

Analisis Multi-Dimensional Model LLM dengan Data Sintetik Pada Pencegahan Pernikahan Dini

Ma'isy Yazid Imaaduddin¹, Rifki Wijaya², Gamma Kosala³,

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹maisyyazidx@student.telkomuniversity.ac.id,

²rifkiwijaya@telkomuniversity.ac.id,

³gammakosala@telkomuniversity.ac.id

Abstrak

Pernikahan dini merupakan isu sosial serius yang berdampak pada kesejahteraan masyarakat, dengan sekitar 12 juta anak perempuan menikah sebelum usia 18 tahun setiap tahunnya. Minimnya media edukatif yang relevan menjadi salah satu penyebab sulitnya pencegahan. Penelitian ini menggunakan data sintetik untuk mengembangkan sistem AI yang menghasilkan respons edukatif dalam konteks pencegahan pernikahan dini. Topik ini penting karena edukasi konvensional kurang efektif dalam menjangkau generasi muda. Sehingga diperlukan pendekatan berbasis teknologi yang komunikatif dan relevan. Penelitian ini menganalisis performa tiga Large Language Models (LLM) yaitu Mistral-7B-Instruct, Llama-3.1-8B-Instruct, dan Qwen2.5-7B-Instruct. Model ini dikembangkan melalui pelatihan lanjutan dengan data khusus dan penggabungan respons AI dengan informasi eksternal yang relevan untuk meningkatkan relevansi dan akurasi teks, yang dievaluasi dengan pendekatan penilaian kualitas respons berdasarkan aspek bahasa dan kesesuaian informasi. Hasil evaluasi menunjukkan bahwa Qwen2.5-7B memperoleh skor tertinggi sebesar 3.893, Llama-3.1-8B menunjukkan kekuatan dalam coherence (hingga 0.998) dan groundedness (0.999) dan Mistral-7B unggul dalam peningkatan naturalness (0.900) setelah fine-tuning. Keseimbangan antara keterbacaan, akurasi fakta, dan daya tarik percakapan menjadi kunci dalam meningkatkan kualitas keluaran asisten AI dalam mendukung edukasi sosial terkait pencegahan pernikahan dini.

Kata kunci: pernikahan dini, large language models (LLM), fine-tuning, retrieval-augmented generation (RAG), chatbot, UniEval

Abstract

Early marriage is a serious social issue that impacts the well-being of society, with around 12 million girls marrying before the age of 18 each year. The lack of relevant educative media is one of the reasons why prevention is difficult. This research uses synthetic data to develop an AI system that generates educational responses in the context of early marriage prevention. This topic is important because conventional education is less effective in reaching the younger generation. Therefore, a technology-based approach that is communicative and relevant is needed. This research analyzes the performance of three Large Language Models (LLM) namely Mistral-7B-Instruct, Llama-3.1-8B-Instruct, and Qwen2.5-7B-Instruct. These models were developed through advanced training with specialized data and fusion of AI responses with relevant external information to improve text relevance and accuracy, which were evaluated with a response quality assessment approach based on language aspects and information appropriateness. The evaluation results showed that Qwen2.5-7B obtained the highest score of 3.893, Llama-3.1-8B showed strength in coherence (up to 0.998) and groundedness (0.999) and Mistral-7B excelled in naturalness improvement (0.900) after fine-tuning. The balance between readability, factual accuracy, and conversational appeal is key in improving the output quality of AI assistants in supporting social education related to early marriage prevention.

Keywords: child married, large language models (LLM), fine-tuning, retrieval-augmented generation (RAG), chatbot, UniEval.

1. Pendahuluan

Latar Belakang

Pernikahan dini merupakan isu sosial global yang berdampak serius terhadap ekonomi, kesehatan, dan kesejahteraan masyarakat, termasuk di Indonesia. Menurut UNICEF, sekitar 12 juta anak perempuan menikah sebelum usia 18 tahun setiap tahunnya [1]. Fenomena ini dipicu oleh berbagai faktor saling terkait, seperti tekanan ekonomi, ketidaksetaraan gender, serta norma sosial yang menganggap pernikahan sebagai cara menghindari stigma “perawan tua” atau perilaku menyimpang [2]. Pernikahan dini tidak hanya merupakan keputusan individu, melainkan hasil dari dinamika sosial dan budaya yang kompleks. Oleh karena itu,

pemahaman yang mendalam mengenai faktor-faktor penyebab sangat penting untuk merancang strategi pencegahan yang efektif.

Di era digital, kemajuan dalam Kecerdasan Buatan (AI) dan Pemrosesan Bahasa Alami (NLP) telah menghasilkan model bahasa besar (LLM) berkinerja tinggi seperti Mistral-7B, Llama-3.1-8B, dan Qwen2.5-7B. Teknik seperti fine-tuning dan Retrieval-Augmented Generation (RAG) digunakan untuk mengoptimalkan model dalam tugas-tugas spesifik, di mana fine-tuning menyesuaikan model untuk tugas tertentu, sementara RAG meningkatkan relevansi jawaban dengan menggabungkan pengambilan informasi eksternal dengan generasi teks [3][4]. Meskipun model proprietary seperti GPT-3 telah terbukti efektif dalam berbagai aplikasi analisis teks dan NLP, memungkinkan eksplorasi data besar secara lebih mendalam [5]. Penelitian ini mengutamakan model open-source seperti Mistral, LLaMA, dan Qwen karena mendukung penyesuaian lokal dan reproduktibilitas. Model open-source ini bersifat open-access dan open-weight, sehingga menawarkan fleksibilitas lebih besar untuk penelitian akademik. Salah satu implementasi yang relevan adalah pengembangan chatbot, yaitu program berbasis NLP yang dapat berinteraksi dengan pengguna secara alami untuk memberikan informasi, dukungan, dan layanan otomatis [6].

Analisis multi-dimensional memberikan kerangka kerja yang komprehensif untuk mengevaluasi berbagai faktor yang mempengaruhi kinerja LLM. Dengan data sintetik, peneliti dapat membandingkan performa tiga model dalam berbagai konteks, mengidentifikasi kekuatan, kelemahan, serta responsivitas model terhadap variasi data. Implementasi fine-tuning dan Retrieval-Augmented Generation (RAG) pada model LLM memungkinkan penyesuaian spesifik tugas dan pengambilan informasi eksternal untuk meningkatkan relevansi jawaban, yang lebih teruji dengan analisis menggunakan UniEval. Melalui penelitian ini, diharapkan tidak hanya berkontribusi pada pengembangan ilmu pengetahuan di bidang AI, tetapi juga memberikan dampak positif bagi masyarakat dalam upaya melindungi hak anak dan meningkatkan kualitas hidup mereka.

Topik dan Batasannya

Penelitian ini berfokus pada analisis performa tiga model Large Language Model (LLM) dalam konteks percakapan edukatif mengenai pencegahan pernikahan dini. Batasan penelitian meliputi:

1. Penggunaan tiga model LLM, yaitu Mistral-7B-Instruct-v0.3, Llama-3.1-8B-Instruct, dan Qwen2.5-7B-Instruct.
2. Dataset yang digunakan merupakan data sintetik yang disusun untuk mensimulasikan interaksi edukatif terkait isu pernikahan dini di wilayah Lombok.
3. Evaluasi dilakukan secara otomatis menggunakan metrik Unified Multi-Dimensional Evaluator (UniEval) tanpa keterlibatan langsung dari pengguna nyata seperti remaja, pendidik, atau konselor.
4. Fokus studi terbatas pada isu pernikahan dini, sehingga temuan tidak dapat digeneralisasi untuk topik edukasi sosial lainnya.

Tujuan

Penelitian ini bertujuan untuk mengevaluasi performa tiga model bahasa besar, yaitu Mistral-7B, Llama-3.1-8B, dan Qwen2.5-7B, dalam menghasilkan respons edukatif yang relevan terhadap isu pernikahan dini. Evaluasi dilakukan melalui empat pendekatan sistem, yaitu *Base*, *Retrieval-Augmented Generation (RAG)*, *Fine-Tuning*, dan gabungan *RAG + Fine-Tuning*. Tujuan utama dari penelitian ini adalah mengetahui sejauh mana pengaruh masing-masing pendekatan terhadap kualitas keluaran model dalam aspek keterbacaan, koherensi, akurasi fakta, dan daya tarik percakapan, yang diukur menggunakan metode evaluasi multi-dimensional.

Organisasi Tulisan

Jurnal Tugas Akhir ini disusun secara sistematis dalam beberapa bagian utama yang menggambarkan alur penelitian. Setelah pendahuluan, bagian Studi Terkait menyajikan teori-teori dan penelitian sebelumnya yang mendukung topik atau pendekatan yang digunakan. Bagian Sistem yang Dibangun menyajikan perancangan dan implementasi sistem yang dikembangkan dalam studi ini. Bagian Evaluasi menyajikan hasil dan analisa hasil pengujian. Dan bagian Kesimpulan, merangkum temuan utama dari penelitian serta memberikan saran untuk pengembangan lebih lanjut dimasa mendatang.

2. Studi Terkait

Kajian ini menganalisis tiga model bahasa besar (LLM) dalam konteks pencegahan pernikahan dini menggunakan data sintetik. Penelitian Zhong et al. menyoroti pentingnya evaluasi multidimensional dalam Natural Language Generation (NLG), yang meliputi akurasi, naturalness, coherence, dan groundedness, aspek-aspek yang krusial dalam komunikasi sensitif terkait isu pernikahan dini [7]. Inan et al. menyoroti peran penting LLM dalam moderasi konten, memastikan informasi tentang pernikahan dini etis, aman, dan sesuai norma sosial [8] [9].

Selain itu, penggunaan LLM sebagai juri untuk menilai kualitas interaksi dapat membantu mengidentifikasi bias dalam penilaian dan meningkatkan relevansi respons dalam upaya pencegahan pernikahan dini. Dengan memanfaatkan LLM sebagai alat evaluasi, peneliti memperoleh wawasan yang lebih objektif