Executable
DL	Deep Learning
FN	False Negative
FP	False Positive
GPS	Global Positioning System
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
IoC	Indicators of Compromise
IP	Internet Protocol
JSON	JavaScript Object Notation
JVM	Java Virtual Machine
LLM	Large Language Model
LTS	Long Term Support

ML	Machine Learning
OS	Operating System
OTP	One-Time Password
RAG	Retrieval-Augmented Generation
RAM	Random Access Memory
RNN	Recurrent Neural Network
SDK	Software Development Kit
SMS	Short Message Service
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
VPS	Virtual Private Server
XML	Extensible Markup Language
TP	Jumlah data positif yang diprediksi benar sebagai positif (Malware terdeteksi Malware)
TN	Jumlah data negatif yang diprediksi benar sebagai negatif (Aman terdeteksi Aman)
FP	Jumlah data negatif yang diprediksi salah sebagai positif (Aman terdeteksi Malware)
FN	Jumlah data positif yang diprediksi salah sebagai negatif (Malware terdeteksi Aman)
%	Persentase

DAFTAR ISTILAH

Analisis Dinamis	Metode analisis perilaku aplikasi saat dijalankan di lingkungan terisolasi (sandbox).
Analisis Hibrida	Penggabungan metode analisis statis dan dinamis untuk meningkatkan akurasi deteksi.
Analisis Statis	Metode pemeriksaan kode dan struktur aplikasi tanpa mengeksekusi programnya.
APK (*Android Package Kit*)	Format file paket aplikasi yang digunakan oleh sistem operasi Android.
Benign	Aplikasi aman yang tidak mengandung kode berbahaya atau niat jahat.
Chain-of-Thought (CoT)	Teknik penalaran bertahap pada AI untuk menghasilkan keputusan yang logis.
Command and Control (C&C)	Server yang dikendalikan penyerang untuk mengirim perintah ke perangkat terinfeksi.
Confusion Matrix	Tabel evaluasi kinerja model klasifikasi yang membandingkan prediksi dengan fakta (TP, TN, FP, FN).
Dekompilasi	Proses mengubah file biner aplikasi kembali menjadi kode sumber yang dapat dibaca.
False Positive	Kesalahan deteksi saat aplikasi aman dianggap sebagai berbahaya oleh sistem.
Large Language Model (LLM)	Model AI yang dilatih pada data teks masif untuk memahami bahasa dan kode pemrograman.
Malware	Perangkat lunak berbahaya yang dirancang untuk merusak sistem atau mencuri data.
Obfuscation	Teknik penyembunyian dan pengacakan kode program agar sulit dianalisis.
Prompt Engineering	Teknik merancang instruksi input (prompt) untuk mengarahkan hasil output model AI.

Recall	Tingkat keberhasilan sistem dalam mendeteksi seluruh sampel malware yang ada.
Sandbox	Lingkungan isolasi aman untuk menjalankan program mencurigakan tanpa risiko infeksi.
Signature-based	Metode deteksi dengan mencocokkan pola unik (tanda tangan) file yang sudah dikenal.
Zero-day	Ancaman malware baru yang belum dikenali dan belum memiliki solusi keamanan.



INTISARI

Dominasi sistem operasi Android menjadikannya target utama serangan *malware* canggih yang menggunakan teknik *obfuscation* dan polimorfik. Metode deteksi tradisional berbasis tanda tangan sering gagal mengenali ancaman *zero-day* karena hanya mencocokkan pola sintaksis tanpa memahami konteks kode. Keterbatasan ini meningkatkan risiko pencurian data dan kerugian finansial, sehingga diperlukan pendekatan deteksi proaktif yang mampu menganalisis niat semantik dari sebuah aplikasi.

Penelitian ini mengimplementasikan *Large Language Model* (LLM) Google Gemini 2.5 Pro sebagai mesin analisis statis otomatis untuk mengatasi masalah tersebut. Metodologi meliputi pengumpulan dataset *malware* dan aplikasi aman, dekompilasi menggunakan Apktool dan Jadx, serta ekstraksi fitur statis krusial seperti izin, API sensitif, dan *string*. Fitur-fitur tersebut dianalisis menggunakan teknik *prompt engineering* terstruktur guna memicu penalaran rantai pikiran (*chain-of-thought*) LLM dalam mengklasifikasikan tingkat bahaya aplikasi.

Hasil pengujian menunjukkan sistem mencapai nilai *Recall* 100%, membuktikan efektivitas LLM dalam mendeteksi seluruh sampel *malware* tanpa kegagalan. Namun, akurasi tercatat 50% akibat tingginya *False Positive* karena model cenderung konservatif terhadap penggunaan izin sensitif pada aplikasi aman. Penelitian ini berkontribusi sebagai alat triase awal yang efisien bagi analis keamanan siber, dengan saran pengembangan lanjut berupa penggabungan analisis dinamis untuk memverifikasi perilaku *runtime*.

Kata kunci: Malware Android, Large Language Model, Analisis Statis, Prompt Engineering, Keamanan Siber.

ABSTRACT

The dominance of the Android operating system makes it a primary target for sophisticated malware attacks utilizing obfuscation and polymorphic techniques. Traditional signature-based detection methods often fail to recognize zero-day threats as they merely match syntactic patterns without understanding code context. This limitation increases the risk of data theft and financial loss, necessitating a proactive detection approach capable of analyzing the semantic intent of an application.

This research implements the Google Gemini 2.5 Pro Large Language Model (LLM) as an automated static analysis engine to address this issue. The methodology includes collecting a dataset of malware and benign applications, decompilation using Apktool and Jadx, and extracting crucial static features such as permissions, sensitive APIs, and strings. These features are analyzed using structured prompt engineering techniques to trigger the LLM's chain-of-thought reasoning in classifying the application's threat level.

Test results show the system achieved a Recall value of 100%, demonstrating the LLM's effectiveness in detecting all malware samples without failure. However, accuracy was recorded at 50% due to a high rate of False Positives, as the model tends to be conservative regarding sensitive permission usage in benign applications. This research contributes as an efficient initial triage tool for cybersecurity analysts, with future development recommendations including the integration of dynamic analysis to verify runtime behavior.

Keywords: *Android Malware, Large Language Model, Static Analysis, Prompt Engineering, Cybersecurity.*