English Language Teaching Educational Journal
Vol. 7, No. 2, August 2024, pp. 90-108
ISSN 2621-6485

OpenAI ChatGPT vs Google Gemini: A study of AI chatbots' writing quality evaluation and plagiarism checking

Siraprapa Kotmungkun [a,1], Wichuta Chompurach [b,2,*], Piriya Thaksanan [c,3]

[a, b, c] Kasetsart University Chalermphrakiat Sakon Nakhon Province Campus, 59/6 Moo 1, Vor Por Ror Or 366 Road, Chiang Khruea, Sakon Nakhon, Thailand
[1] siraprapa.ko@ku.th ; [2] wichuta.c@ku.th*; [3] peereeya.t@ku.th

ARTICLE INFO

Article history
Received 14 September 2024
Revised 21 September 2024
Accepted 24 October 2024

Keywords
AI Chatbot
L2 writing
Plagiarism
Writing Evaluation
Google Gemini
OpenAI ChatGPT

ABSTRACT

This study explores the writing quality of two AI chatbots, OpenAI ChatGPT and Google Gemini. The research assesses the quality of the generated texts based on five essay models using the T.E.R.A. software, focusing on ease of understanding, readability, and reading levels using the Flesch-Kincaid formula. Thirty essays were generated, 15 from each chatbot, and evaluated for plagiarism using two free detection tools—SmallSEOTools and Check-Plagiarism—as well as one paid tool, Turnitin. The findings revealed that both ChatGPT and Gemini performed well in terms of word concreteness but demonstrated weaknesses in narrativity. ChatGPT showed stronger performance in referential and deep cohesion, while Gemini excelled in narrativity, syntactic simplicity and word concreteness. However, a significant concern was the degree of plagiarism detected in texts from both AI tools, with ChatGPT's essays exhibiting a higher likelihood of plagiarism compared to Gemini's. These findings highlight the potential limitations and risks associated with using AI-generated writing.

How to Cite: Kotmungkun, S., Chompurach, W., & Thaksanan, P. (2024). OpenAI ChatGPT vs Google Gemini: A study of AI chatbots' writing quality evaluation and plagiarism checking. *English Language Teaching Educational Journal, 7* (2), 90-108. https://doi.org/10.12928/eltej.v7i2.11572

1. Introduction

Over the past decade, it has become the new era of artificial intelligence (AI). The term "artificial intelligence" (AI) was coined by McCarthy et al. in 1955 to describe machines and processes that can mimic human intelligence and make decisions like humans (McCarthy et al., 1955; Zhai, 2023). AI has increasingly integrated into our everyday lives and routine activities (Lee & Park, 2023). The development of AI models, such as OpenAI ChatGPT (Generative Pre-trained Transformer) and Google Gemini, is one of significant technological breakthroughs that has attracted a lot of attention in various fields including healthcare, education, research, journalism, and industry (Evans et al., 2023; Zhai, 2023). This was because of the high quality of their outputs across a wide range of topics and their language and content accuracy (Stahl & Eke, 2023). Moreover, since the particular AI models use natural language (NL) text generation, the models can generate "human-sounding" (p. 107) texts in different languages, from a word level to a paragraph level, on almost any topic and in any writing style (Jovanovic, 2022).

In terms of education, especially in EFL (English as a Foreign Language) writing classrooms, technologies have been used for various purposes. For example, machine translation i.e. Google Translate has been used in English writing classrooms because it was viewed as a helpful, reliable tool helping learners complete their writing tasks (Chompurach, 2021). Also, AI technologies have been used in completing writing tasks because of their benefits that were time-saving and easy to use as well as access (Suaverdez & Suaverdez, 2023). Therefore, the particular technologies have been used in L2 writing from brainstorming to draft editing (Lingard, 2023); moreover, learners could receive their writing skills' strengths, weaknesses, and concerning comments from AI (Barrot, 2023). However, there have been ethical considerations in the use of AI technologies in writing that were authorships, the potential for biases and discrimination, and the risk of plagiarism (Suaverdez & Suaverdez, 2023).

1.1. Large Language Models: OpenAI ChatGPT and Google Gemini

In November 2022, OpenAI, a research laboratory focused on artificial intelligence, released ChatGPT (Tlili et al., 2023). ChatGPT, an OpenAI-developed web-based AI system that utilizes GPT large language models (LLM), is projected to revolutionize numerous facets of society (Stahl & Eke, 2023). According to OpenAI, ChatGPT dialogue format enables it to respond to follow-up questions, admit errors, challenge faulty assumptions, and refuse inappropriate requests (OpenAI, 2022; Zhai, 2023). In terms of ChatGPT3.5, it is the first version of ChatGPT which is a large language model from OpenAI. It is built on top of InstructGPT, which is a GPT3.5 model that has been fine-tuned on a large dataset of conversational text. ChatGPT3.5 has also been fine-tuned using reinforcement learning from human feedback, which helps it to generate text that is more aligned with human preferences (Espejel et al., 2023).

In December 2023, the evolution of Google's Large Language Models (LLMs) called Google Bard took a significant step forward with the advent of Large Multimodal Models (LMMs) called Google Gemini. For Bard, it has already made a remarkable impact. This means users can access Bard through an easy-to-use interface, which allows them to collaborate directly with a generative AI system (Espejel et al., 2023). For Gemini, LMMs seek to build upon the strengths of LLMs by incorporating multi-sensory processing capabilities. This expansion aims to achieve more robust and general intelligence, enabling computers to interact with humans on a level closer to how humans naturally communicate with each other (Yang et al., 2023 as cited in Lee et al., 2023).

As mentioned earlier, generative AI (GenAI) is used in various fields including education. This is because GenAI is a powerful tool that can be used to enhance students' learning experience in higher education and can respond to user prompts to generate highly original outputs, which can be used for a variety of educational purposes (Chan & Hu, 2023) including second language (L2) writing (Barrot, 2023).

1.2. AI Large Language Models Use in Education Settings

During the first decade of the 21st century, significant progress in machine learning and natural language processing enabled the development of more sophisticated AI-powered educational tools (Zhai, 2023). For example, these tools could tailor their approach to the individual needs of each learner, provide customized instruction, and even assess assignments (Zhai, 2023). With these various benefits, many scholars, instructors, and educators have paid attention to both benefits and drawbacks of these models to education settings.

In 2023, the study of Baidoo-Anu and Ansah was conducted to explore how interactive ChatGPT was and what its potential benefits and drawbacks on education were. In doing so, the researchers analyzed the research articles on ChatGPT from November 2022 - March 2023; the data revealed both these LLMs' pros and cons. On the one hand, ChatGPT was viewed as a personal tutor, an automated essay evaluation, a translator, an interactive virtual tutor, and adaptive learning provider. On the other hand, these LLMs were perceived as a virtual tutor lacking human interaction, a limited-understanding system, a biased model, a less creative chatbot, a model depending on trained data, a limited contextual understanding system, and a limited ability instructor.

These findings in terms of the LLM benefits are in line with the study of Evans et al (2023). Evans et al (2023) have conducted a study to explore the impacts of ChatGPT on access-efficiency, employment, and education. The findings revealed ChatGPT can be used to improve access and efficiency in a variety of industries and for individuals with disabilities or language barriers. ChatGPT

can be used to create translation tools that can help non-native speakers understand and communicate in a foreign language. For employment, ChatGPT created new job opportunities in AI and NL processing but also raising concerns about job displacement. Moreover, for education, ChatGPT can be used to improve language learning and personalized education. It can provide instant feedback, explanations, and examples to help learners improve their grammar, vocabulary, and comprehension. Additionally, ChatGPT can be used to enhance the accessibility and inclusivity of education by providing support for students with disabilities or language barriers. This research showed ChatGPT can help to break down language barriers and give everyone equal access to information, services, and education.

Similarly, Tlili et al (2023) also found that ChatGPT helped users by providing clear and concise explanations of complex topics, using language that was appropriate for the target audience. The researchers have conducted a qualitative research study to investigate chatbots in education. To do this, the study focused on ChatGPT as an example of an advanced chatbot that was being used by early adopters. The findings of the study indicated that ChatGPT has the potential to transform education in a variety of ways. The majority of participants in the study indicated that ChatGPT was effective in increasing the likelihood of educational success by providing users (teachers and students) with basic knowledge of various topics. Furthermore, ChatGPT was acknowledged by the participants as an efficient tool for providing information of a wide range of complex topics in a clear and concise language. Conversely, a few participants held the opposing view that the misuse of ChatGPT by learners could also diminish their innovative capacities and critical thinking skills.

In terms of university students' voices on Gen AI technologies including ChatGPT and Google Bard, Chan and Hu (2023) have conducted a survey with closed-ended and open-ended questions to explore 399 university students' use and perceptions toward GenAI technologies mentioned earlier. The findings showed the students understood the limitations of GenAI abilities, but the overall students still had positive attitude toward these technologies. Moreover, the students found these GenAI technologies were great tools because of 24/7 availability and the technologies can help them save time. However, they had got concerns about the way these technologies can undermine the university education value.

From previous studies on the impacts of the LLMs on education, the findings showed ChatGPT's both benefits and drawbacks. On the one hand, ChatGPT still lacked human interactions (Baidoo-Anu & Ansah, 2023) and could reduce users' innovative and critical thinking ability (Tlili et al., 2023). On the other hand, ChatGPT demonstrated exceptional competencies in various fields including machine translation (MT) (Baidoo-Anu & Ansah, 2023; Evans et al., 2023; Tlili et al., 2023). Students' voices, however, showed that these GenAI technologies were great tools (Chan & Hu, 2023). Nevertheless, the potential impacts of ChatGPT and the other chatbot powered by Google called Gemini on education remain less explored. The following section delves into the use of LLMs in L2 writing, examining how these AI tools have been utilized and their potential implications for language learning and teaching.

1.3. Use of Large Language Model in L2 Writing

Writing is one of the most essential skills that L2 learners have to develop (Ariyanti, 2016; Hyland, 2003; Meiranti, 2012), and it is a complex linguistic skill that requires cognitive processing to effectively convey thoughts and emotions in a written form (Hasnawati et al., 2023). Through this writing activity, L2 students were expected to master grammars, vocabulary words, punctuations (Meiranti, 2012), and they should take these aspects into account that are a clear intention, a well-organized framework, appropriate word selection, effective language use, and coherent thought expression (Hyland & Jiang, 2017). In terms of essay writing models, Ariyanti's study (2016) provided definitions of each essay type.

1. Descriptive Essay: Provide a clear and detailed portrayal of the topic to effectively convey its essence;
2. Classification Essay: Provides concise definitions of their characteristics, allowing readers to understand the distinctions between them clearly;
3. Cause and Effect Essay: Employs the appropriate vocabulary and expressions that effectively convey the relationship between causes and their corresponding effects;

4. Comparison and Contrast Essay: Goes beyond identifying similarities and differences by structuring comparisons effectively and employing appropriate clauses of comparison, contrast, and concession;
5. Argumentative Essay: Focuses on presenting well-reasoned arguments, supported by credible evidence, with careful evaluation of argument strength.

With the complexity of L2 writing and the benefits of AI, several educators have adopted these technologies into their writing classrooms. This following research has reported the use of AI technologies in L2 writing classrooms. The study of Fitria (2023) reported the benefits, especially on English writing. In her research, Fitria (2023) studied how ChatGPT wrote English essays. Adopting a qualitative research paradigm, the researcher collected data from documents and observation and adopted Miles et al's (2018) three steps including data reduction, data display, and conclusion to analyze the obtained data. The data revealed ChatGPT can address a wide range of prompts, including English essays encompassing descriptive narratives on any topics. ChatGPT adhered to the conventions of event sequencing and essay structure, employing both active and passive voice constructions while it also maintained appropriate tense usage aligned with the essay topic.

For academic writing, Mohammed et al (2023) conducted a research study to explore the utilization of ChatGPT among Arab postgraduate students in India, with a particular focus on its impact on their academic writing. Through a questionnaire administered to 40 postgraduates enrolled in five universities across Kerala, the study revealed the perceived benefits and tangible effects of ChatGPT on their academic writing, language proficiency, and overall academic performance. The findings reported more than a half of the participants (n=29) were familiar with ChatGPT. In terms of the potential benefits of ChatGPT use, the highest mean score (mean= 3.07) was for the statement saying "ChatGPT texts are highly reliable and need not be second checked for reliability" (p. 11). Also, the data reported that ChatGPT was viewed as an effective tool for variety tasks because it helped them generate ideas, gather information, and enhance their writing skills. Therefore, most of the students used ChatGPT in these tasks including researching topics, translating texts, and writing assignment.

In terms of second language (L2) writing, Barrot (2023) has explored the several advantages and challenges of ChatGPT. In the study, the researcher reviewed ten studies and suggested the potential benefits of the integration of this technology in L2 writing classrooms. The data revealed since ChatGPT enabled users to revise their requests and rejected any unsuitable ones, L2 learners could draw upon these features to conduct sustained and coherent conversations without the anxiety of being criticized. Also, ChatGPT provided extensive range of writing assistance, from simple to complex tasks. It could produce human-like texts which generally exhibited coherence and grammatical accuracy. This could make it a beneficial resource for users to enhance their writing skills and refine their language usage. Moreover, ChatGPT could generate outlines in various formats, including sentence, topic, alphanumeric, and decimal system structures. This feature provided L2 students with access to both a guiding framework and a foundation for crafting a well-structured outline tailored to their chosen topics. In terms of feedback, ChatGPT provided users their writing strengths, weaknesses, and specific comments by automatically grading following predefined criteria. Therefore, instead of completely prohibiting ChatGPT, educators can investigate strategies for integrating these AI-powered tools into their teaching practices and harness their benefits (Barrot, 2023).

In 2023, Suaverdez and Suaverdez (2023) have conducted a study to examine writing texts or papers by AI generators and test them with AI detectors. The researchers collected writing texts from different sources that were an original man-made paper, papers from AI chatbots that were ChatGPT (both free and paid), Phin (free), and Moonbeam (free). The results showed from a free tool called "Contentatscale", the paper from ChatGPT (free) received the highest percentage (96%) and the paid one received 91% interpreted as "likely to be human" (p. 6). Moreover, OpenAI classifier (free) reported that papers from both free and paid ChatGPT were "very unlikely AI-generated" (p. 6) while ones from Phin (free) and Moonbeam (free) were reported "unclear if it is AI- generated" (p. 6). However, Origibnality.AI detected that papers from both free and paid ChatGPT and Phin (free) were 100% AI while one from Moonbeam (free) were 99% AI. In terms of plagiarism checking from the paid checker, the paper from Phin received the highest percentage (12%) of plagiarized content while ones from ChatGPT free and paid and Moonbeam received 8%, 1% and 0%, respectively. For the free checker, the paper from Moonbeam received the highest percentage (3%) of plagiarized content while ones from ChatGPT paid and free and Phin received 1%, 0% and 0%, respectively. The researchers

also recommended careful evaluation, strategic planning, and establishment of guidelines and regulations for the utilization of AI tools in academic writing and school tasks to minimize the potential negative consequences.

Reviewing the literature, the researchers found some mixed messages. On the one hand, the potential benefits of ChatGPT to English or L2 writing have been reported (Barrot, 2023; Fitria, 2023; Mohammed et al., 2023). On the other hand, there have been worries and concerns about the ethical issues of the AI-model use in L2 writing (Suaverdez & Suaverdez, 2023). Moreover, there has been limited studies exploring Gemini's benefits and drawbacks with using in L2 writing. The present research aims to explore and compare the writing quality of ChatGPT and Google Gemini with the Text Ease and Readability Assessor (T.E.R.A.). The researchers chose to focus on these two tools, ChatGPT and Google Gemini, as they are at the forefront of language model innovation (Rane et al., 2024). Both platforms demonstrate exceptional capabilities in engaging in complex conversations, producing textual content, and generating a wide variety of creative outputs, making them ideal for a comprehensive analysis of AI-generated writing innovation (Rane et al., 2024). Also, three plagiarism checkers, TurnItIn and the other free plagiarism checkers: check-plagiarism.com and smallseotools.com were applied to detect the percentages of plagiarized content of each chatbot.

Research Questions

1. What is the writing quality level of OpenAI ChatGPT and Google Gemini based on five essay models in terms of the easability and readability?
2. What is the percentage of plagiarized content in the generated texts from OpenAI ChatGPT and Google Gemini as detected by plagiarism checkers?

2. Methods

The present study has been conducted following the multi-method qualitative approach (Davis et al., 2011). To do so, the researchers can acquire in-depth understanding regarding the quality of AI chatbots', ChatGPT and Google Gemini, L2 generated writing tasks.

2.1. Data Collection Tools

1) Basic Prompts to Gather Data from AI Chatbots

To enable both AI chatbots to provide expected responses, the researchers adopted the prompt engineering technique from Rayhan (2023). Prompt engineering is an essential tool for refining the accuracy and authenticity of AI language models because it can be crafted to align with the objectives of specific tasks or domains (Rayhan, 2023). To design effective prompts, the researchers have adhered to the design principles for basic prompts which consist of a few words or a short sentence and can be either generalized or task-specific. In doing so, the researchers can ensure that the AI language model generates accurate and relevant outputs for a wide range of tasks and users. The Table 1 shows the basic prompts being used in the present study based on the types of essay models.

Table 1. Basic Prompts Used in the Present Study

Essay Models	Basic Prompts
Descriptive	Write a 300-word essay describing the most beautiful place in Thailand.
Classification	Write a 300-word essay classifying genres of music.
Cause and Effect	Write a 300-word essay about causes and effects of global warming.
Comparison and Contrast	Write a 300-word essay about living in a city vs. living in the country.
Argumentative	Write a 300-word essay about the benefits of social media.

As a result, Table 2 shows the data the researchers obtained.

Table 2. The Overall Data Obtained from the Two AI Chatbots

AI Chatbots	Essays Written by AI Chatbots	Word Count	Average Word Count by Writing Type
ChatGPT	ChatGPT Descriptive 1	357	359.33
	ChatGPT Descriptive 2	368	
	ChatGPT Descriptive 3	353	
	ChatGPT Classification 1	370	333.33
	ChatGPT Classification 2	321	
	ChatGPT Classification 3	309	
	ChatGPT Cause and Effect 1	314	320.67
	ChatGPT Cause and Effect 2	318	
	ChatGPT Cause and Effect 3	330	
	ChatGPT Comparison and Contrast 1	339	331
	ChatGPT Comparison and Contrast 2	412	
	ChatGPT Comparison and Contrast 3	440	
	ChatGPT Argumentative 1	330	330.67
	ChatGPT Argumentative 2	326	
	ChatGPT Argumentative 3	337	
Average Word Count by ChatGPT		335	
Gemini	Gemini Descriptive 1	416	398.67
	Gemini Descriptive 2	367	
	Gemini Descriptive 3	413	
	Gemini Classification 1	426	399.33
	Gemini Classification 2	382	
	Gemini Classification 3	390	
	Gemini Cause and Effect 1	339	397
	Gemini Cause and Effect 2	412	
	Gemini Cause and Effect 3	440	
	Gemini Comparison and Contrast 1	383	385
	Gemini Comparison and Contrast 2	393	
	Gemini Comparison and Contrast 3	379	
	Gemini Argumentative 1	428	435.33
	Gemini Argumentative 2	481	
	Gemini Argumentative 3	397	
Average Word Count by Gemini		403.07	

2) T.E.R.A. (The Text Ease and Readability Assessor) Program

After gathering data from the AI chatbots, the researchers are applying the T.E.R.A. program (McNamara et al., 2014) to assess the generated texts in terms of their easability and readability. To do so, the program analyzes text on five components and Flesch-Kincaid Grade Level.

1. Narrativity: Generally, more story-like texts have higher narrativity scores, making them easier to read, though exceptions exist.
2. Syntactic Simplicity: Higher simplicity scores occur in texts with fewer clauses and words per sentence, and greater structural similarity within paragraphs.
3. Word Concreteness: Texts with more concrete words are easier to read, while abstract words lower readability.
4. Referential Cohesion: Overlapping words or concepts between sentences improve cohesion, though low cohesion may sometimes prompt deeper reader engagement.
5. Deep Cohesion: Refers to the seamless connection of ideas through varied use of connecting words.
6. Flesch-Kincaid Grade Level: This formula estimates readability, providing valuable insights for educational purposes.

Each of these components is evaluated for a specific text by comparing it to a massive collection of other texts in a corpus (Graesser, McNamara, & Kulikowich, 2011 as cited in McNamara et al., 2014). Also, in addition to providing a concise overview, T.E.R.A. also determines the grade level of the text employing the Flesch-Kincaid Grade Level readability formula (Kincaid, Fishburne, Rogers,

& Chissom, 1975 cited in McNamara et al., 2014). Therefore, by adopting the particular program, the researchers could obtain the data concerning the easability and readability of AI chatbots' generated texts following each essay model type.

3) Plagiarism Checkers

To examine the plagiarism percentage of the obtained data, the researchers applied the three plagiarism checkers: one is paid (TurnItIn) and the other two are free (Check-Plagiarism and SmallSEOTools' plagiarism detector). For TurnItIn, it is a paid plagiarism checker that compares submissions against over 47 billion web pages and premium academic content in multiple languages, providing detailed categorized matches for high-stakes writing (Turnitin, n.d.). The two free tools, Check-Plagiarism (check-plagiarism.com) and SmallSEOTools' plagiarism detector (smallseotools.com), offer reliable detection services for students and writers. Check-Plagiarism conducts a deep search to identify potential plagiarism, while SmallSEOTools scans text against billions of webpages for thorough analysis (Small SEO Tools, n.d.; Check Plagiarism, 2024).

2.2. Research Procedures

The researchers conducted a pilot study to determine whether the basic prompts yielded relevant data. The prompts were then refined and used with both OpenAI ChatGPT and Google Gemini, ensuring consistency by using identical prompts for each chatbot. The 30 essays generated were evaluated using the T.E.R.A. writing evaluation program, and plagiarism detection was performed using Turnitin, Check-Plagiarism, and Small SEO Tools. Finally, qualitative document analysis was applied to address the two research questions.

2.3. Data Analysis

To analyze the obtained data, the researchers adopted document analysis. Document analysis entails a systematic examination and interpretation of data to extract meaning, enhance comprehension, and acquire empirical knowledge (Corbin & Strauss, 2008 as cited in Bowen, 2009; Rapley, 2007 as cited in Bowen, 2009). This study adopted document analysis to examine data, specifically the assessed results from the T.E.R.A. program and plagiarism checker outputs. Document analysis, as described by Corbin and Strauss (2008 as cited in Bowen, 2009), involves a process of skimming, thorough reading, and interpretation. This method incorporates elements of content analysis, organizing data into categories that align with the research questions. (Corbin & Strauss, 2008 as cited in Bowen, 2009; Strauss & Corbin, 1998 as cited in Bowen, 2009). The documents include background papers, books, diaries, event programs, program proposals, scripts, reports, survey data, and public records (Bowen, 2009), for example. Therefore, the researchers were able to interpret and assess the quality of the texts generated by the AI chatbots, OpenAI ChatGPT and Google Gemini. Additionally, the researchers were able to determine the percentage of plagiarized content produced by the AI chatbots.

3. Findings and Discussion

The results of the present research are reported following the two research questions.

3.1. RQ 1: What is the writing quality level of OpenAI ChatGPT and Google Gemini based on five essay models in terms of the easability and readability?

To report the results concerning the writing quality of the two AI Chatbots based on five essay models in terms of the easability and readability, the researchers divided them into three parts which are as follows:

1) Overall Analysis of Essays Generated by Two AI Chatbots

From Table 2, there were 30 writing essays – fifteen generated by ChatGPT and the others generated by Gemini. All essays were generated in March 2024. The overall word count in average is 369 words. Comparing ones from ChatGPT to Gemini's, the researchers found when given the prompt to write a 300-word essay about each writing type, the average word count of writing generated by ChatGPT was closer to the specified number. This means that the average word count of the 15 essays by ChatGPT is 335 words, with a minimum of 309 words and a maximum of 370 words. For the average word count of essays generated by Gemini, it was 403 words, with a minimum of 339 words and a maximum of 481 words. These findings indicate that while both AI models were capable of

generating essays in response to the same prompts, ChatGPT was more precise in meeting the specified word count, whereas Gemini generated more verbose content.

2) Easability and Readability of Essays Generated by Two AI Chatbots

The data obtained from the T.E.R.A. program (McNamara et al., 2014), used to assess the easability and readability of 30 writing essays generated by the two AI chatbots, showed the following results. Among the five components (narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion), the essays generated by both ChatGPT and Gemini received the highest percentage for word concreteness and the lowest percentage for narrativity. On average, the 15 essays generated by ChatGPT had a maximum of 75% for word concreteness and a minimum of 3% for narrativity. In contrast, the 15 essays generated by Gemini had a maximum of 80% for word concreteness and a minimum of 4% for narrativity. Higher scores indicate that the essays were easier to read. The findings provide valuable insights into the linguistic and structural characteristics of AI-generated content. Both chatbots scored highest in word concreteness and lowest in narrativity, indicating that while the essays were generally easy to understand, they lacked storytelling or narrative elements. This suggests that both AI models prioritize clarity and straightforward language over engaging, narrative-driven writing.

Comparing these two AI chatbots, Gemini reported higher scores in narrativity (4%), syntactic simplicity (69%), and word concreteness (80%). This indicates that Gemini's writing was slightly more narrative-driven, easier to understand syntactically, and used more concrete language. These qualities could make Gemini's content more suitable for users seeking direct and easily digestible information. In contrast, ChatGPT achieved higher scores in referential cohesion (14%) and deep cohesion (51%), suggesting that its essays were better at linking ideas and maintaining logical consistency throughout the text. This means that ChatGPT's content might be more effective for complex topics that require a clear flow of ideas and a well-structured argument. For narrativity and deep cohesion, the two AI chatbots had very slight differences in scores—only a 1% difference. The minimal difference in scores suggests that while there were distinctions in their linguistic capabilities, the gap was not significant. Therefore, both tools could be considered relatively close in terms of narrative and cohesive quality.

Table 3. Means of Text Ease and Readability Scores Coh-Metrixx Component Scores

AI Chatbot	Word Count	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion	Flesch-Kincaid Grade Level
ChatGPT	335	3%	55%	75%	14%	51%	14
Gemini	403	4%	69%	80%	5%	50%	13

Besides analyzing the easability and readability of the texts, the T.E.R.A. program also reported the Flesch-Kincaid Grade Level to indicate the reading levels (basic, average, and advanced) of the analyzed texts. These levels were aligned with school levels (kindergarten, elementary, middle school, high school, college, and post-grad) and the corresponding age ranges of the readers (5-8, 8-11, 11-14, 14-17, 17-20, and 20). The essays generated by ChatGPT had Flesch-Kincaid Grade Level scores ranging from 12 (minimum) to 17 (maximum), with an average score of 14. This means that these essays were at an advanced reading level, corresponding to college to post-grad school levels, and targeted readers aged 17-20. On the other hand, the essays generated by Gemini showed a slightly wider range of scores, with grade level scores ranging from 10 (minimum) to 16 (maximum) and an average of 13. This means that these essays were at average to advanced reading levels, corresponding to high school to post-grad school levels, and targeted readers aged 14-20. The analysis shows that both ChatGPT and Gemini generated essays aimed at readers with advanced reading abilities, but with some differences. ChatGPT's content was more complex, suited for readers with high literacy skills, making it ideal for advanced learners. In contrast, Gemini produced content accessible to a broader audience, ranging from high school to college students, making it suitable for a wider range of educational settings. These differences highlight ChatGPT's focus on higher-level language and Gemini's flexibility in accommodating various reading levels.

3) Easability and Readability of Essays Generated by Two AI Chatbots Based on Five Essay Models

The analysis of the easability, readability, and reading level of texts generated by ChatGPT and Gemini for each essay model is shown in Tables 4 and 5.

The results of the texts generated by ChatGPT are shown in Table 4 as follows. Firstly, in narrativity, the ChatGPT argumentative essays achieved the highest percentage (5%) while the lowest percentages were found in two essay models: classification (2%) and cause and effect (2%). Secondly, in syntactic simplicity, the highest percentage was achieved by ChatGPT's Comparison and Contrast essays (74%) whereas the descriptive essays had the lowest percentage at 26%. Thirdly, in word concreteness, the descriptive essays obtained the highest percentage at 99% while the cause and effect ones had the lowest at 59%. Fourthly, in referential cohesion, the highest percentage was achieved by the classification model (31%) while the comparison and contrast essays had the lowest at only 4%. For the last component, deep cohesion, the highest percentage was found in the argumentative writing (86%) while the lowest percentage was found in the comparison and contrast essays (19%). The analysis of Flesch-Kincaid Grade Level shows that "ChatGPT Classification" and "ChatGPT Argumentative" essays achieved the highest score of 16 while "ChatGPT Descriptive" had the lowest score of 12.

Table 4. Mean Values of Each Essay Model Generated by ChatGPT

Comparison of Five Essay Models	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion	Flesch-Kincaid Grade Level
Descriptive	4%	26%	99%	15%	41%	12
Classification	2%	50%	91%	31%	58%	16
Cause and Effect	2%	59%	59%	11%	54%	15
Comparison and Contrast	3%	74%	64%	4%	19%	13
Argumentative	5%	67%	61%	8%	86%	16

The results of the texts generated by Gemini are shown in Table 5 as follows. For the first component, narrativity, "Gemini Argumentative" essays had the highest percentage at 10% while the lowest percentage (3%) was found in three essay models: "Gemini Descriptive", "Gemini Cause and Effect", and "Gemini Comparison and Contrast". Secondly, in syntactic simplicity, both "Gemini Classification" and "Gemini Cause and Effect" had the highest percentage at 76% while "Gemini Descriptive" and "Gemini Argumentative" had the lowest percentage at 61%. Thirdly, in word concreteness, the highest percentage was found in "Gemini Descriptive" (97%) whereas "Gemini Argumentative" had the lowest at 64%. Fourthly, in referential cohesion, the essay model with the highest percentage was "Gemini Argumentative" (16%) while "Gemini Comparison and Contrast" had the lowest at only 1%. For the last component, deep cohesion, the highest percentage was shown by "Gemini Argumentative" (87%) while "Gemini Descriptive" had the lowest at 25%. The analysis of Flesch-Kincaid Grade Level shows that "Gemini Argumentative" achieved the highest score of 15 while the lowest score of 11 was found in "Gemini Descriptive".

Table 5. Mean Values of Each Essay Model Generated by Gemini

Comparison of Five Essay Models	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion	Flesch-Kincaid Grade Level
Descriptive	3%	61%	97%	2%	25%	11
Classification	4%	76%	93%	3%	36%	12
Cause and Effect	3%	76%	66%	2%	54%	14
Comparison and Contrast	3%	72%	80%	1%	51%	12
Argumentative	10%	61%	64%	16%	87%	15

These findings suggest that AI-generated content from ChatGPT and Gemini can significantly contribute to enhancing educational practices. However, understanding the specific capabilities and limitations of each AI tool is crucial for maximizing their effectiveness in learning and teaching contexts. ChatGPT is more precise in meeting specified word counts, making it ideal for tasks requiring strict adherence to length, such as academic assignments. In contrast, Gemini produces more detailed and verbose content, which is useful for creative writing or content creation where elaboration is valued. The findings also revealed that both ChatGPT and Gemini similarly and well performed regarding word concreteness. This means both AI chatbots are effective for generating easy-to-understand educational content, useful for creating study materials. However, their low scores in narrativity suggest they are less ideal for storytelling tasks. Furthermore, the findings of this current study suggest that ChatGPT and Gemini have their strengths in a particular way. ChatGPT has a positive tendency to outperform Gemini in terms of referential cohesion and deep cohesion domains. Meanwhile, Gemini tends to perform better than ChatGPT regarding narrativity, syntactic simplicity, and word concreteness. This means ChatGPT is better for assignments involving complex ideas and structured arguments, while Gemini is more suitable for tasks requiring clarity and simplicity. In addition, the two AI tools differ in the reading levels of the texts they produce. ChatGPT tends to create content suitable for college to post-graduate readers, making it ideal for higher education. It's a useful tool for generating challenging materials for advanced learners or as study aids for university students. In comparison, Gemini can generate texts ranging from high school to post-graduate levels, making it a more adaptable tool that fits a wider variety of educational settings.

Although the two chatbots differently outperform from one another, both of these OpenAI tools are recognized for their usefulness in assisting language learning. The outcomes of this study accord with the previous studies relevant to the effective use of chatbot in English writing learning (Fitria, 2023; Barrot, 2023; Zhang et al., 2023; Seyyedrezaei et al., 2022; Yan, 2023). Lee and Lee (2022) conducted a meta-analysis exploring the effect size of using chatbot assisted language learning. The outcomes of the study indicated a positive effect on language learning across all language skills. The differences in linguistic features and performance between ChatGPT and Gemini suggest that educators and students can benefit from selecting AI tools that match their specific needs. By understanding the strengths and limitations of each tool, they can make better informed decisions for their writing objectives.

3.2. RQ 2: What is the percentage of plagiarized content in the generated texts from OpenAI ChatGPT and Google Gemini as detected by plagiarism checkers?

To report the results concerning the plagiarized-content percentage of the generated writing texts from the two AI Chatbots by plagiarism checkers, the researchers divided them into three parts which are as follows:

1) Free Plagiarism Checkers: SmallSEOTools and Check-Plagiarism

Using the two free plagiarism checkers, SmallSEOTools and Check-Plagiarism, identified 12 out of the 30 essays as containing plagiarized content as shown in Table 6. The former checker found that 10 essays (five generated by ChatGPT and the others by Gemini) containing plagiarism while the latter found 11 essays (five generated by ChatGPT and the other six by Gemini) containing it. In terms of the writing model, the average percentage showed that cause and effect essays from the two AI Chatbots contained the highest percentage of plagiarized content.

The analysis revealed that both free plagiarism checkers similarly identified two essays — "ChatGPT Cause and Effect 2" and "Gemini Cause and Effect 3"— as containing the highest percentages of plagiarized content. According to the SmallSEOTools checker, "ChatGPT Cause and Effect 2" had 17% plagiarism while "Gemini Cause and Effect 3" had 12%. For the Check-Plagiarism checker, the data showed "ChatGPT Cause and Effect 2" contained 18% plagiarism, followed by "Gemini Cause and Effect 3" with 13%.

The analyzed data revealed that among the five writing models, cause-and-effect essays contained the highest percentage of plagiarism while the classification essays generated by ChatGPT and the argumentative essays by Gemini had the lowest percentage (0%). The cause-and-effect essays produced by both AI chatbots were most frequently found to contain plagiarized content. Specifically, two out of the three essays generated by ChatGPT were reported to have plagiarism, which was similar to Gemini. Two out of three essays generated by Gemini also contained plagiarized content. In contrast, the analysis identified two writing models with no detected plagiarism (0%). The first was

the classification essays generated by ChatGPT. When checked using SmallSEOTools and Check-Plagiarism, all three classification essays were found to have 0% plagiarism. The second was the argumentative essays generated by Gemini. All argumentative essays, when tested with both free plagiarism checkers, also showed no plagiarized content (0%).

Table 6. The Percentage of Plagiarized Content Found in AI Chatbots' Writing by Free Plagiarism Checkers

Essays Written by AI Chatbots	Free Plagiarism Checkers		Percentage of Plagiarized Content of each Writing Model
	smallseotools .com	*check-plagiarism.com*	
ChatGPT Descriptive 1	5%	6%	
ChatGPT Descriptive 2	0%	0%	1.83%
ChatGPT Descriptive 3	0%	0%	
ChatGPT Classification 1	0%	0%	
ChatGPT Classification 2	0%	0%	0.00%
ChatGPT Classification 3	0%	0%	
ChatGPT Cause and Effect 1	0%	0%	
ChatGPT Cause and Effect 2	17%	18%	7.67%
ChatGPT Cause and Effect 3	5%	6%	
ChatGPT Comparison and Contrast 1	5%	6%	
ChatGPT Comparison and Contrast 2	0%	0%	1.83%
ChatGPT Comparison and Contrast 3	0%	0%	
ChatGPT Argumentative 1	0%	0%	
ChatGPT Argumentative 2	0%	0%	2.83%
ChatGPT Argumentative 3	11%	6%	
Average Percentage	0.03%	0.03%	3%
Gemini Descriptive 1	4%	0%	
Gemini Descriptive 2	0%	5%	1.50%
Gemini Descriptive 3	0%	0%	
Gemini Classification 1	0%	0%	
Gemini Classification 2	4%	4%	1.33%
Gemini Classification 3	0%	0%	
Gemini Cause and Effect 1	0%	0%	
Gemini Cause and Effect 2	4%	9%	6.33%
Gemini Cause and Effect 3	12%	13%	
Gemini Comparison and Contrast 1	0%	0%	
Gemini Comparison and Contrast 2	0%	4%	2.50%
Gemini Comparison and Contrast 3	5%	6%	
Gemini Argumentative 1	0%	0%	
Gemini Argumentative 2	0%	0%	0%
Gemini Argumentative 3	0%	0%	
Average Percentage	0.02%	0.03%	2%

2) Paid Plagiarism Checker: TurnItIn

The analysis of data from the paid plagiarism checker on essays generated by AI chatbots revealed that 21 out of 30 essays contained plagiarized content. In terms of the percentage of plagiarism, three essays generated by ChatGPT were found to have over 50% plagiarism: "ChatGPT Cause and Effect 2" (55%), "ChatGPT Cause and Effect 3" (54%), and "ChatGPT Cause and Effect 1" (50%). For

Gemini, the three essays with the highest plagiarism percentages were "Gemini Cause and Effect 2" (37%), "Gemini Cause and Effect 3" (26%), and "Gemini Argumentative 2" (10%) as shown in Table 7.

Table 7. The Percentage of Plagiarized Content Found in AI Chatbots' Writing by a Paid Plagiarism Checker

Essays Written by AI Chatbots	TurnItIn	Percentage of Plagiarized Content of each Writing Model
ChatGPT Descriptive 1	8%	
ChatGPT Descriptive 2	7%	5.67%
ChatGPT Descriptive 3	2%	
ChatGPT Classification 1	0%	
ChatGPT Classification 2	0%	1.33%
ChatGPT Classification 3	4%	
ChatGPT Cause and Effect 1	50%	
ChatGPT Cause and Effect 2	55%	53.00%
ChatGPT Cause and Effect 3	54%	
ChatGPT Comparison and Contrast 1	13%	
ChatGPT Comparison and Contrast 2	9%	9.33%
ChatGPT Comparison and Contrast 3	6%	
ChatGPT Argumentative 1	17%	
ChatGPT Argumentative 2	33%	30.00%
ChatGPT Argumentative 3	40%	
Average Percentage	19.87%	
Gemini Descriptive 1	0%	
Gemini Descriptive 2	2%	0.67%
Gemini Descriptive 3	0%	
Gemini Classification 1	0%	
Gemini Classification 2	0%	0.00%
Gemini Classification 3	0%	
Gemini Cause and Effect 1	8%	
Gemini Cause and Effect 2	37%	23.67%
Gemini Cause and Effect 3	26%	
Gemini Comparison and Contrast 1	0%	
Gemini Comparison and Contrast 2	3%	2.33%
Gemini Comparison and Contrast 3	4%	
Gemini Argumentative 1	0%	
Gemini Argumentative 2	10%	6.00%
Gemini Argumentative 3	8%	
Average Percentage	6.53%	

For ChatGPT, 13 out of 15 essays were reported to contain plagiarism, with the exceptions of "ChatGPT Classification 1" and "ChatGPT Classification 2." In contrast, eight out of 15 essays generated by Gemini were found to have plagiarism. Additionally, across the writing models, the data showed that essays generated by ChatGPT contained a higher average percentage of plagiarism than those generated by Gemini.

Based on the five writing models, the highest average percentage of plagiarized content was found in the cause-and-effect essays from both ChatGPT ($\bar{x}$ = 53%) and Gemini ($\bar{x}$ = 23.67%). The highest percentage was 55% ("ChatGPT Cause and Effect 2") while the lowest was 8% ("Gemini Cause and Effect 1"). In contrast, the classification writing model had the lowest average percentage of plagiarized content ($\bar{x}$ = 1.33% for ChatGPT and $\bar{x}$ = 0% for Gemini) compared to the other models. Five out of the six classification essays had no plagiarism, except for "ChatGPT Classifying 3," which contained 4%.

For the argumentative writing model, five essays contained plagiarism. The highest percentage was 40% ("ChatGPT Argumentative 3"), followed by 33% ("ChatGPT Argumentative 2"), with the lowest being 0% ("Gemini Argumentative 1"). In the comparison-and-contrast writing model, five out of six essays were also found to contain plagiarism. The highest percentage was in "ChatGPT Comparison and Contrast 1" (13%), while "Gemini Comparison and Contrast 1" showed no plagiarism.

3) Both Free and Paid Plagiarism Checkers: SmallSEOTools, Check-Plagiarism, and TurnItIn

When both free and paid plagiarism checkers were used to detect the content plagiarism in all 30 essays generated by the two AI Chatbots, the analysis identified a significant portion (23 out of 30) containing plagiarized content. In average, the essay with the highest percentage of plagiarism was “ChatGPT Cause and Effect 2” (30%), followed by “ChatGPT Cause and Effect 3” (21.67%) and “ChatGPT Argumentative 3” (19%). This means the essays contained the highest percentage were generated by ChatGPT. In addition, the analysis showed that the average percentage of the fifteen essays generated by ChatGPT was higher than ones by Gemini as shown in Table 8.

Among the 23 essays containing plagiarism, the analysis found that some plagiarized content was detected only by the free plagiarism checkers while some was detected only by the paid software. For plagiarized content detected only by the free software, "Gemini Descriptive 1" showed 4% plagiarism using SmallSEOTools, and "Gemini Classifying 2" showed 4% plagiarism using both free tools. Regarding plagiarism detected only by the paid software, there were 11 essays, as shown in Table 8. Among these, the highest percentages of plagiarism were found in "ChatGPT Cause and Effect 1" (50%), followed by "ChatGPT Argumentative 2" (33%) and "ChatGPT Argumentative 1" (17%).

Furthermore, Table 8 showed that the average percentage of plagiarism across the 15 essays generated by ChatGPT was higher than those generated by Gemini. Specifically, the average plagiarism percentage for ChatGPT essays was 6.64% while for Gemini, it was 2.19%.

In terms of writing models, ChatGPT's essays had a higher mean percentage of plagiarism in four out of five models as shown in Table 9. For descriptive writing, ChatGPT’s essays had an average of 3.11% plagiarism, compared to 1.22% for Gemini. In cause-and-effect writing, ChatGPT's essays averaged 22.78%, while Gemini's averaged 12.11%. ChatGPT’s comparison and contrast essays had an average of 4.33%, compared to 2.44% for Gemini. In argumentative writing, ChatGPT's essays had an average of 11.89%, whereas Gemini's averaged 2%. The only writing model where Gemini had a higher average than ChatGPT was classification, with 0.89% for Gemini and 0.44% for ChatGPT.

Table 8. The Percentage of Plagiarized Content Found in AI Chatbots' Writingby Both Free and Paid Plagiarism Checkers

Essays Written by AI Chatbots	Plagiarism Checkers			$\bar{x}$
	smallseo tools.com	*check-plagiarism .com*	*TurnItIn*	
ChatGPT Descriptive 1	5%	6%	8%	6.33%
ChatGPT Descriptive 2	0%	0%	7%	2.33%
ChatGPT Descriptive 3	0%	0%	2%	0.67%
ChatGPT Classification 1	0%	0%	0%	0.00%
ChatGPT Classification 2	0%	0%	0%	0.00%
ChatGPT Classification 3	0%	0%	4%	1.33%
ChatGPT Cause and Effect 1	0%	0%	50%	16.67%
ChatGPT Cause and Effect 2	17%	18%	55%	30.00%
ChatGPT Cause and Effect 3	5%	6%	54%	21.67%
ChatGPT Comparison and Contrast 1	5%	6%	13%	8.00%
ChatGPT Comparison and Contrast 2	0%	0%	9%	3.00%
ChatGPT Comparison and Contrast 3	0%	0%	6%	2.00%
ChatGPT Argumentative 1	0%	0%	17%	5.67%
ChatGPT Argumentative 2	0%	0%	33%	11.00%
ChatGPT Argumentative 3	11%	6%	40%	19.00%
Average Percentage	**0.03%**	**0.03%**	**19.87%**	**6.64%**
Gemini Descriptive 1	4%	0%	0%	1.33%
Gemini Descriptive 2	0%	5%	2%	2.33%
Gemini Descriptive 3	0%	0%	0%	0.00%
Gemini Classification 1	0%	0%	0%	0.00%
Gemini Classification 2	4%	4%	0%	2.67%
Gemini Classification 3	0%	0%	0%	0.00%
Gemini Cause and Effect 1	0%	0%	8%	2.67%
Gemini Cause and Effect 2	4%	9%	37%	16.67%
Gemini Cause and Effect 3	12%	13%	26%	17.00%
Gemini Comparison and Contrast 1	0%	0%	0%	0.00%
Gemini Comparison and Contrast 2	0%	4%	3%	2.33%
Gemini Comparison and Contrast 3	5%	6%	4%	5.00%
Gemini Argumentative 1	0%	0%	0%	0.00%
Gemini Argumentative 2	0%	0%	10%	3.33%
Gemini Argumentative 3	0%	0%	8%	2.67%
Average Percentage	**0.02%**	**0.03%**	**6.53%**	**2.19%**

Table 9. The Average Percentage of Plagiarized Content Found by both Free and Paid Checkers in Each Writing Model

Five Essay Models	AI Chatbots		$\bar{x}$
	ChatGPT	*Gemini*	
Descriptive	3.11%	1.22%	2.17%
Classification	0.44%	0.89%	0.67%
Cause and Effect	22.78%	12.11%	17.44%
Comparison and Contrast	4.33%	2.44%	3.39%
Argumentative	11.89%	2.00%	6.94%

With a significant portion (23 out of 30) containing plagiarized, the analysis sheds light on the potential of AI-generated content while also highlights critical concerns regarding plagiarism. The study's findings indicated that some written texts generated by ChatGPT were more prone to plagiarism detection than those produced by Gemini. Of the five writing models, the causes and effect essays were mostly plagiarized for both AI tools specially composed by the ChatGPT. The outcomes of this current study differ from those of Khalil and Er (2023) in that essay writing, particularly scientific essays produced by ChatGPT, tended to be recognized for their high originality. In contrast, this current study found that Turnitin detected the cause-and-effect essays as plagiarism at the highest level, similar to the free software plagiarism detector, which also showed that cause and effect writing produced by ChatGPT and Gemini exhibited the highest degree of plagiarism. At this point, cause and effect essays commonly address mostly debated topics, for instance, climate change or the effects of technology. These issues have been widely discussed, resulting in the unintended employment of identical ideas and language features from previous work. This is in line with Aydın and Karaarslan (2022), arguing that ChatGPT seemed to underperform in generating original work after paraphrasing compared to other chatbots.

In comparing free versus paid checkers, this study highlights the differences between the free plagiarism tools and the paid program. As mentioned, Turnitin performed better in detecting plagiarism. These findings align with those reported by Suaverdez and Suaverdez (2023), though there are some differences. In their study, the paid tool identified a higher amount of plagiarized content compared to the free checkers. Similarly, the present study found that the paid tool outperformed the two free programs, particularly in detecting plagiarism in the cause-and-effect essays. However, the findings also indicated that the free checkers performed better in certain writing models generated by Google Gemini, including descriptive, classification, and comparison and contrast essays. While SmallSEOTools and Check-Plagiarism identified a lower percentage of plagiarism compared to Turnitin, they still proved reliable to some extent in detecting plagiarized content.

The findings shed light on the importance of caution when relying on AI tools and encourage users to be cautious when using AI tools for academic or professional writing. Furthermore, this analysis provides valuable insights into the reliability of AI tools in generating original content and the effectiveness of different plagiarism detection systems. It highlights the need for caution and vigilance when using AI-generated texts, particularly in educational contexts where originality and academic integrity are critical.

4. Conclusion and Implications

The present study has explored and compared the writing quality of OpenAI ChatGPT and Google Gemini based on five essay models in terms of the easability and readability using the T.E.R.A program. Moreover, it has revealed the plagiarized content percentage of the texts generated by the two AI chatbots using three plagiarism checkers. The results revealed that both ChatGPT and Gemini similarly and well performed regarding word concreteness; however, they showed limitations in narrativity. ChatGPT was likely to excel over Gemini in areas like referential cohesion and deep cohesion. On the other hand, Gemini outperformed ChatGPT in aspects such as narrativity, syntactic simplicity, and word concreteness. Some essays written by ChatGPT and Gemini were detected for

their plagiarism. Some written texts generated by ChatGPT were more likely to be detected for plagiarism compared to those created by Gemini.

AI has been increasingly used across disciplines. In an educational setting, chatbot has a great impact on how educators and teachers design courses, create teaching and learning materials, and assess students' learning. It is worth noting that chatbot should be incorporated into teaching and learning with good care. As seen in this study, one AI tool performs differently from one another. As a matter of fact that AI tool has its unique ability, it is good for a teacher to select an appropriate tool for a particular teaching and learning objective. Take a material design, for example, where a teacher should take a special care for the levels of the students, choosing the most appropriate AI tool is a key to success. Some AI tools, like Gemini might involve a variety of age ranges more than others. In this case, in a high school context, Gemini could produce some model texts that are appropriate for high school students. In addition, one major concern that should be mentioned regarding the degree of plagiarism found in texts generated by AI tools. Based on the outcomes of this study, a percentage of plagiarism could be found in both ChatGPT and Gemini. This means that teachers and students should not heavily rely on AI tools. To raise students' awareness of plagiarism, they should be encouraged to use AI as language learning assistants to avoid producing writing tasks containing plagiarism which is a violation of ethics. Also, they should practice necessary English skills and be trained how to use AI properly and efficiently in language learning. Additionally, plagiarism-detection software should be incorporated into classroom.

Acknowledgment

The researchers would like to express sincere gratitude to the Division of Research Administration and Academic Services, Kasetsart University Chalermphrakiat Sakon Nakhon Province Campus, Thailand, for providing the necessary resources and financial support. The researchers also extend appreciation to the Head of the Languages Department, Asst. Prof. Dr. Dentisak Dokchandra, as well as departmental colleagues and all administrative officers in the Faculty of Liberal Arts and Management Science, for their invaluable guidance, moral support, and encouragement. Their assistance and suggestions greatly contributed to the successful completion of this research.

Declarations

Author contribution : All authors, Siraprapa Kotmungkun, Wichuta Chompurach, and Piriya Thaksanan contributed equally to this research, participating in all stages, including conceptualization, data collection, analysis, and writing. Wichuta Chompurach, the corresponding author, led the manuscript writing and coordinated the collaboration with Siraprapa Kotmungkun and Piriya Thaksanan. All authors reviewed and revised the manuscript and approved the final version.

Funding statement : The research is funded under the Division of Research Administration and Academic Services, Kasetsart University Chalermphrakiat Sakon Nakhon Province Campus, Thailand.

Conflict of interest : The authors declare no conflict of interest.

Ethics declaration : This research was conducted using AI-generated written data from OpenAI ChatGPT and Google Gemini. No human participants were involved in the study, and as such, the research does not raise ethical concerns regarding human subjects. The AI-generated texts were used solely for the purpose of assessing writing quality and plagiarism detection. All procedures followed were in compliance with ethical guidelines for non-human data research.

Additional information : No additional information is available for this paper.

REFERENCES

Ariyanti, A. (2016). Shaping students' writing skills: The study of fundamental aspects in mastering academic writing. *Indonesian Journal of EFL and Linguistics, 1*(1), 63-77. Retrieved from http://www.indonesian-efl-journal.org

Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT generated literature review: Digital twin in healthcare. *Emerging Computer Technologies, 2*, 22-31. İzmir Akademi Dernegi. http://dx.doi.org/10.2139/ssrn.4308687

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative Artificial Intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52-62. https://doi.org/10.61969/jai.1337500

Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning, 36*(6), 584–607. https://doi.org/10.1080/09588221.2021.1936071

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27-40. https://doi.org/10.3316/QRJ0902027

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://doi.org/10.1186/s41239-023-00411-8

Check Plagiarism. (2024). *Check plagiarism: Free online plagiarism checker*. https://www.check-plagiarism.com/

Chompurach, W. (2021). "Please let me use Google Translate": Thai EFL students' behavior and attitudes toward Google Translate use in English writing. *English Language Teaching, 14*(12), 23-35. https://doi.org/10.5539/elt.v14n12p23

Davis, D. F., Golicic, S. L., & Boerstler, C. N. (2011). Benefits and challenges of conducting multiple methods research in marketing. *Journal of the Academy of Marketing Science, 39*, 467–479. https://doi.org/10.1007/s11747-010-0204-7

Espejel et al. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal, 5*, 100032. https://doi.org/10.1016/j.nlp.2023.100032

Evans et al. (2023). ChatGPT impacts on access-efficiency, employment, education and ethics: The socio-economics of an AI language model. *BizEcons Quarterly, 16*, 1-17.

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching, 12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

Hasnawati, H., Mujahidin, E., & Tanjung, H. (2023). Analyzing students' difficulties in writing English essay. *International Journal of Social Science and Human Research, 6*(10), 5954-5959. https://doi.org/10.47191/ijsshr/v6-i10-22

Hyland, K. (2003). *Second language writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667251

Hyland, K., & Jiang, F. K. (2017). Is academic writing becoming more informal? *English for Specific Purposes, 45*, 40-51. https://doi.org/10.1016/j.esp.2016.09.001

Jovanovic, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer, 55*(10), 107-112. https://doi.org/10.1109/MC.2022.3192720

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. In *Proceedings of the 25th International Conference on Human-Computer Interaction (HCII 2023)* (Vol. 14040, pp. 475–487). https://arxiv.org/pdf/2302.04335

Lee, G. G., Shi, L., Latif, E., Gao, Y., Bewersdorf, A., Nyaaba, M., et al. (2023). *Multimodality of AI for education: Towards artificial general intelligence*. arXiv. https://arxiv.org/abs/2312.06037

Lee, H., & Lee, J. H. (2022). The effects of robot-assisted language learning: A meta-analysis. *Educational Research Review, 35*, 100425. https://doi.org/10.1016/j.edurev.2021.100425

Lee, J. & Park, J. (2023). AI as "Another I": Journey map of working with artificial intelligence from AI-phobia to AI-preparedness. *Organizational Dynamics, 52*, 100994. https://doi.org/10.1016/j.orgdyn.2023.100994

Lingard, L. (2023). Writing with ChatGPT: It illustrates its capacity, limitations & implications for academic writers. *Perspectives on Medical Education, 12*(1), 261–270. https://doi.org/10.5334/pme.1072

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). A proposal for Dartmouth summer research project on artificial intelligence. *AI Magazine, 27*, 12.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Meiranti, R. (2012). Improving students' writing skills through field trip method. *English Review: Journal of English Education, 1*(1), 89-96.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.

Mohammed, A., Al-ghazali, A. & Alqohfa, K. (2023). Exploring ChatGPT uses in higher studies: A case study of Arab postgraduates in India. *Journal of English Studies in Arabia Felix, 2*(2), 8–15. https://doi.org/10.56540/jesaf.v2i2.55

OpenAI. (2022). ChatGPT: Optimizing language models for dialogue. OpenAI. https://openai.com/blog/chatgpt/ (Accessed November 29, 2023)

Rane, N., Choudhary, S., & Rane, J. (2024). Gemini or ChatGPT? Capability, performance, and selection of cutting-edge generative artificial intelligence (AI) in business management. *Studies in Economics and Business Relations, 5*(1), 40-50. https://doi.org/10.48185/sebr.v5i1.1051

Rayhan, A. (2023). *Mastering prompt engineering: Techniques for creating powerful and effective AI language models*. Independently Published.

Seyyedrezaei, M. S., Amiryousefi, M., Gimeno-Sanz, A., & Tavakoli, M. (2022). A meta-analysis of the relative effectiveness of technology-enhanced language learning on ESL/EFL writing performance: Retrospect and prospect. *Computer Assisted Language Learning*, 1–34. https://doi.org/10.1080/09588221.2022.218782

Small SEO Tools. (n.d.). Small SEO tools: Text analysis tools. Small SEO tools. https://smallseotools.com/

Stahl, B. C. & Eke, D. (2023). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management, 74*, 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700

Suaverdez, J., & Suaverdez, U. (2023). Chatbots impact on academic writing. *Global Journal of Business and Integral Security*, (2). Retrieved from https://www.gbis.ch/index.php/gbis/article/view/150

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Turnitin. (n.d.). Turnitin: Empower students to do their best, original work. Turnitin. https://www.turnitin.com/

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Zhai, X. (2023). ChatGPT user experience: Implications for education. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4312418

Zhang, S., Shan, C., Lee, J. S. Y., & et al. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 28, 15223–15243. https://doi.org/10.1007/s10639-023-11805-6