

Editing synthetic text from generative artificial intelligence: two exploratory case studies

Michael Farrell

IULM University, Milan, Italy

michael.farrell@iulm.it

Abstract

As the use of generative artificial intelligence (GenAI) becomes more mainstream, an increasing number of authors may turn to this technology to write directly in a second language, bypassing traditional translation methods. Consequently, professional editors may have to develop new skills: shifting from correcting translation and non-native errors to editing AI-assisted texts. This study includes several stages: participant selection, text planning, prompt engineering, text generation and text editing. The recruited authors provided prompts for GPT-4 to generate texts, edited the output as they desired and then passed them on to professional editors for a final edit. All participants reported their experiences and described the nature of their interactions. The findings reveal that, while GenAI significantly improved the grammatical accuracy of the non-native English texts, it also introduced anomalies. In conclusion, although AI was useful in these two cases, it did not fully replace the human editors, and professional translators — with their language skills — may like to consider offering this additional service. The study also suggests that both authors and editors should be trained in synthetic-text editing to fully harness the benefits of AI-assisted writing, and that further research should be conducted with diverse texts and authors to generalize the findings.

1 Introduction

With the advent of generative artificial intelligence (GenAI), an increasing number of authors may be tempted to bypass traditional translation and craft their texts directly in a second language with the aid of GenAI prompts. This practice might be termed AI-assisted second-language authoring. Some authors have always preferred to write in their second language and many professional editors already make a career out of correcting these texts.

The consequent shift presents new challenges and opportunities for professional editors, who will need to transition from correcting translation and non-native errors to editing synthetic texts (STs) generated by GenAI, hereon referred to as synthetic-text editing (STE). While machine translation (MT) output can also be considered a form of ST since it is created artificially, it is useful to limit the term ST to output generated by systems based on large language models (LLMs) (Farrell, 2024). In contrast, traditional MT output is created by AI systems trained using parallel corpora, such as Google Translate or DeepL Translate.

If the envisaged scenario becomes reality, there may be a slight decrease in traditional translation work and an increase in demand for synthetic-text editors. Professional translators, bilingual post-editors and author's editors, with their language skills, could be ideally positioned to offer this new service.

2 Aim and limitations

The experiment aims to explore the feasibility of using GenAI as a tool to allow authors to write directly in a second language, bypassing traditional translation methods. If the results of this

limited experiment are promising, it should ideally be repeated with different kinds of text on a wider variety of subjects in different languages by authors from diverse backgrounds.

3 Method

Figure 1 shows the method schematically. Refer to the sections for details.

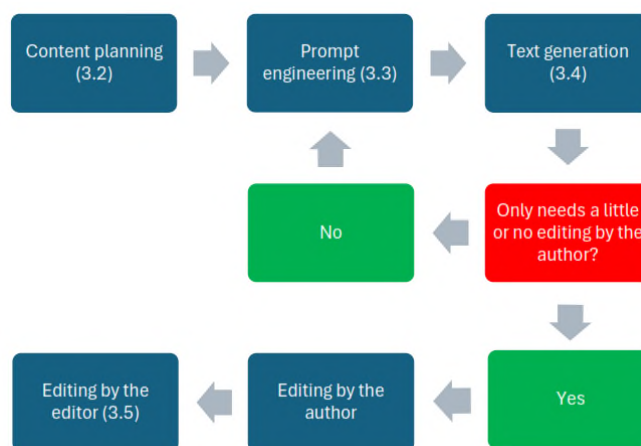


Figure 1. Experimental working method

3.1 Recruitment

A call for participants was published on the internet and distributed through social media channels (LinkedIn and Facebook). The Mediterranean Editors & Translators¹ association was also asked to share the call with its members since they belong to two of the three potentially affected professions.

The applicant authors were asked about their experience, their native language and other languages they knew, the subject areas they would like to write about for the experiment, whether the text they would write would be real or a simulation, and to provide any other information they considered important.

The candidate editors were asked about their experience as English-language editors, particularly with non-native authors, the languages they were proficient in besides English, their preferred subject areas, and any experience they had of post-editing (MTPE), STE, translation and human translation revision, as well as any other information they considered relevant.

The authors were also asked to provide a sample of at least 100 words they had previously written in English without the aid of AI, MT, computer tools (except for dictionaries) or other people, on the same or a similar subject as the text they intended to write. This text served two purposes: firstly, so that the editor could gauge the author's knowledge of English and, secondly, as a sample that could be used during GenAI prompt engineering.

3.2 Content planning

The authors were asked to provide the precise subject or a provisional title for the text they planned to write in English using GPT-4's web interface (500 to 2000 words). They were also asked about their usual approach to planning a text of this kind, which computer tools they usually used to write in English, whether their text would have a particular structure (such as an introduction, discussion, etc.), and if their text needed to comply with a style guide or specific writing conventions.

¹ www.metmeetings.org

3.3 Prompt engineering

A prompt engineering technique based on the automatic prompt engineer (APE) (Zhou et al., 2022) was used. Essentially, GPT-4 was asked to *reverse engineer* its own prompt from the sample text the author had provided.

Firstly, GPT-4 was instructed to correct the English of the sample text. Then, without starting a new chat, it was asked to summarize the sample text as a list of short notes. Lastly, again in the same chat, GPT-4 was asked to write four different prompts, in order from best to worst, which — together with the notes — would cause it to generate the corrected sample text.

The short notes and four prompts were then sent to the author as a model on which to base a prompt which could be used to generate the text they wanted to write. The authors were told that, if they found it easier, they could write their prompt and/or notes in their native language (or any other language), and even mix languages.

If the author wanted to organize their text into sections, they were instructed to divide the notes into the same sections with headings, and if they had to follow a style guide or specific conventions, this too had to be added to the prompt.

3.4 Text generation

The researcher checked the prompt provided by the author for completeness. He then fed it to GPT-4, took the generated text and sent it back to the author, together with the actual prompt used to generate it. The author was also sent a feedback questionnaire asking their opinion of the output and how they wished to proceed. They could choose to edit the prompt to see if better results could be obtained, including by breaking the task down into steps and using prompt chaining techniques (Wu et al., 2021), or they could use the GenAI output as a base for the text they had in mind. The authors were allowed to make as many edits to the AI-generated text as they felt were required to achieve the desired result, including rewriting, deleting or adding entire paragraphs. All edits were marked using *Track Changes* in Microsoft Word.

Once the researcher received the author's final draft, he asked them to give examples of edits they had made and explain why they were necessary. The researcher then sent the file to the editor with all edits hidden (*Accept All Changes*) so that they could not tell which parts the author had edited.

If the authors required, the researcher and editors were willing to sign nondisclosure agreements, but they were warned that it is not advisable to share sensitive or unpublished data on online platforms.

3.5 Text editing

After giving their initial impressions, the editors were asked to do the work they would normally do, while noting the changes they made on an *errors and textual anomalies* form. This questionnaire suggested several error categories, including second-language authoring errors (Corder, 1975; James, 1998), MT errors (Popović, 2018) and some commonly reported hallmarks of GenAI (Dondoni Braz, 2024; Dou, 2022; Gillham, 2024; Gluska, 2023; OpenAI, 2022). The editor and author were asked to interact as they normally would during the editing process, without any interference from the researcher. Once editing was complete and the author approved the final text, the researcher sent the authors and editors final feedback forms.

4 Results

Three authors and three editors answered the call. Each author was paired with an editor based on the subject matter and type of article the author intended to write: Author 1 (A1) with Editor 1 (E1), Author 2 (A2) with Editor 2 (E2), etc. Each pair was considered a separate case study. The third case study did not reach conclusion due to participant dropout.

4.1 Participants

A1 and A2 know each other: A1 suggested that A2 answer the call for participants. However, they did not consult each other during the experiment. None of the other participants had met before.

4.1.1 Author 1

A1 is in his twenties and writes in Italian (his native language) for a local newspaper while studying History of Art at university. He has no professional experience of writing in English, which he knows to B2 level (Council of Europe, 2001). However, he does write in English for his studies and can also speak Spanish. He chose History of Art as his preferred subject area and decided to write an article solely for the purpose of this experiment. He normally writes newspaper articles and academic papers/research reports in his native language for professional or study reasons.

4.1.2 Editor 1

E1 is in her thirties and has over fifteen years of experience as an English-language editor. In addition to her native English, she speaks Dutch. For this experiment, she was willing to edit texts on any subject. She has extensive experience of correcting non-native English and some experience with MTPE. She mentioned having done STE and revised human translations and has considerable experience as a translator. Her typical editing work includes blog posts, business plans or reports, non-fiction books, marketing materials and web copy.

4.1.3 Author 2

A2 writes in her native Italian for a different local newspaper than A1 while studying Art and Literature at the same university. She is in her twenties and has no experience of composing articles, academic papers or other short texts in English, except for blog posts, which she has been writing both professionally and for fun for a year. English is her only second language, and she knows it to B2 level (Council of Europe, 2001). Her preferred subjects are art, literature, cinema and poetry. She initially considered using the article after the experiment but ultimately decided against it. She normally writes newspaper articles, academic papers/research reports, novels, short stories, poetry, and scripts for films, television and theatre in Italian for professional or academic purposes.

4.1.4 Editor 2

E2 is in her thirties with three years' experience as an English-language editor. She normally edits newspaper articles, academic papers, research reports, blog posts, technical manuals, scripts for films, television or theatre, business plans or reports, and essays. Besides her native English, she is fluent in Italian. For this experiment, she was willing to edit texts on professional development, health, career transitions, fashion and AI. She has a lot of experience in correcting non-native English and no experience in MTPE. She has also done a fair amount of STE, translation and human translation revision. She is currently a full-time content writer in Italy and has worked as a teacher of English as a second language for many years.

4.2 Content planning

4.2.1 Author 1

Since his editor was not an expert in the History of Art, A1 planned a blog post suitable for laypeople. He chose the title *How to Guide People to Look at a Work of Art*. He typically plans

such texts through brainstorming and by researching relevant sources. While writing in English, he usually uses DeepL Translate², WordReference.com³, PONS⁴ and Google Translate⁵.

4.2.2 Author 2

A2 chose the title *The Universal Language of Art* for her piece, which she classified as a newspaper article/blog post/short essay. Regarding her approach to planning such texts, she said that she just writes them and fixes them at the end, stating, “I plan only interviews.” She normally consults WordReference.com⁶, the Oxford English Dictionary⁷ and Merriam-Webster⁸ while writing in English.

4.3 Prompt engineering

Neither author chose to organize their text into sections or specified a style guide or writing conventions. None of the participants were asked to sign nondisclosure agreements.

4.3.1 Author 1

A1 said that the structure of the prompt he was asked to write was as he expected and found the process laborious but not overly difficult. He added that it was a helpful way to clarify his thoughts before writing. The notes he provided to GPT-4 were mostly in English and partly in Italian.

4.3.2 Author 2

A2 said that the structure of the prompt was as she expected, and quick and easy to write. The notes she provided to GPT-4 were entirely in English.

4.4 Text generation

Neither author chose to edit the prompt and try again. A2 opted to keep the raw GenAI output exactly as it came, while A1 decided to make some changes.

4.4.1 Author 1

A1 noted that the generated text was better than he expected, awarding it a score of eight out of ten. He observed that there was no content that was not implicit in the prompt and identified no serious errors. He commented that the raw output resembled something a human might write and was surprised by its accuracy, describing it as “a good base, especially for the lexicon,” since only a few things needed editing. One change he made was to replace an example provided by GPT-4 (a painting by Leonardo da Vinci) with one he considered more appropriate (a fresco by Michelangelo Buonarroti). Although A1 stated in the feedback form that there was no missing content, he added a whole sentence which he defined as “the main message of the article” and another to help the reader “understand the fact that art is something close to each of us.”

² www.deepl.com

³ www.wordreference.com

⁴ www.pons.com

⁵ <https://translate.google.com>

⁶ www.wordreference.com

⁷ www.oed.com

⁸ www.merriam-webster.com

4.4.2 Author 2

A2 noted that the generated text was better than she expected and also gave it a score of eight out of ten. She observed that there was no content that was not implicit in the prompt and there was nothing missing. Moreover, she commented that the GenAI output looked like something a human could write and found no serious errors.

4.5 Text editing

4.5.1 First impressions

4.5.1.1 Editor 1

E1 said that the text she received was very much better than the sample of her author's English written without the aid of AI or other tools. She classified it as an academic paper/research report or short essay, gave it a score of eight out of ten, and said it showed no serious issues. However, she added, “on first reading it seems quite high level/vague.” Although the subject matter was different from what she usually edits, she felt comfortable working with it.

4.5.1.2 Editor 2

E2 noted that the sample of her author's English written without the aid of tools displayed a high level of creativity. However, it contained spelling and grammar mistakes typical of native Italian speakers, which were absent in her GPT-4-generated text. She gave the GenAI-assisted text a score of six out of ten and said that it was repetitive and redundant, over-reliant on common phrases and lacked novelty and creativity. She added that the text was typical of GenAI, stating, “a couple of sentences in you begin to think ‘Wow, this has been written well.’ When you reach the second paragraph, it becomes dull — it lacks the human touch. Phrases like the *language of art* are repeated, all sentences are long, in fact of a similar length. What’s more, they are all highly descriptive and fanciful. It doesn’t speak to the reader.” Despite this criticism, she concluded that, on the whole, the information was very interesting and that it just needed tweaking. The subject matter was in line with the kind of thing she normally edits, and she classified the text as an academic paper/research report or short essay.

4.5.2 Errors and textual anomalies detected

4.5.2.1 Editor 1

4.5.2.1.1 Introduced by the author’s edits

E1 found a calque of an Italian expression (*the major part* instead of *the majority*), an incorrect or inconsistent verb tense (*is* instead of *was*) and an improper use of articles (*a* on-site installation). She also flagged a part that might benefit from being made more gender-neutral (*criticism is man's response to man, and we are all human*). There were also some misused prepositions (*the same of* and *see throughout the former*).

4.5.2.1.2 In the GenAI raw output

E1 noted a little redundancy in one expression (*composed of [...] composition*) and the non-existent word *grasitating* (see discussion below). The expression *the journey through art*, which E1 replaced with *looking at a work of art*, also came from GPT-4. She said, “I felt it could be clearer and tie in more with the actual topic of the piece.”

4.5.2.2 Editor 2

4.5.2.2.1 Introduced by the author's edits

The author chose not to edit the raw GenAI output.

4.5.2.2.2 In the raw GenAI output

E2 said that there was an over-reliance on common phrases noting that most sentences started with *this* or *the*, and that the word *through* and the expressions *the language of art* and *the universal language* were used too much. She added, “AI is a giveaway with this text due to the overly long and descriptive sentences that all tend to follow the same structure.”

4.5.2.3 Summary of textual anomalies

Anomaly	Description or effect	Also seen in MT
Excessive repetition of words or phrases	Poor lexical variety	Yes ⁹
Redundancy	Repetition of information without adding new meaning or value	No
Non-existent words	<i>See discussion</i>	Yes ¹⁰
Blandness	Absence of emotion, creativity or engagement	No
Verbosity	Overly long, highly descriptive, fanciful sentences	No
Low burstiness	Most sentences start in the same way and have uniform structure and length	No
Lack of complex analysis	Superficial, vague and lacking specificity	No
Perfect grammar and spelling	Grammatical mistakes and typos are more typical of human-written copy	No

Table 1. ST textual anomalies reported in this study

4.5.3 Author-editor interaction

4.5.3.1 Author 1

A1 was unable to judge if the kind and frequency of interaction with the editor were different due to the use of GenAI, as he does not have sufficient experience working with this type of editor.

4.5.3.2 Editor 1

E1 stated that the interaction with the author during the editing was more or less the same as normal. Regarding differences, she said, “I’ve never worked in a situation where I know that the text was written with the help of AI, so that was the only difference — that the author *blamed* a few things on ChatGPT.” She added, “when editing under these kinds of circumstances, there’s a kind of *third party* involved, which is a bit odd. When I ask an author for clarification, I want to know what it is that *they* (not the AI) meant or wanted to say.”

⁹ Vanmassenhove et al., 2021

¹⁰ Macken, 2019

4.5.3.3 Author 2

A2 said that her level of interaction with the editor during the editing was more or less the same as normal. However, she did not feel that she had sufficient experience of working with this kind of editor to say whether the interaction was in any way different from normal.

4.5.3.4 Editor 2

E2 remarked that there was no real interaction with the author during the editing, which is not the norm.

4.5.4 Authors' and editors' final opinions

4.5.4.1 Author 1

A1 rated the likelihood of using GenAI again as a tool for writing in English at eight out of ten, although he would never use it for academic papers, poems or essays. He found GenAI effective but noted that excessive use might affect a writer's ability to "feel the text." He rated the likelihood of employing an editor again to correct his English at ten out of ten, stating that he would have expected to pay 25 euros for their service in this experiment.

Originally, his preferred method for producing texts in English was to write in his native language and then use a professional translator. However, after this experiment, he said he would now write newspaper articles, blog posts and "light" texts directly in English with the aid of GenAI. He found this method useful for addressing his main difficulty with English (vocabulary) and believed GenAI could be a good tool for editing, although "it cannot replace a human editor."

4.5.4.2 Editor 1

E1 rated her likelihood of accepting future STE assignments as ten out of ten. She said she would have charged 35 euros for her work, had it been a real job. E1 was surprised at how much of the raw GenAI output her author had "taken wholesale." She expected her author to use GenAI more as a starting point and then tweak the output to make it their own. She reiterated her initial impression that the text seemed bland and not concrete enough, especially considering it was about art. She remarked that the text could have been much more engaging if it had focused on a specific work of art as a visual aid example, stating, "but this is — in my opinion — one of the limitations of AI at the moment: it's never specific enough to what you're trying to achieve." Assuming her author were an art expert, she would have recommended he write his article in his native Italian, where he could fully express himself, and then have it translated for a better overall text. "Editing it in this way myself would have been beyond the scope of a copyeditor, but something a development editor — and an art expert — might consider for a longer piece, like a book."

E1 does not believe there would have been much advantage in understanding her author's native language or having experience as a translator because "generally speaking, the quality of the language was very high." She noted one calque where she could immediately tell A1 had translated directly from Italian (see section 4.5.2.1.1), but added, "any monolingual English speaker could have spotted that and worked it out." However, she added that this text was a straightforward example. "Across a longer text where more edits have been made by the author (rather than coming directly from ChatGPT), it might get more annoying for the editor as you'll spend more time trying to figure out what the author meant/asking for clarification."

4.5.4.3 Author 2

A2 rated the likelihood of using GenAI in her English writing process again at only 50%, despite acknowledging its effectiveness in helping her. She rated the probability of employing an editor to correct her English again at ten out of ten and said that she would have expected to pay her editor 20 euros for her service. A2’s preferred method for producing texts in English was already to write directly in English before the experiment. She added that using GenAI as an English writing tool was an interesting approach and “is useful when you have to write an article or an essay with a huge number of words.”

4.5.4.4 Editor 2

E2 rated the likelihood of accepting STE assignments again in the future at ten out of ten. She said she would have charged 50 euros for the editing she did, had it been a real job. In her opinion, the challenge was less about correcting grammar errors and more about making the text more engaging and giving it a more *human* voice.

She added that she believed a monolingual English speaker could edit ST just as well as a translator who knew the author’s language although it might make the process quicker depending on the extent of the author’s errors, “but having knowledge of their language is not a guarantee that the editor will produce great work.”

5 Discussion

Since LLMs and modern MT engines are both artificial neural networks, one might naively imagine that the kinds of errors that occur in GenAI output might be similar to those commonly found in MT output, particularly when mixed-language prompts are used. However, the only two anomalies reported they have in common were the coining of non-existent words (Macken, 2019), and an over-reliance on common phrases, which manifests itself in MT output as poorer lexical variety (Vanmassenhove et al., 2021) and normalization (Toral, 2019). Interestingly, two of the excessively repeated expressions were found in the prompt, specifically in the title the author provided. This phenomenon is reminiscent of what is known in web copywriting as *keyword stuffing*¹¹.

The GenAI-assisted text also exhibited anomalies not normally associated with raw MT output (see section 4.5.2.3), such as redundancy, lack of engagement and complex analysis, and low burstiness, a feature also measured by the automatic AI content detector GPTZero (Chaka, 2023). Consequently, STE differs from MTPE more than one might initially suspect. No cases of hallucination (Xu et al., 2024) were identified, probably because — in this experiment — GenAI was used to write up notes provided by the author rather than create new content.

The authors were impressed with the GenAI output, noting that it closely resembled human-written text, whereas the editors immediately recognized it as different. The authors’ opinion is consistent with the observations of Clark et al. (2021), who noted that “untrained evaluators are not well equipped to detect machine-generated text”. Even with training, Clark et al. found that the detection success rate only marginally improved, reaching about 55%. Dou et al. (2022) proposed a framework that could potentially improve this rate, which was validated in a subsequent study (Dugan et al., 2023). However, the ten error categories identified in their framework do not align with the anomalies reported in this study, except for redundancy. In fact, Clark et al. found that style-related aspects were not reliable detection criteria. Nevertheless, these are issues that an editor has to address. In future studies, it would be

¹¹ https://developers.google.com/search/docs/essentials/spam-policies?hl=en&visit_id=638602790106513155-3530511413&rd=1#keyword-stuffing

interesting to investigate whether GPT-4 can be explicitly prompted to generate text devoid of the reported anomalies.

Both editors acknowledged that GenAI improved their author's English grammar and spelling, which are known to be infrequent error types in ST¹² (Dou, 2022; Gillham, 2024).

Regarding the non-existent word *grasitating*, this error has already been noted several times¹³. The intended word is *grasping* in all previous reports and in this experiment. The precise cause of this error is unclear, but it may be related to the tokenization of non-existent words found in the training data, probably resulting from optical character recognition (OCR) errors. For example, *gravitation* is sometimes read as *grasitation* by OCR software¹⁴. Interestingly, *grasitating* appears to be gaining traction as a neologism, as a simple Google search shows¹⁵, and may one day appear in the dictionary.

The use of GenAI seems to alter the author-editor dynamic. In this experiment, when the author accepted the unaltered GenAI output, interaction with the editor was minimal. In the other case the editor perceived a sort of *third party* whose work the author could not clarify. This experiment showed that authors are likely to use GenAI when writing in a second language, especially if they have relied on translation in the past. However, one editor in this experiment suggested that her author may have obtained better results if he had written in his native language and had his text translated, instead of using GenAI.

Human editing, by both authors and editors, remains crucial in refining and enhancing GenAI output. However, the authors in this experiment would have been willing to pay only 40 to 71% of what the editors would have typically charged for the service they provided. Despite the editors' work, the final texts still have a high likelihood of being recognized as GenAI output (79% and 95%, respectively, according to the Plagamme AI detector¹⁶). Therefore, if the hypothetical publishers had a strict no-AI rule, more extensive editing might be necessary, potentially making this working method uneconomical.

Besides using GPT-4 as a drafting tool, as in this experiment, one author suggested using it as an editing tool too. When prompted to correct the grammatical errors he had unintentionally introduced into his draft, GPT-4 successfully removed all of them. However, it erroneously corrected the previously mentioned calque to *the fact that many*, instead of *the majority*. In this case, the author would probably have reinstated his original error to correct GPT-4's misinterpretation. Interestingly, it also corrected *grasitating*. Upon examining this new output, his editor remarked that it had helped with "the part of editing that takes the smallest chunk of my time." Nevertheless, the partial results of the unconcluded third case study further support this use, suggesting that for complex texts, it might be more fruitful for the author to draft the paper in their second language unaided and then use GPT-4 as an editing tool to refine the rough draft.

The editors did not see any significant advantage in understanding the author's native language or having translation experience. However, since both editors were bilingual, they may not fully appreciate the potential difficulties monolingual editors might face in correcting English as a second language.

The working method presented in this paper can replace translation only when there is no need for an original text in the author's native language. Moreover, if the same text is required in multiple languages, it is clearly more cost-effective to write in one of those languages and

¹² They are not so infrequent in other languages.

¹³ https://www.reddit.com/r/ChatGPT/comments/1ai9cfi/chatgpt_made_up_a_word_typo/

¹⁴ <https://www.govinfo.gov/content/pkg/GPO-CRECB-1916-pt4-v53/pdf/GPO-CRECB-1916-pt4-v53-14.pdf> — Search for the word *grasitation* in the text with CTRL+F (mentioned in discussion at footnote 13).

¹⁵ www.google.com/search?q=grasitating

¹⁶ <https://www.plagamme.com>

translate it into the others. Consequently, any decrease in translation work due to GenAI-assisted second-language authoring is not likely to be substantial.

6 Conclusion

We should be very cautious about generalizing the conclusions of these two exploratory case studies since they both concern similar types of text on very similar topics in the same language by authors with similar profiles. It would be advisable to repeat the experiment with different kinds of text on a wide variety of subjects in different languages by authors from diverse backgrounds.

The errors and textual anomalies found were either human errors introduced by the authors writing in a second language or typical GenAI anomalies, such as verbosity and excessive repetition of words or phrases (OpenAI, 2022), probably excluding hallucination (see discussion above). A summary of the detected anomalies is shown in Table 1 in section 4.5.2.3.

Although the authors found prompt engineering intuitive, providing some basic training in this area might be beneficial. This may lead them to adjust the initial prompt to try to produce better base GenAI output for editing. Moreover, both second-language authors and editors should be trained to discern and enhance AI-generated content through STE. This study reveals the importance of human editors in adding creativity and engagement to AI-generated texts.

References

All hyperlinks last accessed 18 September 2024.

- Chaka, Chaka. 2023. Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning & Teaching*, July 2023. <https://doi.org/10.37074/jalt.2023.6.2.12>
- Clark, Elizabeth, Tal August, Sofia Serrano, Nikita Haduong, Suchin Gururangan, Noah A. Smith. 2021. All That's 'Human' Is Not Gold: Evaluating Human Evaluation of Generated Text. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. <https://aclanthology.org/2021.acl-long.565/>
- Corder, Stephen Pit. 1975. *Error Analysis, Interlanguage and Second Language Acquisition*. Cambridge University Press. <http://dx.doi.org/10.1017/S0261444800002822>
- Council of Europe. 2001. *Council for Cultural Co-operation. Education Committee. Modern Languages Division. Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Dondoni Braz, Ana Clara. 2024. Can You Spot AI-Generated Text? Learn How To Recognise It. *Growth Tribe*. <https://growthtribe.io/blog/spotting-ai-generated-text>
- Dou, Yao, Maxwell Forbes, Rik Koncel-Kedziorski, Noah A. Smith, Yejin Choi. 2022. Is GPT-3 Text Indistinguishable from Human Text? *Scarecrow: A Framework for Scrutinizing Machine Text*. <https://doi.org/10.48550/arXiv.2107.01294>
- Dugan, Liam, Daphne Ippolito, Arun Kirubakaran, Sherry Shi, Chris Callison-Burch. 2023. Real or Fake Text?: Investigating Human Ability to Detect Boundaries between Human-Written and Machine-Generated Text. <https://doi.org/10.48550/arXiv.2212.12672>
- Farrell, Michael. 2024. On the various kinds of post-editing - Machine translation post-editing, translation memory match editing, hybrid post-editing, monolingual post-editing, stealth post-editing and synthetic-text editing. Presented at *Scenari Multimediali e Didattica della Traduzione - Teaching Translation for Multimedia Scenarios*, Milan - 14-15 December 2023, currently undergoing peer review.
- Gillham, Jonathan. 2024. How To Identify AI-Generated Text?. *Blog of Originality.ai AI & Plagiarism Detector*. <https://originality.ai/blog/identify-ai-generated-text>
- Gluska, Justin. 2023. How to Check If Something Was Written with AI. *Gold Penguin*, <https://goldpenguin.org/blog/check-for-ai-content/>
- James, Carl. 1998. *Errors in Language Learning and Use: Exploring Error Analysis*. Routledge. ISBN 9780582257634
- Macken, Lieve, Laura Van Brussel, Joke Daems. 2019. NMT's Wonderland Where People Turn into Rabbits. A Study on the Comprehensibility of Newly Invented Words in NMT Output. *Computational Linguistics in the Netherlands Journal* 9:67–80. <https://www.clinjournal.org/clinj/article/view/93/84>

- OpenAI. 2022. Introducing ChatGPT. <https://openai.com/index/chatgpt/#fn-1>
- Popović, Maja. 2018. Error Classification and Analysis for Machine Translation Quality Assessment. In Translation Quality Assessment. Machine Translation: Technologies and Applications, vol 1. Springer, Cham. https://doi.org/10.1007/978-3-319-91241-7_7
- Toral, Antonio. 2019. Post-edits: an exacerbated translationese. In Proceedings of Machine Translation Summit XVII Volume 1: Research Track, p. 273–281, Dublin, Ireland. European Association for Machine Translation. <https://doi.org/10.48550/arXiv.1907.00900>
- Vanmassenhove, Eva, Dimitar Shterionov, Matthew Gwilliam. 2021. Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, April 19-23, 2021. <https://doi.org/10.48550/arXiv.2102.00287>
- Wu, Tongshuang, Michael Terry, Carrie J. Cai. 2021. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. <https://doi.org/10.48550/arXiv.2110.01691>
- Xu, Ziwei, Sanjay Jain, Mohan Kankanhalli. 2024. Hallucination is Inevitable: An Innate Limitation of Large Language Models. <https://doi.org/10.48550/arXiv.2401.11817>
- Zhou, Yongchao, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, Jimmy Ba. 2022. Large language models are human-level prompt engineers. Published as a conference paper at ICLR 2023. <https://doi.org/10.48550/arXiv.2211.01910>

Appendix: Funding

The research project reported in this paper has received funding from the International Center for Research on Collaborative Translation at the IULM University and a Small Grant for Research from Mediterranean Editors and Translators.

The authors and editors were offered a token participation fee, but only three of them chose to accept it. They received 125 euros each.