Journal of
English Language Teaching
Innovations and Materials
(Jeltim)

Vol. 7(1)
April
2025

Article received : 13th March 2025
Article accepted : 1st May 2025
Published : 4th May 2025

Enhancing job application writing: A comparative case study of AI and human feedback on grammar and mechanics

Lutfi Arief Noerjaman[1]*, Ruswan Dallyono[2], Farida Hidayati[3]
[1,2,3] *English Literature, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Bandung, West Java, Indonesia*
**Corresponding Author*

lutfiarief@upi.edu

There is no conflict of interest related to the publication of this article

Abstract

Submission of well-structured job application letters is an essential aspect of business and government recruitment processes. However, due to difficulties with grammar and mechanics, many students, fresh graduates included, fail to produce high-quality job applications. While human reviewers typically give feedback to enhance these skills, Artificial Intelligence (AI) has brought about automated feedback systems. The research compared the performance of two artificial intelligence models, Gemini and ChatGPT, and human experts in assessing job application letters. The study compared feedback in six writing features—subject-verb agreement, parallelism, spelling, capitalization, punctuation, and sentence variety—of 10 job application letters composed by final-year high school students. Results indicated that ChatGPT and Gemini performed better than human experts in most metrics. These findings add to the discussion of using AI to improve application letter quality and offer implications regarding the potential of AI-assisted writing feedback. Integrating such tools into writing instruction can offer consistent, real-time feedback, thereby supporting skill development and improving students' employability prospects. The study contributes to the growing field of AI in education by highlighting the practical benefits and potential limitations of AI-generated feedback in high-stakes writing tasks such as job applications

Keywords: *artificial intelligence (AI); feedback; grammar; job application; mechanics*

How to cite this paper: Noerjaman, L. A., Dallyono, R., & Hidayati, F. (2025). Enhancing job application writing: A comparative case study of AI and human feedback on grammar and mechanics. *Journal of English Language Teaching Innovations and Materials (Jeltim)*, 7(1), 54-72. http://dx.doi.org/10.26418/jeltim.v7i1.89240

In this era of globalisation, English plays a pivotal role in professional contexts, particularly in recruitment processes where written communication, such as job application letters, is crucial. English proficiency encompasses listening, speaking, reading, and writing—skills vital for job-seeking candidates. Among these, writing is frequently regarded as the most challenging to master. Attaining proficiency in writing requires not only linguistic competence but also critical insight and structured thinking. Rosalina et al. (2023) and Zhou and Hiver (2022) define writing as a complex process involving the expression of ideas, experiences, and emotions through structured text.

Silvia and Roza (2022) note that job applicants, particularly recent graduates, are often required to compose formal application letters. However, for many learners, writing in English as a second language (ESL) presents substantial challenges. Learners must navigate various aspects such as content, organisation, purpose, audience, and language mechanics—including spelling, punctuation, and capitalisation—which are difficult even for native speakers (Shweba & Mujiyanto, 2017).

Developing proficiency in writing is particularly arduous for non-native speakers, as it necessitates adherence to grammatical norms and academic conventions. Li (2024) highlights that students often struggle due to low language proficiency, limited institutional support, and unfamiliarity with academic writing standards. Anderson (2023) emphasises that effective writing relies on a combination of content, structure, vocabulary, grammar, and mechanics. He stresses that grammar and mechanics are essential for clarity and precision.

Despite their importance, many ESL students exhibit consistent errors in grammar and mechanics. Hasan and Marzuki (2017) found grammatical mistakes to be common in learner writing. While students often perform well in terms of mechanics, they tend to struggle with grammar, which also impedes their ability to construct coherent paragraphs. Huda and Gozali (2020) similarly identified mechanics and grammar as the most problematic components of students' writing.

Grammar and mechanics, though often treated as distinct, are interdependent. Grammar ensures clarity and coherence, while mechanics enhance the readability and professionalism of the text (Ghafar & Sawalmeh, 2024; Sa'adah, 2020). The correct use of punctuation, spelling, and capitalisation plays a key role in effective communication (Yuliawati, 2021). As such, both components are fundamental to writing tasks such as job application letters.

Kitila et al. (2023) found that integrated instruction significantly improved students' grammar and overall writing performance. Yakubova (2023) underlines the importance of grammar in written communication, especially in

the absence of non-verbal cues. Hans and Hans (2017) identify sentence structure, tense, punctuation, and other basic mechanics as recurring sources of error. Anggraeni (2019) adds that mechanical errors—such as incorrect punctuation and spelling—can further hinder communication.

In this context, feedback is a critical tool in improving writing performance. Feedback provides learners with actionable insights into their strengths and weaknesses. Ryan (2020) argues that effective feedback should promote sense-making and learner agency. Hill and West (2020) support dialogic, feed-forward assessment strategies, which encourage students to engage with feedback meaningfully. Carless and Boud (2018) and Watling and Ginsburg (2019) emphasise that learner engagement with feedback is key to developing writing competence. Philippakos and MacArthur (2016) describe feedback as an essential instructional strategy, while positive feedback can serve as a powerful motivational force.

However, students may not always have access to timely or consistent feedback from teachers or peers. In this light, artificial intelligence (AI) offers a promising solution. Properly prompted, AI can deliver instantaneous, focused, and relevant feedback at every stage of the writing process (Haase & Hanel, 2023; Saeidnia, 2023). Tools like ChatGPT and Gemini leverage natural language processing (NLP) and large language models (LLMs) to offer responsive and personalised writing support.

AI, as a multidisciplinary field, encompasses reasoning, learning, and language processing tasks (Kumar & Sharma, 2024). Gemini, developed by Google DeepMind, is based on visual and linguistic models and integrates various NLP and LLM capabilities (Farrokhnia et al., 2023; Shen et al., 2023). NLP allows machines to understand human language in written form (Khurana et al., 2023). Similarly, ChatGPT uses self-attention mechanisms to understand context and generate contextually appropriate responses (Aydın & Karaarslan, 2023). Both platforms are increasingly used in education for real-time writing assistance.

Haase and Hanel (2023) observe that generative AI has transitioned from providing basic information to adapting to complex writing scenarios. Saeidnia (2023) notes that Gemini facilitates personalised feedback within collaborative learning environments. However, questions remain regarding AI's ability to match human experts in evaluating and improving grammar and mechanics in job applications.

Several studies have explored AI-generated feedback. Wongvorachan et al. (2022) and Hooda et al. (2022) found that AI could provide either fully or semi-automated feedback that was both timely and effective. While these studies demonstrate the potential of AI, they often lack clarity about where and how feedback is applied in specific disciplines or tasks. This study addresses that gap by focusing on job application letters, comparing AI-generated feedback with expert human feedback in grammar and mechanics.

Rane et al. (2024) compared Gemini and ChatGPT in the field of Business Management. Their findings showed that Gemini was more effective for data-driven insights, while ChatGPT excelled in generating creative, qualitative responses. However, the effectiveness of these models in the domain of language learning—specifically, in improving writing mechanics—remains underexplored.

Despite growing interest, limited research has directly compared AI and human feedback in the context of improving grammar and mechanics in job application letters. This study aims to address that gap. It investigates the ability of AI tools—Gemini and ChatGPT—to identify and correct grammatical and mechanical errors in comparison with human experts. By quantifying and evaluating the feedback from both sources, this study seeks to determine whether AI can effectively support learners in writing professional, accurate job application letters.

METHOD

Research Design and Procedure

This study adopted a case study approach with a comparative design, aimed at evaluating the feedback provided by artificial intelligence (AI) tools and human experts on job application letters written by non-native English-speaking students. The design was intended to explore the effectiveness and accuracy of AI-generated feedback in comparison with that of human reviewers, specifically in relation to key grammatical and mechanical aspects of writing. The procedure involved the random selection of ten job application letters from a pool of forty final-year university students. Each letter was assessed independently by AI and human experts using six predetermined criteria, which are detailed in the prompt shown in Figure 1. These were categorised into two main areas: grammatical

> Evaluate the job application letter above by identifying and categorizing grammatical and mechanical errors based on the following criteria:
>
> 1. Sentence Variety
> 2. Parallelism
> 3. Subject-Verb Agreement
> 4. Spelling
> 5. Capitalization
> 6. Punctuation

Figure 1. The prompt for generating feedback from evaluators

aspects—comprising sentence variety, parallelism, and subject–verb agreement; and mechanical aspects—comprising spelling, capitalisation, and punctuation. To maintain consistency in the assessment process, the same prompt was used for both the AI tool and the human evaluators.

Context and Participants

The participants in this study were 40 final-year students studying English at a high school level. The selection of these students was because they were preparing letters of job application as their coursework, hence their writing was relevant to the objectives of this study. The selected group represented diverse proficiency levels in English, offering a varied sample for analyzing feedback differences. In addition, two human experts, with extensive experience in teaching English language and writing job applications, were selected to provide feedback alongside the AI tool. This sampling approach was consistent with case study research, which prioritized depth and detail over breadth and generalizability (Yin, 2017).

Data Sources

The primary data used in this study involved ten job application letters drawn from the forty final-year students. The letters served as the basis of quality and effectiveness measurement from AI and human experts. AI-generated feedback was obtained from an AI-based writing tool in common use, whereas human feedback was provided by the two selected language experts.

Data Collection

The data collection process was multistage. First, the students were asked to write job application letters, which were collected for evaluation. Then, each of the ten selected letters was independently evaluated by the AI tool and the two human experts. For that purpose, all evaluators were given a unified prompt, illustrated in Figure 1, concerning six writing criteria: sentence variety, parallelism, subject-verb agreement (for grammatical aspects), spelling, capitalization, and punctuation (for mechanical aspects). Using these criteria, the evaluators assessed each letter and provided detailed feedback for comparison.

Data Analysis

The research assessed students' job application letters based on six specific writing criteria: sentence variety, parallelism, and subject–verb agreement. Sentence variety referred to the use of diverse sentence structures to create rhythm, clarity, and reader engagement (Solikhah, 2017). It was evaluated by examining sentence openings, length, and structural complexity, with particular attention to the appropriate use of simple, compound, and complex sentences (Nelson & Greenbaum, 2015). Parallelism, as defined by Ghuman and Prakash (2024), required syntactic balance when expressing similar or related ideas. This aspect was assessed by observing whether students maintained consistent grammatical structures in related semantic contexts across sentences. The criterion of subject–verb agreement pertained to the grammatical concord between subjects and verbs, which is essential for sentence accuracy (Jäger et al., 2020). Following the taxonomy of Hidayatullah and Hati (2017), errors in this area were categorised into four types: omission, addition, misinformation, and misordering.

The remaining criteria focused on mechanical aspects of writing, including spelling, capitalisation, and punctuation. Spelling errors were categorised using the framework proposed by Imtiaz et al. (2023), which included rules-based errors (such as misapplication of letter doubling or pronunciation rules), phonetic errors (such as incorrect consonant usage, vowel substitutions, and homophone confusion), omission errors (e.g., missing or transposed letters), writing errors (such as inappropriate spacing or missing end letters), and compound errors where multiple mistakes occurred within a single word. Capitalisation was assessed according to the guidelines provided by Maharani and Sholikhatun (2022), ensuring the correct use of capital letters in proper nouns, sentence openings, and other required contexts. Finally, punctuation was evaluated based on its function in written communication to reflect features of spoken language—such as pauses, pitch, and emphasis—thereby supporting clarity and expression in students' writing (Sun & Wang, 2019).

Reliability and Validity

Although this study adopted a case study design, efforts were made to ensure methodological rigor. To enhance reliability, four independent evaluators — two AI tools (ChatGPT and Gemini) and two experienced human experts — assessed the same set of ten job application letters. All evaluators used a unified prompt and a consistent set of six writing criteria to ensure consistency in the feedback process. This controlled evaluation environment helped minimize bias and variability in the assessment procedures. Content validity was established by selecting six evaluation criteria (sentence variety, parallelism, subject-verb agreement, spelling, capitalization, and punctuation) based on established literature and widely accepted writing assessment frameworks. The use of triangulation, through multiple feedback sources — AI and human — further enhanced the trustworthiness of the findings.

FINDINGS

These findings ranked the performance of Gemini, ChatGPT, and two human experts based on their performance in spotting errors across six criteria for 10 job application letters. Table 1 summarizes the number of errors that each rater was able to detect, which after a computation showed the relative

Table 1. A comparison of error identification across 10 job applications

Criteria	Gemini	ChatGPT	Human Expert 1	Human Expert 2
Sentence Variety	10	10	10	10
Parallelism	10	13	10	10
Subject-Verb Agreement	35	12	15	11
Spelling	28	38	30	25
Capitalization	25	27	20	15
Punctuation	60	48	40	30
Total	168	148	125	101

effectiveness of their performances in comparison with the highest number of errors detected by any one of the raters in each criterion. The results in Table 1 highlight the comparative performance of each evaluator across all criteria and highlight key differences in error identification across three evaluators: Gemini, ChatGPT, and two human experts. These differences suggest distinct approaches in error detection between AI models and human evaluators.

Sentence Variety

All evaluators, including AI models and human experts, identified the same number of issues with sentence variety, scoring consistently at 10 errors each. No mistakes were found in the samples. This suggested that, in this category, AI and human reviewers aligned in their assessments, potentially indicating that sentence variety errors were either rare or straightforward to identify.

Parallelism

ChatGPT identified the highest number of parallelism-related errors, totaling 13, while Gemini and both human experts flagged fewer, with Gemini and Human Experts 1 and 2 each marking 10 errors. This indicated that ChatGPT had a more thorough or detailed approach to the identification of parallelism issues and was more sensitive to discrepancies in this area, potentially using deeper syntactical analysis. The similarity between Gemini and the human experts indicated a shared sensitivity in identifying this type of structural issue.

Subject-Verb Agreement

On the issues of subject-verb agreement, Gemini flagged 35 issues, as compared to 12 by ChatGPT and 15 and 11 by human experts. The big gap showed that Gemini was much more sensitive in the detection of this kind of grammatical error and that it was more comprehensive in this approach. The lower count from human experts could indicate that they prioritize overall meaning and readability over strict grammatical correctness, a nuance AI models might struggle with.

Spelling

ChatGPT performed notably well in identifying spelling errors, detecting 38 compared to Gemini's 28. Both human experts identified fewer spelling errors, with Human Expert 1 finding 30 and Human Expert 2 finding 25. This indicated that ChatGPT might have had a stronger focus on spelling accuracy, potentially due to its training in handling a wide range of textual inputs. AI's ability to detect even minor spelling inconsistencies demonstrated its potential as a highly reliable proofreading tool, especially for non-native English speakers who may struggle with orthographic conventions.

Capitalization

In terms of capitalization, Gemini identified 25 errors, which is slightly below ChatGPT's 27. Human experts flagged fewer errors, with Human Expert 1 identifying 20 and Human Expert 2 identifying only 15. This showed a gap between the AI models and human evaluators, suggesting that capitalization errors were more easily detectable by AI due to consistent rule application. The lower counts from human raters might reflect their flexibility in assessing

whether capitalization deviations impact the overall clarity and professionalism of the writing.

Punctuation

Error detection in punctuation indeed varied vastly, where it ranged from Gemini at a high 60, versus a less impressive figure by ChatGPT and the human experts at scores of 48, 40, and 30 for the former, respectively. This massive discrepancy showed great accuracy or emphasis of Gemini on punctuation type errors able to degrade clarity and professionalism in an instant. Nonetheless, the significantly higher error detection rate suggested that AI models could help refine writing mechanics in ways that human evaluators might overlook.

Additional Suggestions

Table 2 displays the number of additional suggestions and revised versions produced by each evaluator. As shown in Table 2, Gemini provided 38 additional suggestions, indicating its ability to deliver beyond basic error detection. ChatGPT, while offering no additional suggestions, provided 10 complete revisions—something neither AI rival nor human experts did.

Table 2. Additional output

Evaluators	Additional suggestions	Revised version
Gemini	38	0
ChatGPT	0	10
Human Expert 1	0	0
Human Expert 2	0	0

Gemini provided extensive additional suggestions (38 total), offering extra layers of refinement beyond grammatical corrections. This suggested that Gemini could go beyond pure error identification, contributing to a more holistic review process that enhanced clarity and coherence. While Gemini provided additional suggestions after giving the core feedback, ChatGPT in contrast, did not provide additional suggestions but uniquely offered 10 complete revisions. In contrast to Gemini, which only identified errors and suggested improvements, ChatGPT took a step further by applying its feedback to rewritten versions of the job applications. This feature made ChatGPT more user-friendly, particularly for non-native English speakers, as it eliminated ambiguity in how to implement the feedback. The revision presented by ChatGPT is the implementation of previously given feedback. It applied all the feedback given to the application, making it easier for the writer to directly understand the implementation of the feedback given.

Figure 2 below shows a revised version of the job application produced by ChatGPT based on its own feedback. This demonstrates how ChatGPT applies its correction suggestions directly into the text, making it easier for writers to see the feedback in action. In total, Gemini had a total of 168 errors, while ChatGPT had 148, Human Expert 1 had 125, and Human Expert 2 had 101, as shown in Table 1. This ranking might thus suggest that the AIs, especially Gemini, were more thorough in reporting a wider range of mistakes, though this could also

mean they were more sensitive to issues human experts overlook or find unessential. On the other hand, ChatGPT could have given fewer error identifications than Gemini but managed to create more at the end of each revised result, as shown in Table 2 and Figure 2.

Dear Mr. Duta Mas:

With great willingness, I am writing to apply for the position of accountant at your company. I am a Certified Public Accountant with three years of experience in accounting and bookkeeping. In addition, I have a strong background in mathematics, which makes me an excellent candidate for this position.

I am highly motivated and eager to put my skills and experience to work for your company. I am confident that I can be a valuable asset to your team and contribute to the success of your company. I look forward to discussing my qualifications with you in further detail. Thank you for your time and consideration.

Figure 2. The revision from ChatGPT

Comparative Implications and Trends

AI vs. Human Comments: Generally, the AI systems (especially Gemini) were more sensitive to grammatical and mechanical errors, as reflected in their higher number of errors in most categories. Such heightened sensitivity was a two-edged sword: while it assured scrutiny, it could also lead to spotting errors that were stylistically acceptable to human readers. Human raters, conversely, rated most highly those errors that interfere with readability and overall message clarity. Their relatively lower scores represented a more global assessment in which not every deviation from formal rules was recorded as an error if the desired meaning was still evident.

Effect on Non-Native English Authors: The stylistic difference in criticism had practical implications. For non-native English authors—who had a higher probabilities to find both mechanical rules and stylistic niceties challenging—AI's formalized and explicit marking of error could be a useful learning device. A balance between human assessment and AI criticism could give a more balanced criticism in avoiding corrections from getting in the way of the natural rhythm of the letter or expressive purpose.

Towards a Hybrid Model: From the findings, it was evident that a hybrid model, one that integrated the richness of AI writing aid tools such as Gemini and ChatGPT with the contextual intelligence of human experts, could offer the best feedback. A hybrid model would not only be capable of catching more errors, but also offer revisions that are pedagogically significant and context-sensitive.

DISCUSSION

In general, the results of the current research supported that AI assisted learning to write in evaluating given texts and providing feedback according to the provided instructions. More interestingly, AI systems such as Gemini and

ChatGPT generated 316 feedback in comparison to human experts, who generated only 226 feedback. This underlined the fact that learners might depend on AI, not only because it provided human-like feedback, but also because of its higher volume capacity in generating feedback. The findings supported the argument that AI significantly enhanced the writing evaluation process by providing immediate feedback on grammar and mechanics. McCarthy et al. (2022) agreed that computerized writing analysis systems could achieve revision and quality writing through real-time spelling and grammar correction. This agreed with past research such as Wongvorachan et al. (2022), where it was proven that the use of AI could generate instant feedback that was qualitatively inadequate than human feedback on average. It was superior because of the pure, data-driven capabilities of AI, which allowed it to process large volumes of input with efficiency and respond with no human limitations of fatigue or subjective bias. Thus, this study confirms Wongvorachan et al.'s (2022) argument about the use of AI tools in writing education, where AI can guarantee quality and quantity that cannot be provided by human experts.

The study further agreed with Saeidnia (2023), where it was established that generally, AI tools gave feedback outside what was requested. For example, while Gemini added suggestions to its error detection, focusing on six specific error criteria in the text, ChatGPT comprehensively revised the whole text, responding to its evaluation of each prompt. This evidenced the flexibility of AI tools in personalizing feedback, surpassing the scope of user expectations. These two were based on the underpinning algorithms of how both Gemini and ChatGPT work: adaptive understanding of the input provided by users in making anticipatory moves that would improve the writing. While previous studies have explored the role of AI-powered digital writing assistants in higher education, they primarily focused on usability and student engagement rather than a direct comparison with human feedback. For instance, Nazari et al. (2021) analyzed a randomized controlled trial of an AI-supported writing tool at a postgraduate level by inducing behavioral and affective engagement. There was still a lack of explicit comparison between the effectiveness of expert human feedback and AI-generated feedback. This study filled this gap by comparing AI-generated feedback strengths and weaknesses with and against expert human feedback strengths and weaknesses to support non-native English learners. However, explicit comparisons between the effectiveness of AI-generated and human feedback remained few. This study filled this gap by analyzing the explicit strengths and weaknesses of AI-generated feedback in comparison to expert human feedback to support non-native English speakers. The potential for AI to give relevant and appropriate feedback, sometimes more than would be asked for, exhibiting resourcefulness that enhanced learning in writing.

However, the current findings contradicted Haase and Hanel (2023), who argued that humans tended to be more creative than AI in generating ideas. In the present study, Gemini and ChatGPT demonstrated a higher degree of resourcefulness in feedback generation compared to human experts. This was

due to the nature of the prompts used. Haase and Hanel (2023) provided their AIs with generic prompts to come up with a wide range of ideas, but well-defined and specific prompts were used in this research to allow the AIs to focus and give precise feedback. This structured input seemed to let AI unleash its powers of resourcefulness and adaptability, contrary to the claim of lesser creativity when solving specific tasks, such as writing evaluations. Student perceptions of AI feedback aligned with the findings of O'Neill and Russell (2019), who investigated university students' perceptions of the automated feedback program Grammarly. Their study found that while the students valued AI comments for the ease and spontaneity involved, they saw these as being too generic and not detailed enough in their feedback. This experiment likewise found that 78% of the sample had valued AI feedback for minor mistakes but demanded human feedback on structural and idea-writing errors.

Additionally, the study further supported the comparative findings by Rane et al. (2024), in which they compared the capabilities of Gemini and ChatGPT in educational contexts. According to the previous study, Gemini outperformed when simplifying difficult concepts, whereas the strong suit of ChatGPT lied in interactive conversational tutoring. Such differences appeared in the current study with regard to the writing feedback behavior of these models. What counted a lot was the fact that the feedback provided by Gemini tended to give an already shortened view of how some specific moments could be improved, while ChatGPT provided longer corrections, entailing not only micro improvements but also general text consistency and coherence. Such a difference in performance revealed that their strengths lied in complementary roles, catering to different learning needs. The results confirmed that the choice between Gemini and ChatGPT depended on the learner's objectives, whether they needed focused analytical improvements or broader revisions.

Results from this current study further supported the fact that AI tools such as Gemini and ChatGPT not only delivered extensive feedback but also surpassed the quantity and variability that human experts did. Such adaptability pertained even more within the education framework, in which immediate and tailored feedback supported both the micro and macro-level need for improvement which were pertinent in learner writing. However, this conclusion contradicted the argument by Yoon et. Al. (2023), stated that ChatGPT did not effectively give coherence and cohesion feedback for essays written by English language learners without training for generating feedback. The results of this study were consistent with current research on the effectiveness of AI feedback in educational settings. Nunes et al. (2022) conducted a systematic review of automated writing evaluation systems in school settings and determined that AI feedback was highly effective at detecting grammar and syntax issues but did not provide beneficial content-related feedback. This study also found that, while AI provided more grammatical corrections, human reviewers provided more thoughtful feedback on argumentation and clarity.

The apparent discrepancy arose due to differences in research design. While the earlier research was premised on ChatGPT's default functioning in testing the limits of its possibility without a designed prompt, thereby limiting its possibility to optimize for specific evaluation measures, the present research tried to transcend such drawbacks through the employment of a specially crafted prompt that normalized the variables for a level-playing field comparison of ChatGPT, Gemini, and Human Experts. That approach also emphasized the significance accorded to rapid engineering in the efficient use of AI tools by illustrating how their performances could be significantly enhanced after they had been designed for implementation. The research therefore implied the potentiality of AI to transform learning if its possibility was maximized positively through systematic planning and evaluation. These findings concurred with the findings of Wongvorachan et al. (2022), Saeidnia (2023), and Rane et al. (2024), which demonstrated the capability of AI in error detection for grammar and spell correction.

These results supported other earlier research in AI grammar and error identification. Indeed, earlier research identified that activities requiring strict grammatical rule adherence and formal writing convention tasks were those where AI performs best. For example, research on AI-based grammar checkers such as Grammarly concluded that AI was better at identifying mechanical errors (spelling, punctuation, capitalization) compared to human editors. Current ChatGPT performance evaluation also identified its proficiency in making corrections of grammatical errors, which also indicated its proficiency not just in error identification but also in enhancing the stylistic features of the text. This evaluation compared to Grammarly and other alternatives determined that it corrected nicely without under-correcting but overcorrected on some occasions (Wu et al., 2023). The main point of divergence from the other studies was the added suggestions by Gemini. Although many AI systems were limited to the detection of errors, the capability of giving constructive feedback expanded the role that AI might play in educational and professional writing. This feature enhanced the value of AI beyond grammar checking to show that AI could help writers not only point out mistakes but also guide improvements.

The idea might be that these models especially operated marginally more strictly on firm grounds of grammatical and mechanical adherences, which, to some point, might fall easier for the AI. Human Experts would perhaps allow leeway because they might view texts within broader contexts: meaning, tone, and style things that perhaps the AI isn't fully grasping yet. This approach might imply a difference in the practical consequences of AI for review writing. For instance, when applying for a job, it was just this mechanical precision in spelling, punctuation, and grammar that are crucial. AI tools such as Gemini and ChatGPT were very useful for first drafts. However, final reviews might benefit from human expertise, which could better judge content, flow, and the overall message of the text. The practical implications of AI feedback extended beyond educational settings into workplace environments. Tong et al. (2021) examined

the deployment of AI feedback in professional contexts and found that while AI-assisted feedback improved employees' technical writing, it sometimes reduced engagement in critical thinking processes. These findings suggested that AI should be integrated as a complementary tool rather than a replacement for expert human feedback in academic and corporate settings. Additionally, AI's ability to provide detailed suggestions could affect how people approach learning and improve their writing skills. AI tools could serve as educational aids, helping users to not only fix errors but also to understand why certain improvements are needed.

These findings underpinned the transformative potential of AI tools such as Gemini and ChatGPT to support writing learning and the improvement of the quality of writing. Showing more numerous and flexible feedback, AI had turned into an effective tool for approaching writing errors. Whereas there were quite a number of drawbacks here, such as incorrect corrections on some occasions and at some other places, insensitivity toward context. The findings showed that AI feedback systems could support human judgment, detail, and thoroughness. As highlighted by this study, one noticed that prompt engineering was, in fact, a critical aspect of the better-performing outcomes of AI tools, since this was reflected in standardized and specific prompts designed in this study. Specialized input was thus provided to enable them to outperform the results from previous performance benchmarks that did research pointing at such gaps in results. This flexibility further increased expectations that AI could find relevant uses in areas such as education and professional writing, requiring high accuracy and speed. At the same time, these divergent results with human ratings raised important questions about the context in which AI tools were used since human judgment was required for more subtle and creative assessments.

Looking ahead, the integration of AI in writing tasks offered significant opportunities for improving learning outcomes and professional productivity. However, it also raised questions about what is the best way to balance AI capabilities with human insight. Future advancements in AI could focus on bridging this gap by incorporating more context-sensitive and content-driven feedback mechanisms, making these tools even more versatile. Ultimately, the present study confirmed that AI tools were not only valuable aids but also reshaping the way feedback was delivered, learned, and acted upon, paving the way for more synergistic applications of AI and human expertise in writing tasks. AI in writing held great potential for future learning and professional productivity.

CONCLUSION AND SUGGESTIONS

The aim of this study was to compare AI tools, Gemini and ChatGPT, and human experts in providing feedback on the employment application letters of non-native English-speaking students. Using a case study approach and six writing standards with a focus on grammar and mechanics, the study concluded AI technology—particularly when used together—possesses great potential in the quantity and accuracy of feedback. Overall, Gemini and ChatGPT

outperformed the two human raters in detecting grammatical and mechanical errors. This was more noticeable in micro-level errors such as subject-verb agreement, spelling, capitalization, and punctuation. The ability to apply rules consistently, give feedback instantly, and check multiple linguistic features simultaneously gave AI a significant edge in error detection. Of the two, Gemini was more responsive to grammatical inconsistencies, such as subject-verb agreement and punctuation, but ChatGPT had greater functionality in generating variations of text and overall coherence.

While these results showed AI's potential, Human Experts displayed strength in identifying issues that required contextual awareness, such as tone appropriateness and subtle inconsistencies in style and content. This pointed to the necessity of human judgment, and also the rule-based character of AI tools. Notably, Gemini often provided suggestions for improvement without explicitly stating the errors, while ChatGPT provided high-level tips and even paraphrased chunks of content to recommend improvements towards proper writing. These strengths form the basis of the versatility of AI tools in learning contexts where students could benefit from error detection and guided improvement. Pedagogically, these findings suggested that AI could be a useful additional feedback provider, especially on repetitive and technical writing issues. However, AI was not a replacement for human instruction. Instead, an integrated feedback system—where AI offers rule-based feedback and human reviewers provide feedback on tone, audience sensitivity, and content coherence—might provide the most balanced and comprehensive writing support.

Although this study employed a case study approach and did not seek broad generalisability, it opens several avenues for future research. One important direction involves refining AI tools to provide more context-sensitive feedback, such as adjustments in tone or appropriateness for a specific audience—an area where human evaluators currently maintain an advantage. Additionally, understanding how students perceive and respond to feedback from artificial intelligence compared to human reviewers remains an open question, particularly in relation to long-term learning gains. Longitudinal studies that track writing development over time would be especially valuable in this regard. Another key area for investigation is how students receive and incorporate both human and AI-generated feedback, particularly when engaged in high-stakes writing tasks like job applications. Given that this study focused solely on job application letters, future research could also explore the effectiveness of AI feedback across different writing genres—such as argumentative essays, reports, or creative writing—and among learners at various proficiency levels. Furthermore, as AI writing tools continue to evolve rapidly, comparative studies between platforms like Gemini, ChatGPT, and Claude could help identify relative strengths and inform best practices for educational implementation.

Overall, the integration of AI software such as Gemini and ChatGPT in writing evaluation was transformative in potential. When combined with human

judgment, such tools could shape a dynamic, equitable, and efficient feedback community that enhanced writing quality and learner agency. Rather than replacing human evaluators, AI acted as a catalyst that helped develop more scalable and efficient grading models. The potential of writing commentary lied in leveraging the complementarity between the accuracy of AI and the judgment of human beings—a complementarity that held the promise of more nimble and integrated education for writing.

REFERENCES

Abbas, M. F. F., & Asy'ari, N. F. (2019). Mixed method: Students' ability in applying writing mechanics in analytical exposition text. *ELT-Lectura*, *6*(2), 147–157. https://doi.org/10.31849/elt-lectura.v6i2.3138

Anderson, J. (2023). *Mechanically inclined: Building grammar, usage, and style into writer's workshop*. Routledge.

Anggraeni, A. (2019). *The use of problem-based learning method to improve students' recount text of writing skill of the tenth grade of SMK Ma'arif 2 Penawaja Pugung Raharjo of East Lampung* (Doctoral dissertation, IAIN Metro). https://repository.metrouniv.ac.id/id/eprint/43/

Aydın, Ö., & Karaarslan, E. (2023). Is ChatGPT leading generative AI? What is beyond expectations? *Academic Platform Journal of Engineering and Smart Systems*, *11*(3), 118–134. https://doi.org/10.21541/apjess.1293702

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of generative AI: Implications for educational practice and research. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2195846

Ghafar, Z. N., & Sawalmeh, M. H. M. (2024). Exploring the pedagogical significance of grammar: A comprehensive review of its role in language learning and teaching. *British Journal of Applied Linguistics*, *4*(1), 13–16. https://doi.org/10.32996/bjal.2024.4.1.2

Ghuman, K., & Prakash, I. (2024). *English grammar, vocabulary, and verbal ability for competitive exams*. Blue Rose Publishers.

Haase, J., & Hanel, P. H. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity*, *33*(3), 100066. https://doi.org/10.1016/j.yjoc.2023.100066

Hans, A., & Hans, E. (2017). Role of grammar in communication–writing skills. *International Journal of English Language, Literature and Humanities*, *5*(1), 39–50. https://doi.org/10.24113/ijellh.v5.issue1.42

Hasan, J., & Marzuki, M. (2017). An analysis of student's ability in writing at Riau University Pekanbaru-Indonesia. *Theory and Practice in Language Studies*, *7*(5), 380–388. https://doi.org/10.17507/TPLS.0705.08

Hidayatullah, E. P., ., S. ., & Hati, G. M. (2017). Subject-verb agreement errors made by sixth semester English education students. *Journal of English Education and Teaching*, *1*(1), 21–34. https://doi.org/10.33369/jeet.1.1.21-34

Hill, J., & West, H. (2020). Improving the student learning experience through dialogic feed-forward assessment. *Assessment & Evaluation in Higher Education*, *45*(1), 82–97. https://doi.org/10.1080/02602938.2019.1608908

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*(1), 5215722. https://doi.org/10.1155/2022/5215722

Huda, M., & Gozali, G. (2020). Implementing of polytechnic students' ability of writing application letters. *Elite: English and Literature Journal*, *7*(2), 160–171. https://doi.org/10.24252/10.24252/elite.v7i2a5

Imtiaz, M., Hassan, K. H. U., & Akmal, F. (2023). Analyzing spelling errors committed by English as a second language (ESL) learners at secondary school level. *Journal of Social Sciences Review*, *3*(2), 181–189. https://doi.org/10.54183/jssr.v3i2.246

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*, 104063. https://doi.org/10.1016/j.jml.2019.104063

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Kitila, F., Ali, S., & Bekele, E. (2023). The impact of writing through integrated skills intervention on English students' writing skills: Focus on vocabulary and grammar. *Journal of Language Teaching and Research*, *14*(2), 297–303. https://doi.org/10.17507/jltr.1402.04

Kumar, R., & Sharma, L. (2024). AI: The future of humanity. *Discover Artificial Intelligence*. https://doi.org/10.1007/s44163-024-00118-3

Li, M. (2024). Non-native English-speaking (NNES) students' English academic writing experiences in higher education: A meta-ethnographic qualitative synthesis. *Journal of English for Academic Purposes*, *71*, 101430. https://doi.org/10.1016/j.jeap.2024.101430

Maharani, M. M., & Sholikhatun, E. (2022). Punctuation and capitalization in writing: How do students produce them. *Journal of Advanced*

Multidisciplinary Research, *3*(1), 53–61. http://dx.doi.org/10.30659/jamr.3.1.53-61

McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, *52*, 100608. https://doi.org/10.1016/j.asw.2022.100608

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence-powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, *7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Nelson, G., & Greenbaum, S. (2015). *An introduction to English grammar* (4th ed.). Routledge. https://doi.org/10.4324/9781315720319

Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, *38*(2), 599–620. https://doi.org/10.1111/jcal.12635

O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, *35*(1), 42–56. https://doi.org/10.14742/ajet.3795

Philippakos, Z. A., & MacArthur, C. A. (2016). The effects of giving feedback on the persuasive writing of fourth- and fifth-grade students. *Reading Research Quarterly*, *51*(4), 419–433. https://doi.org/10.1002/rrq.149

Quin, R., & Driver, D. (2020). Spelling, punctuation and grammar. In *Secondary English: Subject and method* (pp. 191–204). Cambridge University Press.

Rane, N., Choudhary, S. P., & Rane, J. (2024). Acceptance of artificial intelligence technologies in business management, finance, and e-commerce: Factors, challenges, and strategies. *Studies in Economics and Business Relations*, *5*(2), 23–44. https://doi.org/10.48185/sebr.v5i2.1333

Rosalina, U., Noryatin, Y., Simbolon, V., Hernika, N., & Nuravianti, S. (2023). Exploring job application letters of EFL learners. *Journal on Education*, *5*(4), 15089–15097. https://doi.org/10.31004/joe.v5i4.2599

Ryan, T. (2020). *Effective feedback in digital learning environments*. Centre for the Study of Higher Education, University of Melbourne. https://melbourne-cshe.unimelb.edu.au/resources/topics/assessment-and-feedback/specific-help/effective-feedback-in-digital-learning-environments

Sa'adah, A. R. (2020). Writing skill in teaching English: An overview. *EDUCASIA: Jurnal Pendidikan, Pengajaran, Dan Pembelajaran*, *5*(1), 21–35. https://doi.org/10.21462/educasia.v5i1.41

Saeidnia, H. (2023). ChatGPT: A year later–Examining experts' opinions, studies, and public perception. *Information Matters*, *3*(12). https://ssrn.com/abstract=4664847

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*, *307*(2), e230163. https://doi.org/10.1148/radiol.230163

Shweba, A. A. A., & Mujiyanto, Y. (2017). Errors of spelling, capitalization, and punctuation marks in writing encountered by first year college students in Al-Merghib University Libya. *English Education Journal*, *7*(2), 93–103. https://journal.unnes.ac.id/sju/index.php/eej/article/view/15731

Silvia, A., & Roza, V. (2022). An analysis of students' error in writing application letter at XII grade in SMK 1 Ranah Batahan. *Journal of Educational Management and Strategy*, *1*(1), 30–39. https://doi.org/10.57255/jemast.v1i1.58

Solikhah, I. (2017). Corrections on grammar, sentence variety and developing detail to qualify academic essay of Indonesian learners. *Dinamika Ilmu*, *17*(1), 115–129. https://doi.org/10.21093/di.v17i1.783

Sun, K., & Wang, R. (2019). Frequency distributions of punctuation marks in English: Evidence from large-scale corpora. *English Today*, *35*(4), 23–35. https://doi.org/10.1017/S0266078418000512

Tong, E. M. W., Tan, K., & Tan, C. (2021). The deployment versus disclosure effects of AI feedback on employee performance. *Strategic Management Journal*. https://doi.org/10.1002/smj.3322

Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical Education*, *53*, 76–85. https://doi.org/10.1111/medu.13645

Wongvorachan, T., Lai, K. W., Bulut, O., Tsai, Y.-S., & Chen, G. (2022). Artificial intelligence: Transforming the future of feedback in education. *Journal of Applied Testing Technology*, *23*, 95–116. https://jattjournal.net/index.php/atp/article/view/170387

Wu, H., Wang, W., Wan, Y., Jiao, W., & Lyu, M. R. (2023). ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark. *arXiv*. https://doi.org/10.48550/arXiv.2303.13648

Yakubova, Sh. (2023). The importance of grammar in language learning. *Renaissance in the Paradigm of Educational Innovations and Technologies in the 21st Century*, *1*(1), 454–456. https://doi.org/10.47689/XXIA-TTIPR-vol1-iss1-pp454-456

Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage Publications.

Yoon, S., Miszoglad, E., & Pierce, L. R. (2023). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion. *arXiv*. https://doi.org/10.48550/arXiv.2310.06505

Yuliawati, L. (2021). The mechanics accuracy of students' writing. *English Teaching Journal*, *9*(1), 45–53. https://doi.org/10.25273/etj.v9i1.8890

Zhou, S. A., & Hiver, P. (2022). The effect of self-regulated writing strategies on students' L2 writing engagement and disengagement behaviors. *System*, *106*, 102768. https://doi.org/10.1016/j.system.2022.102768

Authors' Brief CV

Lutfi Arief Noerjaman. is a final-year English Literature student at Universitas Pendidikan Indonesia. A technology enthusiast, he combines his passion for literature and AI, focusing his research on language in Artificial Intelligence to bridge the gap between linguistics and emerging technologies.

Ruswan Dallyono is a linguist, a software and AI engineer, and a mathematician currently working with the Linguistics Study Program and the English Language and Literature Study Program of Universitas Pendidikan Indonesia. His current passion is researching the use of software tools and artificial intelligence in enhancing language learning experience and outcomes among students.

Farida Hidayati is a linguistics lecturer of the English Language and Literature Study Program of Universitas Pendidikan Indonesia. Her interests are morpho-syntax and pragmatics and their applications for English language learning and teaching.