

Pengembangan dan Evaluasi Chatbot Asisten Dosen untuk Layanan Konsultasi: Studi Kasus Mata Kuliah Dinamika Tim Perangkat Lunak MTI UI = Development and Evaluation of a Teaching Assistant Chatbot for Consultation Services: A Case Study of MTI UI Software Team Dynamics Course

Papilaya, Ferdinand Amos, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920575067&lokasi=lokal>

Abstrak

Penelitian ini bertujuan mengembangkan dan mengevaluasi chatbot asisten dosen yang memberikan layanan konsultasi akademik secara otomatis dan responsif kepada mahasiswa. Studi ini mengambil topik Scrum dalam mata kuliah Dinamika Tim Perangkat Lunak sebagai kasus uji, yang diajarkan di Program Magister Teknologi Informasi Universitas Indonesia (MTI UI). Pengembangan sistem dilakukan secara kolaboratif oleh dua penulis dengan peran berbeda: penulis pertama membangun aplikasi front-end dan back-end serta melakukan pengujian melalui User Acceptance Testing (UAT) dan API Testing; sedangkan penulis kedua mengembangkan model language model (LLM) yang mencakup persiapan data, fine-tuning model, evaluasi kinerja, dan deployment menggunakan FastAPI. Evaluasi chatbot dilakukan secara kuantitatif menggunakan metrik BLEU, ROUGE, dan BERTScore, serta secara kualitatif melalui wawancara dengan pengguna. Hasil evaluasi menunjukkan bahwa model berbayar ChatGPT unggul dalam seluruh metrik evaluasi, sementara model open-source seperti LLaMA dan Bloom masih menunjukkan potensi meskipun perlu peningkatan kualitas melalui penambahan data fine-tuning. Menariknya, proses fine-tuning tidak selalu memberikan peningkatan performa yang konsisten, bahkan pada beberapa kasus justru menurunkan akurasi model. Namun demikian, penggunaan model open-source tetap menjanjikan dari sisi efektivitas biaya, karena tidak bergantung pada token seperti model komersial.

.....This study aims to develop and evaluate a teaching assistant chatbot designed to provide automated and responsive academic consultation services to students. The case study focuses on the Scrum topic within the Software Team Dynamics course at the Master of Information Technology program, Universitas Indonesia (MTI UI). The development process was a collaboration between two authors with distinct roles: the first author was responsible for building the front-end and back-end applications and conducting tests using User Acceptance Testing (UAT) and API Testing; the second author focused on developing the large language model (LLM), including data preparation, model fine-tuning, performance evaluation, and deployment using FastAPI. The chatbot's performance was evaluated using quantitative metrics such as BLEU, ROUGE, and BERTScore, along with qualitative evaluation through user interviews. The results show that the commercial ChatGPT model achieved the best performance across all evaluation metrics. In contrast, open-source models such as LLaMA and Bloom showed promising but still suboptimal results, requiring larger and more comprehensive fine-tuning datasets to improve. Interestingly, fine-tuning did not always lead to consistent performance improvement; in some cases, it even degraded the model's quality. Nevertheless, open-source models remain advantageous in terms of cost-effectiveness, as they do not incur token-based usage fees like ChatGPT.