

AI Editing: Are We There Yet?

Rachel Baron

Abstract

Recently emerged artificial intelligence (AI) tools, such as ChatGPT, have the potential to facilitate all aspects of academic research and publication. Over the past year, we have rigorously tested several generative AI models to explore how we can use the technology to refine our editing. This article presents a summary of our findings, identifying the strengths and weaknesses of AI tools for editing and highlighting some of the ongoing challenges we have encountered in incorporating these tools into our editing process. The insights provided by our testing should help both authors and editors make decisions about which editing tasks AI can be effectively used for and which tasks are best left for human editors.

The introduction of ChatGPT (from OpenAI) in November 2022, followed by other generative AI models such as BERT (from Google), has revolutionized many aspects of our lives. The initially intense debate around its use in academia has abated somewhat, as the early ethical issues have been mostly resolved and the answers to various questions have become clearer. AI has various potential uses for almost every aspect of academic research, from scoping initial research ideas to analyzing data, proofreading manuscripts, and identifying suitable journals for article submission.¹⁻³

Along with the initial hype came numerous reports of AI's spectacular "fails," some of which were amusing, and others more troublesome. The main problem with its use for academic research is its tendency to "hallucinate," whereby it presents facts and even references that look plausible but are completely invented (e.g., Bhattacharyya et al⁴). Another issue with generative AI is that it tends to reproduce inherent biases and stereotypes that exist in the training dataset.⁵ While these problems are of concern, they are not the focus of this paper. Here, we aim to provide insights into the utility of AI tools for editing and proofreading academic text.

Dr Rachel Baron (<https://www.linkedin.com/in/rachel-baron-75bb1817/>; <https://orcid.org/0009-0000-6502-9734>) is the Managing Editor for Social Sciences and the Head of Training & Development at AsiaEdit.

Opinions expressed are those of the authors and do not necessarily reflect the opinions or policies of the Council of Science Editors or the Editorial Board of Science Editor.

<https://doi.org/10.36591/SE-4703-23>

We recently published the findings of a survey we conducted with our clients at 2 timepoints in 2023, asking them about their attitudes toward, and use of, AI-based editing and writing tools.⁶ In that article, we cautioned against using AI without postediting by a human editor. Here, we elaborate on the findings of our extensive testing of ChatGPT and other AI-based editing tools for editing academic papers written by multilingual researchers. As a company providing editing services to researchers and authors across Asia, we consider AI to be both a threat and an opportunity. As AI becomes increasingly sophisticated and accepted, it poses a risk to many people's jobs, not just those in the editing industry. However, it is also an opportunity that we anticipate will help us to improve our services and reduce our costs, which we can then pass on to our clients. Our mission is to help multilingual writers to disseminate their research, and AI has great potential to further level the playing field by eliminating or reducing language barriers for authors.

Some of the challenges associated with using AI editing tools may well be resolved as the technology develops. However, there seems to be an inherent limitation in large language models such as ChatGPT, which are pretrained on a massive corpus of data and work by predicting the most likely next word based on the probability of words occurring together. Whereas a human editor retains an overall view of the paper in mind when reading each paragraph or section, AI tools are only trained to find the word with the highest probability of appearing next. In fact, they do not even work at the word level, but split up words into shorter chunks referred to as "tokens," so they are actually predicting the next most likely token to appear. This involves no analysis or understanding of the text, which can lead to the kinds of distortions of meaning and errors that we report here.

Obtaining the result you are looking for with AI tools is not an easy task. It is simple enough to provide the tool with an instruction such as "copyedit the following text," but the result is unlikely to be satisfactory. Writing useful instructions, or "prompts," requires a lot of trial and error combined with expertise in analyzing the output. This is what we have focused our efforts on over the past 18 months, and we report our findings here.

We expect this article to provide useful insights for copy editors, journal editors, authors, and others interested in the use of AI for editing purposes, which will help them to understand the strengths of AI, while being aware of its shortcomings. We do not aim to dissuade people from using AI tools, as they undoubtedly have a growing place

CONTINUED

in every writer's toolbox; rather, we aim to highlight areas where caution should be applied in their use. We also do not aim to review the use of AI tools for the numerous other academic tasks that it may be useful for, such as designing studies, summarizing the literature, analyzing data, and helping to write the initial draft of a paper, as these are outside the scope of this article.

Practicalities and Ease of Use

Although AI tools, such as ChatGPT, are readily available and simple to use, their limitations present various practical challenges for authors and editors. A major limitation is that when using a browser-based interface, there is a limit to the amount of text that can be copied into the chat window for editing. This limit is currently 4,096 characters in ChatGPT 3.5, which is far short of a typical research paper. This means the tool cannot edit the paper as an entire document, which creates various practical problems with breaking up the file into smaller parts and reassembling them. A greater problem, however, is that without the full context, the editing cannot take account of the content and style in the other parts of the paper, which leads to inconsistencies and repetition.

The limit for ChatGPT 4 is much higher, at around 25,000 words, but this is a subscription service that still does not overcome the more serious problems encountered with the free version. There are various ways to get around the word limit in the free version, yet none of them are capable of ensuring consistency across a document, a task that is second nature to human editors. While this can be overcome by incorporating some additional editing time after the AI edit, it does not solve the more serious problem that we have consistently encountered when trying to edit whole papers: After processing more than a few hundred words, ChatGPT seems unable to cope and starts deleting large chunks of text and replacing them with single-sentence summaries.

Other issues include losing formatting when copying and pasting between Microsoft Word and the AI editor. None of the versions we have tested are able to distinguish textual elements, such as headings, which it tends to incorporate into the main text. These changes have to be identified and the headings reinstated when the text is copied back into Word. Similar problems occur with footnotes, tables, and other nontextual elements, such as equations, meaning that the main text must be copied in short chunks to avoid these elements.

Rather than copying and pasting text into a browser-based chat window such as ChatGPT, an increasing number of add-in tools are being developed for Word that are based on ChatGPT and perform the editing or make suggestions directly within the document. The suggestions appear in a pane alongside the main text, which can then be accepted or rejected. This method has the advantage of viewing the

suggested changes in the familiar environment of Word with the changes tracked. The tools usually offer various prompts to vary the level of editing, although we have not yet found one that offers the right balance between under- and over-editing. We cover this in more detail in the following section. A great disadvantage of these tools is that the editing is done at the sentence level, which again, creates problems with inconsistency and repetition. For example, abbreviations are redefined each time they appear because the AI tool does not recognize they have already been defined. Concepts and definitions of terms are also re-explained and redefined, leading to a great deal of redundancy. In the following example, the four groups had already been defined and referred to earlier in the paper, but as the AI editor does not remember this contextual information, it attempts to provide a definition:

Section II examines four different groups of people who want to make changes to the fourway reformthingsumpsare (done. These groups are called reform rejecters, incrementalists, systemic reformers, and IIA quitters). inThe termssectionoftalks theirabout respectivewhatpositions each group thinks and thewhat measures they want to bedotaken-to advance make their proposedideas reforms happen.

These suggested changes can, of course, be rejected, but because they are enmeshed with the more useful changes, it takes a lot of effort to distinguish necessary from unnecessary changes and diminishes any time savings that might have been gained from AI editing.

Another problem with Word-based add-ins is that they do not always allow the changes to be transferred from the viewing pane to the document itself without creating problems with the formatting. This can occur if the text contains footnotes, reference fields, equations, and other nonstandard text. Again, this is not a fatal flaw, but it does require additional time to implement the suggested changes in the document. In documents containing many such elements (e.g., any document in which references have been inserted using referring software), this might mean making every change manually.

Editing Quality

AI can produce excellent quality text in grammatically perfect English. However, it does not do so reliably and consistently, and its performance in some areas can be quite poor. Editing with AI is not as prone to some of the more well-known problems encountered with text generation, such as the tendency to "hallucinate" and invent nonexistent citations. However, it throws up sufficient problems to make us wary of using it without a high level of oversight. We outline some of these problems in the following sections.

Depth and Accuracy of Editing

Human editors should aim to edit text in such a way that it retains the author's voice and preserves the intended meaning. We try

CONTINUED

not to make unnecessary changes, first, because doing so takes additional time that clients must eventually pay for, and second, because we should always respect the author's own style and word choice, where appropriate. AI takes neither of these things into consideration. Prompts that ask the AI to edit lightly tend to do little more than correct gross errors, but allowing it free reign by asking it to, for example, "edit the following text," inevitably results in a complete rewrite that leaves little trace of the author's original style. Perhaps some authors do not object to this; however, our experience suggests that many authors do not welcome the wholesale deletion and rewriting of text and consider the unnecessary substitution of close synonyms (e.g., shows → indicates, meaningful → significant) pointless at best. We have yet to find an AI editor or prompt that prevents this kind of rewriting; even including instructions such as "do not substitute synonyms" in the prompt does not stop it from doing so.

The following excerpt illustrates these problems. The first sentence, along with its references, is deleted entirely. The second sentence contains an unnecessary synonym substitution (examine → investigate). The change in tense from past to present is incorrect because the study has been completed. The change from "the co-influence" to "the relationship between" alters the intended meaning, as the author really did mean "co-influence": the focus in the edited version is on the relationship between "team manager's interpretation and inconsistent mental models" rather than their combined effect on team effectiveness.

Editing that changes the author's intended meaning in this

As the core process of cognition, interpretation and sense-making are crucial to mental models and deviance of organization, organizational supports and colleagues interpretation would reduce the deviant behaviors and negative effects of inconsistent mental models (Weick, 1995; Tuzun et al., 2017). This study's purpose was to examine the co-influence of relationship between team manager's interpretation and inconsistent mental models on team effectiveness, especially with a focus on examining the moderating effect of interpretation on the relationship between inconsistent mental models and performance. Based on the questionnaire of interpretation mode developed by this study firstly, after extensive The study employs a questionnaire survey, examined to gather data and determine the co-effect of interpretation and inconsistent mental models on team performance. Figure 1 depicts our overall The research model is depicted in Figure 1.

way is highly problematic. Scientific writing must be accurate, above all else. Multilingual writers understandably struggle to express their meaning clearly, and the editor's job involves deciphering meaning from ambiguous and unclear writing. We are not infallible, but our years of expertise and field-specific academic knowledge mean that we get it right most of the time. When we are unsure, we leave a comment explaining why the meaning is unclear and offering suggestions for improvement. In contrast, given a string of words that do not immediately express a clear meaning, an AI editor will make a best guess based on the probability of particular words following one another, rather than on some inherent sense or understanding

of the text. Sometimes this will result in a perfectly written sentence that clearly expresses the author's intended meaning. At other times, it will result in superficially correct nonsense, or at least a distortion of the intended meaning. Unfortunately, AI tools aim to please by always offering a suggestion, which does not necessarily equate to an improvement.

Some authors are able to easily spot such changes and correct them. However, to do so can require a command of English that many multilingual writers do not possess. If they were able to express the meaning clearly in the first place, there would have been no ambiguity in the original text, and the change of meaning would not have occurred. Indeed, the task of spotting changes in meaning is not trivial, even for our expert editors. Given that AI editing often produces superficially well-written sentences, the editor doing the post-AI checking needs to continually go back and forth between the original and the edited text to spot any infelicitous changes. Some of these may be relatively trivial, but given the nature of scientific research, they have the potential to be disastrous. We have found the task of checking for changes of meaning to be so time consuming that it often eliminates any gains from the initial AI editing.

To avoid the potential for changes of meaning, it is possible to use alternative prompts, such as "proofread" or "edit the text for grammar only," that avoid excessive editing and make it much easier to check for inappropriate changes. However, this creates the opposing problem of under-editing, requiring further human editing to fix awkward syntax and ambiguous phrasing.

Grammatical Correctness

AI is very good at correcting basic English grammar, and we have encountered few serious problems in this area. However, a few things have given us cause for concern, some of which are described here.

Articles

We find many errors in AI's use of definite and indefinite articles, often because it does not recognize the context. For example, whereas we would always use the definite article before common nouns such as "participants" and "results" when referring to the present study, AI editors generally do not. Multilingual writers also struggle to understand the nuances of article usage, so we assume they would find it difficult to identify such errors. However, it is not only incorrect but also confusing to omit articles where they are needed. For example, we should assume that a sentence that begins "Results show that..." refers to results in general, perhaps from numerous studies. However, "The results show that..." changes the meaning so that the reader now knows "the results" are those of the present study. The following is an example of where the AI editor incorrectly deletes "the," creating a grammatical error:

CONTINUED

In our study, the SRMR = 0.044, NFI = 0.899, GOF = 0.758, and R2 (AM medium = 0.548, AP large = 0.858, ASE medium = 0.580, AB medium = 0.558) values for academic performance showed good predictive relevance.

Subject–Verb Agreement

Although AI editors rarely make agreement errors in simple sentences, they are prone to do so in sentences where the subject (noun) does not immediately precede the verb. In the following example, the verb “to be” belongs to the subject “governance quality” and should be in the singular form “is”; however, in the AI-edited sentence it is changed to “are,” presumably because it perceives the verb as belonging to the plural “processes and institutions.”

Governance quality refers to how well the processes and institutions of public governance function, and are widely recognized as having an important influence on human well-being.

Verb Tense

Correct and consistent use of verb tense is one of the most difficult aspects of English grammar for multilingual writers, and the problem is compounded by differences in norms across academic fields. Overall, we are not particularly impressed by AI’s ability to select the correct verb tense for the context. For example, in the following excerpt from an abstract, the verb tense for describing the study should be in the simple past tense, but the AI editor changed it to present tense. Given the study has been completed, it does not make sense to refer to its aims in the present tense. Furthermore, we would at least hope to see all associated verbs in the same tense, yet the subsequent sentence switches to past tense when stating what was investigated.

Original: This study compares the mental health of medical, nursing and administrative staff in the UK emergency department and the comparative orthopaedic department. The study investigated the impact of coping strategies and the support people received from their colleagues (i.e., social support).

AI edited: The study aims to investigate the pressure emergency physicians face. The aim of the study is to determine whether there are any differences in the mental health of staff in these two departments. The study investigated the impact of coping strategies and social support on people.

As the excerpt shows, the AI editing does not improve on the relatively well-written original. It needed only a light edit, yet the edited version reads less well than the original. This example again illustrates how AI can delete important information for no apparent reason. In this case, the reference to the emergency department and orthopaedic department is replaced with “these two departments,” but without naming them, the sentence has no meaning.

In the following example, the AI editor assumes the sentence in the introduction of a paper refers to a specific past event, rather than makes a general statement:

Original: Policy in Hong Kong is developed by ministers appointed politically to head government bureaus.

AI Edited: The policy was developed by ministers appointed politically to head government bureaus in Hong Kong.

“Lost” Elements of Text

Because Word-integrated AI editing tools tend to edit line-by-line, they generally do not take account of the surrounding context. This method has various disadvantages, among which is the loss of the narrative elements of text. As each sentence is treated as a stand-alone statement, AI editors tend to delete words and phrases that link sentences and paragraphs, such as “however,” “therefore,” “in contrast,” and “moreover.” Markers, such as “first ... second ... third” are also deleted and sometimes replaced with “additionally” as the AI editor does not recognize the numbered sequence.

Consistency

An important aspect of copyediting is to ensure consistency of style, terminology, and formatting throughout a document. For example, a human editor will check that all repeated terms are the same, without switching back and forth between near-synonyms. It is not uncommon to find papers that start off, for example, using the term “company,” then drift off into using firm, business, enterprise, venture, corporation, organization, and more, leaving the reader to wonder whether the author really means to imply differences between them or is just attempting to make the text more interesting by providing variety. In almost every paper, we find variations, for example, in the use of singular vs. plural (e.g., “parents and their children” vs. “the parent and their child”); the use of the possessive (e.g., “tribunals’ treaty interpretation process” vs. “the treaty interpretation process of tribunals”); the capitalization and italicization of terms and headings; the presentation of numbers as words or digits; and the reporting of statistics. None of the AI editing tools we have used have been able to ensure consistency in all, or sometimes any, of these aspects of style, which leaves yet another task for the human editor.

Paraphrasing Quotations and Deleting Citations

A rather alarming tendency of AI editors is to take quotations out of quotation marks, lightly paraphrase them, and delete the citation. The following is one such example, although here, at least, the citation was retained:

CONTINUED

This **outcome metric is consistent metric aligns** with the traditional Chinese **narrative belief** that the officialdom **was** a natural **outlet path** for outstanding students and scholars² in imperial China (Zhang, 1920).

It hardly needs pointing out that this could put authors at risk of unintentional plagiarism.

Failure to Alert the Author to Unclear or Missing Information

Human editors write frequent comments to the author, querying various aspects of the writing. For example, they might need to ask the author for clarification when the intended meaning is unclear; they will point out missing important information, such as from where a sample was collected; they will prompt the author to add missing references; they will suggest moving sentences or paragraphs to improve the flow; and they will point out inconsistencies between figures reported in the text and in tables, or at different places in the text. None of these issues will be highlighted by an AI editor, and it is much harder for the human editor to add such comments post-AI editing, as they will not engage with the material in as great a depth and may not identify the problems as easily.

AI Signature

Recently, we have started to receive enquiries from customers asking us to make the text “more human,” presumably because they have used AI to assist with the writing. Consequently, we have collected a set of “signature” phrases that alert our editors to the potential use of AI in papers sent to us for editing. The following are a few examples of such phrases, which can often make the text sound more like marketing material than academic scholarship:

- delve/delve into
- tapestry
- leverage
- it’s important to note/consider (and similar phrases with “dummy” pronouns)
- remember that
- navigating
- landscape
- in the world of/in the realm of
- embark
- dynamic
- testament
- embrace
- intricate
- excessive use of flowery adjectives (e.g., “meticulously commendable”)
- complex and multifaceted

Of course, these phrases are not exclusively used by AI, but it is worth mentioning them because it helps to be aware

of the possibility that a paper has been written or edited with AI assistance. Editors need to take particular care with such papers because they are more likely than usual to contain distortions of meaning and errors, such as those described above.

Conclusion

Our findings suggest that AI tools are not yet ready to take on the task of editing academic papers without extensive human intervention to generate useful prompts, evaluate the output, and manage the practicalities. Our concerns echo those of previous studies (e.g., Meyer et al⁵, Lingard et al⁷), suggesting that despite the hype and promise, pure AI editing is still some way off.

When we began experimenting with AI editing tools, we were cautiously optimistic that they would soon be able to reduce our editing times and cut the cost of our services for clients. Despite the limitations identified in our testing, we have recently launched our hybrid AI-human editing service. Although our main focus remains on fully human editing, AI-assisted editing is now an option, especially for early-career researchers who find it particularly difficult to access full-priced editing services. Nevertheless, whatever improvements are made to AI editing models, we believe that intervention by a human editor will continue to be an essential step in maintaining the high-quality service that academic editors offer their clients. We hope that our findings help editors and authors to refine the outputs of their own AI-assisted writing and editing.

References and Links

1. Marchandot B, Matsushita K, Carmona A, Trimaille A, Morel O. ChatGPT: The next frontier in academic writing for cardiologists or a Pandora’s box of ethical dilemmas. *Eur Heart J Open*. 2023;3:orad007. <https://doi.org/10.1093/ehjopen/oead007>.
2. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*. 2023;11:887. <https://doi.org/10.3390/healthcare11060887>.
3. Chen T-J. ChatGPT and other artificial intelligence applications speed up scientific writing. *J Chinese Med Assoc*. 2023;86:351-353. <https://doi.org/10.1097/JCMA.0000000000000900>.
4. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus*. 2023;15:e39238. <https://doi.org/10.7759/cureus.39238>.
5. Meyer J, Urbanowicz R, Martin P, O’Connor K, Li R, Peng P, Bright T, Tatonetti N, Won K, et al. ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining*. 2023;16:20. <https://doi.org/10.1186/s13040-023-00339-9>
6. Sebastian F, Baron R. AI: What the future holds for multilingual authors and editing professionals. *Sci Ed*. 2024;47:38-42. <https://doi.org/10.36591/SE-4702-01>.
7. Lingard L. Writing with ChatGPT: an illustration of its capacity, limitations and implications for academic writers. *Persp Med Educ*. 2023;12:261-270. <https://doi.org/10.5334/pme.1072>.