

## ABSTRAK

*Feedback generation* adalah proses pembuatan umpan balik (*feedback*) yang bertujuan memberikan respon, saran, koreksi, atau penilaian terhadap sesuatu, misalnya teks, proses belajar siswa, atau sebuah karya. *Feedback generation* telah menjadi salah satu pendekatan penting dalam meningkatkan pengalaman belajar mahasiswa, khususnya dalam tugas terbuka yang membutuhkan evaluasi naratif. Penelitian ini bertujuan untuk mengevaluasi kualitas *feedback* yang dihasilkan oleh model *Large Language Model* (LLM), yaitu GPT-4o mini dan LLaMA 3.1, terhadap jawaban mahasiswa dari dua mata kuliah: Sistem Cerdas dan Metodologi Penelitian dan Penulisan Ilmiah (MPI). Pendekatan yang digunakan adalah *prompt engineering* dengan empat strategi *prompting* (*Zero-Shot*, *One-Shot*, *Few-Shot*, dan *Chain-of-Thought*) dan tiga jenis persona (Base, Dosen, dan Ahli). Evaluasi dilakukan dengan empat metrik otomatis: BLEU, ROUGE-L, METEOR, dan BERTScore, terhadap total 6.426 entri *feedback* yang dihasilkan. Hasil menunjukkan bahwa strategi *Few-Shot* dan *One-Shot* memberikan performa terbaik secara konsisten, sementara penggunaan persona Base menghasilkan skor tertinggi dibanding persona lainnya. Model LLaMA unggul pada aspek struktural, sedangkan GPT menunjukkan performa lebih tinggi dalam kemiripan semantik. Kombinasi dengan strategi *Few-Shot*, persona Base, dan model GPT menjadi konfigurasi paling optimal dengan hasil evaluasi sebesar 0.0795 untuk BLEU, 0.2419 untuk ROUGE-L, 0.3239 untuk METEOR, dan 0.6013 untuk BERTScore-F1. Penelitian ini menunjukkan bahwa kombinasi strategi *prompting* dan persona yang tepat dapat meningkatkan kualitas *feedback* LLM secara signifikan dalam konteks pendidikan tinggi.

**Kata kunci** : *feedback generation*, *Prompt Engineering*, LLM.