

ChatGPT vs Human Writing: A Qualitative and Quantitative Study of IELTS Task 2 Authorship

Aida Akif Mirzayeva¹

¹ *Baku Higher Oil School*

Yeni Salyan Highway 3rd km, 25, Sabail district, Bibiheybat settlement, Baku, 1023, Azerbaijan

DOI: [10.22178/pos.122-97](https://doi.org/10.22178/pos.122-97)

LCC Subject Category: L7-991

Received 27.08.2025

Accepted 27.09.2025

Published online 30.09.2025

Corresponding Author:

[Aida Akif Mirzayeva](#)

© 2025 The Author. This article is licensed under a [Creative Commons Attribution 4.0 License](#)



Abstract. Artificial intelligence tools, such as ChatGPT, have recently become increasingly prevalent in the academic world due to their remarkable capabilities for generating essays, including logical structure, contextual relevance, and advanced vocabulary. Undoubtedly, these tools outperform humans in many aspects; however, despite their language-driven strengths, machine-generated content exhibits apparent shortcomings that make it easily identifiable as machine-authored. This article examines the critical deficiencies and limitations of AI writing, with a focus on the significant qualitative and quantitative differences between human-written and automated texts. It emphasises the originality and emotional colouring of human writing while highlighting the absence of these qualities in AI-generated essays despite their linguistic advantages. These findings may guide educators in helping learners refine their writing skills, identify common errors in human writing, and recognise the unique style that distinguishes it from machine-generated writing. The outcomes can be helpful not only in the academic environment but also in AI development, making the generated texts more closely resemble human writing.

Keywords: academic writing; AI tools; ChatGPT; higher education; IELTS; teaching English

INTRODUCTION

Recent advancements in language technologies have led to the development of intelligent automation tools, such as ChatGPT, which has been identified as one of the most frequently used ones [1]. These tools can produce essays with coherent and accurate responses to various topics, utilising a diverse range of lexical resources. At first glance, these models impress educators and learners alike with their fluency, coherence, relevant contextual language, and formal tone, leaving the average human writer at a significant disadvantage. However, as an individual becomes more familiar with AI-generated essays, it becomes increasingly challenging to differentiate them stylistically or semantically. Further analysis reveals a similar pattern and style in these writings. This shift prompts a critical examination of the core features of human writing while simultaneously comparing them to those of machine-generated content.

The participants in this research consisted of 15 foundation-year students, aged 17 to 19, learning English as a second language (ESL) at Baku Higher Oil School. Their academic essays were produced under supervision during graded writing assignments, with a strict prohibition against the use of AI tools. Primarily, students with a C1 level of English were selected to minimise the likelihood of linguistic errors. Researchers presented the same IELTS essay topics to ChatGPT and analysed the responses, revealing notable similarities and differences through both quantitative and qualitative comparisons. On a larger scale, these findings may help refine machine-generated writing to better resemble human writing and assist higher education students in improving their writing skills by identifying significant areas for improvement.

METHOD

This research employed qualitative and quantitative approaches to compare ChatGPT-generated

essays with those written by students in higher education. The methodology integrated text-based analysis, comparative evaluation, and error categorisation to explore the linguistic, stylistic, and structural differences between human and machine-generated academic essays.

Textual Analysis with Commentary. The study conducted an in-depth textual analysis to explore the written discourse produced by the two groups. The researchers focused specifically on the features outlined in the BC band descriptors — coherence and cohesion, lexical resources, and syntactic structures — as well as on stylistic originality and emotional colouring, and identified and proposed additional features that extended beyond these criteria. A commentary-based approach involved interpretive reflection at both rhetorical and linguistic levels, grounded in the written text.

Comparative Analysis. The researchers employed a comparative approach throughout the study to analyse and contrast human-written essays with those generated by ChatGPT, using the same IELTS Task 2 prompts. They compared 15 essays produced by foundation-year students at Baku Higher Oil School with responses generated by ChatGPT. To ensure the authenticity of the student-written material, invigilators supervised the students as they wrote the essays during the official assessment.

Error and Feature Analysis. A quantitative error analysis was conducted across student essays to identify deficiencies of human writing, including lexical errors related to semantic choice and spelling, as well as grammatical structures and components. The errors were evaluated against ChatGPT's results, highlighting its linguistic precision while also highlighting its grammatical limitations.

Prompting Protocol for AI Text Generation. For consistent comparison, the same IELTS prompts were presented to ChatGPT and analysed according to the same standard criteria, without any modifications. Additionally, the researchers compared these responses to human-written essays to conduct a thorough textual and structural analysis.

RESULTS AND DISCUSSION

Assessment Criteria Analysis. The assessment criteria for the analysed essays are based on the

British Council public band descriptors. These descriptors include task response, coherence and cohesion, lexical resource, and grammatical range and accuracy, along with the proposed additional features that extend beyond these parameters. Simultaneously, the researchers provided the same topics to ChatGPT and analysed the results to identify similarities and differences between human-written and AI-generated essays.

Task Response. The students in the Foundation Course were assigned a Task 2 essay on the problems caused by cybercrime and potential solutions for individuals and businesses. Below are the main ideas presented by one of the students, who later achieved a band score of 7 at the IELTS exam, as well as those generated by ChatGPT:

	Student 1	AI
Problems	not being tech-savvy; anonymity of Internet	financial loss; reputational damage
Solutions	providing accessible digital training; improving cybersecurity infrastructure	adopting proactive cybersecurity measures; investing in cybersecurity infrastructure

Upon analysing the essay's task achievement, it becomes evident that the ambiguous prompt "What problems are caused ..." led the student to focus on causes rather than problems. In contrast, the machine-generated essay effectively addressed the prompt, as its ideas closely aligned with the requirements and focused solely on the issues arising from the specified problem. To broaden the scope of the research, an additional 15 essays were examined, revealing that only 3 of 15 students made a similar error; this indicates that, on a cognitive level, approximately 20% of students may misinterpret the prompt despite their efforts to be as accurate and attentive as possible when faced with multiple options. It could also be influenced by exam-related stress or time constraints, which are significant factors contributing to errors or failures.

The second crucial element of the topic above was identifying the problems faced by both individuals and businesses. It is noteworthy that, while 13 students addressed personal threats, only two emphasised the business matters re-

quired by the question. Consequently, multiple components in a question tend to deteriorate people's attention spans, causing them to overlook the final aspect in approximately 75% of cases. In contrast, the ChatGPT essay effectively highlighted all key points, covering both personal and business issues. Thus, a machine author can efficiently handle multiple tasks simultaneously.

The same features were identified in the study by Revell T, Yeadon W, and Cahilly-Bretzin G et al., who noted that the primary differentiating factor between ChatGPT and authentic writing is the latter's lack of detailed focus. The researchers also emphasise that a lack of nuance and incisiveness is not uncommon among weaker students [2]. However, as observed in our case, this issue may also arise in the writing of upper-intermediate and advanced learners.

Coherence and cohesion. Further analysis and comparison of student and AI writing focus on coherence and cohesion. In the first case, the student demonstrated a clear overall progression with ideas logically developed and well-supported. The student effectively employed cohesive devices such as "in my opinion, additionally, consequently, first, furthermore, in conclusion". As shown, the writers avoided repeating these devices and instead used a variety of transitional signals. In contrast, coherence in the digitally-generated essay is not as smooth as that of human writing. For instance, while the students' writing presents two well-developed ideas in each main body paragraph with clear impact or outcomes, the ChatGPT-created essay introduces two main ideas in each section, with a third idea added in the last sentence of the main body, leaving it unsupported and lacking a concluding sentence.

Regarding the cohesive devices utilised in the AI variant, the essay included terms such as "at a broader scale, additionally, firstly, for example, secondly, moreover, at an individual level, on a policy level, in conclusion, however". Both AI tools and human writers commonly employ these cohesive devices. Furthermore, the example provided by ChatGPT appears overly personal as it discusses his friend's stolen card, which is not advisable in academic writing.

Lexical Resource. AI-generated texts are impressive for their extensive vocabulary, particularly their use of uncommon words. However, the researchers [3, 4] point out limitations in AI writing's vocabulary, highlighting a lack of flexibility

in favour of clarity compared to human writing. Despite this, the precision tailored to the context and the breadth of lexical resources are sufficient, even though there is repetitive usage across different essays. The effective collocations found in essay include: a formidable threat, digitised world, corporate data, destabilise the public, pragmatic solutions, commercial entities, compromised credentials, corporate level, data breaches, regulatory fines, posing threats, imperative, swift responses, undermine financial stability, cultivate a culture of cyber-awareness, embrace robust technological safeguards, mitigate threats, enhance digital resilience, unsolicited emails.

In the human-written essay, several uncommon and effective collocations were utilised, including: pressing issue, adversely affect, run a business, anonymity of the Internet, provide accessible digital tracing, incorporate specific programs, become aware of cyberbullying, confront problems, tighten law enforcement, and ensure a reduction in online offences. When compared, the quantity of such vocabulary is nearly equal, with 11 instances for AI and 10 for humans, demonstrating that students can incorporate uncommon words and collocations as effectively as AI, provided they have memorised and used them in the appropriate context.

The researchers identified the following lexical errors in the learners' writings:

Semantic errors:

Error	Correct Variant
lead to	stem from
thief	steal
parolas	passwords

Spelling mistakes:

Error	Correct Variant
threating	threatening
technology	technoloy
weather	whether
consitantly	consistently
encrypting	encrypting
especiall	especially
quiete	quite
truthfull	truthful
companis	companies
fullfil	fulfill

The identified deficiencies in lexical resources are relatively few: only 12 errors were found

across 15 essays, including three semantic mistakes and 10 spelling errors. Given their advanced understanding of most words, students are more likely to make spelling errors than semantic ones.

To assess the vocabulary range of the artificial intelligence system, 15 essays on the topic of "Crime", each with various IELTS-based titles, were prompted to ChatGPT. Although the researchers instructed the system to write Task 2 IELTS essays, which typically require 250–300 words, it produced overly long essays that often exceeded 350 words. To fully evaluate its potential, the prompts specified a Band 9 essay. A thorough analysis revealed 41 effective collocations at the C1-C2 level. Notably, hardly any of these collocations were repeated within a single essay, which contrasts with the common issue of student writing.

For a more in-depth analysis of AI-generated essays, the following information could help examine their specific writing features [5].

Essays No	Word Count	No of collocations
Essay 1	426	57
Essay 2	424	56
Essay 3	355	56
Essay 4	369	53
Essay 5	408	54
Essay 6	423	54
Essay 7	331	49
Essay 8	277	48
Essay 9	324	51
Essay 10	333	52
Essay 11	357	50
Essay 12	323	48
Essay 13	356	49
Essay 14	377	46
Essay 15	399	47

The observed range of word count is as follows:

400 < – 4 essays (27%)

350- 400 – 5 essays (50%)

300- 350 – 4 essays (40%)

250- 300 – 1 essay (10%)

The number of effective C1- C2 level collocations is between 45 and 59:

55-59 – 3 essays (20%)

50-54 - 6 essays (40%)

45-49 – 6 essays (40%)

Furthermore, nearly all the essays generated by ChatGPT contained three main body paragraphs. At the same time, students generally prefer to write only two paragraphs to adhere to the assignment guidelines and avoid exceeding the recommended word limit.

Another characteristic of ChatGPT's writing is its adherence to British English spelling standards. For example, it uses behavioural, self-defence, specialised, favour, and centre.

As for the frequency of commonly used words by the AI system on the topic of "Crime", it ranks the following way:

Ranking	Word	Frequency
1	crime/ criminals/ crime	150
2	punishment/ sentences	75
3	behaviour	65
4	juvenile/ youth/ teenagers	65
5	social/ socio-economic	60
6	rehabilitation/ reform/ reintegration	55
7	justice/ judicial/ legal	55
8	prevention/preventive	50
9	education	45
10	family/ parental	40

The analysis reveals similarities and differences in vocabulary usage between human writing and AI-generated text. The most distinctive features include the total word count, with ChatGPT significantly exceeding it, and the presence of spelling errors in human writing. Conversely, the researchers observed similarities between the collocations that students had memorised and those generated by various AI platforms [6].

Grammar. Regarding the grammatical structures employed in both generated and authentic essays, researchers present conflicting views, either pointing to a lack or an abundance of these structures. While previous years saw scholars highlighting the grammatical deficiencies of machine-generated writing, recent studies have shifted the focus to its positive aspects, advocating the use of AI tools to teach grammar and correct human errors. According to J. Jahan and his co-authors, who conducted research in a similar academic environment involving university education students in the experiment, human writing can substantially improve through AI corrections of their work [7]. In our experiment, we will ana-

lyse generated essays alongside those produced by students to identify key differences and determine how students can learn from them to enhance their writing. The S1 essay included the following grammatical patterns:

... that may affect ... - adjective clause

... as they would be reluctant... - adverbial clause of reason

... provided that ... - conditional sentence

... those who ... - adjective clause

Despite the complex vocabulary, the majority of the sentences had a simple structure—11 out of 15. Similarly, ChatGPT produced a limited range of grammatical structures, with only two variations.

While – an adverbial clause used to indicate contrast

That – Adjective clause

Therefore, both writings failed to impress with their range of grammatical structures, showing only a narrow range of statistical variation.

When examining other grammatical features, the machine-generated text included six sentences in a single essay using the Present Participle. In contrast, the human learner used it only once, highlighting a significant difference between the two.

Another distinctive characteristic of ChatGPT writing is its use of the gerund.

Student: By incorporating, by doing this, by investing, addressing this issue requires raising awareness and tightening law enforcement (5 times)

ChatGPT: By staying vigilant and embracing training, employees can update passwords, and avoiding suspicious links can reduce vulnerability (6 times).

In comparison, both of the essays included extensive usage of gerund or gerundial constructions, which were employed either after prepositions or as a subject or a predicative.

Furthermore, other grammatical features include modal verbs:

Student: can (5 times), would (3 times), should (2 times), must (1)- total 11 times

ChatGPT: can (7 times), would (0), should (once), must (2 times)- total 10 times.

The analysis revealed that although the number of uses of modal verbs was approximately the same, their selection varied.

Furthermore, both pieces of writing included the infinitive:

Student: to tackle cybercrime (subject), to ensure (purpose), to avoid theft (purpose), would not be able to benefit (be able to)- 4 times.

ChatGPT: to combat (purpose), to identify (purpose), learn how to recognise (between two verbs), to address these issues (purpose), force to spend (between two verbs), used to make (between two verbs)- 6 times.

So, once again, there is minimal difference in the quantitative use of the infinitive. Therefore, studying ChatGPT essay samples will not enhance students' writing unless they evaluate their own work using AI tools. By doing so, they can easily identify their mistakes, receive instant feedback, and even generate the content to support improvement of their writing skills [8].

Regarding the students' grammatical mistakes, seven of the eight essays contained errors. They were as follows:

Grammar	Error	Correct Variant
Demonstrative Pronouns	this	these
Plural/ singular noun	datas; informations	data information
Noun/ verb agreement	another factor that cause; cards that is not; people has become	another factor that causes; cards that are not; people have become
Prepositions	from a variety of reasons	for a variety of reasons
Articles	Internet	the Internet
Passive Voice	people may expose to cyberattacks	people may be exposed to cyberattacks
Adjective	more safe	safer
Pronouns	it	them

As is observed, 15 student essays contained 11 grammatical mistakes, most of which were related to the use of plural and singular forms. Given the high level of language proficiency, these errors may stem from inattention caused by multi-tasking while writing, making it difficult for non-

native learners to concentrate on vocabulary and grammar simultaneously.

Originality and Emotional Colouring. Besides the standard criteria mentioned, emotional colouring and originality are key factors that primarily distinguish human writing from AI-generated writing. These features cannot be replicated in any machine-generated essay, regardless of how closely it attempts to mimic human expression.

Originality and stylistic innovation are core features of human writing that convey the writer's voice with great vividness. As observed in the students' essays, the tone varied across topics, with each essay presenting a different level of risk and articulating strong opinions on issues they had rarely encountered due to their young age. Conversely, although machine-generated essays are designed to mimic human writing and are often perceived as sounding natural and acceptable [9], their neutral and balanced approach distances them from human style, which is characterised by its variation and confusion at times, yet being strongly opinionated and leaning towards specific points emphasised in the essay.

Another aspect to consider is the emotional tone present in the academic writings of both human and machine authors. By emotional colouring, we refer to the writer's attitude, which can evoke either positive or negative feelings in the reader, or, conversely, maintain a neutral or emotionless tone.

For instance, in the same essay about crime, the student articulates her concerns, highlighting the elderly or less tech-savvy individuals who may be easy targets for cybercriminals. These examples could be drawn from real-life experience to generate ideas. Another example can be illustrated as follows: "These incidents, in turn, can lead individuals to lose trust in online companies, making them hesitant to share personal information to avoid theft or online fraud."

As can be observed, the human writer emphasises loss of trust, which carries an emotional connotation. In contrast, ChatGPT conveyed a similar idea but with different wording: "The frequency of such crimes weakens public confidence in online platforms." This suggests that the information derived from any online platform ad-

ressing this issue is akin to mere statistical survey results. Consequently, it fails to express any personal feelings as one might expect, or even to mimic human emotions.

CONCLUSIONS

The comparative analysis of ChatGPT-generated and human-written IELTS Task 2 essays reveals that, while the AI system excels in linguistic precision, contextual relevance, and lexical sophistication, it still falls short in crucial human attributes such as originality, emotional depth, and stylistic individuality. These findings indicate that although AI-generated writing may meet academic criteria, the primary factors that render it less authentic are a lack of personal engagement, cognitive variability, and human nuance.

By contrast, students' essays, despite their typical grammatical and lexical errors, often demonstrate a deeper understanding of the task, particularly when they resonate emotionally or experientially with human beings. Human writing encompasses not only linguistic knowledge but also the writer's personal attitude, perspective, inner voice, and situational awareness, qualities that are often absent in machine-generated writing, which tends to maintain a neutral tone and provide generalised responses.

From a pedagogical perspective, this research suggests that AI writing tools can serve as supplementary resources for language learning, particularly in expanding vocabulary and correcting grammar. However, overreliance on ChatGPT for writing may undermine students' ability to express their own voice, leading them to adopt a more neutral tone. This reliance can result in writing that prioritises sophisticated vocabulary and error-free sentences at the expense of personal expression.

Finally, this study emphasises that linguistic accuracy alone cannot substitute or compensate for the absence of human-like qualities such as personal expression, emotional resonance, and situational sensitivity. Therefore, a long-term approach would strive to bridge the gap between algorithmic precision and irreplaceable cognitive and emotional depth of human writing.

REFERENCES

1. Durak, H. Y., Eğin, F., & Onan, A. (2025). A comparison of Human-Written versus AI-Generated text in discussions at educational settings: investigating features for ChatGPT, Gemini and BingAI. *European Journal of Education*, 60(1). doi: [10.1111/ejed.70014](https://doi.org/10.1111/ejed.70014)
2. Revell, T., Yeadon, W., Cahilly-Bretzin, G., Clarke, I., Manning, G., Jones, J., Mulley, C., Pascual, R. J., Bradley, N., Thomas, D., & Leneghan, F. (2024). ChatGPT versus human essayists: an exploration of the impact of artificial intelligence for authorship and academic integrity in the humanities. *International Journal for Educational Integrity*, 20(1). doi: [10.1007/s40979-024-00161-8](https://doi.org/10.1007/s40979-024-00161-8)
3. Islam, N., Sutradhar, D., Noor, H., Raya, J. T., Maisha, M. T., & Farid, D., MD. (2023). Distinguishing human-generated text from ChatGPT-generated text using machine learning. *arXiv (Cornell University)*. doi: [10.48550/arxiv.2306.01761](https://doi.org/10.48550/arxiv.2306.01761)
4. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1). doi: [10.1007/s40979-023-00146-z](https://doi.org/10.1007/s40979-023-00146-z)
5. Fariello, S., Fenza, G., Forte, F., Gallo, M., & Marotta, M. (2024). Distinguishing Human From Machine: A Review of Advances and Challenges in AI-Generated Text Detection. *International Journal of Interactive Multimedia and Artificial Intelligence*, In press, 1. doi: [10.9781/ijimai.2024.12.002](https://doi.org/10.9781/ijimai.2024.12.002)
6. Alsaweed, W., & Aljebreen, S. (2024). Investigating the accuracy of ChatGPT as a writing error correction tool. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1–18. doi: [10.4018/ijcallt.364847](https://doi.org/10.4018/ijcallt.364847)
7. Jahanzeb, J., Bisma, A., & Uzma, M. (2024). Enhancing Practical English Grammar Skills Through AI: A Study on the Impact of ChatGPT-Assisted Feedback on Student Writing. *Contemporary Journal of Social Science Review*, 2(4).
8. Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. doi: [10.1016/j.asw.2023.100745](https://doi.org/10.1016/j.asw.2023.100745)
9. Sardinha, T. B. (2023). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1), 100083. doi: [10.1016/j.acorp.2023.100083](https://doi.org/10.1016/j.acorp.2023.100083)