

Abstrak

Penilaian esai bahasa Inggris di Indonesia seringkali lambat dan tidak konsisten, terutama bagi pelajar *English as a Foreign Language* (EFL), karena guru kesulitan memberikan umpan balik detail pada sejumlah besar esai, seperti yang diajukan ke festival bahasa universitas. Kecerdasan buatan menawarkan solusi, namun kinerja model bahasa dalam menilai esai belum dibandingkan secara komprehensif. Studi ini menggunakan 71 esai dengan tema peluang internasional melalui bahasa sebagai masukan, menghasilkan skor, umpan balik, dan daftar kesalahan bahasa. Namun, umpan balik dari model saat ini seringkali terlalu umum, gagal memenuhi kebutuhan pembelajaran bahasa Inggris sebagai bahasa asing yang efektif. Studi ini bertujuan untuk membandingkan tiga model kecerdasan buatan dalam penilaian esai, dengan fokus pada konsistensi, kualitas umpan balik, dan deteksi kesalahan tata bahasa. Topik ini penting untuk meningkatkan efisiensi penilaian, misalnya dalam kompetisi bahasa, tetapi model saat ini seringkali tidak konsisten atau tidak jelas, menciptakan kesenjangan dengan kebutuhan pendidikan. Sebanyak 71 esai dari festival bahasa di universitas di Jawa Barat dievaluasi oleh tiga model kecerdasan buatan berdasarkan kriteria seperti tata bahasa, pilihan kata, logika argumen, gaya penulisan, dan kesesuaian konten. Sebuah prompt spesifik memandu evaluasi, dan output dianalisis menggunakan rubrik untuk konsistensi, kualitas umpan balik, dan deteksi kesalahan. Penelitian ini menghasilkan pedoman untuk memilih model terbaik dan meningkatkan evaluasi otomatis. Hasil dan kontribusi Gemini unggul dalam umpan balik berkualitas (70,42% Sangat Membantu) dan deteksi kesalahan (78,87% Sangat Bisa), diikuti oleh ChatGPT (49,30% Sangat Membantu, 54,93% Sangat Bisa), sementara LLaMA 4 konsisten (85,92% Konsisten Dengan Catatan) tetapi kurang spesifik (43,66% Tidak Bisa). Penelitian ini mendukung penggunaan kecerdasan buatan sebagai asisten penilaian esai untuk EFL di Indonesia.

Kata kunci: ChatGPT, Gemini, LLaMA, penilaian esai, kecerdasan buatan, umpan balik, kesalahan tata bahasa, konsistensi.