IMPLEMENTATION OF LLAMA MODEL FOR AUTOMATIC IMAGE CAPTION GENERATION IN ANDROID-BASED APPLICATION

By

Joe Aqilla Vandyta, NIM 2115101027

Informatics Engineering Department

ABSTRACT

This research aims to design and build an Android application named Descripix that utilizes the LLaMA Vision model for automatic image caption generation. The research background is the many content creators, photographers, and social media users experience creative blocks when writing captions, *which hinders* the consistency of their content publication. An initial survey showed 90% of respondents experienced difficulty writing captions, often delaying uploads or even posting without captions during creative block periods. The Waterfall development method is used with Django REST Framework backend and MySQL database that integrates LLaMA Vision from Groq.com via REST API. The system processes image metadata as additional input for In-Context Learning (ICL) prompting, producing consistent contextual captions that match the expected linguistic patterns. Comparison of caption generation results with and without ICL proves improved accuracy and output consistency. The Android frontend uses Jetpack Compose with MVVM architecture and Retrofit to communicate with the backend. The application has two user modes: guest can upload images and generate captions, while authenticated users can save, edit, and manage caption history. System testing includes integration testing with 19 scenarios showing 100% success on all main functions, as well as SUS usability testing on 20 respondents with an average score of 84.25, indicating good usability above the global average of 68, suggesting the Descripix application is easy to use and acceptable to users.

Keyword: Caption Generator, LLaMA Vision, In-Context Learning, Android, Django Rest Framework

IMPLEMENTASI MODEL LLAMA UNTUK PEMBUATAN CAPTION GAMBAR OTOMATIS PADA APLIKASI ANDROID

Oleh

Joe Aqilla Vandyta, NIM 2115101027

Jurusan Teknik Informatika

ABSTRAK

Penelitian ini bertujuan merancang dan membangun aplikasi Android bernama Descripix yang memanfaatkan model LLaMA Vision untuk *generate caption* gambar otomatis. Latar belakang penelitian adalah banyaknya *content creator*, fotografer, dan pengguna media sosial yang mengalami *creative block* dalam membuat *caption* gambar, sehingga menghambat konsistensi publikasi konten mereka. Survei awal menunjukkan 90% responden mengalami kesulitan menulis caption, sering menunda unggahan atau bahkan mengunggah tanpa *caption* selama periode creative block. Metode pengembangan Waterfall digunakan dengan backend Django REST Framework dan database MySQL yang mengintegrasikan LLaMA Vision dari Groq.com melalui REST API. Sistem memproses metadata gambar sebagai input tambahan untuk In-Context Learning (ICL) *prompting*, menghasilkan caption kontekstual yang konsisten dan sesuai pola Linguistik yang diharapkan. Perbandingan hasil *generate caption* dengan dan tanpa ICL membuktikan peningkatan akurasi dan konsistensi *output*. Frontend Android menggunakan Jetpack Compose dengan arsitektur MVVM dan menggunakan Retrofit untuk berkomunikasi dengan *backend*. Aplikasi memiliki dua mode pengguna: *guest* dapat mengunggah gambar dan menghasilkan caption, sementara *authenticated user* dapat menyimpan, mengedit, dan mengelola riwayat caption. Pengujian sistem mencakup integration testing dengan 19 skenario yang menunjukkan keberhasilan 100% pada semua fungsi utama, serta usability testing SUS pada 20 responden dengan skor rata-rata 84.25, menunjukkan usability baik di atas rata-rata global 68, mengindikasikan aplikasi Descripix mudah digunakan dan dapat diterima pengguna.

Kata Kunci: Caption Generator, LLaMA Vision, In-Context Learning, Android, Django Rest Framework