Pengambangan Tolok Ukur Evaluasi LLM dalam Tugas Klasifikasi, Ekstraksi, dan Ekstraksi Konten Definisi pada Teks Hukum = Development of LLM Evaluation Benchmarks in the Tasks of Classification, Extraction, and Content Extraction of Definitions in Legal Texts

Muhammad Ilham Ghozali, author

Abstrak

Penelitian ini mengevaluasi performa berbagai model Large Language Models (LLM) dalam tugas klasifikasi definisi, ekstraksi definisi, dan ekstraksi konten definisi dari teks hukum berbahasa Indonesia. Model yang digunakan meliputi LLaMa-2-7b, LLaMa-3-8b-Instruct, Mistral-7B, Gemma-7b, Qwen1.5-7B, dan Zephyr-7b-beta, dengan teknik prompting zero-shot dan few-shot. Teknik prompting few-shot menunjukkan efektivitas dalam beberapa kasus, namun belum konsisten meningkatkan akurasi.
.....This study evaluates the performance of various Large Language Models (LLMs) in the tasks of definition classification, definition extraction, and definition content extraction from legal texts in Indonesian. The models used include LLaMa-2-7b, LLaMa-3-8b-Instruct, Mistral-7B, Gemma-7b, Qwen1.5-7B, and Zephyr-7b-beta, with zero-shot and few-shot prompting techniques. The few-shot prompting technique showed effectiveness in some cases, but has not consistently improved accuracy.