

Pengembangan Chatbot Berbasis Komodo-7B dengan Low-Rank Adaptation dan Retrieval-Augmented Generation untuk Pemahaman dan Tanya Jawab pada Dokumen PDF Bahasa Indonesia = Komodo-7B Based Chatbot Development with Low-Rank Adaptation and Retrieval-Augmented Generation for Understanding and Question-Answering on Indonesian PDF Documents

Muhammad Naufal Faza, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920544032&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk mengembangkan sistem chatbot yang mampu menjawab pertanyaan seputar akademik Teknik Komputer UI. Sistem ini memanfaatkan teknologi Large Language Model (LLM) Komodo-7B yang telah di-fine-tuning dengan teknik Low-Rank Adaptation (LoRA) dan diintegrasikan dengan Retrieval Augmented Generation (RAG). Dataset Ultrachat yang diterjemahkan ke Bahasa Indonesia digunakan untuk fine-tuning model Komodo-7B, sementara dokumen PDF Kurikulum Teknik Komputer UI 2020 v4 digunakan sebagai sumber informasi untuk model RAG.

Pengujian performa model Komodo-7B menunjukkan bahwa LoRA efektif dalam meningkatkan kemampuan model dalam memahami dan menghasilkan teks percakapan Bahasa Indonesia. Namun, pengujian performa chatbot menggunakan dua dataset pertanyaan, yaitu dataset custom yang dihasilkan menggunakan Giskard dan API ChatGPT, dan dataset Fathurrahman Irwansa yang telah diadaptasi, menunjukkan bahwa sistem chatbot masih memiliki ruang untuk peningkatan. Tingkat akurasi yang rendah pada kedua dataset (32% pada dataset custom dan 24,1% pada dataset Fathur) mengindikasikan bahwa sistem retrieval yang digunakan kurang akurat dalam menemukan konteks yang relevan. Meskipun demikian, ketika model RAG dapat mengambil konteks yang relevan, model Komodo-7B menunjukkan akurasi yang cukup tinggi (80% pada dataset custom dan 91,29% pada dataset Fathur, dihitung dari jumlah ketika kedua Komodo-7B dan konteks benar, kemudian dibagi dengan jumlah ketika konteks benar).

Hasil penelitian menunjukkan bahwa model Komodo-7B memiliki potensi yang baik untuk digunakan pada sistem chatbot jika dikombinasikan dengan sistem retrieval yang lebih akurat. Penelitian ini memberikan kontribusi dalam pengembangan sistem chatbot berbasis LLM untuk menjawab pertanyaan seputar akademik, dan membuka peluang untuk penggunaan yang lebih luas di lingkungan Universitas Indonesia.

.....This research aims to develop a chatbot system capable of answering questions regarding the academic curriculum of Computer Engineering at Universitas Indonesia. The system utilizes the Komodo-7B Large Language Model (LLM), fine-tuned with Low-Rank Adaptation (LoRA) and integrated with Retrieval Augmented Generation (RAG). The Ultrachat dataset, translated into Indonesian, is used for fine-tuning the Komodo-7B model, while the 2020 v4 Computer Engineering Curriculum PDF document serves as the information source for the RAG model. Performance evaluation of the Komodo-7B model demonstrates that LoRA effectively enhances the model's ability to understand and generate Indonesian text. However, chatbot performance testing using two question datasets, a custom dataset generated using Giskard and the ChatGPT API, and the Fathur dataset adapted from prior research, reveals that the chatbot system still has room for improvement. The low accuracy on both datasets (32% on the custom dataset and 24.1% on the Fathur dataset) indicates that the retrieval system employed is not sufficiently accurate in finding relevant context.

Nevertheless, when the RAG model successfully identifies relevant context, the Komodo-7B model exhibits relatively high accuracy (80% on the custom dataset and 91.29% on the Fathur dataset, calculated from the total of when Komodo-7B and the context are both correct, then divided by the total of when the context is correct). The research findings suggest that the Komodo-7B model holds significant potential for chatbot systems when combined with a more accurate retrieval system. This study contributes to the development of LLM-based chatbot systems for answering academic-related questions and opens up opportunities for broader applications within Universitas Indonesia.