ABSTRACT

Feedback generation is the process of creating feedback that aims to provide responses, suggestions, corrections, or assessments of something, for example text, student learning processes, or a work. Feedback generation has become a critical approach in enhancing student learning, especially in open-ended assignments that require narrative evaluation. This study aims to evaluate the quality of feedback generated by two Large Language Models (LLMs), GPT-4o mini and LLaMA 3.1, in response to student answers from two university courses: Intelligent Systems and Research Methodology and Scientific Writing. The method involves prompt engineering with four prompting strategies (Zero-Shot, One-Shot, Few-Shot, and Chain-of-Thought) and three types of personas (Base, Lecturer, and Expert). Evaluation was conducted using four automatic metrics: BLEU, ROUGE-L, METEOR, and BERTScore, on a total of 6,426 feedback outputs. The results show that Few-Shot and One-Shot strategies consistently outperformed others, and the Base persona yielded the highest scores across all models. LLaMA performed better in structural metrics, while GPT was superior in semantic similarity. The combination of Few-Shot strategy, Base persona, and GPT model is the most optimal configuration with evaluation results of 0.0795 for BLEU, 0.2419 for ROUGE-L, 0.3239 for METEOR, and 0.6013 for BERTScore. This study demonstrates that the right combination of prompting strategy and persona can significantly enhance the quality of LLM-generated feedback in higher education contexts.

Keywords : feedback Generation, Prompt Engineering, LLM.